

Estudo sobre conversão de sinal de áudio para MIDI

Matheus Bitarães de Novaes

Este relatório tem por intuito descrever o desenvolvimento de um algoritmo para conversão de um determinado sinal de áudio para o formato de dados MIDI. Este trabalho faz parte da disciplina de Processamento de Sinais do Programa de Pós-Graduação da Universidade Federal de Minas Gerais

I. Introdução

Durante a década de 70, os dispositivos musicais eletrônicos foram se tornando cada vez mais populares no mercado musical. Com isto começou a nascer um desejo de modificar os sintetizadores, que eram monofônicos e analógicos, para sintetizadores digitais, capazes de controlar outros instrumentos e realizar troca de dados. Alguns fabricantes desenvolveram técnicas próprias, porém isto impedia a comunicação entre instrumentos de outras marcas.



The Back Cover of Herbie Hancock's 1978 record Sunlight

Visando melhorar este panorama, o fundador da Roland Corporation, Ikutaro Kakehashi, propôs uma padronização entre as principais marcas de instrumentos digitais do mundo. E foi assim que nasceu o protocolo MIDI (Musical Instrument Digital Interface), lançado oficialmente em outubro de 1982 e com as especificações lançadas em agosto de 1983.

II. O sinal MIDI

O sinal MIDI é um sinal digital que transmite em tempo real informações de como um sinal de áudio deve ser excitado, bem como outros dados de estado do instrumento, de acordo com as entradas tocadas em um controlador.

As mensagens MIDI são informações de 8 bits transmitidas em série à 31,25 kbit/s. Esta taxa foi escolhida porque é a divisão exata de 1 MHz, a velocidade a que muitos dos primeiros microprocessadores operam [1]. O primeiro bit de cada dado identifica se a informação é um byte de estado ou de um byte de dados, e é seguido por sete bits de informação.[2] Um bit de partida e um bit de parada são adicionados a cada byte para fins de enquadramento, de modo que um byte MIDI requer dez bits para a transmissão.[1]

Uma mensagem MIDI pode conter 16 canais, ou seja, pode informar mensagens para 16 diferentes “instrumentos” ou 16 posições do mesmo instrumento. Há cinco tipos de mensagens: de Voz do Canal (Channel Voice), de Modo do Canal (Channel Mode), Comum do Sistema (System Common), em Tempo Real do Sistema (System Real-Time), e Exclusiva do Sistema (System Exclusive).[3]

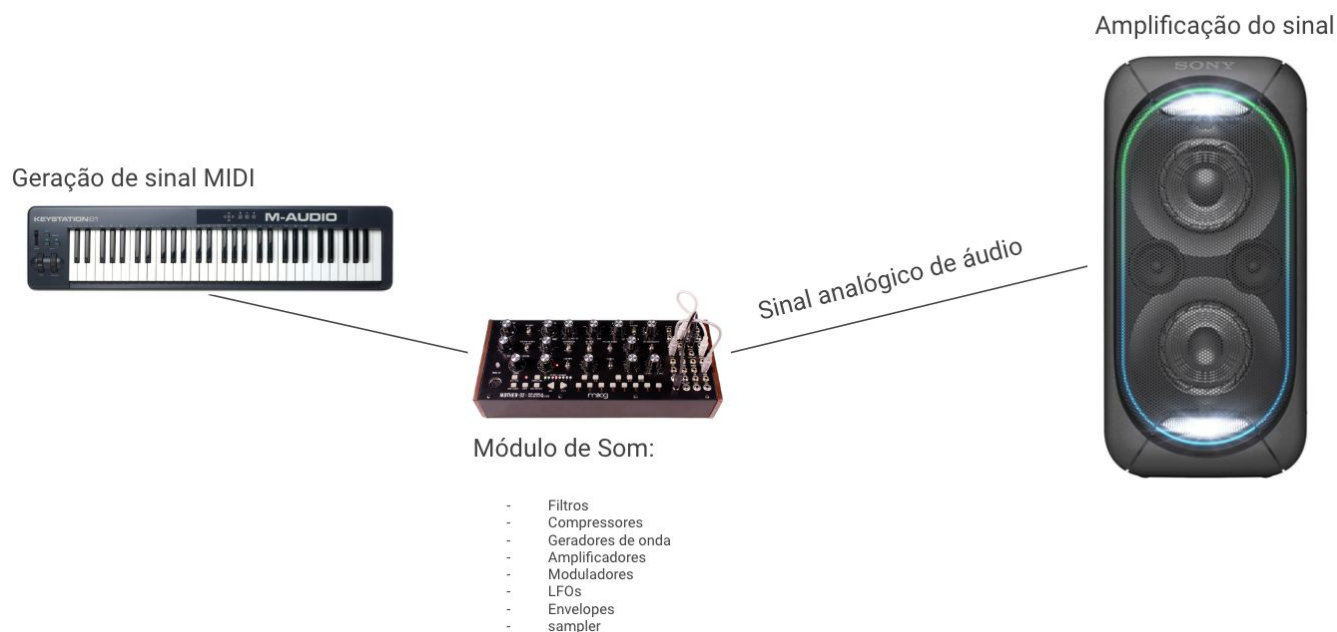
As principais mensagens MIDI estão descritas abaixo:

Voice Message	Status Byte	Data Byte1	Data Byte2
Note off	8x	Key number	Note Off velocity
Note on	9x	Key number	Note on velocity
Polyphonic Key Pressure	Ax	Key number	Amount of pressure
Control Change	Bx	Controller number	Controller value
Program Change	Cx	Program number	None
Channel Pressure	Dx	Pressure value	None
Pitch Bend	Ex	MSB	LSB

Tais sinais podem ser utilizados de diversas maneiras. A mais popular é a utilização em um teclado, onde o próprio teclado é um controlador (emissor de sinais MIDI) e gerador de som (receptor de sinais MIDI)



Outro uso popular é a utilização de um teclado controlador, que é responsável apenas pela emissão de sinais MIDI, combinado com um(ou mais) módulo(s) gerador(es) de som.



III. Conversão de um sinal de áudio para MIDI

Com o intuito de estudar mais sobre o sinal MIDI, foi proposto o objetivo de transformar um sinal de áudio (uma gravação de um solo de oboé) em informação MIDI, para depois transmitir esta mesma informação para um reprodutor MIDI e comparar a conversão. Desta forma, é possível comparar a conversão, bem como executar o mesmo solo de oboé com outros instrumentos.

Para esta conversão, foi necessário extrair algumas informações do sinal de entrada:

- **Frequência fundamental em determinados espaços de tempo:** A aproximação da frequência fundamental nos dá a informação de qual nota o instrumento digital deve reproduzir.
- **Energia do sinal:** A energia do sinal nos informa qual é o volume que cada nota deve alcançar.
- **Tempo de permanência em uma frequência fundamental:** determina a duração da nota.

Foi definida uma divisão igual da trilha em blocos de 0.2s. Ete tempo foi definido empiricamente. Menos que isso gerava muita instabilidade na aproximação da frequência fundamental e mais que isso era prejudicial para a aproximação da dinâmica de transição das notas. Esta estratégia de janela fixa pode ser aprimorada para detectar a variação da frequência fundamental em um espectrograma e assim atuar com maior assertividade, separando os blocos onde as notas são as mesmas.

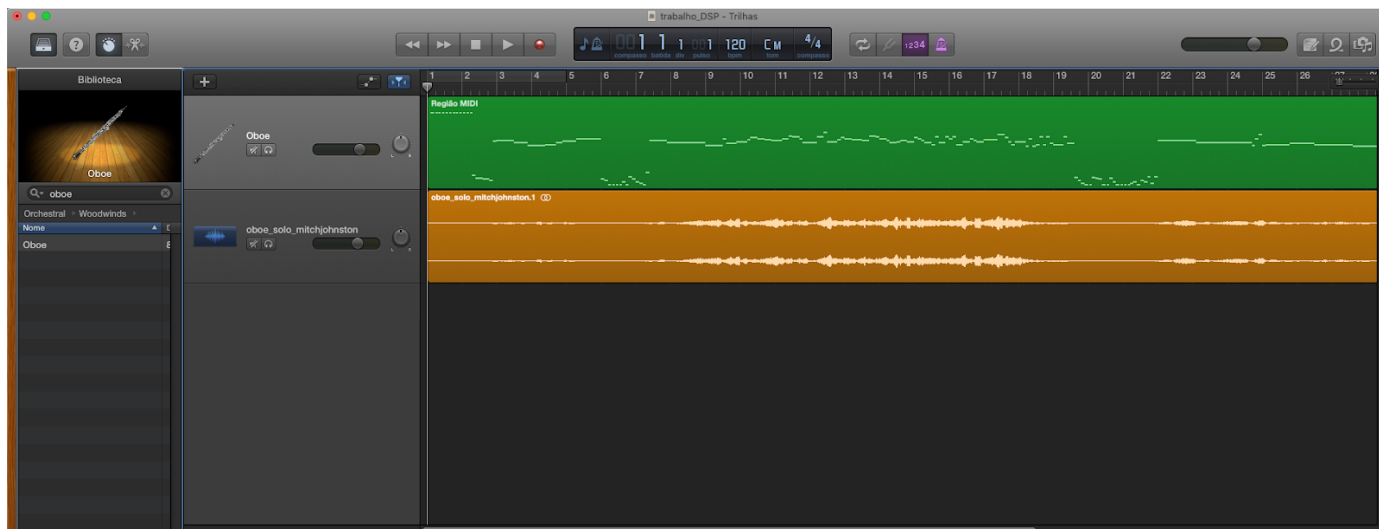
Para a estimativa da frequência fundamental em cada bloco de dados, foi utilizada a seguinte estratégia:

- 1) Transformada de Fourier (através da função `fft()` do Matlab).
- 2) `findPeaks`, para encontrar os principais picos.
- 3) Estes picos eram ordenados em ordem decrescente
- 4) Calculada a moda dos 20 maiores picos para estimar a frequência fundamental

Para conversão da energia em intensidade sonora (velocity), foram realizados alguns testes e foi definido a proporção $\text{velocity} = 50 + \text{energia do sinal}$.

Foi implementado um algoritmo no Matlab que fazia as devidas conversões e salvava um arquivo .mid (extensão para arquivos MIDIs) utilizando uma toolbox de conversão adquirida em <https://github.com/miditoolbox/1.1> [4].

O resultado é a trilha em verde abaixo. As trilhas original e MIDI foram abertas utilizando o programa GarageBand, que possui também um módulo digital com alguns sons que podem ser executados por MIDI (inclusive oboé). O formato da trilha MIDI é parecido com o de um espectrograma do sinal. Este formato é popularmente conhecido no meio da música como piano roll.

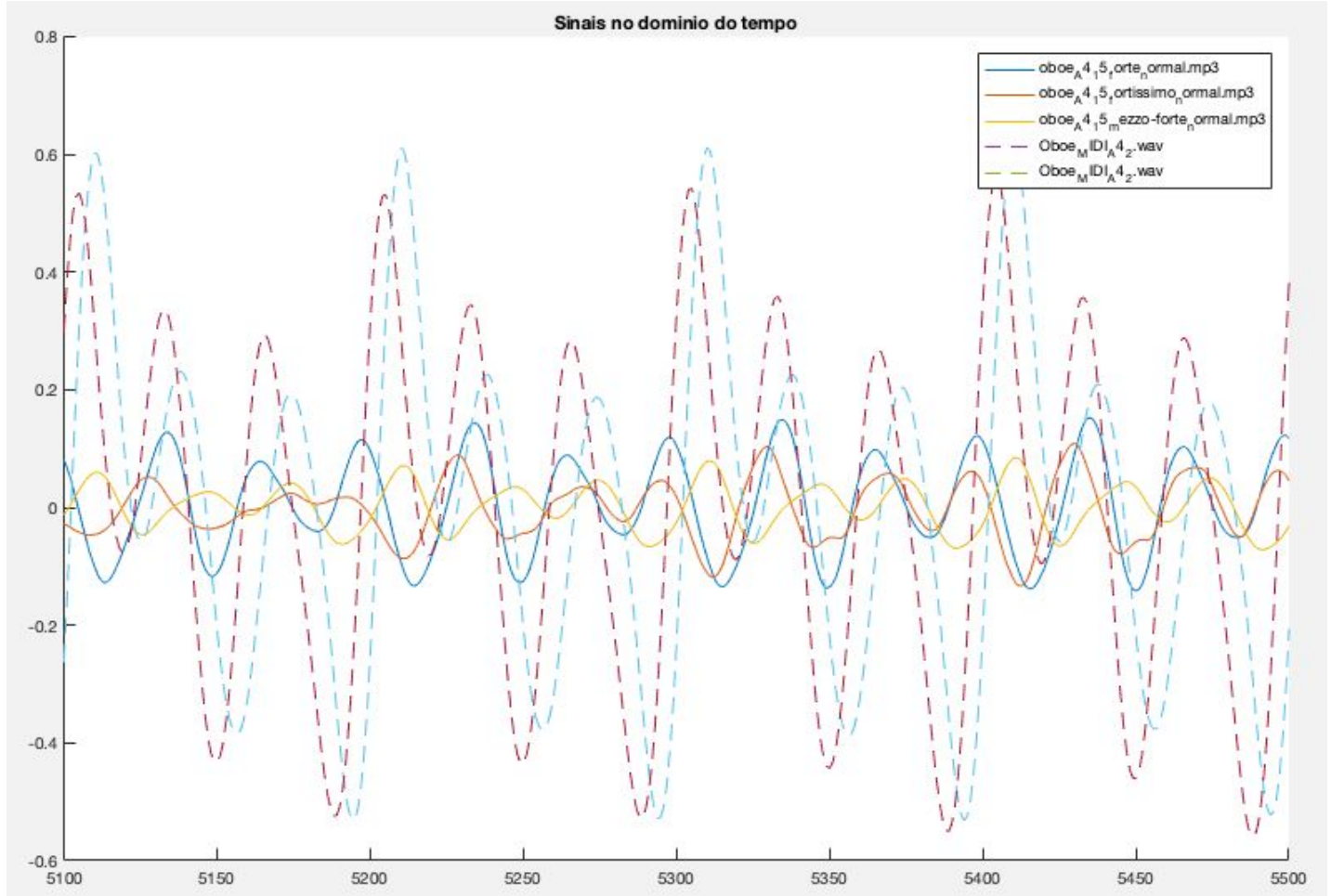


Os resultados da conversão apresentaram boas aproximações tonais para janelas de 0.2s. O solo conseguiu ser acompanhado inclusive em momentos de uma maior velocidade de transição entre notas. Porém, este janelamento fixo pode pegar o exato momento onde ocorrem transições de nota, o que prejudica a estimativa da frequência fundamental. Uma melhoria seria identificar regiões de transições de notas e intensidade e criar janelas baseadas nisso.

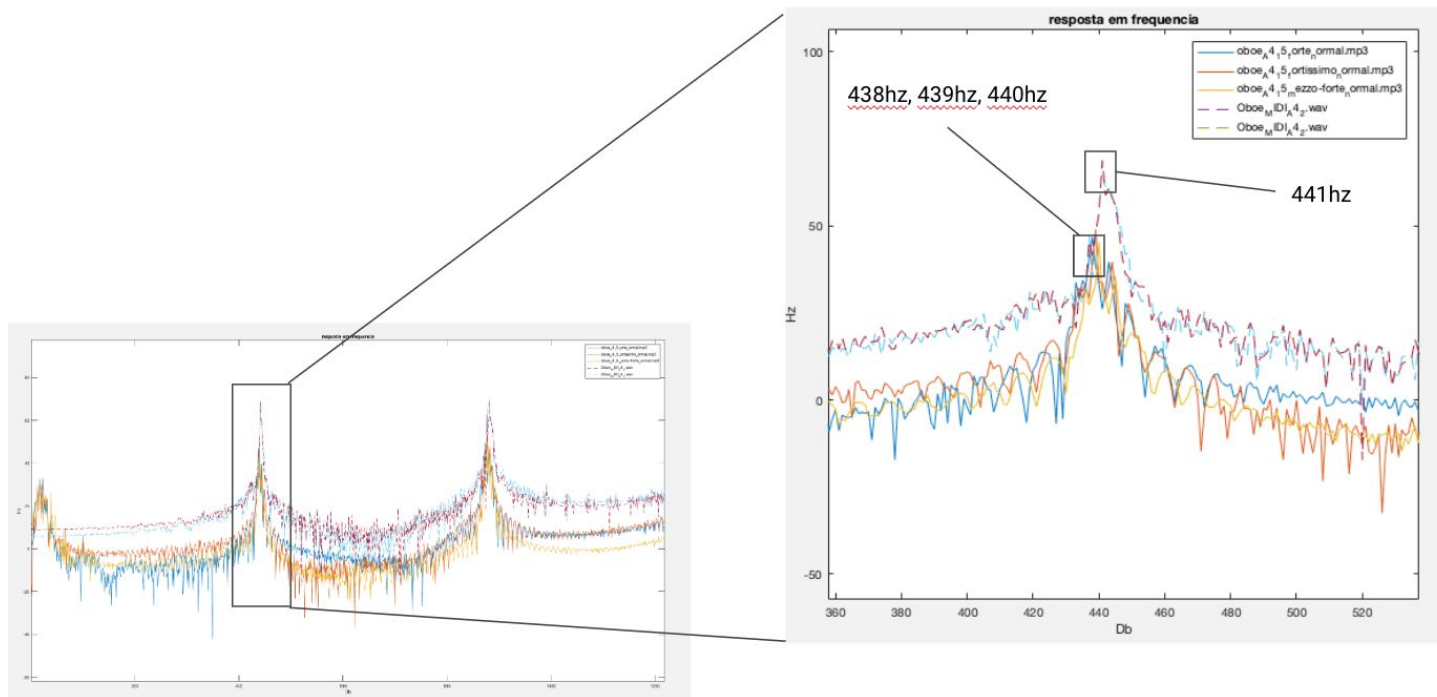
IV. Comparação entre som original e sinal sintetizado

Para esta parte, foram utilizadas três gravações de oboés (todos em A4) e dois sinais sintetizados de um oboé (através de envios de sinais MIDI para um gerador de som)

Fazendo comparações no domínio do tempo, podemos observar que os sinais sintetizados possuem um comportamento exatamente periódico e previsível, enquanto as gravações representam ruídos, sinais de não serem exatamente periódicas e uma certa defasagem entre elas.



No domínio da frequência podemos observar que as duas notas sintetizadas apresenta a exata mesma frequência fundamental, enquanto as três gravações dos oboés variam entre 438hz, 439hz e 440hz. Tal variação não é perceptível para a nota La(A).



Portanto, como era esperado, as informações extraídas das ondas sonoras de sons sintetizados são bem mais coincidentes entre si do que de sons reais. Apesar destas características, a sonoridade do oboé sintetizado apresenta uma boa semelhança com as gravações originais.

V. Conclusão

Através deste trabalho foi possível obter um maior entendimento sobre o protocolo MIDI e sua importância para o avanço da tecnologia da síntese sonora na música. Foi possível também estudar e realizar uma implementação de um conversor de ondas sonoras para MIDI que, apesar de utilizar estratégias “ingênuas” conseguiu uma boa aproximação do som original, como foi demonstrado na apresentação prática. Ao final, foi possível também elencar algumas características que diferenciam as ondas sintetizadas das ondas reais, comparando assim o som gerado por uma informação MIDI com o som de um instrumento real.

VI. Referências

- [1] Manning, Peter. Electronic and Computer Music. 1985. Oxford: Oxford University Press, 1994. Print.
- [2] Huber, David Miles. "The MIDI Manual". Carmel, Indiana: SAMS, 1991.
- [3] Hass, Jeffrey. "Chapter Three: How MIDI works 9". Indiana University Jacobs School of Music. 2010. Web. 13 de agosto de 2012.
- [4] Toiviainen, P., & Eerola, T. (2016). MIDI Toolbox 1.1. URL: <https://github.com/miditoolbox/1.1>