

MAC0300 - Métodos Numéricos da Álgebra Linear

Professor Walter Mascarenhas

1º EP - Transcrição Automática de Música Monofônica

Data de Entrega: 11/09/2012

1 Especificações

1.1 Objetivo

O objetivo deste EP é, partindo da representação digital de uma onda sonora monofônica (um arquivo formato WAV com taxa de amostragem 44100 Hz), analisar suas características e transformá-la em uma representação simbólica, um arquivo formato MIDI. Os arquivos WAV serão apenas sons monofônicos simples. Exemplos serão disponibilizados no site.

De uma forma geral, para realizar essa tarefa, serão necessários os seguintes passos:

- decodificar as informações contidas no arquivo WAV;
- aplicar a Transformada de Fourier para cada segmento da onda, com intervalos de tempo determinados;
- analisar o espectro de frequências e obter a frequência fundamental do intervalo;
- determinar as notas e sobretons correspondentes a cada frequência fundamental; e
- transcrever a música para um arquivo MIDI, criando eventos MIDI correspondentes a cada intervalo analisado.

A detecção da frequência fundamental deverá ser feita por análise espectral do sinal do áudio com os seguintes métodos de Transformada Discreta de Fourier (DFT): Multiplicação pela Matriz de Fourier e FFT (Fast Fourier Transform). Deve-se fazer comparações entre os tempos de execução dos métodos e apresentar gráficos dos resultados obtidos.

1.1.1 Sistema Operacional

O programa deve compilar e funcionar no Linux, devendo ser executado pela linha de comando.

1.1.2 Linguagem de Programação

A implementação do programa pode ser realizada na linguagem a sua escolha, contanto que sejam usados apenas recursos básicos da linguagem para a implementação da Transformada Discreta de Fourier pelo método de Multiplicação pela Matriz de Fourier.

Deve-se usar um algoritmo de FFT (Fast Fourier Transform) já disponível pela linguagem ou por alguma biblioteca a sua escolha. As linguagens recomendadas são Octave, R e Python.

1.1.3 Arquivo de Entrada

O programa deve receber, pela entrada padrão, um arquivo de áudio em formato WAV.

O nome do arquivo será da forma “nome.WAV”. Exemplo: lullaby.WAV

Por exemplo, a execução de um programa em Octave, por um terminal, seria:

```
$ octave EP1.m lullaby.WAV
```

1.1.4 Arquivo de Saída

O programa deve devolver, na pasta de execução do programa, um arquivo em formato MIDI, de nome igual ao arquivo de entrada, só que com extensão MIDI. Exemplo: lullaby.MIDI

1.1.5 Algoritmos

A análise espectral do som, para detecção da frequência fundamental e de suas harmônicas, deve ser feita, para efeito de comparação, por dois algoritmos de DFT: Multiplicação pela Matriz de Fourier e FFT (Fast Fourier Transform).

Implemente o método de Multiplicação pela Matriz de Fourier. Veja seção sobre o algoritmo na parte de Leitura Complementar.

Para a FFT, use o algoritmo já disponibilizado pela sua linguagem ou use uma biblioteca. Por exemplo, em Octave, $y = fft(x)$.

Dica: após implementar o método da multiplicação, teste se a matriz de DFT, aplicada a um sinal qualquer x , ou seja, $y = Wx$, resulta no mesmo sinal dado pela função de FFT (Fast Fourier Transform) já disponibilizada.

Além disso, devem ser mostrados os gráficos da onda no domínio tempo e no domínio da frequência e devem ser coletados os tempos de execução para uma posterior análise de complexidade e comparação dos dois algoritmos.

1.1.6 Manual

Junto com o arquivo que contém seu código-fonte, submeta uma breve descrição de como compilar e usar o seu programa, inclusive com as especificações já ditas anteriormente.

1.1.7 Relatório

Elabore um relatório contendo os seguintes itens:

- Explique sucintamente a modelagem de um som usando Série de Fourier;
- Explique como a Transformada de Fourier fornece informações para a transcrição musical feita em seu programa;
- Faça a comparação do tempo de execução dos dois métodos de cálculo da Transformada Discreta de Fourier: Multiplicação pela Matriz de Fourier e FFT. Mostre dois gráficos comparativos com

os tempos médios dos métodos para o mesmo número de execuções. Tome o cuidado de não incluir o tempo de criação da matriz DFT no método da multiplicação.

- Descreva vantagens e desvantagens entre os métodos Multiplicação pela Matriz de Fourier e FFT.
- Mostre os gráficos das ondas de áudio no domínio do tempo e no domínio da frequência (frequência no eixo das abscissas e amplitude em decibéis no eixo das ordenadas). Mostre os espectros obtidos para ambos os métodos.

1.2 Dicas

- Use o Google! ☺
- É permitido copiar da Internet código para as tarefas auxiliares, como ler e decodificar WAV e escrever arquivos MIDI, pois o foco do EP está nos algoritmos.

2 Leitura Complementar

2.1 Transcrição Automática de Música Monofônica

A transcrição automática de música é uma das áreas mais ativas em computação musical. Ela é também conhecida como reconhecimento de música e consiste na análise matemática de um sinal de áudio, normalmente em formato WAV, e na sua conversão em notação musical, normalmente em formato MIDI. Os sistemas de transcrição musical possuem aplicações como extração de melodia, pesquisa em banco de dados musical e sistemas musicais interativos.

Os formatos WAV e MIDI diferem na representação do som. O formato WAV registra digitalmente qualquer som, incluindo fala, enquanto o formato MIDI é principalmente uma sequência de notas (ou eventos MIDI).

A transcrição de WAV para MIDI é uma pequena parcela da transcrição automática musical, que é um problema de inteligência artificial ainda não resolvido. Apenas como comparação, o problema de transcrição de texto (OCR - Optical Character Recognition) está resolvido com acurácia de 95%, os programas de reconhecimento de fala funcionam com acurácia de 70-80% e os sistemas de reconhecimento de músicas monofônicas funcionam com acurácia de 60-70%.

As músicas monofônicas são aquelas tocadas por um único instrumento, uma nota por vez, como em um solo de trombeta. Elas são bem menos comuns que as músicas polifônicas, que são tocadas por um ou mais instrumentos, várias notas de uma vez, como em uma orquestra. No entanto, ainda não conseguimos uma boa acurácia nos sistemas de transcrição de músicas polifônicas.

Para criar um arquivo MIDI a partir de um arquivo em formato WAV, um músico deve determinar a frequência fundamental, a velocidade e a duração de cada nota e salvar esses parâmetros em uma sequência de eventos MIDI. Um programa de transcrição automática de música deve fazer a mesma coisa. Mesmo para sons tocados por um único instrumento, essa não é uma tarefa fácil, pois o sinal gravado no arquivo não contém dados específicos da música. A variedade de timbres, construções harmônicas e transições dificulta a criação de um algoritmo matemático que reconstrua precisamente a música. É por esse motivo que nos restringiremos a reconhecer sons monofônicos simples.

2.2 Conceitos Preliminares

O som é uma onda que se propaga em um meio material, como o ar, com uma certa frequência. O ouvido humano é capaz de perceber sons em frequências de 20 a 20.000 Hz. As ondas com frequência baixa (entre 20 e 100 Hz) soam de forma grave e sons com frequência elevada (acima de 400 Hz) soam de forma aguda.

Uma música é uma composição de sons com frequências estáveis ao longo do tempo. Um som pode ser visto como uma série harmônica composta pela frequência fundamental e seus sobretons. A frequência fundamental é a frequência mais baixa da série harmônica, sendo considerada a frequência da nota musical. As frequências mais altas são chamadas harmônicas ou sobretons, que são múltiplos inteiros da frequência fundamental.

A correspondência entre nota e frequência depende da escala musical usada.

Na música ocidental é comum ver a escala de 7 notas (escala diatônica), que pode estar em modo maior ou menor. Na escala de Dó maior, por exemplo, as notas correspondentes às alturas são: Dó, Ré, Mi, Fá, Sol, Lá e Si.

Atualmente, costuma-se adotar a escala de 12 tons, que divide a oitava em 12 partes iguais a largura de um semitom. Nesta escala, todas as notas derivam da décima segunda raiz de 2 (aproximadamente 1,0595). Ela é usada para a criação de teclados musicais virtuais, onde a frequência da n -ésima tecla é dada pela fórmula:

$$f(n) = (\sqrt[12]{2})^{n-49} \times 440 \text{ Hz}$$

A tabela de correspondência em notação científica pode ser na página da Wikipedia: [Scientific pitch notation](#)

2.3 Análise Espectral de uma Onda Sonora

Os conceitos de frequência fundamental e sobretons são descritos nos coeficientes de Fourier e os sons produzidos são modelados pela Série de Fourier. A Transformada de Fourier produz o espectro de um sinal sonoro, que permite a análise no domínio da frequência, não do tempo. Quando trata-se de um sinal quantizado (discreto), usa-se a Transformada Discreta de Fourier (DFT).

Como exemplo, se tocássemos em um sistema de som de computador um segmento de uma senoide do tipo $100 \sin(2\pi vt)$, com frequência $v = 400$ ciclos/seg (Hz), um tom conhecido como Lá de diapasão seria produzido. O tom de frequência fundamental de 440 Hz corresponde à nota A4 em notação científica. Um tom puro tem uma única frequência fundamental, logo está associado a uma única frequência, no caso 440 Hz. Calculando o espectro de Fourier desta senoide, um único pico de 440 Hz ficará evidente. Veja figura 1.

Os sons de instrumentos musicais são, ao longo do tempo, superposições de vários tons puros, que podem ser interpretadas matematicamente como soma de ondas senoidais.

Por exemplo, o espectro de Fourier de uma nota de um piano contém picos localizados, aproximadamente, nas frequências 330 Hz, 660 Hz, 990 Hz, 1320 Hz e 1620 Hz. Essas frequências são todas múltiplas inteiras da menor frequência, 330 Hz. Esta frequência base, 330 Hz, é a que chamamos de fundamental. Neste caso, ela é aproximadamente igual à frequência padrão de 329.628 Hz para a nota E4. Os múltiplos inteiros desta frequência fundamental são chamadas sobretons. Veja figura 2.

O método ingênuo de implementação da Transformada Discreta de Fourier (DFT) é o que usa multiplicação pela Matriz de Fourier. Um método mais eficiente, muito usado na prática em Processamento de Sinais, é a Transformada Rápida de Fourier, conhecida como FFT (Fast Fourier Transform).

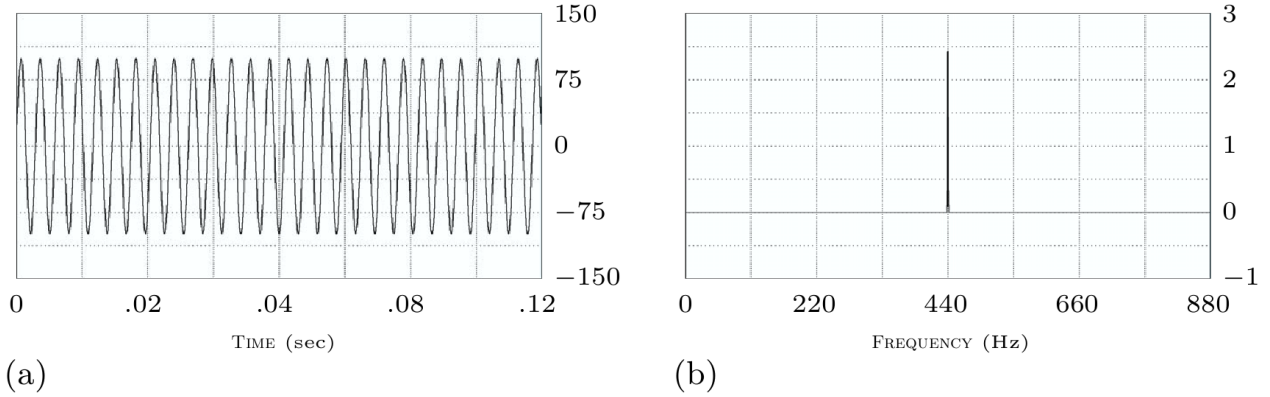


Fig. 1: Análise de Fourier de um tom puro. (a) Gráfico de um segmento finito de um tom puro de 440 Hz. (b) Espectro de Fourier calculado computacionalmente.

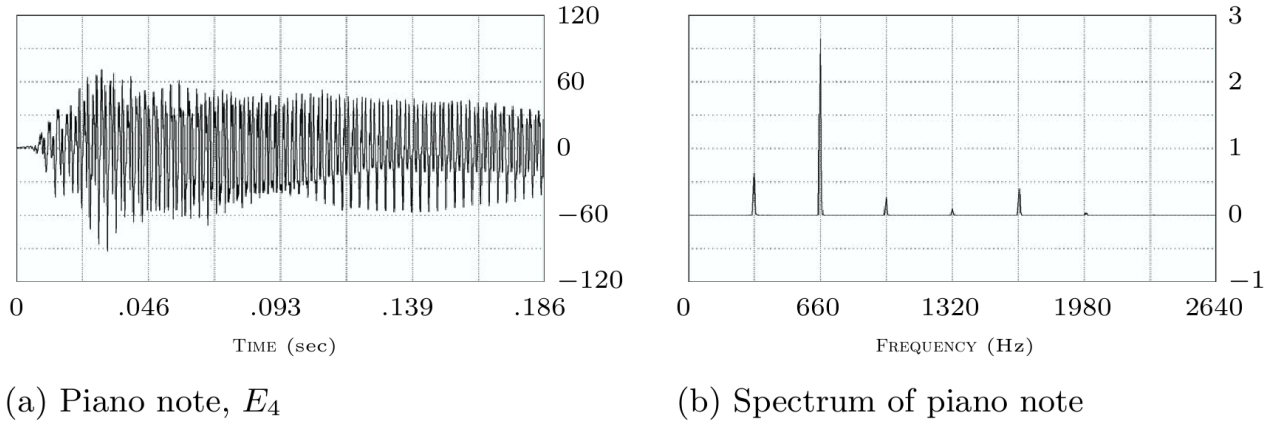


Fig. 2: Análise de Fourier da nota E_4 do piano (E acima do C médio). (a) Nota E_4 do piano. (b) Espectro da nota do piano.

2.3.1 DFT por Multiplicação pela matriz de Fourier

O método consiste em obter a representação no domínio das frequências de uma função no domínio do tempo. Essa é uma transformação linear, logo é possível obtê-la através de uma multiplicação matricial. Essa transformação W , de tamanho $N \times N$ pode ser definida como:

$$W = \left(\frac{\omega^{jk}}{\sqrt{N}} \right)_{j,k=0,\dots,N-1}, \text{ onde } \omega = e^{-\frac{2\pi i}{N}}$$

Assim, a DFT X de um sinal x é dada por: $X = Wx$.

Repare que as entradas dessa matriz são complexas. Para realizar a multiplicação, decompõe a matriz W como soma de duas matrizes, uma com a parte real e a outra com a parte imaginária. Assim, para realizar $X = Wx$, faça $X = Ux + i(Vx)$, onde U contém as componentes reais e V contém as componentes da parte imaginária.

Cada posição j do vetor X representará uma das senoides da série harmônica, assim, diz quanto do sinal tem uma frequência fracionária de $\frac{j-1}{N}$. Por exemplo, a primeira posição mede a componente

DC do sinal, a segunda posição mede quanto do sinal tem frequência fracionária $\frac{1}{8}$, e assim por diante.

Para o cálculo da amplitude de cada onda basta calcular $A = \sqrt{u^2 + v^2}$, onde u denota a parte real e v a parte imaginária.

Veja, para uma referência mais detalhada do algoritmo, a página da Wikipedia: [DFT Matrix](#)

2.4 Conversão de WAV para MIDI

O formato WAV – Waveform Audio File Format – é o padrão para arquivos de áudio em computadores. Um arquivo em formato WAV contém áudio em formato de modulação de pulsos (PCM), que, diferentemente do mp3, não contém compressão. PCM é o método padrão para representar digitalmente (quantizar) amostras dos sinais analógicos. Quando quantizado, o sinal do áudio deixa de ser analógico (contínuo) e passa a ser digital (discreto), possibilitando seu processamento digital. A taxa de amostragem das gravações é normalmente 44100 Hz.

O formato MIDI – Musical Instrument Digital Interface – contém uma sequência de comandos que controlam programas ou hardwares musicais, como sintetizadores ou sequenciadores. Esses comandos não são sons, são instruções que, em geral, geram um som. Por exemplo: selecione instrumento #1 (Acoustic Grand Piano), toque a nota #60 (C5) com velocidade #127. Dessa forma, não é possível representar, por exemplo, uma fala humana em um formato MIDI, mas é possível editar uma nota ou trocar o instrumento em uma música neste formato.

Resumidamente, o processo de conversão de WAV para MIDI consiste nos seguintes passos:

- Leitura do sinal quantizado em um arquivo de formato WAV (Decodificação);
- Análise espectral do sinal usando Transformada de Fourier para obtenção da frequência fundamental e das harmônicas;
- Determinação das notas e tempos;
- Transcrição das notas criando eventos MIDI.

3 Fontes e Referências

- [Time-Frequency Analysis of Musical Instruments](#), Jeremy F. Alm; James S. Walker
- [Notação Padrão para as notas musicais](#)
- [DFT](#)
- [FFT](#)
- [Matriz de DFT](#)
- [Monografia de Adriano Brito Mitre \(BCC 2006\)](#)