

Computação Musical – 1º Exercício Programa

Prof. Marcelo Queiroz

Data de entrega: **29/5/2009**

Instruções: Os EPs podem ser feitos em grupos de até 3 pessoas. A entrega será feita pelo PACA até 23h55 do dia 29/5/2009.

Osciloscópio Digital

Neste EP vamos desenvolver um analisador de sinais de áudio semelhante a um osciloscópio analógico daqueles que a gente vê em laboratórios de Física. A entrada será obtida de um microfone. A saída do programa será atualizada constantemente e consiste de 3 displays: uma janela mostrando a forma de onda, outra o espectro simplificado e a saída padrão textual que sinalizará os eventos musicais (notas) detectados.

Aspectos técnicos de implementação são detalhados nas seções a seguir.

Leitura do Microfone

No PureData podemos ler informações do microfone usando o objeto `adc~`. Este produzirá um fluxo regular de dados, normalmente à taxa de $R=44100\text{Hz}$ (isso é configurado na comunicação do Pd com o servidor de som), com valores em ponto flutuante entre -1 e +1, que alimentarão as demais etapas do processamento: cálculo do espectro, análise do espectro para obtenção da estimativa de frequência fundamental, e formatação dos dados para representação da forma de onda e do espectro harmônico.

Cálculo do Espectro

Para o cálculo do espectro utilizaremos o objeto `rfft~` aplicado a janelas de $N=4096$ amostras, que correspondem a $\frac{N}{R} = \frac{4096}{44100} \approx 0.093$ segundos. O tamanho da janela é definido por um objeto `block~`, e a resolução do espectro correspondente será de $\frac{R}{N} \approx 10.767\text{Hz}$. Esta será também a taxa de atualização dos displays do osciloscópio, pois a cada espectro calculado serão obtidas as informações para preenchimento das respectivas saídas.

Em cada janela de análise os valores das amostras devem ser multiplicados por uma função de Hamming antes do cálculo da FFT. Lembre que a função de Hamming pode ser calculada numa janela indexada de 0 até $N-1$ pela expressão

$$0.54 - 0.46 \cos\left(2\pi \frac{n}{N-1}\right)$$

onde n é o índice da amostra na janela.

Análise do Espectro

Após o cálculo do espectro, consideraremos principalmente (mas não exclusivamente) os valores absolutos $|F_0|, \dots, |F_{\frac{N}{2}}|$ que correspondem ao espectro de magnitude do fragmento analisado (estes são obtidos como $|F_k| = \sqrt{a_k^2 + b_k^2}$ onde a_k e b_k são as partes real e imaginária de F_k).

Observe que os valores $(F_0, F_1, F_2, \dots, F_{\frac{N}{2}})$ possuem a informação espectral referente às frequências 0Hz (constante d.c.), 10.767Hz, 21.533Hz, \dots , $\frac{N}{2} \frac{R}{N} = 22050\text{Hz}$; os índices maiores que $\frac{N}{2}$ repetiriam essa informação na forma de valores conjugados na ordem inversa (porém o objeto `rfft~` não calcula estes valores).

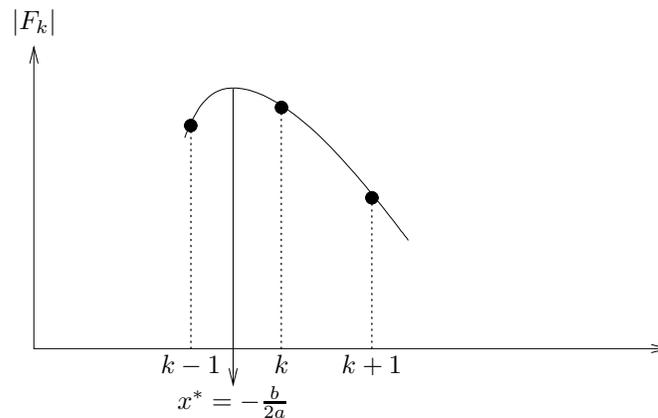
Para extrair as informações pretendidas do espectro temos que ser capazes de discernir entre espectros ruidosos e espectros “musicais”. Neste contexto, consideraremos “musicais” os espectros que possuem um padrão de harmonicidade, ou seja, que possuem a energia concentrada em picos associados a frequências que são múltiplos inteiros de uma certa frequência fundamental f_0 .

A identificação da frequência fundamental f_0 para cada janela analisada é um problema instigante devido à ocorrência, tanto em sons naturais quanto artificiais, de fenômenos tais como harmônicos ausentes, maior pico do espectro associado a um harmônico de ordem superior a 1, ligeira inarmonicidade dos parciais, etc. Faz-se necessário a adoção de alguma heurística que, esperamos, será capaz de encontrar a frequência certa na maioria dos casos considerados.

Considere a frequência de pico f_p que maximiza o valor de $|F_k|$ para $k = 3, \dots, \frac{N}{2}$ (isto é, desconsideraremos os valores de k para os quais a frequência correspondente é inferior a 22Hz). A fim de aumentar a precisão da análise, consideraremos as duas amostras vizinhas à amostra que maximiza $|F_k|$ e estimaremos a frequência de pico fazendo uma interpolação quadrática dos valores do espectro. O polinômio $p(x) = ax^2 + bx + c$ que satisfaz $p(k-1) = |F_{k-1}|$, $p(k) = |F_k|$ e $p(k+1) = |F_{k+1}|$ tem os coeficientes a , b e c dados por

$$\begin{aligned} a &= \frac{|F_{k+1}| - 2|F_k| + |F_{k-1}|}{2} \\ b &= |F_{k+1}| - |F_k| - a(2k+1) \\ c &= |F_k| - ak^2 - bk \end{aligned}$$

e o ponto x^* que maximiza $p(x)$ é $x^* = -\frac{b}{2a}$, correspondendo à frequência $f_p = x^* \frac{R}{N}$ Hz:



Utilizaremos esta frequência de pico f_p para estimar a frequência fundamental f_0 tentando averiguar se cada uma das frequências $f_p, \frac{f_p}{2}, \frac{f_p}{3}, \dots, \frac{f_p}{HMAX}$ é ou não forte candidato a ser fundamental (isso corresponde à hipótese de que f_p é um dos HMAX primeiros harmônicos do sinal considerado). Os candidatos a frequência fundamental serão comparados levando em consideração sua “energia harmônica total” (soma dos valores do espectro na série harmônica do candidato considerado) bem como sua “energia fundamental” (relação da energia do primeiro harmônico em relação à energia harmônica total).

Observe porém que o espectro todo está disposto sobre uma série harmônica com fundamental em 10.767Hz; isso não tem relação nenhuma com o sinal de entrada mas sim é uma consequência da escolha dos parâmetros da análise. Frequentemente encontramos em espectros ruidosos uma concentração casual de energia sobre os índices pares, ou múltiplos de 3, o que indica ser interessante excluir de consideração frequências muito baixas (pois a soma EHT delas incluirá muitos índices). Considerando que a entrada será utilizada por uma voz humana podemos tranquilamente excluir de consideração candidatos a fundamental abaixo de FRQMIN=65Hz.

Mais especificamente, para o candidato $f_j = \frac{f_p}{j}$ podemos definir

$$\text{EHT}_j = \sum_{i=1}^{if_j \leq \frac{R}{2}} |F(if_j)| \quad \text{e} \quad \text{EF}_j = \frac{|F(f_j)|}{\text{EHT}_j},$$

e definiremos $f_0 = f_{j^*}$ onde o índice j^* é aquele que satisfaz

$$\text{EHT}_{j^*} = \max\{\text{EHT}_j \mid f_j > \text{FRQMIN} \text{ e } \text{EF}_j > \text{EFMIN}\}.$$

Estas fórmulas envolvem o cálculo de $|F(\omega)|$ para valores de ω que não estão necessariamente na série harmônica de 10.767Hz. Para obtê-los você pode calcular os índices fracionários correspondentes e usar um objeto de leitura interpolada como o `tabread4~`.

De posse de uma frequência fundamental f_0 poderemos testar se o espectro é harmônico ou ruidoso medindo a relação entre o EHT daquela frequência e a energia total do espectro $\text{ET} = \sum_{i=0}^{\frac{N}{2}} |F_i|$. Espectros que satisfaçam $\frac{\text{EHT}}{\text{ET}} < \text{EHTMIN}$ serão considerados ruidosos, os outros definirão eventos musicais.

Formato da Saída

Para cada espectro produzido devemos atualizar a informação sobre a forma de onda, o espectro harmônico e detecção de eventos musicais.

Forma de Onda

A janela contendo a forma de onda mostra sempre uma lista de valores entre -1 e +1; ela deve se atualizada da seguinte maneira: se o espectro é ruidoso o gráfico mostrará os 4096 valores lidos da entrada, do contrário o gráfico mostrará exatamente 3 períodos completos da forma de onda. Cada evento musical possui uma frequência fundamental f_0 definida, e o número de amostras que corresponde a 3 períodos completos é $K = 3\frac{R}{f_0}$.

Considerando que um evento musical normalmente dura mais do que 0.093 segundos, devemos garantir que de uma janela de análise para a seguinte nós atualizamos a “forma gráfica” da onda sem deixar com que a fase inicial da janela de análise atrapalhe a visualização, fazendo com que a onda pareça estar “andando” aleatoriamente para a frente ou para trás; em outras palavras, queremos estabilizar a forma de onda. Uma maneira de estabelecer um ponto inicial preciso para a forma de onda é tomar os K pontos a partir de uma amostra k_0 que corresponda ao ponto onde o primeiro harmônico está em “fase de seno”. Lembre que este harmônico é dado pela expressão $H_1 = 2\alpha \cos(2\pi f_0 t + \varphi)$ onde α e φ correspondem à magnitude e à fase da informação espectral (do espectro original, complexo) associadas à frequência fundamental f_0 ; este harmônico está em fase de seno quando $2\pi f_0 t_0 + \varphi = \frac{3\pi}{2}$ ou equivalentemente quando $t_0 = \frac{\frac{3\pi}{2} - \varphi}{2\pi f_0}$, de onde $k_0 = R t_0$. O valor correto de φ pode ser obtido pela leitura interpolada dos valores a_k e b_k do espectro no índice fracionário k que corresponde à frequência fundamental f_0 , e lembrando que a fase de F_k pode ser calculada com o objeto `atan2(a_k, b_k)`.

Espectro Harmônico

A janela contendo o espectro deve mostrar valores entre 0 e 1, que representam os extremos mínimo e máximo do valor absoluto de qualquer amostra do espectro. Lembrando da expressão da DFT temos que $|F_n| = |\sum_{k=0}^{N-1} f_k e^{-i2\pi n \frac{k}{N}}| \leq \sum_{k=0}^{N-1} |f_k e^{-i2\pi n \frac{k}{N}}| = \sum_{k=0}^{N-1} |f_k| \leq N \cdot \max\{|f_k|\}$, de onde concluímos que, no nosso caso, $|F_n| \leq N$, $\forall n$. Na prática os valores são bem menores que esse, então consideraremos $\text{MAGNMAX} = \frac{N}{4}$ e apresentaremos todos os valores relativos a este máximo.

Se o espectro em consideração é ruidoso, devemos mostrar simplesmente o gráfico do espectro normalizado. Se o espectro possui um padrão de harmonicidade, representaremos o espectro harmônico do sinal, que mostra apenas a energia da série harmônica a partir da fundamental f_0 , ou seja, os valores $|F(f_0)|, |F(2f_0)|, |F(3f_0)|, \dots$, relativos a MAGNMAX, na forma de um histograma.

Eventos Musicais

Em relação aos eventos musicais, considere a seguinte representação numérica para as notas: podemos obter o número d de semitons acima ou abaixo da frequência de referência 440Hz a partir da expressão $440 \cdot 2^{\frac{d}{12}} \approx f_0$ cuja solução inteira é $d = \text{round}(12 \log_2 \frac{f_0}{440})$. O início de uma nota deve ser notificado na janela do Pd (com objeto print), através deste valor de d e sua frequência em Hz, juntamente com o instante de início (relativo ao início da execução do programa, medido em segundos); seu fim deve ser notificado com o instante da detecção do término da nota.

Observe que a frequência fundamental f_0 de janelas sucessivas correspondentes a uma mesma nota poderá flutuar ligeiramente, mas a nota será considerada única se os valores arredondados de d coincidirem. Uma desvantagem desta heurística é que se a frequência fundamental estiver bem no meio do caminho entre duas “notas” inteiras então qualquer flutuação será detectada como uma mudança de nota: por exemplo se f_0 flutua em torno de $452.89\text{Hz} = 440 \cdot 2^{\frac{0.5}{12}}$ então d ficará oscilando aleatoriamente entre 0 e 1. Esta instabilidade pode ser corrigida considerando-se que uma mudança de nota depende não apenas da mudança do valor de d mas também de uma distância considerável entre as frequências fundamentais sucessivas, por exemplo pela condição $\left| \frac{f_{0\text{atual}} - f_{0\text{anterior}}}{f_{0\text{anterior}}} \right| > 0.03$ (0.03 é metade de um semitom).

Considerações Finais

Se você está se perguntando quais são os valores de HMAX, EFMIN e EHTMIN, parabéns! Isso indica que você leu o enunciado com atenção... ;-)

Você pode utilizar HMAX=10, pois com exceção de sons artificiais o pico de energia sempre está nos primeiros harmônicos. Estimar empiricamente os limiares EFMIN e EHTMIN é parte do trabalho deste EP: use a garganta!

Você pode testar o seu código em entradas diferentes do microfone através do objeto readsf~. Este objeto varre o arquivo de áudio produzindo um fluxo de sinais de áudio análogo ao do objeto adc~.

Além do help dos objetos mencionados neste enunciado, outros objetos úteis para a implementação das fórmulas serão: timer, metro, trigger, delay (para controles temporais) e select, moses e until (para condições e laços). Veja também a listagem em <http://www.flexatone.net/docs/pdg.pdf>.

Alguns exemplos da documentação oficial do Pd também poderão ajudar bastante. No menu “Help” escolha a opção “Browser...”, selecione o diretório “2.control.examples” e olhe os patches 05, 06, 14, 15 e 18. Depois disso, abra “3.audio.examples” e olhe os patches E01, I01 e I02.

Bom Trabalho!