

Computação Musical – 1º Trabalho Maior

Prof. Marcelo Queiroz

Data de entrega: 8/5/2013

Instruções: Os EPs devem ser feitos individualmente. A entrega será feita pelo PACA até 23h55 do dia 8/5/2013.

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 do microfone ou de um arquivo. A saída do programa será atualizada constantemente e consiste de 4 partes: uma janela mostrando a forma de onda, outra o espectro de magnitude, uma saída numérica para as frequências detectadas e uma saída textual que sinalizará eventos musicais.

Aspectos técnicos de implementação são detalhados nas seções a seguir.

Leitura da Entrada

No PureData podemos ler informações do microfone usando o objeto [adc~] e de um arquivo de áudio usando [readsf~] (ou [soundfiler] e [tabplay~]). Estes objetos produzirão um fluxo regular de dados com $R = 44100$ amostras por segundo (taxa de amostragem usual do Pd, não precisa ser alterada). Este fluxo alimentará 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 amostras, parâmetro este configurável pelo usuário. O tamanho da janela é definido por um objeto [block~]¹, e a resolução do espectro correspondente será de $\frac{R}{N}$ 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. Um objeto útil para sincronizar processamentos de controle ao ciclo de blocos do DSP é o [bang~], que sinaliza o fim de cada bloco de processamento.

Em cada janela de análise o sinal de entrada deve ser multiplicado por uma função de Hamming antes do cálculo da FFT. A função de Hamming, que é definida numa janela indexada de 0 até $N-1$ pela expressão $H[n] = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N-1}\right)$, pode ser obtida no Pd com o objeto [hamming~].

Análise do Espectro

Após o cálculo do espectro, consideraremos os valores absolutos $|F_0|, \dots, |F_{\frac{N}{2}}|$ (espectro de magnitude) e os ângulos $\angle F_0, \dots, \angle F_{\frac{N}{2}}$ (espectro de fase) correspondentes ao fragmento analisado. Estes são obtidos como $|F_k| = \sqrt{a_k^2 + b_k^2}$ e $\angle F_k = \text{atan2}(b_k, a_k)$, onde a_k e b_k são as partes real e imaginária de F_k produzidas pelo objeto [rfft~].

¹Observação: os objetos [adc~] e [block~] não podem estar no mesmo patch, o que obriga a análise a ser realizada em um subpatch.

²Observe a ordem dos argumentos do objeto [atan2~].

Observe que os valores $(F_0, F_1, F_2, \dots, F_{\frac{N}{2}})$ possuem a informação espectral referente às frequências 0 Hz (constante d.c.), $\frac{R}{N}$ Hz, $2\frac{R}{N}$ Hz, \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~` nem chega a calcular estes valores).

A identificação da frequência fundamental f_0 para cada janela analisada de um som que possui altura musical bem definida é um problema instigante devido à ocorrência, tanto em sons naturais quanto artificiais, de fenômenos tais como harmônicos ausentes ou harmônicos mais fortes do que a fundamental, ligeira inarmonicidade dos parciais, etc. Para simplificar o nosso trabalho usaremos uma heurística simples que consiste em encontrar o maior pico do espectro de magnitude usando interpolação quadrática dos bins adjacentes ao maior valor de $|F_k|$. Este subproblema foi resolvido no segundo trabalho menor, e vocês devem aproveitar aquela solução neste trabalho. Esta solução será capaz de encontrar a frequência certa na maioria dos casos simples (mas você é convidado a explorar também situações onde esta heurística produz estimadores incorretos, para entender melhor o problema de rastreamento de pitch).

Formato da Saída

Para cada espectro produzido devemos atualizar a forma de onda estabilizada e o espectro de magnitude normalizado, produzir informação sobre a frequência fundamental e relatar na saída textual do Pd informações sobre eventos musicais.

Forma de Onda Estabilizada

A janela contendo a forma de onda deverá mostrar exatamente 3 períodos completos da forma de onda (ou seja, $K = 3\frac{R}{f_0}$ amostras). Além disso, 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 instante em que 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 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 (usando `[tabread4]`) do espectro de fase no índice fracionário $f_0 \frac{N}{R}$, correspondente à frequência fundamental f_0 obtida pela interpolação quadrática.

Para melhorar ainda mais a visualização da saída, utilizaremos como ponto inicial da visualização não exatamente o índice k_0 , mas o índice mais próximo de k_0 onde o sinal cruza o valor zero de forma ascendente. Ou seja, sendo $k_+ \geq k_0$ o primeiro índice (depois de k_0) tal que $x[k_+ - 1] < 0$ e $x[k_+] \geq 0$ e sendo $k_- < k_0$ o último índice (antes de k_0) tal que $x[k_- - 1] < 0$ e $x[k_-] \geq 0$, definiremos o ponto inicial como $\hat{k} = k_+$ se $k_+ - k_0 \leq k_0 - k_-$ e $\hat{k} = k_-$ caso contrário.

Para mostrar o fragmento do sinal da entrada entre os índices \hat{k} e $\hat{k} + K$, uma dica que evita escrever laços explícitos é usar o objeto `[phasor~]` com frequência $\frac{R}{N}$ e fase 0 para produzir um sinal com N valores igualmente espaçados entre 0 e 1, jogar estes valores em um `[expr~ \hat{k} + $v1 * K]` e fazer a leitura interpolada destes índices com `[tabread4~]`.

Espectro de Magnitude Normalizado

A janela contendo o espectro deve mostrar valores normalizados entre os extremos 0 e 1. Lembrando da expressão da DFT temos que $|F_n| = \left| \sum_{k=0}^{N-1} f_k e^{-i2\pi n \frac{k}{N}} \right| \leq \sum_{k=0}^{N-1} |f_k e^{-i2\pi n \frac{k}{N}}| = \sum_{k=0}^{N-1} |f_k| \leq N \cdot \max\{|f_k|\}$, de onde concluímos que, no nosso caso, $|F_n| \leq N$, $\forall n$. Assim basta dividir o espectro de magnitude por N . Na prática os valores encontrados serão bem menores que esse, então convém configurar a escala vertical do display para 0.25 ou 0.5.

Detecção de Frequência Fundamental e Pitch

A estimativa da frequência fundamental será obtida pela interpolação quadrática do maior pico do espectro de magnitude, como foi explicado anteriormente e realizado no segundo trabalho menor. Essa informação, que deverá aparecer em uma caixa de número, será atualizada uma vez a cada janela de análise (mecanismo controlado por um [bang~]). Além disso, queremos relatar na janela principal do Pd (com o objeto [print]) os inícios e fins de notas, bem como uma representação numérica na escala MIDI.

A conversão de frequências em Hz para MIDI usa a fórmula $f_0 = 440 \cdot 2^{\frac{d-69}{12}}$ (por convenção a nota A4=440 Hz está associada ao número MIDI 69), cuja solução inteira é $d = 69 + \text{round}(12 \log_2 \frac{f_0}{440})$. Você pode usar o objeto [ftom] para fazer essa conversão.

O início de um evento deve ser notificado na janela do Pd através do valor MIDI e da frequência em Hz, juntamente com o instante de início, medido em segundos e relativo ao início da execução do programa (use o objeto [timer] para isso); o fim do evento deve ser notificado com o instante da detecção do término da nota, conforme segue.

A frequência fundamental f_0 de janelas sucessivas correspondentes a uma mesma nota poderá flutuar ligeiramente, até mesmo em sons sintéticos, devido à imprecisão de nossa análise. Estimativas obtidas em janelas sucessivas serão consideradas como pertencentes a um mesmo evento se

$$\left| \frac{f_{0\text{atual}} - f_{0\text{anterior}}}{f_{0\text{anterior}}} \right| < 2^{1/12} - 1$$

($2^{1/12} - 1 \approx 0.0595$ é um semitom). Sempre que essa condição for violada consideraremos que um novo evento se inicia, e relataremos a nova frequência fundamental, o pitch correspondente e o instante em segundos do início do novo evento.

Para que a saída não fique poluída com estimativas absurdas de f_0 quando o sinal estiver muito próximo do silêncio (ruído de fundo), colocaremos um limiar inferior no nível de volume do sinal para que a saída textual seja de fato produzida: apenas janelas com valor de [rms~] maior do que RMS_MIN=0.01 deverão ser consideradas para a detecção de eventos (este valor sugerido pode depender de ajuste em função do equipamento e do nível de ruído do ambiente).

Considerações Finais

Alguns exemplos da documentação oficial do Pd poderão ajudar bastante. No menu “Ajuda” escolha a opção “Navegador de ajuda do Pd”, selecione o diretório “Pure Data/2.control.examples” e olhe os patches 05, 06, 07, 14, 15, 18 e 24. Depois disso, abra “3.audio.examples” e olhe os patches A06, E01, I01 e I02.

Bom Trabalho!