Computação Musical – $2^{\underline{o}}$ Exercício Programa

Prof. Marcelo Queiroz

Data de entrega: **3/6/2005**

Instruções: Os EPs devem ser feitos individualmente. A entrega será feita por e-mail para mim até a data indicada.

Osciloscópio Digital

Neste EP vamos escrever 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 saida padrão textual que sinalizará os eventos musicais (notas) detectados.

O programa não terá parâmetros de linha de comando: após iniciado, ele deve ler continuamente dados da entrada, processá-los e atualizar a saída, até ser interrompido pelo usuário. A implementação será ligeiramente dependente do sistema: todos os programas devem estar escritos em linguagem C e funcionar em ambiente Linux com arquitetura ALSA.

Aspectos técnicos de implementação são detalhados nas seções a seguir.

Leitura do Microfone

No ambiente ALSA/Linux, usando a configuração default, o microfone pode ser considerado como um arquivo (/dev/dsp) sem cabeçalho contendo uma lista de amostras de 8bits unsigned (valores entre 0...255 representando os inteiros entre -128...+127), amostrado a `R`=8KHz; você pode imaginar este driver como um "usuário" que digita os valores no teclado a intervalos regulares, e portanto a informação só está disponível a cada $\frac{1}{8000}$=0.128seg. Este arquivo pode ser lido em C usando `fscanf(mic,"%c",&amostra)` (obs: o mesmo arquivo pode ser utilizado para escrita, acessando os alto-falantes).

Considerando que as janelas para cálculo do espectro só serão lidas após o processamento e atualização da saida referente à janela anterior, a "resolução temporal" do osciloscópio será menor do que 0.128 segundos, e vários dados de entrada serão perdidos durante o processamento; não nos preocuparemos com isso.

Cálculo do Espectro

Para o cálculo do espectro utilizaremos o algoritmo FFT aplicado a cada janela de `N`=1024 amostras, que corresponde a $\frac{\mathtt{N}}{\mathtt{R}} = \frac{1024}{8000}$ =0.128 segundos. A resolução do espectro será de $\frac{\mathtt{R}}{\mathtt{N}}$ =7.8125Hz.

Em cada janela os valores das amostras serão multiplicados por uma função de Hamming. Lembre que a função de Hamming pode ser calculada numa janela indexada de 0 até `N-1` pela expressão

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

onde n é o índice da amostra na janela.

Análise do Espectro

Após o cálculo do espectro, consideraremos principalmente os valores absolutos $|F_0|, \ldots, |F_{N-1}|$ que correspondem ao espectro de magnitude do fragmento analisado. Observe que os valores $(F_0, F_1, F_2, \ldots, F_{\frac{N}{2}})$ possuem a informação espectral referente às freqüências 0Hz (constante d.c.), 7.8125Hz, 15.625Hz, ..., $\frac{N}{2}\frac{R}{N}$=4000Hz; os índices superiores a $\frac{N}{2}$ repetem de certa forma essa informação (possuem valores conjugados na ordem inversa).

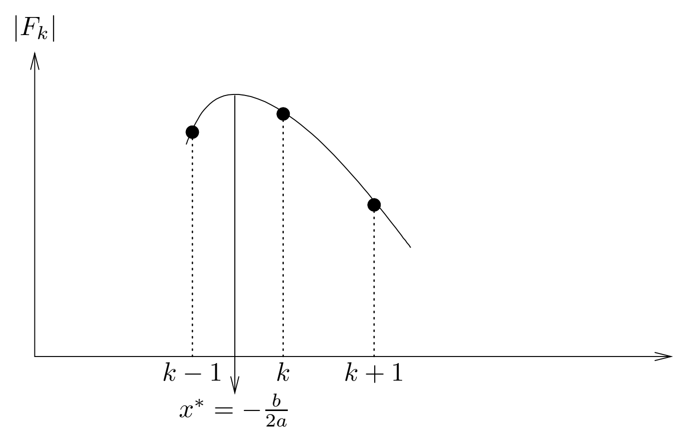
Para extrair as informações pretendidas do espectro temos que ser capazes de discernir entre espectros ruidosos e espectros "musicais"; neste contexto, consideraremos "musicais" os espectros que possuem um padrão de harmonicidade, ou seja, que possuem a energia concentrada em picos associados a freqüências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

A identificação da freqüência 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 inharmonicidade dos parciais, etc; faz-se necessário a adoção de alguma heurística que, esperamos, será capaz de encontrar a freqüência certa na maioria dos casos considerados.

Considere a freqüência de pico f_p que maximiza o valor de $|F_k|$ para $k = k_0, \ldots, \frac{\mathtt{N}}{2}$, onde $k_0 = \min\{k \mid k\frac{\mathtt{R}}{\mathtt{N}} \geq 20\}$ (isto é, desconsideraremos todos os valores de k para os quais a freqüência correspondente é inferior a 20Hz). A fim de aumentar a precisão da análise, consideraremos as duas amostras vizinhas à amostra que maximiza $|F_k|$ e estimaremos a freqüência de pico fazendo uma interpolação quadrática dos valores do espectro. Considere que $|F_k| \geq |F_n|$, $\forall k = k_0, \ldots, \mathtt{N-1}$. 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 freqüência de $x^*\frac{\mathtt{R}}{\mathtt{N}}$Hz:



Utilizaremos esta freqüência de pico f_p para estimar a freqüência fundamental f_0 tentando averiguar se cada uma das freqüências $f_p, \frac{f_p}{2}, \frac{f_p}{3}, \ldots, \frac{f_p}{HMAX}$ 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 freqüência 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 7.8125Hz; isso não tem relação nenhuma com o sinal de entrada mas sim é uma conseqüência da escolha dos parâmetros da análise. Freqüentemente 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 freqüências muito baixas (pois a soma EHT delas incluirá muitos índices). Considerando que a entrada será utilizada por uma voz humana podemos tranqüilamente excluir de consideração candidatos a fundamental abaixo de FRQMIN=65Hz.

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

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

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

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

Estas fórmulas envolvem o cálculo de $|F(\omega)|$ para valores de ω que não estão necessariamente na série harmônica de 7.8125Hz. Utilize o seguinte critério: sejam k e $k+1$ os índices do espectro tais que $\omega \in [k\frac{R}{N}, (k+1)\frac{R}{N}]$; então

1. Se $|F_k|$ é máximo local ($|F_k| = \max\{|F_{k-1}|, |F_k|, |F_{k+1}|\}$) então use a interpolação quadrática centrada no índice k;
2. Se $|F_{k+1}|$ é máximo local então use a interpolação quadrática centrada no índice $k+1$;
3. Caso contrário, use interpolação linear entre k e $k+1$ (ou seja, defina $k_\omega = \frac{\omega}{R/N}$ e faça $|F(\omega)| = (k+1-k_\omega)|F_k| + (k_\omega - k)|F_{k+1}|$.

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

Gráficos com Gnuplot

Podemos abrir janelas e mostrar gráficos no Gnuplot a partir do nosso programa em C utilizando a biblioteca gnuplot-i disponível em `http://ndevilla.free.fr/gnuplot/`. A biblioteca e alguns exemplos simples de uso estão em `http://ndevilla.free.fr/gnuplot/gnuplot_i-2.10.tar.gz`. A documentação online é bem direta e há exemplos curtos das chamadas em fragmentos de programas em C.

Além do `#include "gnuplot_i.h"` no início de nosso código, as principais funções que utilizaremos são `gnuplot_init`, `gnuplot_setstyle`, `gnuplot_plot_x` e `gnuplot_close`. Além destes, a função `gnuplot_cmd` permite passar um comando genérico usando a sintaxe do modo interativo do Gnuplot: por exemplo, a chamada `gnuplot_cmd(window1,"set yrange [-128:128]")` prepara a janela associada ao apontador `window1` (do tipo `gnuplot_ctrl *`, veja a documentação) para mostrar gráficos associados a formas de onda.

Formato da Saída

Para cada espectro produzido devemos atualizar informação sobre a forma de onda, espectro de magnitude e detecção de notas musicais.

A janela contendo a forma de onda mostra sempre uma lista de valores entre -128 e +128; ela deve se atualizada da seguinte maneira: se o espectro é ruidoso o gráfico mostrará os 1024 valores lidos da entrada, do contrário o gráfico mostrará 3 períodos completos da forma de onda. Cada evento musical possui uma freqüência fundamental f_0 definida, e o número de amostras que corresponde a 3 períodos completos é $K = 3\frac{1}{f_0}\mathtt{R}$. Considerando que um evento musical pode durar mais do que $\frac{1}{8}$ de segundo, 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 freqüência 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 $k_0 = \mathtt{R}t_0$.

A janela contendo o espectro mostra 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 concluimos que, no nosso caso, $|F_n| \leq 128\cdot\mathtt{N}$, $\forall n$. Na prática os valores são bem menores que esse, então consideraremos MAGNMAX$=\frac{128\cdot\mathtt{N}}{4}$ e apresentaremos todos os valores relativos a este máximo. Se o espectro em consideração é ruidoso, devemos mostrar simplesmente o gráfico do espectro normalizado. Do contrário, utilizaremos uma versão simplificada do espectro que mostra apenas a energia da série harmônica a partir da fundamental f_0, ou seja, os valores $|F(f_0)|, |F(2f_0)|, |F(3f_0)|, \ldots$, relativos a MAGNMAX.

Em relação aos eventos musicais, considere a seguinte representação numérica para as notas: podemos obter o número d de semitons acima ou abaixo da freqüência de referencia 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 saida padrão, com este valor de d e sua freqüência em Hz, juntamente com o instante de início (relativo ao início da execução do programa); seu fim deve ser notificado com o instante da detecção do término da nota. Observe que a freqüência fundamental f_0 de janelas sucessivas correspondentes a uma mesma nota poderá flutuar ligeiramente, mas a nota será considerada única se os valores integrais de d coincidirem. Uma desvantagem desta heurística é que se a freqüência fundamental estiver bem no meio de 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 452.89Hz$= 440 \cdot 2^{\frac{0.5}{12}}$ 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 freqüências 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 do EP.

Vou disponibilizar em breve um executável para Linux e alguns arquivos de entrada. Você pode testar o seu código em arquivos de entrada diferentes do `/dev/dsp` introduzindo um pequeno delay na leitura para garantir que cada janela leva $\frac{1024}{8000}$seg para ser obtida.

Bom Trabalho!