1 2 9 0
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA
DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES

Processamento Digital de Sinal

Trabalho Prático Nº 6 – Banco de filtros em escala Mel

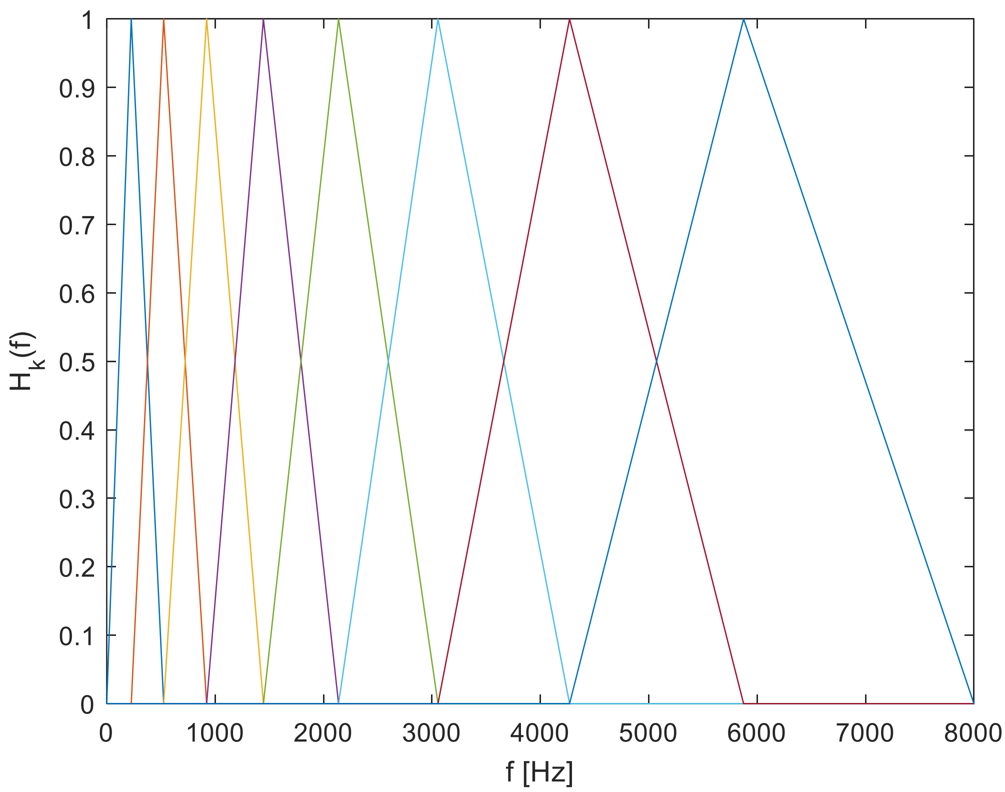
Descrição

Neste trabalho pretende-se implementar um banco de filtros em escala "mel", onde a largura dos filtros cresce com a frequência central dos filtros. Trata-se de um sistema muito usado para extração de "features" de áudio, uma vez que mimetiza o funcionamento do sistema auditivo humano.

A escala de mel (de *mel*odia) é uma escala perceptual de tonalidade, baseada em comparações de afinação (tom mais alto ou mais baixo que outro). O ponto de referência é definido com 1000 mel relativo a um tom de 1000 Hz. Acima de 700 Hz, intervalos cada vez maiores são necessários para produzir incrementos de afinação iguais. A relação entre a escala em mel, φ, e a escala em Hertz, f, é a seguinte:

$$\varphi = 1127\ln\left(1 + f/700\right).$$

Todas as operações são definidas na frequência, através das DFTs das tramas de sinal de entrada. Os filtros são triangulares com vértices de tal forma que a sua soma é constante, exceto no 1º e último filtro. A figura seguinte mostra o caso de um banco de 8 filtros, para uma frequência de amostragem de 16kHz. Estes filtros são vistos como *pesos* aplicados ao módulo quadrado da DFT da trama, e, portanto, são filtros que integram na sua banda o espetro do sinal de entrada.



Para uma situação com tramas de N amostras e avanço de trama de M amostras (com $M<N$ tal que existe sobreposição de N-M amostras), temos a situação da figura seguinte.
Nesta figura a trama m é dada pelas seguintes N amostras:

$$x_m[n] = x[mM - n] \ , \ n = 0 : N-1,$$

o que corresponde a um sistema não maximamente decimado se $M<N$.

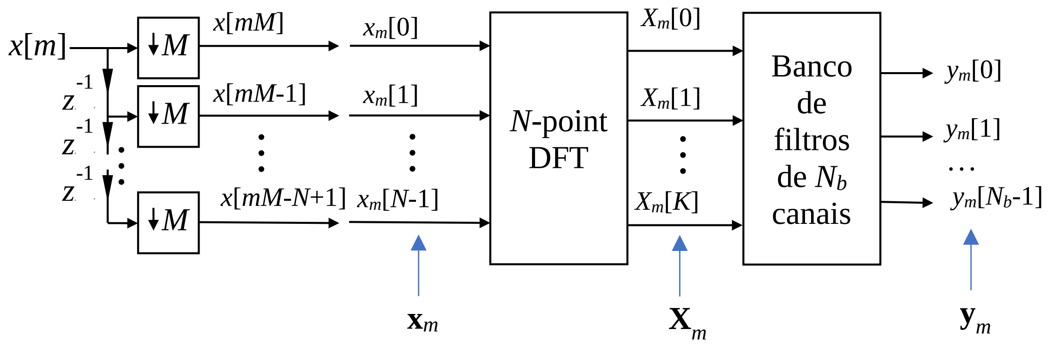
DEPARTAMENTO DE
ENGENHARIA ELECTROTÉCNICA
E DE COMPUTADORES

Processamento Digital de Sinal

Trabalho Prático Nº 6 – Banco de filtros em escala Mel

Uma DFT de N pontos é aplicada a esta trama, resultando nos N coeficientes $X_m[k]=\sum_{n=0}^{N-1} x_m[n]\cdot e^{-jk\frac{2\pi}{N}n}$ e onde k=0:K, K=N/2 (índices dos "bins" de interesse da DFT até $\omega=\pi$ ou $f=f_s/2$). O bloco do banco de filtros de N_b canais corresponde a uma matriz, **H**, de dimensão $N_b\times(K+1)$ onde cada linha corresponde a um filtro triangular e onde as N_b saídas são dadas por $\mathbf{y}_m=\mathbf{H}\cdot|\mathbf{X}_m|^2$. De notar que este produto matricial corresponde ao cálculo das energias à saída do banco de filtros. Para cada canal i, a energia da trama m é [1]:

$$y_m[i]=\frac{1}{K}\sum_{k=0}^{K} H[i,k]\left|X_m[k]\right|^2 .$$



Trabalho Prático

a) Leia um sinal de áudio com a função `audioread`. Faça tramas de 512 amostras de 160 em 160 amostras (N=512; M=160; janela retangular). Se for N_x o comprimento do sinal, deve gerar $N_f=\left\lfloor\frac{N_x-N}{M}\right\rfloor+1$ tramas inteiras. Pode usar a função `buffer`. Experimente com o seguinte comando:

```
buffer(1:10,5,5-2,'nodelay')
```

b) Considere um banco de filtros de 40 canais (N_b=40). Use o código seguinte para calcular a matriz **H** de dimensão 40×257.

```
Nb=40;N=512;K=N/2; k=0:K; f=k*fs/N;
mu0=0;mu1=1127*log(1+fs/2/700); %mel máxima em fs/2.
melCF=linspace(mu0,mu1,Nb+2); fCF=700*(exp(melCF/1127)-1); %CF: characteristic frequencies
%Cáculo de H e plot dos triângulos:
H=zeros(Nb,K+1);
for i=1:Nb
    fa=fCF(i); fc=fCF(i+1); fp=fCF(i+2); %freq. dos vértices do triângulo i.
    H(i,:)= (f-fa)/(fc-fa).*(f>=fa & f<=fc) + (f-fp)/(fc-fp).*(f>fc & f<=fp);
    plot(fCF,fCF==fCF(i+1)); hold on
end
hold off; figure; plot(f,H) %porquê a diferença?
```

---

[1] A energia de um sinal real $z[n]$ com DFT $Z[k]$ é $E_z=\sum_{n=0}^{N-1} z^2[n]=\frac{1}{N}\sum_{k=0}^{N-1}\left|Z[k]\right|^2=\frac{1}{N}\left(\left|Z[0]\right|^2+2\sum_{k=1}^{N/2-1}\left|Z[k]\right|^2+\left|Z[\tfrac{N}{2}]\right|^2\right)$. Se $\left|Z[0]\right|=\left|Z[\tfrac{N}{2}]\right|=0$, esta energia pode ser dada por $E_z=\frac{2}{N}\sum_{k=1}^{N/2-1}\left|Z[k]\right|^2=\frac{2}{N}\sum_{k=0}^{N/2}\left|Z[k]\right|^2=\frac{1}{K}\sum_{k=0}^{K}\left|Z[k]\right|^2$, com K=N/2. No presente caso, todos os filtros triangulares têm peso zero em ω=0 (k=0) e em $\omega=\pi$ (k=N/2), verificando-se esta condição.

Processamento Digital de Sinal

Trabalho Prático Nº 6 – Banco de filtros em escala Mel

Interprete o código e verifique o resultado nas duas figuras geradas. Note que uma figura é gerada usando os vértices dos triângulos e outra em que a abcissa são os pontos correspondentes aos “bins” da DFT (que nem sempre coincidem com os vértices dos triângulos).

c) Calcule as saídas (energias) para todas as tramas.

d) Faça um sonograma em dB com o resultado que obteve. Compare este sonograma em escala mel com o sonograma DFT. Interprete as diferenças.
Notas:
Use a função `db()` para gerar sonogramas em dB. Use a instrução `axis xy` e `colorbar`.
Não é fácil colocar no sonograma mel as frequências respetivas em Hz, uma vez que a escala não é linear. Contudo, podemos alterar as etiquetas verticais do gráfico de acordo com a escala mel inversa. Para isso, faça o seguinte:

1. gerar o gráfico com a escala mel (usar nº de tramas como abcissa e `melCF` como ordenada);
2. ler os números em mel das etiquetas: `umel=get(gca,'YTick')`
3. alterar para Hz: `fmel=700*(exp(umel/1127)-1)'`
4. passar a strings e colocá-las no gráfico:

```
zz=num2str(round(mu)); set(gca,'YTickLabel',cellstr(zz))
```

No final da aula entregue o script que produziu, na “Submissão de Trabalhos” do Nónio.