高等數位訊號處理(Advanced Digital Signal Processing)
Homework 03 – 2017/05/19

R05546030 **彭新翔**

1. Write a Matlab program that can convert a numbered musical notation (簡譜) into a music file (*.wav).

   Example: (Twinkle twinkle little stars)

   ```
   score = [1,1,5,5,6,6,5];        % 1:Do, 2:Re, 3:Mi...
   beat = [1,1,1,1,1,1,2];         % 拍子
   name = ‘twinkle’;
   getmusic(score, beat, name)     % Generate the music file twinkle.wav
   ```

   The Matlab file should be mailed to displab531@gmail.com.

   With basic requirement (score, beat, name): 28 scores

   程式的功能越多，考慮的因素越多，分數越高

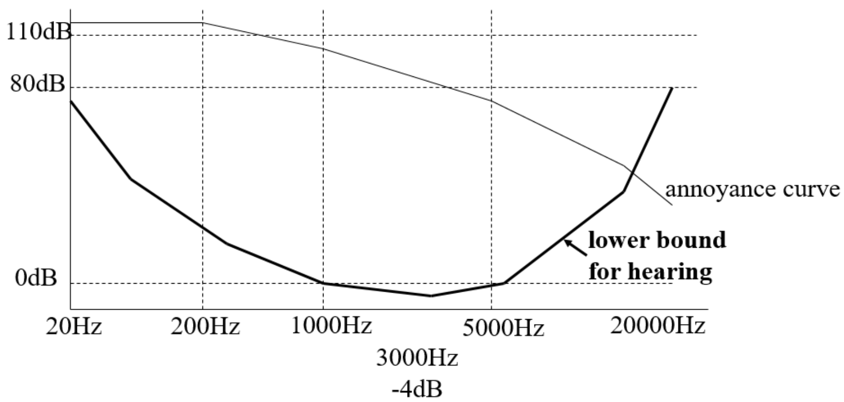
Code 已寄至實驗室信箱。

2. Suppose that there are three vocal signals: (i) $\cos(20\pi t)$, (ii) $\sin(100\pi t)$, (iii) $\sin(4000\pi t)$.

(a) Which voice sounds louder?

(b) Which voice signal can be propagated in a longer distance?

(a) 人耳普遍可以辨識的頻率範圍為 20Hz – 20000Hz，但對於不同頻率的聲音有不同的敏感程度，如下圖所示：



在一般狀況下，人類對於 3000Hz 左右頻率的聲音最為敏感，因此聽起來所感受到聲音最大的應為 (iii) $\sin(4000\pi t)$。

(b) 一般來說，頻率越高、波長越短，繞射能力越弱（但穿透能力較高），因此在傳播過程中容易散射耗損能量，聲音在空氣中傳播時由於在正常狀態下容易存在許多障礙物，因此低頻率且波長大的 (i) $\cos(20\pi t)$ 能夠相較於其他高頻率的訊號傳遞得更遠。

3. In addition to the DCT, which is adopted by MP3, write at least three possible ways that can compress a music signal more efficiently.

離散餘弦變換(DCT, Discrete Cosine Transform)在進行音樂訊號壓縮時，包括但不限於以下特點可以使得壓縮時更有效率：

(1) 處理音訊訊號時，常將其能量大部分集中於頻域的一個小範圍內，對於不重要的訊號分量可以有效減少所需的 bit 數。
(2) 具有較高的抵抗干擾能力，因此傳輸過程中錯誤編碼的影響遠小於預測編碼。
(3) 採用了 Entropy Coding 可以有效壓縮數據。
(4) 考量音訊訊號・對於重複的旋律進行了壓縮。
(5) 考量音訊訊號，針對特定和弦頻率的倍數進行壓縮。
(6) 考量音訊訊號，對於特定音符固定性進行了壓縮。

4. In the JPEG process, (a) why the DCT is used instead of the DFT and the KLT for transformation? (b) Why the input image is separated into several 8×8 blocks before using the DCT?

(a) 相較於離散傅立葉變換(DFT, Discrete Fourier Transform)而言，離散餘弦變換(DCT, Discrete Cosine Transform)具有以下特點：

- 能量更為集中
- output 為實數
- 一樣都具備較快速的演算法

而相較於 Karhunen-Loeve Transform (KLT) 來說，離散餘弦變換(DCT, Discrete Cosine Transform)具有以下特點：

- 獨立於 input
- output 為實數
- 對大部分影像而言，鄰近峰值處能夠近似 KLT。
- 尤其當 $\text{corr}\{f[m,n], f[m+\tau, n+\tau]\} = \rho^{\tau}\rho^{\eta}$ 其中 $\rho \to 1$ 時有較快速的演算法

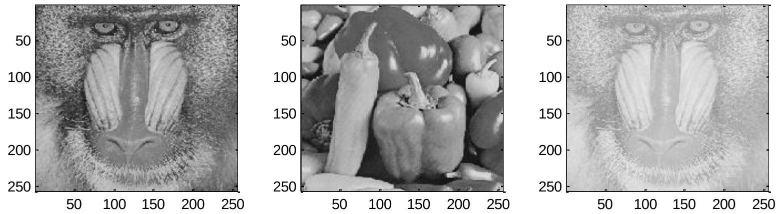
(b) 在進行離散餘弦變換(DCT, Discrete Cosine Transform)必須將影響切分成 8×8 方格的考量因素如下：

- 定義 JPEF 規範時，經由實驗最佳化所得到的結果。
- 具有較低的計算複雜度
- 可以有效降低 buffer size
- 切分後反映了影像不同區域的特性，區分區塊內的一致性

5. (a) Why the normalized root mean square error (NRMSE) may not reflect the similarity between two images?

(b) Can the NRMSE measure the similarity between two audio signals? Why?

(a) 如下圖所示：



上圖中實際上為圖一與圖三較為相似，但若以正規化方均根差(NRMSE, Normalized Root Mean Square Error)考慮時，會得到圖一與圖二較為接近的結果。其主要原因在於人眼對於圖像辨識在結構性差異上，並不是僅有亮度和振幅差異，反而更為敏感，這也導致了即使 NRMSE 在理論上合理，但由於無法辨別出其他特徵而無法反映實際結果。

(b) NRMSE 在聲音訊號上並不如影像訊號那般能夠明顯辨識差異，最主要的原因在於音樂訊號上可能存在有延遲或由於振幅大小的不同，導致無法有效區隔兩段聲音訊號。

6. Suppose that $P(x=n)=\frac{e^{-\lambda}\lambda^n}{n!}$ for $n=0,1,2,3,\cdots$ where $\lambda=0.98$. Also suppose that $\mathrm{Length}(x)=10000$. Estimate the range of the total coding lengths in the binary system when using (i) the Huffman code and (ii) the arithmetic code.

已知 $N=10000, k=2$，由 $\text{Entropy}=\sum_{j=1}^{J}P(S_j)\log\frac{1}{P(S_j)}$ 可知：

$$\text{Entropy}=\sum_{n=0}^{\infty}\frac{e^{-0.98}0.98^n}{n!}\log\frac{n!}{e^{-0.98}0.98^n}\approx 1.29327$$

(i) The Huffman Code

將上述已知條件代入 $N\frac{\text{Entropy}}{\log k}\le b\le N\frac{\text{Entropy}}{\log k}+N$ 可得：

$$10000\frac{1.29327}{\log 2}\le b\le 10000\frac{1.29327}{\log 2}+10000$$
$$\Longrightarrow\ 18658\le b\le 28658$$

(ii) The Arithmetic Code

將上述已知條件代入 $\left\lceil N\frac{\text{Entropy}}{\log k}\right\rceil\le b\le\left\lfloor N\frac{\text{Entropy}}{\log k}+\log_k 2\right\rfloor+1$ 可得：

$$\left\lceil 10000\frac{1.29327}{\log 2}\right\rceil\le b\le\left\lfloor 10000\frac{1.29327}{\log 2}+\log_2 2\right\rfloor+1$$
$$\Longrightarrow\ 18658\le b\le 18659$$