ADSP HW3

R10942152 謝家樺

Homework 3 (Due: 4/29th)

(1) Write a Matlab or Python 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 / Python code should be handed out by NTUCool.

With basic requirement (score, beat, name): 24 scores

程式的功能越多，考慮的因素越多，分數越高 (30 scores)

額外實作的功能：①調整 base frequency ②調整歌曲速度(BPM)

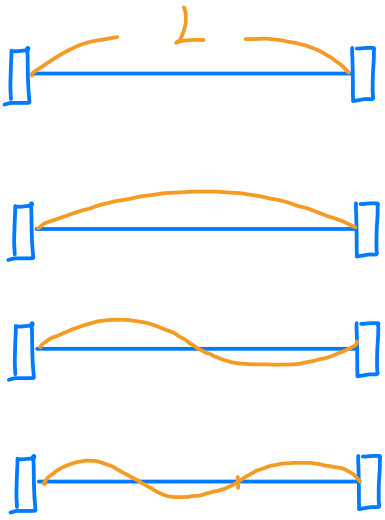
③音符漸弱功能(exponential / linear decay)

(2) (a) In the noiseless case, in what condition we cannot use the variation of amplitude to separate a speech signal into several syllables?

子音的能量小於母音，所以一般情況能用amplitude將音節分開，但當遇到雙母音或是沒有子音的字的時候，信號上看起來能量都一樣，就無法分辨

(b) Why a music signal always has the chord (和弦) phenomenon? (10 scores)

考慮一條2端固定的弦，長度為L



共振會發生在

$L=\frac{\lambda}{2}$

$L=\lambda$

$L=\frac{3}{2}\lambda$

$\lambda=\frac{2L}{n}\ (n\in\mathbb{N})$

可得在基頻的整數倍會發生共振

若基頻為Do，$f=330Hz$，則在

$f=660Hz$ (Mi) 也會發生共振

$f=990Hz$ (Sol)

形成和弦

(3) (a) Why a music signal is easier to compress than other vocal signals? (Write at least 3 reasons) (b) Why a cartoon / mark image is easier to compress than other images? (Write at least 2 reasons) (10 scores)

(a) ① 音樂信號的頻率分佈比較固定，只會在固定音高及其倍頻有能量

② 音樂信號的拍子是固定的，固定的時間間隔才有信號

③ 在單一音符內的頻率是固定不變的

以上三個特點使音樂訊號具有一致性，容易壓縮

(b) ① 卡通或標誌的同個區域內的顏色大多一致

② Edge 往往是較簡單的線條 e.g. 圓，方形，直線

上述兩個一致性使得卡通比一般影像容易壓縮

(4) (a) Why the YCbCr color space is applied instead of RGB in the 4:2:2 and 4:2:0 techniques? (b) What is the compression ratio of 4:2:0? (10 scores)

(a) 因為 RGB 3個 channel 都差不多重要，轉成 YCbCr 之後，只有 Y channel 是重要的，剩下的 CbCr 的部份可以拿來壓縮，結果不會差太多

(b) 4:2:0 是將 Cb, Cr 的 column and row 壓縮成原來的一半大小

使 compression ratio = 2

(5) (a) Why we always use the DCT instead of the DFT and the KLT to image compression? (Write two reasons). (b) Which of the following compression techniques are lossless? (i) 4:2:0; (ii) DC difference; (iii) zigzag; (iv) quantization table; (v) the Huffman code. (10 scores)

(a) ① DFT的結果會出現實數and虛數，使壓縮的过程还要額外記錄虛數部，誤效能下降

② DCT更擅於將頻譜轉成低頻區域上，誤压縮过程更有效果。

③ DCT independant of input compare to KLT

(b) (ii) DC difference

(iii) zigzag

(v) Huffman code

(6) Suppose that $P(x = \text{'a'}) = 0.5$, $P(x = \text{'b'}) = 0.3$, $P(x = \text{'c'}) = 0.1$, $P(x = \text{'d'}) = 0.1$.

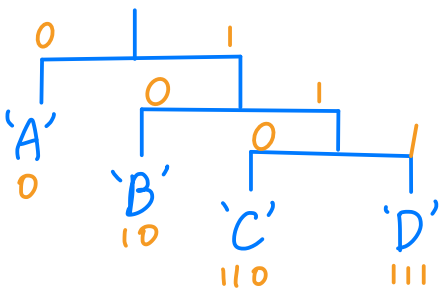
(a) Determine the coding tree of x when using the Huffman code in the binary (二進位) system.

(b) What is the entropy of x?

(c) What is the result of the arithmetic coding if x = 'aba'?

(d) Suppose that length(x) = 100,000. Estimate the range of the total coding lengths in the binary system when using the arithmetic code. (20 scores)

(a)



(b) $\text{entropy} = P(x='a')\cdot\log\left(\frac{1}{P(x='a')}\right)\ P(x='b')\cdot\log\left(\frac{1}{P(x='b')}\right)$

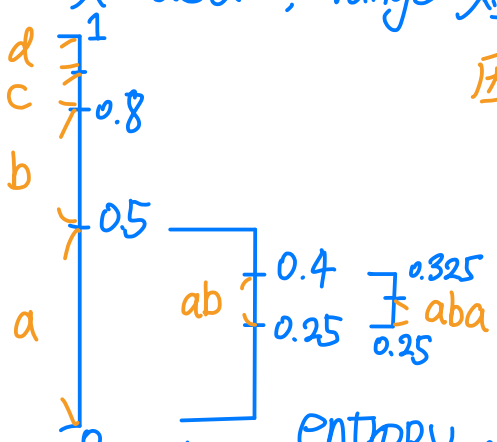
$P(x='c')\cdot\log\left(\frac{1}{P(x='c')}\right)\ P(x='d')\cdot\log\left(\frac{1}{P(x='d')}\right)$

$= 0.5\cdot\log\left(\frac{1}{0.5}\right) + 0.3\cdot\log\left(\frac{1}{0.3}\right) + 0.1\cdot\log\left(\frac{1}{0.1}\right) + 0.1\cdot\log\left(\frac{1}{0.1}\right)$

$= 1.168$ #

(c) x='aba', range 為 [0.25, 0.325], $0.25 \le \frac{4}{16} < \frac{5}{16} \le 0.325$

因此 b=4, C=4, 編碼結果為 0100 #



(d) $\text{ceil}\left(N\cdot\frac{\text{entropy}}{\log k}\right) \le b \le \text{floor}\left(N\cdot\frac{\text{entropy}}{\log k} + \log_k 2 + 1\right)$

$= \text{ceil}\left(100000\cdot\frac{1.168}{\log 2}\right) \le b \le \text{floor}\left(100{,}000\cdot\frac{1.168}{\log k} + \log_2 2 + 1\right)$

$= 116800 \le b \le 116802$ #

(7) Write at least three conditions that applying the NRMSE may not well reflect the similarity of two vocal signals. (10 scores)

① 兩個頻率差一点点的聲音信號即便聽起來一樣，但NRMSE會差很多

② 相位不同的信號聽起來也一樣，但NRMSE差很多

③ 聲音出現位置不同聽起來相同，NRMSE差很多