第四章 實驗結果與討論分析

本章節會對 FFT1~FFT5 共有五種整數型態 FFT 作誤差與辨識率分析。

4.1 誤差分析

平均絕對誤差值(average absolute error ,簡稱 AAE)定義：

$$AAE = \frac{\sum_{k=0}^{N-1} \left| \frac{FFT[k]_{(int)}}{SF} - FFT[k]_{(float)} \right|^2}{N}$$

音框溢位(frame overflow)定義：

只要音框中 FFT 運算過程中的值不在整數 $-2^{31} \sim 2^{31}-1$ 範圍之間，則稱為 frame overflow。

音框溢位率(frame overflow rate)定義：

$$\text{frame overflow rate} = \frac{\text{all frame overflow number}}{\text{all frame number}}$$

表格 4- 1 為 TCC300 語料資訊，我們取 TCC300 中男女聲音各 112 個 wav 檔，共 224 個 wav 檔來作誤差分析。這 224 個 wav 檔包含了 TCC300 中所有的音節。音框取樣點 N=320，Overlap=160 及預強調參數 a=0.975。

	TCC300 Corpus
Speaker	150 males and 150 females
Sampling rate	16 kHz
Bits per sample	16 bits
Total files	8913
Total Time	26.34 hours

表格 4- 1　TCC300 語料資訊

以下圖 4- 1~圖 4- 3 為整數型態 DFT 和 FFT1~FFT5，利用 TCC300 取樣的 224 個 wav 檔，所得到平均絕對誤差值，分別和音框溢位率(0%~100%)、(0%~10%)和(0%~0.5%)的分析圖：

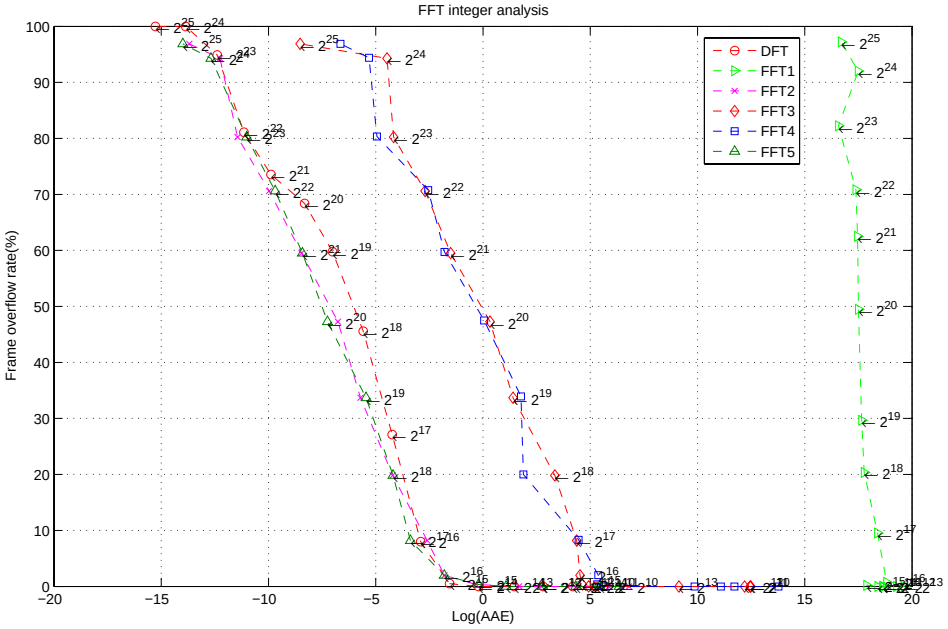

圖 4- 1 整數型態 FFT 音框溢位率(0%~100%)和平均絕對誤差值分析圖

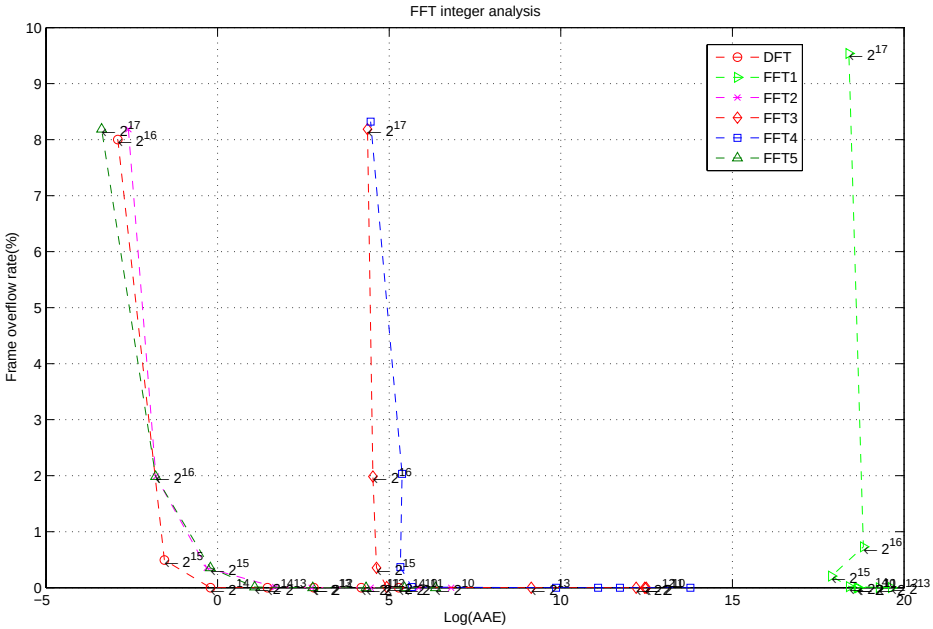

圖 4- 2 整數型態 FFT 音框溢位率(0%~10%)和平均絕對誤差值分析圖

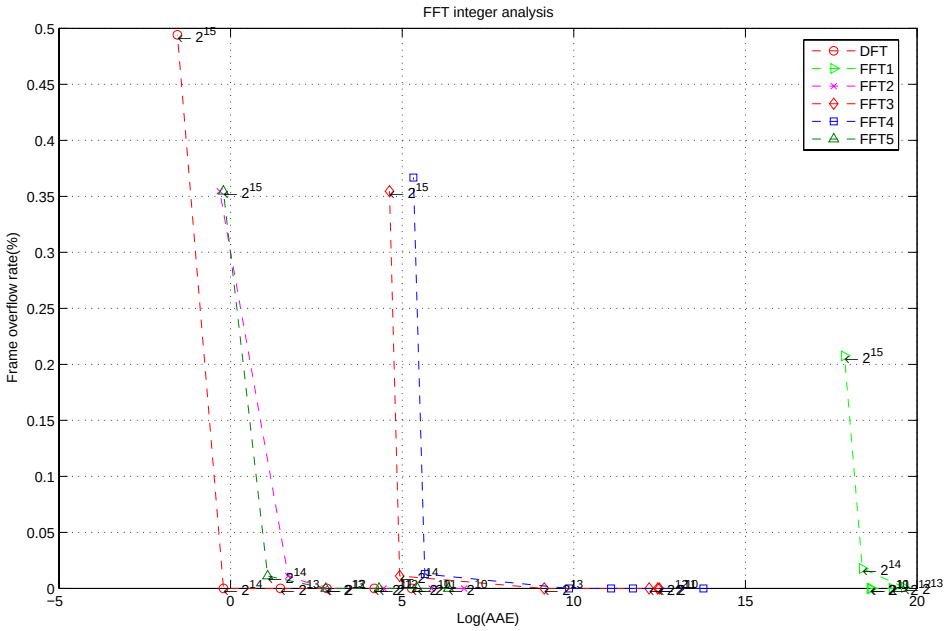

圖 4- 3 整數型態 FFT 音框溢位率(0%~0.5%)和平均絕對誤差值分析圖

在圖 4- 1~圖 4- 3 中，X 軸為 Log(AAE)愈左邊代表誤差愈小，反之誤差愈大。Y 軸為音框溢位率(%)愈下方代表愈少音框溢位，反之音框溢位愈多。其中圖 4- 2 和圖 4- 3 為了方便分析，將圖 4- 1 放大音框溢位率(0%10%)和(0%~0.5%)的部分。由圖 4- 3 中可明顯看出，整數型態 DFT 誤差為最低，其次是本篇所提出的 FFT5，接下來是 FFT2、FFT4 和 FFT3 最差為 FFT1。

由上圖可知在 SF=2^{10}~2^{13}， FFT1~FFT5 音框溢位率為 0，不過這只是取 TCC300 中 224 個 wav 檔，共 600790 個音框作分析的結果。為了確定 SF=2^{10}~2^{13} 對整個 TCC300 語料的音框溢位率為 0。我們用 TCC300 所有的語料 8913 個 wav 檔，共 9481954 個音框，針對 SF=2^{10}~2^{13} 再作一次 FFT1~FFT5 的音框溢位率分析。其結果如圖 4- 4 所示，由圖中可知當 SF = 2^{13} 會有音框溢位發生，所以先取 SF=2^{12} 來作辨識率分析。

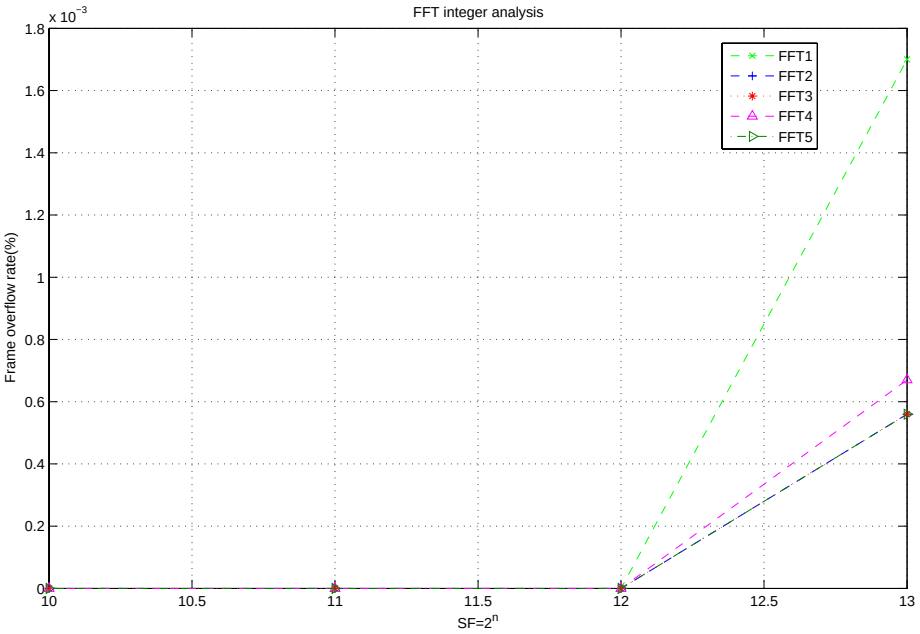

圖 4- 4 整數型態 FFT SF= 2^{10} ~ 2^{13} 和音框溢位率分析圖

另外在圖 4- 3 中，DFT 和 FFT2~FF4 的 SF 愈大則平均絕對值誤差愈小，但是 FFT1 SF= 2^{10} 卻比 SF= 2^{11} 的平均絕對誤差值還要小，這是因為建表時誤差太大而造成的。舉一 FFT1 簡單例子作說明：

假設 N=3，取 SF1= 2^{11} <SF2= 2^{12} 在經過整數化運算過程如下：

$$\sin(\pi/4) = 0.707$$

$$SF1 = 2^{11}$$

$$SF2 = 2^{12}$$

$$sf1 = 12.6992$$

$$sf2 = 16$$

$$x1 = \text{int}(0.707 \times sf1) = 9$$

$$x2 = \text{int}(0.707 \times sf2) = 11$$

$$\sin(\pi/4) \times \sin(\pi/4) \times \sin(\pi/4) = 0.3536$$

$$AAE1 = \left| \frac{x1 \times x1 \times x1}{SF1} - 0.3536 \right|^2 = \left| 0.3506 - 0.3536 \right|^2 = 9 \times 10^{-6}$$

$$AAE2 = \left| \frac{x2 \times x2 \times x2}{SF2} - 0.3536 \right|^2 = \left| 0.3250 - 0.3536 \right|^2 = 817.9 \times 10^{-6}$$

由運算結果 AAE2>AAE1 可知這是整數化過程中所產生的誤差，使得 SF 愈大而平均絕對誤差值也愈大。

4.2 辨識率分析

辨識率定義：

$$\text{辨識率} = \frac{\text{辨識正確的句數}}{\text{唐詩全部3210句}}$$

唐詩基本資訊如下所示：

	Testing Data
Content	唐詩
Speakers	8 Males and 2 Females
Sample Rate	16 kHz
Bits Per-sample	16 bits
Total	3210 Files = 4.5 hours

表格 4- 2 唐詩資訊

下表為整數型態 MFCC 各步驟之 SF 值，數據由參考資料[9]所得。其中 FFT 的 SF= 2^{12} (音框溢位率為 0 時，最大 SF 值)。

	預強調	漢明窗	FFT	三角帶通濾波器	DCT	對數能量
SF	2^{14}	2^{14}	2^{12}	2^{11}	1846	e^{6}

表格 4- 3 整數型態 MFCC 各步驟之 SF 值(TCC300 語料)

我們將所得到的辨識率、建表元素個數、記憶體容量和辨識時間作比較，如下表所示。(測試機器 CPU 為 Intel® Pentium® M processor 1.73GHz)

	Float FFT	FFT1	FFT2	FFT3	FFT4	FFT5
辨識率	95.58%	0.031%	80.53%	78.89%	79.19%	81.34%
建表元素個數		256	256	512	18	128
記憶體容量		1.5k bytes	1.7k bytes	3.29k bytes	0.29k bytes	0.95k bytes
辨識時間	1934 sec	2312 sec	1942 sec	1941 sec	1942 sec	1940 sec

表格 4- 4 FFT1~FFT5 辨識率、建表元素個數、記憶體容量和辨識時間比較表

由表格 4- 4 的數據結果來作兩點比較：

1. FFT1 和 FFT2 的建表元素個數一樣為 256 個，但辨識率卻相差約 80%和記憶體容量相差約 0.2k bytes。

   舉例說明其原因：假設 SF=4096 和 N=8，則 FFT1 的建表放大係數為 $sf = 2^{\frac{\log_2 4096}{\log_2 8}} = 2^4 = 16$，而 FFT2 的建表放大係數為 SF=4096。在比較之下很容易得知，FFT1 建表值和 FFT2 相差 $\frac{FFT2_{SF}}{FFT1_{sf}} = \frac{4096}{16} = 256$ 倍。另外假設其建表後的值 FFT1 和 FFT2 分別為 $\text{int}\left(sf \times \sin\left(\frac{\pi}{4}\right)\right) = 11$ 和 $\text{int}\left(SF \times \sin\left(\frac{\pi}{4}\right)\right) = 2896$，由此可知兩者所占記憶體就相差 2 bytes(一個數字占 1 byte)。又因為 FFT1 建表的放大係數比 FFT2 少，使得在整數型態 FFT 運算時誤差很大，也就造成 FFT1 的辨識率只有 0.031%。

2. FFT5 比 FFT4 的建表元素的數量多 110 個，但是辨識率 FFT5 卻比 FFT4 要高約 2%。

   因為 FFT4 利用建表值，動態運算求出其它 *W* 值。原本 FFT4 在建表時的值已有誤差，再用有誤差的值去求其它 *W* 值，誤差會更大。反觀 FFT5 將所用到的 *W* 值皆建入表中，雖然比 FFT4 多了 110 個建表元素，但可使 *W* 值誤差比 FFT4 要低，在辨識率上當然 FFT5 會比 FFT4 要來的高。

接下來將整數型態 FFT 辨識率和絕對誤差值作分析，如下圖所示。由圖中可看出當平均絕對誤差值愈小，則辨識率會愈高。

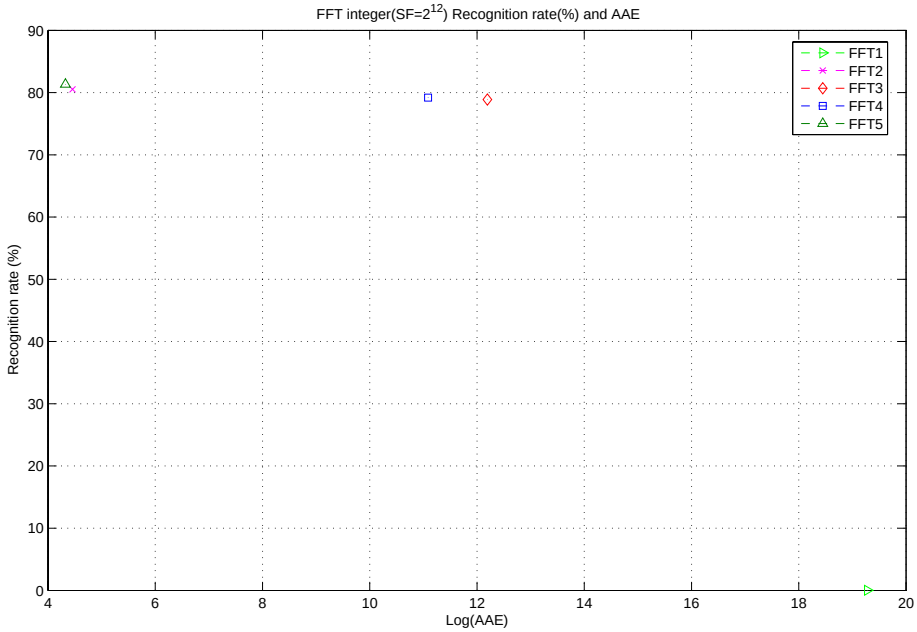

圖 4- 5 整數型態 FFT SF= 2^{12} 辨識率與平均絕對誤差值分析圖

最後我們取 FFT5 的 SF 值由 2^{10} ~ 2^{16}，作辨識率和音框溢位率分析，如圖 4- 6 和圖 4- 7 所示。由兩圖可知當 SF= 2^{13} 時，雖然音框溢位率為 0.00056%，但辨識率可到達 81.65%，比 SF= 2^{12} 的辨識率還要高 0.31%，代表 SF= 2^{13} 時的音框溢位率，還不會對辨識率產生太大的影響。所以本篇論文取 SF= 2^{13} ，能得到最佳的辨識率為 81.65%。

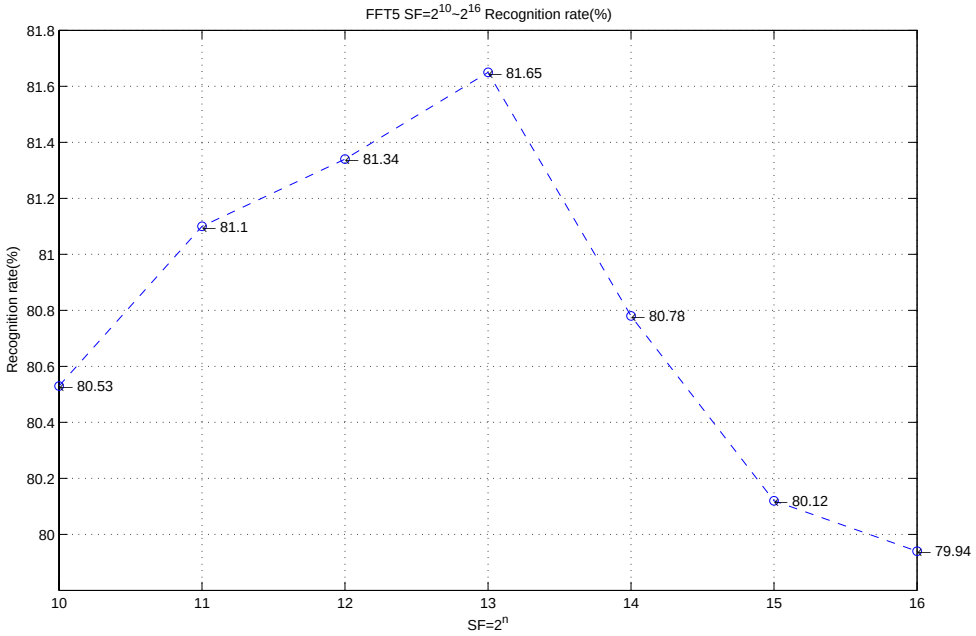

圖 4- 6 FFT5 SF= 2^{10} ~ 2^{16} 辨識率分析圖

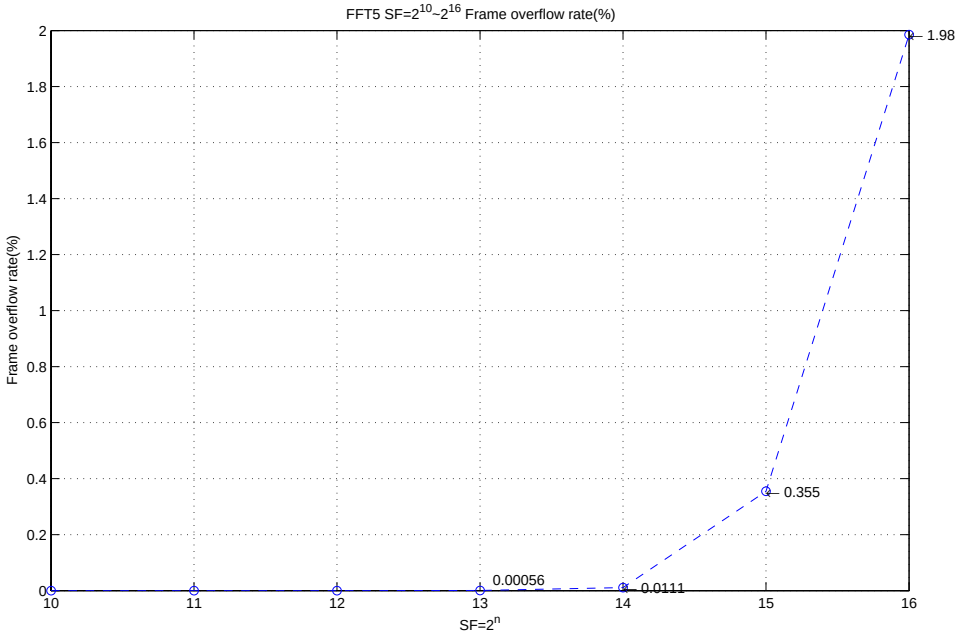

圖 4- 7 FFT5 SF= 2^{10} ~ 2^{16} 音框溢位率分析圖