組合式倒頻譜統計正規化法於強健性語音辨識之研究
Associative Cepstral Statistics Normalization Techniques for Robust Speech Recognition

杜文祥 Wen-hsiang Tu
暨南國際大學電機工程學系
Dept of Electrical Engineering, National Chi Nan University, Taiwan
aero3016@ms45.hinet.net

吳光杰 Kuang-chieh Wu
暨南國際大學電機工程學系
Dept of Electrical Engineering, National Chi Nan University, Taiwan
s95323529@ncnu.edu.tw

洪志偉 Jeih-weih Hung
暨南國際大學電機工程學系
Dept of Electrical Engineering, National Chi Nan University, Taiwan
jwhung@ncnu.edu.tw

摘要

一套自動語音辨識系統，在雜訊環境下其辨識效果通常會受到明顯影響，該如何有效地克服這樣的問題，一直以來都是此領域研究的重點，本論文即是針對此問題加以研究，而提出幾種改進技術。在過去的研究中，有一系列的改進技術，是藉由正規化語音特徵的統計特性來降低雜訊的影響，例如：倒頻譜平均消去法、倒頻譜平均值與變異數正規化法與統計圖等化法等，這些方法被證明皆有明顯的效能，可以有效提升語音特徵在雜訊環境下的強健性。本論文即是以這三種倒頻譜特徵參數正規化技術爲背景，發展一系列改進之強健性方法。

前面所提到的三種特徵參數正規化技術中所須用到的特徵統計值，通常是由整段的語句或片段的語句所包含的特徵求得，而在過去本實驗室的研究中，曾運用以碼簿(codebook)爲基礎的方式來求取這些統計值，發現相對於之前的作法能有明顯進步。在本論文第一部分，我們提出一改良式的碼簿建構程序，其中使用語音偵測(voice activity detection, VAD) 技術來分隔訊號中的語音成分與非語音成分，然後利用語音部分的特徵來建構碼簿，同時對所建立之碼簿中的每個碼字(codeword)賦予權重(weight)，此程序所建構的碼簿，經實驗證實，可以提升原始碼簿式(codebook-based)特徵參數正規化法的效能。而在第二部份，我們則是整合上述之碼簿式(codebook-based)與整段式(utterance-based)兩類方法所得到之特徵統計資訊，發展出所謂的組合式(associative)特徵參數正規化法。此類組合式的新方法相較於整段式與碼簿式的方法，能得到更好的效果，更有效地提升加成性雜訊環境下語音的辨識精確度。

Abstract

The noise robustness property for an automatic speech recognition system is one of the most important factors to determine its recognition accuracy under a noise-corrupted environment. Among the various approaches, normalizing the statistical quantities of speech features is a

very promising direction to create more noise-robust features. The related feature normalization approaches include cepsral mean subtraction (CMS), cepstral mean and variance normalization (CMVN), histogram equalization (HEQ), etc. In addition, the statistical quantities used in these techniques can be obtained in an utterance-wise manner or a codebook-wise manner. It has been shown that in most cases, the latter behaves better than the former.

In this paper, we mainly focus on two issues. First, we develop a new procedure for developing the pseudo-stereo codebook, which is used in the codebook-based feature normalization approaches. The resulting new codebook is shown to provide a better estimate for the features statistics in order to enhance the performance of the codebook-based approaches. Second, we propose a series of new feature normalization approaches, including associative CMS (A-CMS), associative CMVN (A-CMVN) and associative HEQ (A-HEQ). In these approaches, two sources of statistic information for the features, the one from the utterance and the other from the codebook, are properly integrated. Experimental results show that these new feature normalization approaches perform significantly better than the conventional utterance-based and codebook-based ones. As the result, the proposed methods in this paper effectively improve the noise robustness of speech features.

關鍵詞：自動語音辨識、碼簿、強健性語音特徵

Keywords: automatic speech recognition, codebook, robust speech feature

一、緒論

本論文所討論及提出的強健式技術，主要是在加成性雜訊環境下，對訓練與測試二者的語音特徵參數的統計特性加以正規化，以降低兩環境的不匹配。其中我們利用梅爾倒頻譜係數(mel-frequency cepstral coefficients, MFCC)做爲語音特徵，結合語音偵測技術(voice activity detection, VAD)[1]與特徵統計值正規化的諸多技術，來提升語音特徵在加成性雜訊環境下的強健性。本論文中所討論的特徵參數正規化法分別爲：

（一）整段式(utterance-based)特徵參數正規化法

即傳統的整段式倒頻譜平均消去法(utterance-based cepstral mean subtraction, U-CMS)[2]、整段式倒頻譜平均值與變異數正規化法(utterance-based cepstral mean and variance normalization, U-CMVN)[3]與整段式統計圖等化法(utterance-based histogram equalization, U-HEQ)[4]。此類方法是以一整段語句爲基準去估算每一維特徵參數的統計特性，並執行特徵參數正規化。

（二）碼簿式(codebook-based)特徵參數正規化法

此類方法是藉由碼簿來幫助我們估算出代表訓練語音特徵與測試語音特徵的統計值，藉此執行語音特徵正規化。在過去的研究裡[5][6][7]，發現此類的方法，包括碼簿式倒頻譜平均消去法(codebook-based cepstral mean subtraction, C-CMS)與碼簿式倒頻譜平均值與變異數正規化法(codebook-based cepstral mean and variance normalization, C-CMVN)等，其效果都比前一類之整段式特徵正規化法來的好。

本論文根據以上所述的二類方法提出一系列改進的技術，分述如下：

① 在過去碼簿式特徵正規化法中[5-7]，碼簿取得方式是將全部的訓練語料轉換的特徵參數作向量量化，這樣的方式可能會使其中許多碼字是對應到非語音的靜音(silence)

或雜訊成份，而使這些碼字較缺乏語音特徵的代表性，同時，每個碼字的權重被設爲相等，這樣可能會使之後所欲計算的特徵統計值較不精確。在本論文中，我們應用端點偵測(voice activity detection, VAD)技術偵測出一段訊號的語音(speech)成分與非語音(silence)成分，然後只使用語音成分的特徵去製作碼簿，同時，不同的碼字根據其對應的原始特徵數目多寡設定其權重(weight)，這種新的碼簿建構程序應可以改善上述之缺點，進而提升各種碼簿式特徵正規化法的效能。

② 我們提出了一新方法，稱爲組合式(associative)特徵正規化法，其主要程序是我們整合前述之碼簿式與整段式兩方所使用的特徵統計資訊，來計算特徵的統計值，藉此來執行特徵的正規化。實驗結果發現此類組合式的方法比碼簿式與整段式的兩類方法，能達到更佳的效果。可能原因在於，組合式的方法降低了碼簿式方法中只取每段訊號前幾個音框作爲純雜訊估測的不準確效應，而使所得的特徵統計值更爲精確。

在之後的第二章裡，我們將簡單介紹整段式(utterance-based)特徵正規化技術。第三章將說明新的虛擬雙通道碼簿的建立程序，藉此改進碼簿式(codebook-based)特徵正規化法的效能。在第四章中，我們敘述本論文所新提出的組合式(associative)特徵正規化法。第五章包含了本論文之實驗所使用之語料庫介紹與本論文所提到的各種特徵正規化技術之實驗結果與相關的討論分析。最後，第六章爲結論與未來展望。

二、整段式(utterance-based)特徵參數正規化技術

本章我們簡要介紹三種在強健性語音辨識中，常被應用的特徵參數正規化技術，分別爲整段式倒頻譜平均消去法(utterance-based cepstral mean subtraction, U-CMS)[2]、整段式倒頻譜平均值與變異數正規化法(utterance-based cepstral mean and variance normalization, U-CMVN)[3]與整段式倒頻譜統計圖等化法(utterance-based cepstral histogram equalization, U-HEQ)[4]。

（一）整段式倒頻譜平均消去法 (U-CMS)

倒頻譜平均消去法(CMS)的目的是希望一語音特徵序列中，每一維度的倒頻譜係數長時間平均值爲0。假設其值不爲0時，我們就將此視爲通道雜訊而加以扣除，此種方法對於降低通道雜訊效應是一種簡單且有用的技術，但是有時對於降低加成性雜訊上也有一定的效果。在多數的作法上，首先我們將整段語音每一維的倒頻譜係數取平均值，然後將每一維的係數減掉其平均值，如此即得到補償後之新特徵，此稱爲整段式倒頻譜平均消去法(utterance-based cepstral mean subtraction, U-CMS)。根據這樣的原則，我們假設 $\{X[n], n = 1,2,...,N\}$ 爲一段語音所擷取到的某一維倒頻譜特徵參數序列，在經過整段式倒頻譜平均消去法(U-CMS)處理後，得到新的經過補償的特徵參數序列 $\{X_{U-CMS}[n], n = 1,2,...,N\}$，其數學式如下所示：

$$X_{U-CMS}[n] = X[n] - \mu_X, \quad n = 1,2,...,N. \qquad \text{式(2.1)}$$

其中 $\mu_X = \frac{1}{N}\sum_{n=1}^{N} X[n]$，$N$ 爲整段語音的音框個數。

因此，在 U-CMS 法中，用以正規化的平均值 μ_X 是由原始整段的特徵序列所得。

（二）整段式倒頻譜平均值與變異數正規化法 (U-CMVN)

語音訊號在經過加成性雜訊的干擾之後，其倒頻譜之平均值和原本乾淨語音倒頻譜平均值之間通常會存在一偏移量(bias)，同時其變異數相對於乾淨語音倒頻譜參數的變異數而言則通常會有縮小的現象，如此便造成了訓練與測試特徵的不匹配，而降低辨識

效果。使用倒頻譜平均值與變異數正規化法(CMVN)的目的是把每一維的倒頻譜特徵參數之平均值正規化爲 0，並將其變異數正規化爲 1，如此便能降低上述的失真，以達到提升倒頻譜特徵參數的強健性。

在倒頻譜平均值與變異數正規化法(CMVN)的作法上，我們是先利用倒頻譜平均消去法(CMS)去作處理(使處理過後的每一維倒頻譜係數平均值爲0)，然後再將處理後的每一維倒頻譜係數除以其標準差，如此得到新的特徵序列。在U-CMVN(utterance-based cepstral mean and variance normalization)法中，假設$\{X[n], n=1,2,...,N\}$是一段語音的某一維倒頻譜特徵參數序列，在經過U-CMVN處理後，得到新的特徵參數$\{X_{U-CMVN}[n], n=1,2,...,N\}$，其數學式如下所示：

$$X_{U-CMVN}[n]=\frac{X[n]-\mu_X}{\sigma_X},\quad n=1,2,...,N. \qquad \text{式(2.2)}$$

其中 $$\mu_X=\frac{1}{N}\sum_{n=1}^{N}X[n],\ \sigma_X=\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(X[n]-\mu_X\right)^2}$$

因此，在U-CMVN中，所用的平均值μ_X與標準差σ_X皆由整段語音的特徵序列而得。

（三）整段式統計圖等化法(U-HEQ)

統計圖等化法(HEQ)的目的，是希望用以訓練與測試之語音特徵兩者能夠具有相同的統計分佈特性，藉由此匹配的轉換過程，降低測試特徵與訓練特徵之間由於雜訊影響所造成的不匹配情形。其作法是將測試語音特徵與訓練語音特徵的機率分佈同時逼近一參考機率分佈。在本論文中所使用的參考機率分佈爲一標準常態分佈。

根據上述，我們假設$\{X[n], n=1,2,...,N\}$爲一段語音某一維倒頻譜特徵參數序列；$F_X(x)$爲$X[n]$的機率分佈$\left(F_X(x)=P(X\le x)\right)$，它是由整段之特徵$\{X[n], n=1,2,...,N\}$求得；$F_N(x)$爲參考機率分佈。則整段式統計圖等化法(utterance-based histogram equalization, U-HEQ)的數學轉換式如下所示：

$$X_{U-HEQ}[n]=F_N^{-1}\left(F_X\left(X[n]\right)\right), \qquad \text{式(2.3)}$$

其中$X_{U-HEQ}[n]$即爲經過整段式統計圖等化法處理後的新特徵參數。

三、改良式碼簿式特徵參數正規化技術

運用所謂的虛擬雙通道碼簿(pseudo stereo codebooks)來估算乾淨語音與含雜訊語音之特徵統計特性，進而執行特徵參數正規化技術，能有效提升雜訊環境下語音辨識率。在過去研究中[5-7]所提出之倒頻譜統計補償法(cepstral statistics compensation)，是對含雜訊之語音倒頻譜係數做轉換，使得經過轉換後的語音倒頻譜特徵之統計值更相似於乾淨訓練語音倒頻譜的統計值，這種方式只針對雜訊語音特徵作倒頻譜正規化補償。而在本論文所提出之方式，則是同時針對乾淨語音與雜訊語音倒頻譜特徵參數作正規化處理。另外，在之前的倒頻譜統計補償法中，所用的每個碼字(codeword)是利用未處理的乾淨語音特徵訓練而得，且每個碼字的比重相同，而在我們改進的方法上，我們應用了語音偵測技術(voice activity detection, VAD)[1]處理乾淨語音訊號，將訊號中的語音區段與非語音區段區隔出來，然後利用純語音區段的語音特徵來訓練碼字，此外，這些碼字根據其涵蓋的特徵數目賦予不同的權重(weight)，因此，由這些碼字所計算的語音特徵統計值，應該更爲精確、更能代表語音特徵的特性。實驗證明，這樣的修正方式能帶來更好的辨識率。

在上一章，我們介紹了三種整段式 (utterance-based)特徵參數正規化技術，分別爲：U-CMS、U-CMVN與U-HEQ。在這裡，我們將利用新修正的碼簿建立方法，建立虛擬雙通道碼簿，執行一系列改良的碼簿式(codebook-based)特徵參數正規化技術。

（一）虛擬雙通道碼簿之建立方式

在原始的碼簿式特徵參數正規化法 [5-7] 中，碼簿之建立方式是將訓練語料庫裡所有的乾淨語音訊號，在轉換至梅爾倒頻譜特徵參數之過程中，保留下語音與雜訊具備線性相加特性的中介特徵參數(intermediate feature)，並且將這些乾淨語音之中介特徵參數訓練成一組碼簿(codebook)，此一乾淨語音碼簿，大致上可以代表乾淨語音在中介特徵參數的特性。在測試語音方面，對於每一句含雜訊的測試語音，假設其前端部分爲純雜訊，然後將這段純雜訊轉換至上述的中介特徵參數，由於乾淨語音與純雜訊在中介特徵參數域具有線性相加(linearly additive)的特性，因此將這些純雜訊的中介特徵參數直接線性相加於先前訓練好的乾淨語音的每個碼字上，便得到了代表雜訊語音(noisy speech)在中介特徵參數的碼簿。最後，將這兩組分別代表乾淨語音與雜訊語音在中介特徵參數域中的碼字轉換至倒頻譜域，所得的兩組倒頻譜特徵碼簿，就稱爲虛擬雙通道碼簿。

在本論文中所提出改良式的碼簿建立法，與原始的碼簿建立法的兩個不同點在於：
(1) 將訓練語料庫裡所有的乾淨語音訊號，先利用文獻[1]所提之語音偵測技術(voice activity detection, VAD)偵測出語音(speech)與靜音(silence)成分，然後只使用語音部分的中介特徵參數來訓練乾淨語音的碼簿。而在原始的方法裡，僅是使用未上述處理的乾淨語音訊號之中介特徵訓練碼簿。
(2) 不同的碼字根據其涵蓋的特徵量，指定不同的權重(weight)，亦即涵蓋較多量特徵的碼字，所佔的權重也就愈大，此意味著每個碼字的出現機率並不相同。這些權重可以用來幫助後續的特徵統計正規化法裡，估測更精準的特徵統計量。而在原始的方法裡，每個碼字未被賦予權重，其隱含了每個碼字的出現機率是均等(uniform)的。

以下，我們詳述此虛擬雙通道碼簿之建立過程：

我們先將語料庫中每一句乾淨語料，透過語音偵測技術[1]區隔出乾淨訓練語料中，屬於語音區段的部份，然後經由梅爾倒頻譜特徵參數(mel-frequency cepstral coefficients, MFCC)擷取流程的前半部，將此屬於語音區段的部份，轉換成一中介特徵向量(intermediate feature vector)的序列，此中介特徵爲梅爾濾波器之輸出值，也就是平緩化後之線性頻譜(linear spectrum)，這些由乾淨語料所得的中介特徵向量，透過向量量化(vector quantization, VQ)後，建立一組包含 M 個碼字的集合，以 $\{\tilde{x}[n] \mid 1 \le n \le M\}$ 來表示，同時，其對應的權重爲 $\{w_n \mid 1 \le n \le M\}$。這組在中介特徵參數域上的乾淨語音碼簿之所有碼字，再由 MFCC 擷取流程的後半部轉換至倒頻譜域，如下式所示：

$$x[n] = f\left(\tilde{x}[n]\right) \qquad \text{式(3.1)}$$

其中 $f(.)$ 代表轉換程序，因此，$\{x[n], w_n \mid 1 \le n \le M\}$ 爲轉換至倒頻譜的碼簿及權重值，這就是乾淨語音的倒頻譜碼簿及權重值。

在雜訊語音方面，我們藉由乾淨語音在中介特徵參數域上的碼字，來建立對應至該段含雜訊之測試語音的碼簿。我們將每一測試語音估測到的純雜訊，在中介特徵參數域（線性頻譜域）上用一組向量 $\{\tilde{n}[p] \mid 1 \le p \le P\}$ 來表示，由於乾淨語音與純雜訊在中介特徵參數域上具有線性相加的特性，因此雜訊語音的碼字可表示成下式：

$$\tilde{y}[m]\big|_{m=(n-1)P+p} = \{\tilde{x}[n] + \tilde{n}[p]\}, \qquad \text{式(3.2)}$$

最後，類似式(3.1)，我們將 $\tilde{y}[m]$ 經由 MFCC 擷取流程後半部轉換至倒頻譜域，如下式所示：

$$y[m] = f(\tilde{y}[m]), \qquad \text{式(3.3)}$$

此外，每個$y[m]$的權重值v_m則設定爲：

$$v_m\Big|_{m=(n-1)P+p}=\frac{w_n}{P}, \qquad \text{式(3.4)}$$

因此，$y[m]$之權重（即v_m）是其對應的乾淨語音碼字$x[n]$之權重w_n的$\frac{1}{P}$，其中P是純雜訊向量$\{\tilde{n}[p]\}$的個數。故$\{y[m],v_m \mid 1\le m\le MP\}$便是代表此句雜訊語音在倒頻譜域上的碼簿及權重值。$\{x[n],w_n\}$與$\{y[m],v_m\}$這兩組分別代表乾淨訓練語音與雜訊測試語音的碼字，我們稱之爲虛擬雙通道碼簿。所謂虛擬的意思，是因爲雜訊語音的碼簿並不是直接由雜訊語音得到，而是經由乾淨語音碼簿與純雜訊估算值所間接得到的。

（二）碼簿式特徵參數正規化技術

這一節中，我們將介紹碼簿式特徵參數正規化技術。在前面曾提到，此類正規化技術，是同時針對乾淨語音與雜訊語音倒頻譜特徵參數作處理。而在這裡的碼簿式特徵參數正規化技術，是藉由在前一節中描述的虛擬雙通道碼簿，來建立特徵之統計量，進而對特徵做正規化。這三種特徵參數正規化技術分別爲：倒頻譜平均消去法(CMS)、倒頻譜平均值與變異數正規化法(CMVN)、與倒頻譜統計圖等化法(HEQ)。對於CMS與CMVN而言，我們利用前一節所述之碼簿與權重$\{x[m],w_m\}$與$\{y[m],v_m\}$，計算出分別代表乾淨語音與雜訊語音特徵的近似統計值，如下式所示：

$$\mu_{X,i}\approx\sum_{n=1}^{N}w_n(x[n])_i,\quad \sigma_{X,i}^2\approx\sum_{n=1}^{N}w_n\left(x[n]\right)_i^2-\left(\mu_{X,i}\right)^2. \qquad \text{式(3.5)}$$

$$\mu_{Y,i}\approx\sum_{m=1}^{NP}v_m(y[m])_i,\quad \sigma_{Y,i}^2\approx\sum_{m=1}^{NP}v_m\left(y[m]\right)_i^2-\left(\mu_{Y,i}\right)^2. \qquad \text{式(3.6)}$$

其中$(u)_i$代表任意向量u之第i維，$\mu_{X,i}$與$\sigma_{X,i}^2$分別代表乾淨語音特徵向量x第i維的平均值與變異數；$\mu_{Y,i}$與$\sigma_{Y,i}^2$分別代表雜訊語音特徵向量y第i維的平均值與變異數，和之前文獻[5-7]中的方法明顯差異在於，此刻我們所用的統計值(平均值與變異數)是以加權平均(weighted average)的形式所測得，而非[5-7]中之均勻平均(uniform average)的形式。

碼簿式倒頻譜平均消去法(codebook-based cepstral mean subtraction, C-CMS)，是對倒頻譜特徵之平均值作正規化處理，其數學表示式如下：

$$(\overline{x})_i=(x)_i-\mu_{X,i},\quad (\overline{y})_i=(y)_i-\mu_{y,i}. \qquad \text{式(3.7)}$$

其中$\overline{x}$與$\overline{y}$分別爲乾淨語音特徵x與雜訊語音特徵y在經過 C-CMS 處理後的新特徵值。

而碼簿式倒頻譜平均值與變異數正規化法(codebook-based cepstral mean and variance normalization, C-CMVN)，是針對倒頻譜特徵之平均值與變異數做正規化處理，其數學表示式如下：

$$(\overline{x})_i=\frac{(x)_i-\mu_{X,i}}{\sigma_{X,i}},\quad (\overline{y})_i=\frac{(y)_i-\mu_{Y,i}}{\sigma_{Y,i}}. \qquad \text{式(3.8)}$$

其中$\overline{x}$與$\overline{y}$分別爲乾淨語音特徵x與雜訊語音特徵y經過 C-CMVN 處理後的新特徵值。

最後，碼簿式倒頻譜統計圖等化法(codebook-based cepsteral histogram equalization, C-HEQ)，其基本作法是利用$\{x[n],w_n\}$與$\{y[m],v_m\}$兩套碼簿分別計算出乾淨語音特徵與雜訊語音特徵之每一維之近似的機率分佈(probability distribution)，然後求一轉換函數，使二者之每一維特徵參數之機率分佈皆逼近於某一事先定義之參考機率分佈。具體

作法如下描述：

假設我們現在藉由碼簿$\{x[n], w_n\}$建立第i維乾淨語音特徵$(x)_i$的累積密度函數，由於碼簿本身意味著離散的形式，若我們假設第$(x)_i$本身對應之隨機變數爲X_i，則X_i的機率質量函數(probability mass function)可用下式表示：

$$P\left(X_i = \left(x[n]\right)_i\right) = w_n, \qquad \text{式(3.9)}$$

而X_i的機率密度函數(probability density function, pdf)，即可以下式表示：

$$f_{X_i}(x) = \sum_{n=1}^{M} w_n \delta\left(x - \left(x[n]\right)_i\right); \qquad \text{式(3.10)}$$

其中$\delta(\cdot)$爲單位脈衝(unit impulse)函數，故X_i之機率分佈，或稱爲累積機率密度函數(cumulative density function)，爲上式$f_{X_i}(x)$之積分，表示如下：

$$F_{X_i}(x) = P\left(X_i \le x\right) = \sum_{n=1}^{M} w_n u\left(x - \left(x[n]\right)_i\right); \qquad \text{式(3.11)}$$

其中$u(x)$爲單位步階函數(unit step function)，定義爲：

$$u(x) = \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases} \qquad \text{式(3.12)}$$

因此，第i維乾淨語音特徵$(x)_i$之機率分佈則可由式(3.11)的$F_{X_i}(x)$表示，同理，藉由碼簿$\{y[m], v_m\}$建立之第i維雜訊語音特徵$(y)_i$的機率分佈可由下式表示：

$$F_{Y_i}(y) = P\left(Y_i \le y\right) = \sum_{m=1}^{MP} v_m u\left(y - \left(y[m]\right)_i\right); \qquad \text{式(3.13)}$$

由上述作法得到$F_{X_i}(x)$與$F_{Y_i}(y)$之後，根據倒頻譜統計圖等化法(HEQ)的原理，我們利用下面兩式分別正規化第i維之訓練乾淨語音特徵$(x)_i$與測試雜訊語音特徵$(y)_i$：

$$(\bar{x})_i = F_N^{-1}\left(F_{X_i}\left((x)_i\right)\right), \qquad \text{式(3.14)}$$

$$(\bar{y})_i = F_N^{-1}\left(F_{Y_i}\left((y)_i\right)\right). \qquad \text{式(3.15)}$$

其中$F_N(\bullet)$爲一參考機率分佈(通常爲標準常態分佈)，$F_N^{-1}(\bullet)$爲$F_N(\bullet)$的反函數，$\bar{x}$與$\bar{y}$則爲經C-HEQ正規化後的新特徵值。

綜合以上所述，在過去的碼簿式特徵參數正規化技術中，所用的碼字是利用原始未分段之乾淨訊號特徵訓練而得，且每個碼字的比重皆相同，而在這裡所提出的改良式碼簿建立法上，我們應用語音偵測技術先將乾淨語音訊號中的語音區段與非語音區段區隔出來，然後利用語音區段的特徵訓練碼字。接著，根據不同的碼字所涵蓋的特徵數目賦予相對之權重(weight)，因此，這些碼字所計算出的語音特徵統計值或機率分佈，應當更爲精確而具代表性。在第四章的實驗結果中，將證明藉由此改良式碼簿所發展的碼簿式特徵參數正規化法，能獲得更好的辨識效果。

四、組合式特徵參數正規化技術

前一章提到，雖然碼簿式特徵參數正規化法之表現普遍比整段式的方法來的好，且

具備了即時運算的優點，但其可能的缺點在於純雜訊資訊不足，導致所得的雜訊語音碼簿不夠精準。因此，本章我們針對上述缺點，提出組合式的特徵參數正規化技術，簡單來說，我們在這些方法中，整合了之前所介紹的碼簿式與整段式方法所求取之特徵統計特性，希望得到更精確的統計值來執行各種特徵正規化法。這些方法，我們統稱爲組合式(associative)特徵參數正規化法。以下兩小節，我們便對組合式倒頻譜平均消去法(associative CMS, A-CMS)、組合式倒頻譜平均值與變異數正規化法(associative CMVN, A-CMVN)與組合式倒頻譜統計圖等化法(associative HEQ, A-HEQ)分別作介紹。

（一）組合式倒頻譜平均消去法(associative CMS, A-CMS)與組合式倒頻譜平均值與變異數正規化法(associative CMVN, A-CMVN)

這一節中將分別介紹 A-CMS 與 A-CMVN 兩種特徵參數正規化法。我們藉由一參數值α的調整，適當地整合碼簿與整段特徵之統計資訊，希望能達到較佳之辨識效果。就整段語句(utterance)的特徵而言，假設$X=\{X_1,X_2,...,X_N\}$爲一段訓練用或測試用語音在所擷取到的某一維倒頻譜特徵參數序列，則其整段式之特徵的平均值與變異數可由下兩式計算而得：

$$\mu_u=\frac{1}{N}\sum_{i=1}^{N}X_i, \qquad \text{式(4.1)}$$

$$\sigma_u^2=\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu_u)^2, \qquad \text{式(4.2)}$$

其中μ_u爲整段式之特徵平均值，σ_u^2爲整段式之特徵變異數，N爲整段語音的音框數。

而在碼簿上的特徵方面，假設$C=\{C_1,C_2,...,C_M\}$爲同一段語音對應到的各碼字(codewords)的某一維(與前一段所述之維值相同)之集合，則此段語音特徵之碼簿式的平均值與變異數可由下兩式計算而得：

$$\mu_c=\sum_{j=1}^{M}w_jC_j, \qquad \text{式(4.3)}$$

$$\sigma_c^2=\sum_{j=1}^{M}w_jC_j^2-\mu_c^2, \qquad \text{式(4.4)}$$

其中μ_c爲碼簿式之特徵平均值，$\sigma_c{}^2$爲碼簿式之特徵變異數，w_j爲每一碼字所對應到的權重，M爲碼字數目。

因此，組合式倒頻譜平均消去法(associative CMS, A-CMS)中，所使用的特徵參數之平均值μ_a，可由下式計算而得：

$$\mu_a=\alpha\cdot\mu_c+(1-\alpha)\cdot\mu_u, \qquad \text{式(4.5)}$$

其中μ_u與μ_c分別如式(4.1)與式(4.3)所示，而α爲一權重值，$0\le\alpha\le1$。
因此，A-CMS 處理後的新特徵參數，可表示爲：

$$\text{A-CMS:}\qquad \tilde{X}_i=X_i-\mu_a,\qquad 1\le i\le N. \qquad \text{式(4.6)}$$

而組合式倒頻譜平均值與變異數正規化法(associative CMVN, A-CMVN)中，所使用的特徵參數之平均值μ_a與變異數σ_a^2，可由下面兩式計算而得：

$$\mu_a=\alpha\cdot\mu_c+(1-\alpha)\cdot\mu_u, \qquad \text{式(4.7)}$$

$$\sigma_a^2 = \left(\alpha\left(\sigma_c^2 + \mu_c^2\right) + (1-\alpha)\left(\sigma_u^2 + \mu_u^2\right)\right) - \mu_a^2, \qquad \text{式(4.8)}$$

其中μ_u、μ_c、σ_u^2與σ_c^2分別如式(4.1)、式(4.3)、式(4.2)與式(4.4)所示，而α爲一權重值，$0 \le \alpha \le 1$。

A-CMVN 處理後的新特徵參數，可表示爲：

$$\text{A-CMVN:} \qquad \tilde{X}_i = \frac{X_i - \mu_a}{\sigma_a} \qquad \text{式(4.9)}$$

由式(4.5)、式(4.7)與式(4.8)可明顯看出，α的大小決定了組合式方法中，使用碼簿式統計量與整段式統計量的比例。當$\alpha = 1$時，A-CMS或A-CMVN即爲原始之碼簿式CMS(C-CMS)或碼簿式CMVN(C-CMVN)，相反地，當$\alpha = 0$時，A-CMS或A-CMVN即爲原始之整段式CMS(U-CMS)或整段式CMVN(U-CMVN)。

（二）組合式倒頻譜統計圖等化法(associative HEQ, A-HEQ)

在這一節中，我們將介紹組合式統計圖等化法(associative histogram equalization, A-HEQ)，類似之前的觀念，我們試著整合單一語句(utterance)特徵及其對應之碼字組合(codebook)兩方的統計資訊，然後建構出一代表此語句特徵的機率分佈$F_X(x) = P(X \le x)$，以作爲 HEQ 法等化特徵所用。以下，我們描述 A-HEQ 執行步驟：

假設某一待正規化的原語句之特定一維的特徵序列爲$\{X_1, X_2, ..., X_N\}$，其中N爲此序列之特徵總數，而其對應到之同一維的碼字，表示爲$\{C_1, C_2, ..., C_M\}$，權重爲$\{w_1, w_2, ..., w_M\}$，其中M爲碼字數目。首先，我們設定一參數β（$0 \le \beta \le \infty$），此參數代表了使用碼簿式資訊相對於使用整段式資訊的比例。接著，我們產生一組數目爲βN的新特徵$\{\tilde{C}_k\}$，此組新特徵是由碼字$\{C_m\}$根據其權重值$\{w_m\}$所建立的，新特徵$\{\tilde{C}_k\}$中有$[\beta N \times w_m]$個特徵的值和C_m完全相同，($[\beta N \times w_m]$代表$\beta N \times w_m$取四捨五入後的值)，換言之，新特徵$\{\tilde{C}_k\}$爲一組整合了權重值的新碼字，當原碼字C_m其權重值爲w_m時，它就會在新特徵$\{\tilde{C}_k\}$中出現$[\beta N \times w_m]$次，例如，假設原碼字集合爲$\{3,5,7\}$，對應之權重爲$\{0.2, 0.5, 0.3\}$，則當假設新特徵$\{\tilde{C}_k\}$的總數爲 20 時，$\{\tilde{C}_k\}$就包括了 4 個 3 $(20 \times 0.2 = 4)$，10 個 5 $(20 \times 0.5 = 10)$與 6 個 7 $(20 \times 0.3 = 6)$，因此，$\{\tilde{C}_k\}$即爲$\{\underbrace{3,3,3,3}_{4個}, \underbrace{5,5,5,...,5}_{10個}, \underbrace{7,7,7,...,7}_{6個}\}$（實際上，由於四捨五入的關係，最後得到的新特徵$\{\tilde{C}_k\}$其總數可能不會恰好是$\beta N$，即恰好爲原語句特徵數目$N$的$\beta$倍）。

接下來，我們就將原語句特徵$\{X_1, X_2, ..., X_N\}$與代表碼字的新特徵$\{\tilde{C}_1, \tilde{C}_2, ..., \tilde{C}_{\beta N}\}$串聯起來，共同決定一組代表此語句特徵的機率分佈：

$$F_X(x) = \frac{1}{(\beta+1)N}\left(\sum_{n=1}^{N} u(x - X_n) + \sum_{k=1}^{\beta N} u\left(x - \tilde{C}_k\right)\right), \qquad \text{式(4.10)}$$

最後，利用 HEQ 的原理，我們將原語句特徵正規化，如下式所示：

$$\text{A-HEQ:} \qquad \bar{x} = F_N^{-1}\left(F_X(x)\right) \qquad \text{式(4.11)}$$

其中F_N爲參考之機率分佈，x爲原始特徵參數(即前面提到的$\{X_1, X_2, ..., X_N\}$)，$\bar{x}$即爲 A-HEQ 法所得之新特徵參數。

由式(4.10)可看出，原語音特徵之機率分佈 $F_X(x)$ 由整句特徵 $\{X_n\}$ 與新碼字特徵 $\{\tilde{C}_k\}$ 共同決定，前者數目爲 N ，後者數目約爲 βN ，因此參數 β 大小決定了A-HEQ中，新碼字特徵 $\{\tilde{C}_k\}$ 對 $F_X(x)$ 的影響程度，當 $\beta = 0$ 時，相當於碼字方面的資訊完全被忽略，A-HEQ即變爲原先所介紹之整段式HEQ法(U-HEQ)，而當 β 很大($\beta \to \infty$)時，原先語句的特徵 $\{X_n\}$ 之資訊則幾乎被省略，則此時A-HEQ即趨近於原先所介紹之碼簿式HEQ (C-HEQ)。

在這一章中，我們介紹了組合式特徵正規化技術，這類技術同時整合了前兩章所介紹之整段式與碼簿式技術所用的特徵統計資訊，透過式(4.5)、式(4.7)、式(4.8)與式(4.10)中之參數 α 與 β 的調整，我們可以彈性地決定兩方所得之統計資訊的比例。在下一章的實驗結果，我們將看到這類組合式特徵正規化技術能帶來更好的語音辨識精確度。

五、辨識實驗結果與相關討論

本章開始是介紹在本論文上所使用的語音資料庫與系統效能的評估方式，而後的內容爲本論文所提及之各種強健性語音特徵參數技術之辨識實驗，其相關結果與討論。

（一）語音資料庫簡介

本論文使用的語音資料庫爲歐洲電信標準協會(European Telecommunication Standard Institute，ETSI)發行的AURORA 2語音資料庫[8]，內容是連續的英文數字字串，其中是以美國成年男女所錄製的乾淨環境連續數字語音，然後加上了八種不同的加成性雜訊與通道效應。這些加成性雜訊分別爲：地下鐵(subway)、人的嘈雜聲(babble)、汽車(car)、展覽會(exhibition)、餐廳(restaurant)、街道(street)、機場(airport)、火車站(train station)等環境雜訊共計八種，而通道效應有兩種，分別爲G712與MIRS[9]。

在AURORA 2資料庫裡有兩種不同的訓練環境及三種不同的測試環境，由於本論文只針對加成性雜訊做討論，因此在這裡，只使用到表一之一種訓練環境與兩種測試環境。

（二）實驗設定

本論文中所使用的特徵參數爲13維（第0維至第12維）的梅爾倒頻譜係數(mel-frequency cepstral coefficients, MFCC)，加上其一階和二階差量，總共爲39維的特徵參數。模型的訓練是使用隱藏式馬可夫模型工具(Hidden Markov Model Toolkit，HTK)[10]來訓練，產生11個數字模型(oh, zero, one, …., nine)與一個靜音模型，每個數字模型包含16個狀態，每個狀態包含20個高斯密度混合。

（三）各種強健性技術之辨識結果與討論

1. 改良之碼簿式特徵正規化法的辨識結果

在這一節中，我們將介紹本論文所提出之新的碼簿建立程序，分別應用於碼簿式倒頻譜平均消去法(C-CMS)、碼簿式倒頻譜平均值與變異數正規化法(C-CMVN)與碼簿式倒頻譜統計圖等化法(C-HEQ)的辨識結果。我們變動所運用的碼字數目 M ，分別設爲16、64與256，來觀測其效應。對於純雜訊的估測值 $\{\tilde{n}[p], 1 \le p \le P\}$ ，我們是以每一段測試語音的前10個音框作爲純雜訊音框的代表，即 $P = 10$ 。以下，表二、表三與表四分別爲新的碼簿建立程序所得之C-CMS、C-CMVN與C-HEQ在不同碼簿數 M 之下所得的平均辨識率（20dB、15dB、10dB、5dB與0dB五種訊雜比下的辨識率平均），AR與RR分別爲相較於基礎實驗結果之絕對錯誤降低率(absolute error rate reduction)和相對錯誤降低率(relative error rate reduction)。在這些表中加"＊"標記者(C-CMS*或C-CMVN*)，則爲原始碼簿建立程序[5-7]所對應之C-CMS或C-CMVN法，而U-CMS、U-CMVN與U-HEQ分別爲整段式CMS、CMVN與HEQ。附帶一提的是，由於原始碼簿特徵正規化法的文獻[5-7]裡，只提及C-CMS與C-CMVN，並未介紹C-HEQ，因此在表四中，我們

只將新的C-HEQ與整段式HEQ(U-HEQ)的效能作比較。

表一、實驗所用之Aurora-2語料庫相關資訊

<table>
<tr><td colspan="3">AURORA2 語音資料庫</td></tr>
<tr><td>取樣頻率</td><td colspan="2">8kHz</td></tr>
<tr><td>語音內容</td><td colspan="2">英文數字 0～9(zero, one, two, three, four, five, six, seven, eight, nine, oh)，共 11 個音。</td></tr>
<tr><td>語音長度</td><td colspan="2">每一段語音包含不超過七個的英文數字</td></tr>
<tr><td>訓練語料</td><td colspan="2">句數：8440 句
摺積性雜訊：G712 通道；加成性雜訊：無加成性雜訊</td></tr>
<tr><td rowspan="2">測試語料</td><td>A 組雜訊環境</td><td>B 組雜訊環境</td></tr>
<tr><td>句數：28028 句
摺積性雜訊：G712 通道
加成性雜訊：
地下鐵雜訊(subway)
人的嘈雜聲雜訊(babble)
汽車雜訊(car)
展覽館雜訊(exhibition)
雜訊強度(signal-to-noise ratio, SNR)：clean、20dB、15dB、10dB、5dB、0dB</td><td>句數：28028 句
摺積性雜訊：G712 通道
加成性雜訊：
餐廳雜訊(restaurant)
街道雜訊(street)
機場雜訊(airport)
火車站雜訊(train station)
雜訊強度(signal-to-noise ratio, SNR)：clean、20dB、15dB、10dB、5dB、0dB</td></tr>
</table>

從這三個表格的結果，我們可觀察到下列幾點：

①就 CMS 法而言，原始之 C-CMS(C-CMS*)相對於基礎實驗結果進步較小(如在 N =256 下，在 Set A 下提升了 6.00%，在 Set B 下提升了 7.41%)，其效果甚至比整段式 CMS(U-CMS)來的差，然而，我們所提出的新 C-CMS，則帶來顯著的進步(如在 N =256 下，在 Set A 下提升了 9.54%，在 Set B 下提升了 13.70%)，由此證實，我們所用的新的碼簿建構程序確實能有效提升 C-CMS 的效果，而且其效果並不會隨著碼字數目的大小，而有明顯的變化。其效果在 Set A 下優於 U-CMS，在 Set B 下則略遜於 U-CMS，這可能原因在於，C-CMS 使用一段語音前幾個音框作雜訊估測，這在 Set B 此非穩定(non-stationary)雜訊環境中是比較不精確的。

②就 CMVN 法而言，原始之 C-CMVN(即 C-CMVN*)相對於基礎實驗結果雖已有了不錯的辨識率提升(如在 M =256 下，在 Set A 下提升了 14.75%，在 Set B 下提升了 18.46%)，但是相較於整段式 CMVN(U-CMVN)而言，在 M =16 與 M =64 下，其效果都比 U-CMVN 還要差，然而，我們所提出之新的 C-CMVN，則有明顯的進步，無論在 M =16、 M =64 或 M =256 下，其效果都比原始的 C-CMVN 還要好，且幾乎都優於 U-CMVN(僅在 M =16 時，Set B 之平均辨識率略遜於 U-CMVN)，由此證實，我們所用的新的碼簿建構程序確實能有效提升 C-CMVN 的效果，而且其效果並不會隨著碼字的大小，而有明顯的變化。

③就 HEQ 法而言，C-HEQ 同樣也能有效提昇辨識率，但無論在 A 組雜訊環境下或

B 組雜訊環境下，其平均辨識率都比 U-HEQ 來得差，我們推測其原因可能在於，C-HEQ 在純雜訊的估測上，是以每一段測試語音的前幾個音框作爲純雜訊音框的代表，因而造成純雜訊資訊不足，導致所得的雜訊語音碼簿不夠精準，最終造成 C-HEQ 辨識率比 U-HEQ 還要差的結果。

表二、U-CMS、原始C-CMS(C-CMS*)、與新C-CMS的平均辨識率(%)

Method	Set A	Set B	average	AR	RR
Baseline	71.92	67.79	69.86	—	—
U-CMS	79.37	82.47	80.92	11.07	36.71
C-CMS*(*M*=16)	74.21	70.81	72.51	2.65	8.81
C-CMS*(*M* =64)	74.03	70.74	72.39	2.53	8.39
C-CMS*(*M* =256)	77.92	75.20	76.56	6.71	22.24
C-CMS(*M* =16)	79.04	79.56	79.30	9.45	31.33
C-CMS(*M* =64)	80.79	80.19	80.49	10.64	35.28
C-CMS(*M*=256)	81.46	81.49	81.48	11.62	38.55

表三、U-CMVN、原始C-CMVN(C-CMVN*)、與新C-CMVN的平均辨識率

Method	Set A	Set B	average	AR	RR
Baseline	71.92	67.79	69.86	—	—
U-CMVN	85.03	85.56	85.30	15.44	51.22
C-CMVN*(*M* =16)	84.44	82.40	83.42	13.57	45.00
C-CMVN*(*M*=64)	84.13	81.53	82.83	12.98	43.04
C-CMVN*(*M*=256)	86.67	86.25	86.46	16.61	55.08
C-CMVN(*M*=16)	85.41	85.21	85.31	15.46	51.27
C-CMVN(*M*=64)	86.92	86.81	86.87	17.01	56.43
C-CMVN(*M*=256)	87.10	87.32	87.21	17.36	57.57

表四、U-HEQ與新C-HEQ的平均辨識率

Method	Set A	Set B	average	AR	RR
Baseline	71.92	67.79	69.86	—	—
U-HEQ	87.00	88.33	87.67	17.81	59.08
C- HEQ(*M*=16)	84.03	84.46	84.25	14.39	47.74
C- HEQ(*M*=64)	86.32	85.90	86.11	16.26	53.92
C-HEQ(*M*=256)	86.22	86.07	86.15	16.29	54.04

2.組合式特徵參數正規化法之辨識結果

在這一節中，我們將介紹本論文所提出之組合式(associative)特徵參數正規化技術之辨識結果，這三種技術分別爲組合式倒頻譜平均消去法(associative CMS, A-CMS)、組合式倒頻譜平均值與變異數正規化法(associative CMVN, A-CMVN)與組合式統計圖等化法(associative histogram equalization, A-HEQ)。在 A-CMS、A-CMVN 與 A-HEQ 三種正

規化技術中，由於在不同的碼字數目 N 下，產生最佳辨識率的 α 值(如式(4.5)、(4-7)與(4.8)中所示)或 β 值(如式(4.10)中所示)不盡相同，因此在以下的實驗辨識結果中，我們只呈現在不同的 N 值時，所產生最佳平均辨識率之 α 值或 β 值之結果。

首先，表五爲 A-CMS 在碼字數目 N 分別爲 16、64 與 256 下，所得到的最佳辨識結果，爲了比較起見，我們也將表二中的基本實驗、C-CMS(M =256)與 U-CMS 的平均辨識率列在表中。從此表中，我們可以觀察到以下幾種情形：

①組合式倒頻譜平均消去法(A-CMS)相較於基本實驗而言，無論在碼字數 M =16、64 與 256 下，其平均辨識率皆有大幅的進步，三者在 A 組雜訊環境下分別有 11.86%、11.30%與 10.98%的辨識率提升，在 B 組雜訊環境下分別有 17.76%、16.82%與 16.83%的辨識率提升，由此可看出 A-CMS 具有不錯之特徵強健化效果。

② A-CMS 在各種不同的碼字數 N 之下，其平均辨識率皆比 C-CMS 與 U-CMS 來得好，其中在 N =16 時能有最佳的效果，在 A 組雜訊環境與 B 組雜訊環境下之平均辨識率分別爲 83.78%與 85.55%，相較於 C-CMS 取 M =256 所得之最佳辨識率，A-CMS 在 A 組雜訊環境與 B 組雜訊環境下分別進步了 2.32%和 4.06%，這些進步都顯示了 A-CMS 優於 C-CMS。最後相較於 U-CMS，A-CMS 在 A 組雜訊環境與 B 組雜訊環境下其辨識率分別可以提升 4.41%和 3.08%。因此由實驗數據中可以證明，相對於 C-CMS 與 U-CMS 而言，A-CMS 都可以得到較好的辨識結果，這可能是因爲 A-CMS 同時整合了 C-CMS 與 U-CMS 所用的統計資訊，所以它更能有效改善語音在雜訊下的強健性。

表五、U-CMS、新C-CMS與A-CMS的平均辨識率

Method	Set A	Set B	average	AR	RR
Baseline	71.92	67.79	69.86	—	—
U-CMS	79.37	82.47	80.92	11.07	36.71
C-CMS(M=256)	81.46	81.49	81.48	11.62	38.55
A-CMS(M=16, α=0.5)	83.78	85.55	84.67	14.81	49.13
A-CMS(M=64, α=0.6)	83.22	84.61	83.92	14.06	46.64
A-CMS(M=256, α=0.6)	82.90	84.62	83.76	13.91	46.13

接著，表六爲A-CMVN在碼字數目 M 分別爲16、64與256下，所得到的最佳辨識結果，在表中，我們也列出原表三中的基本實驗、C-CMVN(M =256)與U-CMVN的平均辨識率以供比較。從此表中，我們可以觀察到以下幾種情形：

①組合式倒頻譜平均值與變異數正規化法(A-CMVN)在碼字數目 M =16、64 與 256 下，相較於基本實驗而言，其平均辨識率皆有大幅的改進，這三種 A-CMVN 在 A 組雜訊環境下分別有 16.19%、16.08%與 15.43%的辨識率提升，在 B 組雜訊環境下分別有 21.18%、20.77%與 20.26%的辨識率提升，由此可以發現 A-CMVN 確實能降低加成性雜訊對語音特徵的干擾，而提升辨識精確度。

② A-CMVN在各種碼字數N的情形下，其平均辨識率皆比C-CMVN、U-CMVN來得好，其中以 N =16時表現爲最佳，在A組雜訊環境與B組雜訊環境下之平均辨識率分別爲88.11%和88.97%，相較於C-CMVN取 M =256所得之最佳辨識率，A-CMVN在A組雜訊環境與B組雜訊環境則分別進步了1.01%與1.65%，這些進步都顯示了A-CMVN優於C-CMVN；而跟U-CMVN比較時，A-CMVN在A組雜訊環境與B組雜訊環境下，其辨識率分別可以提升3.08%和3.41%，其相對改善率分別爲20.55%與23.62%。類似之前的

A-CMS，A-CMVN同時整合了C-CMVN與U-CMVN所用的統計資訊，因此我們預期它具備了更佳的語音特徵強健化的效果，實驗數據也確實驗證了A-CMVN的表現明顯優於C-CMVN與U-CMVN。

表六、U-CMVN、新C-CMVN與A-CMVN的平均辨識率

Method	Set A	Set B	Average	AR	RR
Baseline	71.92	67.79	69.86	—	—
U-CMVN	85.03	85.56	85.30	15.44	51.22
C-CMVN(M=256)	87.10	87.32	87.21	17.36	57.57
A-CMVN(M=16, α=0.7)	88.11	88.97	88.54	18.69	61.98
A-CMVN(M=64, α=0.8)	88.00	88.56	88.28	18.43	61.12
A-CMVN(M=256, α=0.8)	87.35	88.05	87.70	17.85	59.20

最後，表七爲A-HEQ在碼字數目 M 分別爲16、64與256下，所得到的最佳辨識結果，爲了比較起見，我們也將表四中的基本實驗、C-HEQ(M=256)與U-HEQ的平均辨識率列在表中。從此表中，我們可以觀察到以下幾種情形：

①對於組合式統計圖等化法(A-HEQ)而言，無論在碼字數 M =16、64與256下，其平均辨識率相較於基本實驗而言，都有大幅的改進，三者在A組雜訊環境下分別有18.15%、17.28%與15.76%的辨識率提升，在B組雜訊環境下分別有23.08%、22.36%與21.10%的辨識率提升，顯示了A-HEQ在語音特徵強健性的效能，且相較於之前所述的兩種組合式特徵正規化法A-CMS與A-CMVN，A-HEQ的表現更爲優異。

②A-HEQ在各種碼字數M的情形下，其平均辨識率皆比C-HEQ與U-HEQ來得好，其中以M=16所得的平均辨識率爲最佳，在A組雜訊環境與B組雜訊環境下之辨識率分別爲90.07%和90.87%，相較於C-HEQ取 M =256所得之最佳辨識率，A-HEQ在A組雜訊環境與B組雜訊環境下其辨識率則分別進步了3.85%與4.80%，這些進步都顯示了A-HEQ優於C-HEQ；而跟U-HEQ比較時，A-HEQ在A組雜訊環境與B組雜訊環境下其辨識率分別提升了3.07%與2.54%，其相對改善率分別爲23.62%與21.76%。類似之前的結果，這裡我們再次驗證了組合式的方法優於碼簿式與整段式的方法，即A-HEQ比C-HEQ與U-HEQ更能提升雜訊環境下語音辨識的精確度。

表七、U-HEQ、新C-HEQ與A-HEQ的平均辨識率

Method	Set A	Set B	Average	AR	RR
Baseline	71.92	67.79	69.86	—	—
U-HEQ	87.00	88.33	87.67	17.81	59.08
C-HEQ(M=256)	86.22	86.07	86.15	16.29	54.04
A-HEQ(M=16, β=0.9)	90.07	90.87	90.47	20.62	68.39
A-HEQ(M=64, β=0.9)	89.20	90.15	89.68	19.82	65.75
A-HEQ(M=256, β=1)	87.68	88.89	88.29	18.43	61.14

六、結論與未來展望

在本論文中，我們主要討論的特徵參數正規化技術，分別爲倒頻譜平均消去法(CMS)、倒頻譜平均值與變異數正規化法(CMVN)與倒頻譜統計圖等化法(HEQ)，這三種技術皆須使用到特徵的統計量。傳統上，這些統計量是經由一整段的語音特徵估測而得。因此，其對應的技術，我們統稱爲整段式(utterance-based)特徵參數正規化技術。在

近年來，本實驗室發展了碼簿式(codebook-based)特徵參數正規化技術，分別爲C-CMS與C-CMVN。顧名思義，在這些方法中，所使用的特徵統計量是由碼簿計算而得，實驗證實這些碼簿式特徵參數正規化技術其表現大致上皆優於整段式特徵參數正規化技術。然而我們發現，它們仍然有進一步的改善空間。因此，本論文中我們提出了一套改良式的碼簿建立程序，相對於原程序的不同之處，在於我們應用了語音偵測技術處理乾淨語音訊號，然後利用純語音區段的語音特徵來訓練碼字；此外，這些碼字根據其涵蓋的特徵數目賦予不同的權重(weight)，此改良法在第三章有詳細的說明。

除了提出上述改良式的碼簿建立程序之外，本論文另一重點在於，我們提出了一系列組合式(associative)特徵參數正規化技術，分別爲A-CMS、A-CMVN與A-HEQ，這些技術中，我們整合了整段式技術與碼簿式技術所用的特徵統計資訊，用此整合後之統計量來執行CMS，CMVN或HEQ，其詳述於第四章中，這樣的技術可以有效地補償碼簿式技術中，純雜訊資訊不足的缺點，第五章中的實驗結果證實，組合式的特徵參數正規化技術比整段式與碼簿式特徵參數正規化技術，均能更明顯地提升辨識精確度。

雖然組合式特徵參數正規化技術效果十分顯著，但其最佳表現有賴於某些自由參數(即式(4.5)、式(4.8)中的α及式(4.10)中的β)的手動調整來整合碼簿式與整段式之統計資訊，因此在未來的發展上，我們希望能自動地求取出最佳的α與β等參數值，來對兩方的統計資訊作更精確的整合，同時，在建構雜訊語音碼簿的程序上，我們也希望能參考許多雜訊估測的方法，更精確測得一段語音中純雜訊的統計特性，期待更有效地提升碼簿式特徵參數正規化技術的效能。

參考文獻

[1] Chung-fu Tai and Jeih-weih Hung, "Silence Energy Normalization for Robust Speech Recognition in Additive Noise Environments", 2006 International Conference on Spoken Language Processing (Interspeech 2006—ICSLP)

[2] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification", IEEE Trans. on Acoustics, Speech and Signal Processing, 1981

[3] S. Tiberewala and H. Hermansky, "Multiband and Adaptation Approaches to Robust Speech Recognition", 1997 European Conference on Speech Communication and Technology (Eurospeech 1997)

[4] A. Torre, J. Segura, C. Benitez, A. M. Peinado, and A. J. Rubio, "Non-Linear Transformations of the Feature Space for Robust Speech Recognition", 2002 International Conference on Acoustics, Speech and Signal Processing (ICASSP 2002)

[5] Tsung-hsueh Hsieh, "Feature Statistics Compensation for Robust Speech Recognition in Additive Noise Environments" , M.S. thesis, National Chi Nan University, Taiwan, 2007

[6] Tsung-hsueh Hsieh and Jeih-weih Hung, "Speech Feature Compensation Based on Pseudo Stereo Codebooks for Robust Speech Recognition in Additive Noise Environments", 2007 European Conference on Speech Communication and Technology (Interspeech 2007—Eurospeech)

[7] Jeih-weih Hung, "Cepstral Statistics Compensation and Normalization Using Online Pseudo Stereo Codebooks for Robust Speech Recognition in Additive Noise Environments", IEICE Transactions on Information and Systems, 2008

[8] H. G. Hirsch and D. Pearce, "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions," Proceedings of ISCA IIWR ASR2000, Paris, France, 2000

[9] ITU recommendation G.712, "Transmission Performance Characteristics of Pulse Code Modulation Channels," Nov. 1996

[10] http://htk.eng.cam.ac.uk/