Exploring the Use of Speech Features and Their Corresponding Distribution Characteristics for Robust Speech Recognition

Author : Shin-Hsiang Lin, Berlin Chen,
Yao-Ming Yeh

Professor:陳嘉平
Reporter :吳柏鋒

Abstract—**The performance of current automatic speech recognition (ASR) systems often deteriorates radically when the input speech is corrupted by various kinds of noise sources. Several methods have been proposed to improve ASR robustness over the last few decades. The related literature can be generally classified into two categories according to whether the methods are directly based on the feature domain or consider some specific statistical feature characteristics. In this paper, we present a polynomial regression approach that has the merit of directly characterizing the relationship between speech features and their corresponding distribution characteristics to compensate for noise interference. The proposed approach and a variant were thoroughly investigated and compared with a few existing noise robustness approaches. All experiments were conducted using the Aurora-2 database and task. The results show that our approaches achieve considerable word error rate reductions over the baseline system and are comparable to most of the conventional robustness approaches discussed in this paper.**

Index Terms—**Clustering, histogram equalization, polynomial regression, robustness, speech recognition.**

摘要

- 簡介

- 群集式為基礎之多項式擬合統計圖法
  Cluster-based polynomial-fit histogram(CPHEQ)

- 多項式擬合統計圖等化法
  Polynomial-fit histogram Equalization(PHEQ)

- 實驗

簡介

- 現今的ASR系統往往因為輸入的語音遭受到噪音的破壞，而造成效能降低

- 廣泛來說，改善語音強健性的方法可分為兩大類:

  (1)直接以特徵域為基礎

  (2)考慮一些特定的統計特徵特性

簡介

- 提出群集式為基礎之多項式擬合統計圖法(CPHEQ)方法
- 它使用語音特徵以及其相對應的分布特性來對語音作補償

群集式為基礎之多項式擬合統計圖法 (CPHEQ)

- 在CPHEQ的背後主要概念源自於兩大方向:
  - 立體聲源為基礎分段線性補償
    Stereo-based piecewise linear compensation for environments(SPLICE)
  - HEQ

- SPLICE 主要目的是使用GMM來將噪音特徵空間特性化，而每個高斯組件皆表示一個特定的失真條件

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 理論上當高斯組件數量無限增加時，SPLICE可以處理線性或非線性失真

- 為避免此缺點，加入了HEQ的概念

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- HEQ是使用非線性轉換函式來補償非線性失真

- 所以在此提出使用CPHEQ，其結合SPLICE和HEQ兩者的來克服各自的缺點

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 一開始先利用語音資料來訓練GMM

- GMM 表示法如下：

$$p(Y_t) = \sum_{k=1}^{K} p(k) p(Y_t | k) = \sum_{k=1}^{K} p(k) N(Y_t; \mu_k, \Sigma_k)$$

  - K ：GMM混合數
  - Y_t ：噪音特徵向量

  μ_k, Σ_k ：平均向量，對角共變矩陣

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 假設已補償的特徵向量 $\tilde{X}_t$ ：

$$\tilde{X}_t = E\left[X_t \mid Y_t\right] = E\left[E\left[X_t \mid Y_t, k\right]\right] = \sum_{k=1}^{K} p\left(k \mid Y_t\right) E\left[X_t \mid Y_t, k\right]$$

  - $p\left(k \mid Y_t\right)$ 機率

$$p\left(k \mid Y_t\right) = \frac{p\left(Y_t \mid k\right) p\left(k\right)}{\sum_{k'=1}^{K} p\left(Y_t \mid k'\right) p\left(k'\right)}$$

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 在特徵域裡，可以合理假設 Y_t 的特徵向量組件 y_t 彼此間都是獨立的

- 由給定的第k個混合模型得到的 y_t 重新儲存值定義如下:

$$\tilde{x}_{t,k} = E\left[x_t \mid Y_t, k\right] \approx E\left[x_t \mid y_t, k\right]$$

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 提出一個源自於HEQ的想法來近似條件期望值:

$$\tilde{x}_{t,k} \approx G_k\left(CDF\left(y_t\right)\right)$$

  - $CDF\left(y_t\right)$是特徵組件 y_t 的CDF值
  - $G_k\left(\bullet\right)$個表示CDF值對應到事先定義的特徵值的反函式

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 一特定維度的特徵向量組件序列 $y_1, \cdots, y_t, \cdots, y_L$ 可以經由以下兩步驟計算得到:

  step1:根據特徵向量組件將序列依遞增排序

  step2:給定特徵向量組件的CDF值為

$$CDF(y_t) \approx \frac{S_{pos}(y_t) - 0.5}{L}$$

其中 $S_{pos}(y_t)$ 是一個回傳 y_t 階層之函式

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 使用多項式函式來近似CDF反函式達到效率與儲存空間的需求
- 在訓練階段，$G_k\left(CDF\left(y_t\right)\right)=\sum_{m=0}^{M} a_{km}\left(CDF\left(y_t\right)\right)^m$

  多項式函式的係數 a_{km} 可以經由MMSE估算得到MMSE定義如下：

$$E_k^2=\sum_{t=1}^{T-1}\left(p\left(k\,|\,Y_t\right)\times\left(x_t-\sum_{m=0}^{M} a_{km}\left(CDF\left(y_t\right)\right)^m\right)\right)^2$$

其中 T:訓練資料中音框總數

y_t，x_t :分別為噪音和乾淨部分的特徵向量組件

群集式為基礎之多項式擬合統計圖法（CPHEQ）

- 在測試階段，重新儲存值可以由下列式子獲得:

$$\tilde{x}_t = \sum_{k=1}^{K}\left(p\left(k \mid Y_t\right)\times\left(\sum_{m=0}^{M} a_{km}\left(CDF\left(y_t\right)\right)^m\right)\right)$$

- 為了減少運算時間，使用最大事後機率準則(MAP)

$$E_k^2 = \sum_{t=1}^{T-1}\left(\delta\left(k \mid Y_t\right)\times\left(x_t - \sum_{m=0}^{M} a_{km}\left(CDF\left(y_t\right)\right)^m\right)\right)^2$$

$$\tilde{x}_t = \sum_{k=1}^{K}\left(\delta\left(k \mid Y_t\right)\times\left(\sum_{m=0}^{M} a_{km}\left(CDF\left(y_t\right)\right)^m\right)\right)$$

$$\delta\left(k \mid Y_t\right) = \begin{cases} 1, & if \quad k = \arg\max_{k'} p\left(k' \mid Y_t\right) \\ 0, & otherwise. \end{cases}$$

多項式擬合統計圖等化法（PHEQ）

- 我們提出一個CPHEQ的變形，為多項式統計等化法polynomial histogram equalization(PHEQ)
- 利用唯一全域轉換來獲得噪音特徵向量組件y_t 的重新儲存值$\tilde{x}_t$

$$\tilde{x}_t = G_k\left(CDF\left(y_t\right)\right) = \sum_{m=0}^{M} a_{km}\left(CDF\left(y_t\right)\right)^m$$

$$E_k^2 = \sum_{t=1}^{T-1}\left(x_t - \sum_{m=0}^{M} a_{km}\left(CDF\left(x_t\right)\right)^m\right)^2$$

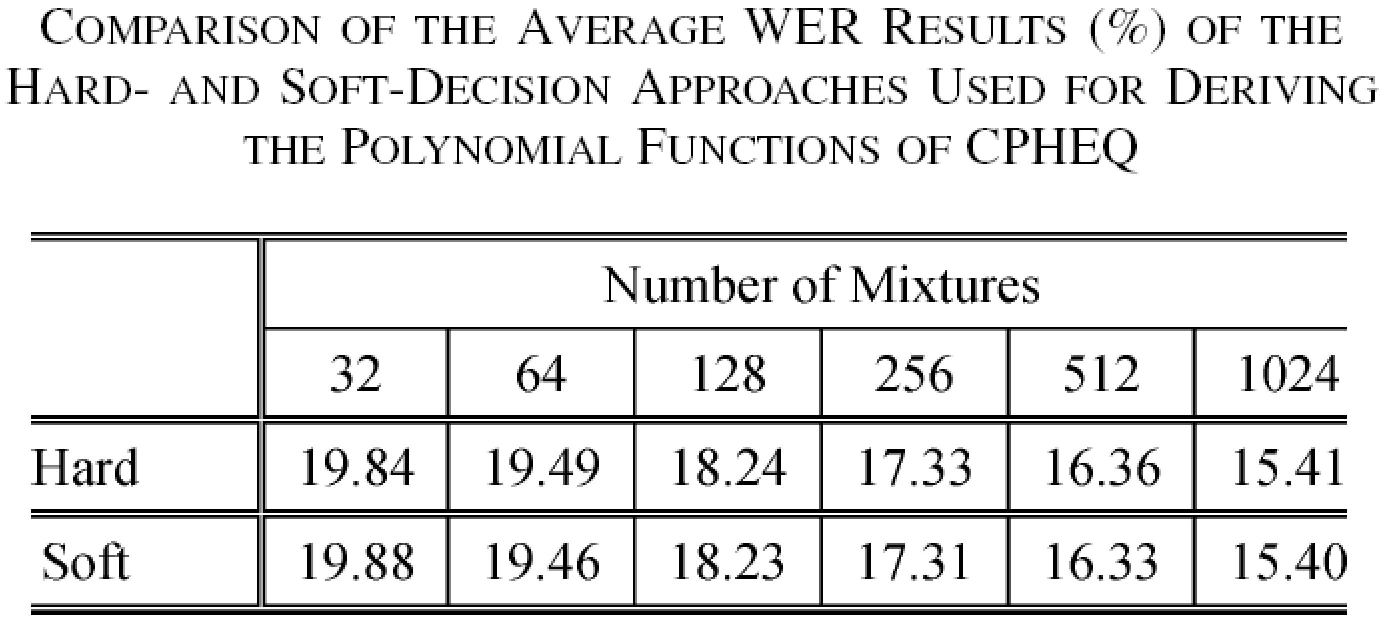
實驗

- 使用AURORA2語料庫
- 以連續可變長度英文數字串在靠近麥克風下錄音
- 人工加入了8種不同的噪音到測試集A和B
- 兩種訓練方案:乾淨訓練和多重環境訓練

實驗

- 一開始我們先計算在不同準則下獲得的多項式在CPHEQ的效能
- 多重環境訓練集
- 在測試集A中噪音型態SNR範圍5~20dB
- GMM數設定32~1024，多項式集階層設3

CPHEQ實驗



Comparison of the Average WER Results (%) of the Hard- and Soft-Decision Approaches Used for Deriving the Polynomial Functions of CPHEQ

	Number of Mixtures					
	32	64	128	256	512	1024
Hard	19.84	19.49	18.24	17.33	16.36	15.41
Soft	19.88	19.46	18.23	17.31	16.33	15.40

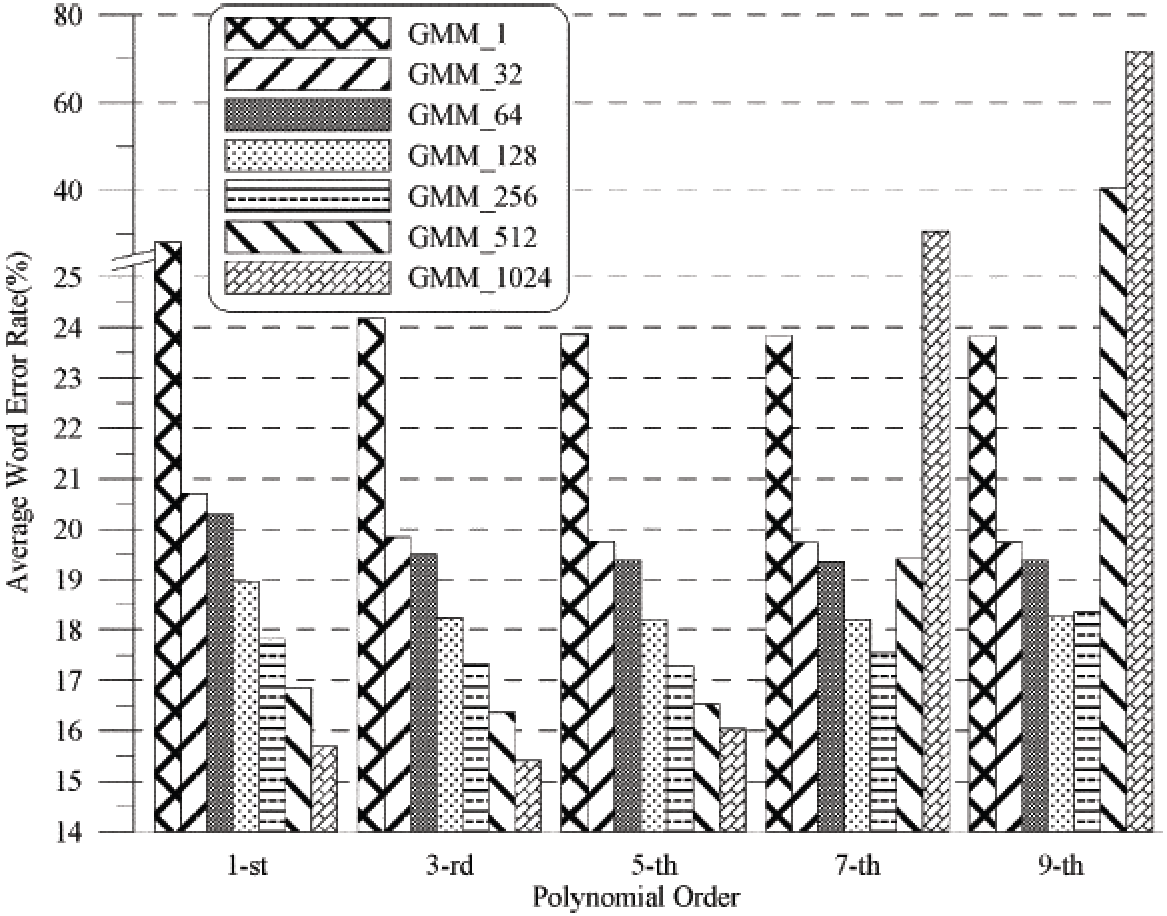


Fig. 1. Average WER results (%) of CPHEQ with respect to different numbers of mixtures and different orders of polynomial functions.

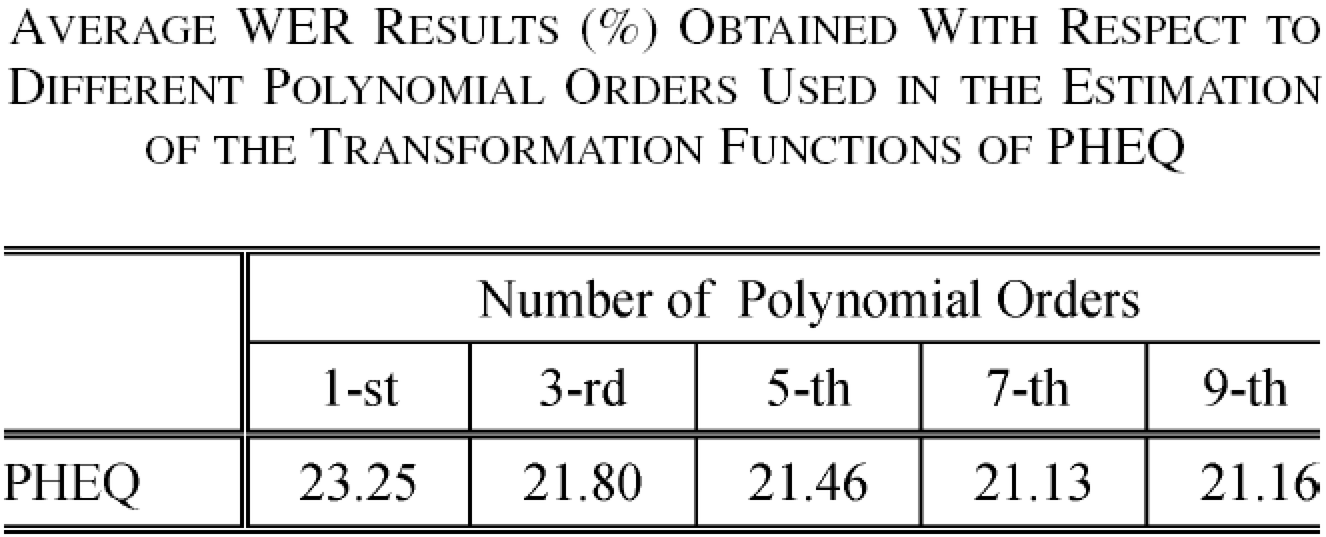
CPHEQ實驗

- 往往在階層太大時候效能會降低

- 在此研究上，產生這些現象可以解釋的原因就是使用**一組有限的訓練資料**

- 使用高階多項式函式可能導致精確擬合值間的振盪

PHEQ實驗



AVERAGE WER RESULTS (%) OBTAINED WITH RESPECT TO DIFFERENT POLYNOMIAL ORDERS USED IN THE ESTIMATION OF THE TRANSFORMATION FUNCTIONS OF PHEQ

	Number of Polynomial Orders				
	1-st	3-rd	5-th	7-th	9-th
PHEQ	23.25	21.80	21.46	21.13	21.16

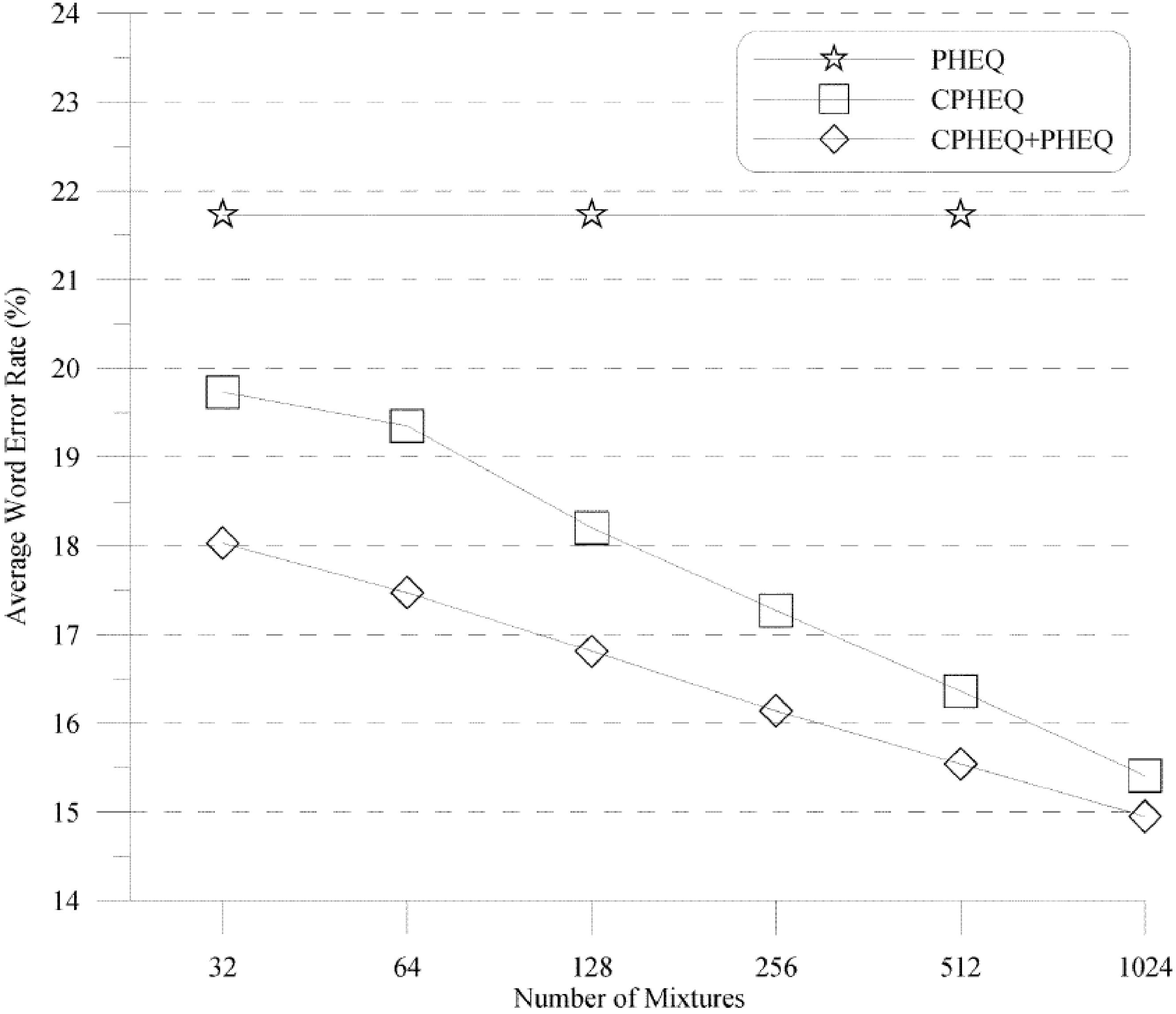
PHEQ實驗

- 數量較小的混合數較不足以來表示噪音特性

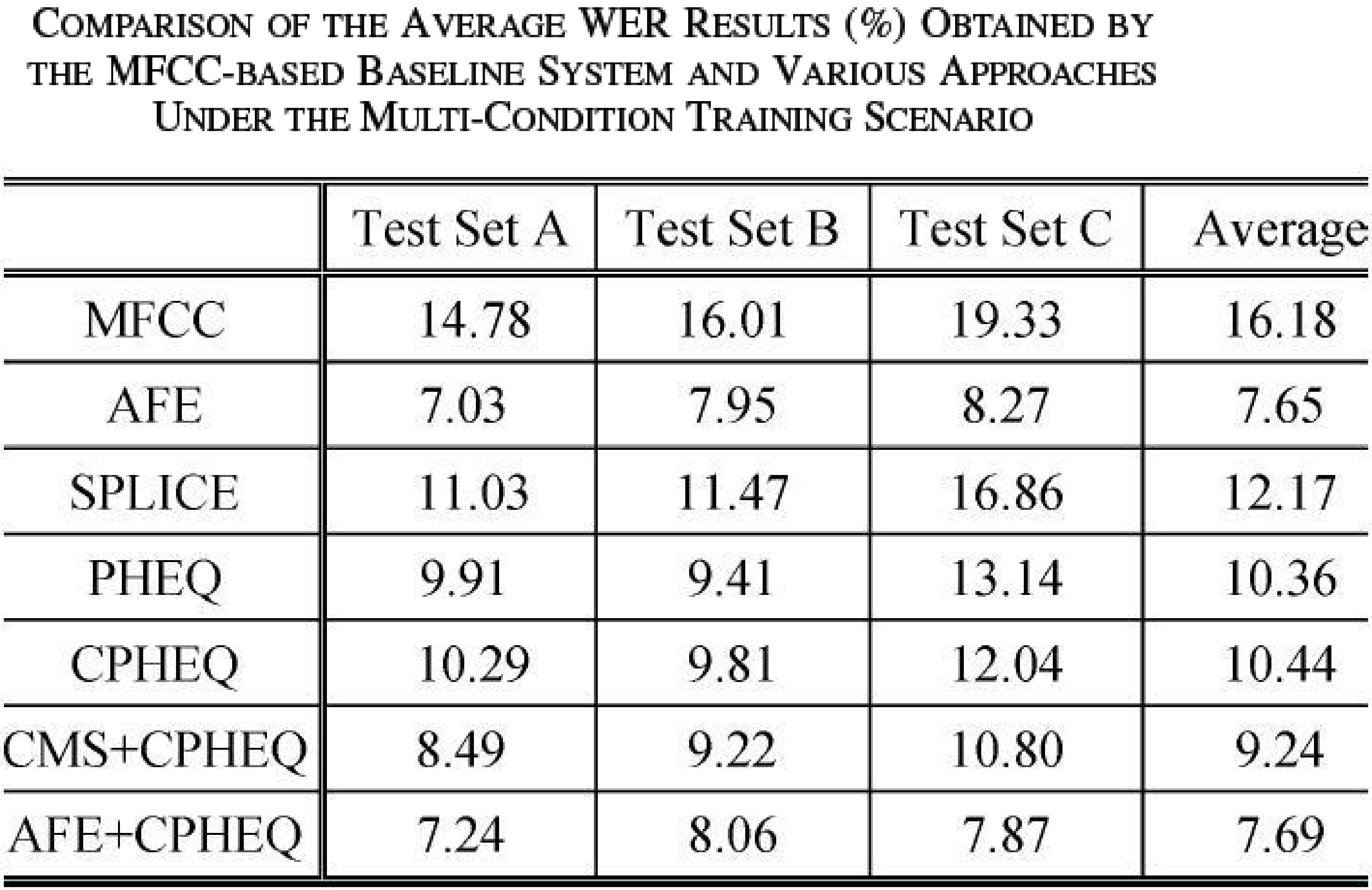
- 透過簡單的線性內插結合CPHEQ和PHEQ兩種方法來解決上之問題

- 初步設定內插權重為0.5且在測試階段是固定的

PHEQ實驗



24
23
22
21
20
19
18
17
16
15
14
Average Word Error Rate (%)
☆ PHEQ
□ CPHEQ
◇ CPHEQ+PHEQ
32
64
128
256
512
1024
Number of Mixtures



Comparison of the Average WER Results (%) Obtained by the MFCC-based Baseline System and Various Approaches Under the Multi-Condition Training Scenario

	Test Set A	Test Set B	Test Set C	Average
MFCC	14.78	16.01	19.33	16.18
AFE	7.03	7.95	8.27	7.65
SPLICE	11.03	11.47	16.86	12.17
PHEQ	9.91	9.41	13.14	10.36
CPHEQ	10.29	9.81	12.04	10.44
CMS+CPHEQ	8.49	9.22	10.80	9.24
AFE+CPHEQ	7.24	8.06	7.87	7.69