Robust Feature Extraction for Continuous Speech Recognition Using the MVDR Spectrum Estimation Method

SatyaDharanipragada, Senior Member,IEEE,
Umit H. Yapanel,Member, IEEE, and Bhaskar D. Rao, Fellow, IEEE

Reporter:黃重翔

Professor:陳嘉平

Abstract

- This paper describes a robust feature extraction technique for continuous speech recognition.
- Central to the technique is the minimum variance distortionless response (MVDR) method of spectrum estimation.
- We consider incorporating perceptual information in two ways: 1) after the MVDR power spectrum is computed and 2) directly during the MVDR spectrum estimation.

Abstract

- We show that incorporating perceptual information directly in to the spectrum estimation improves both robustness and computational efficiency significantly.
- We analyze the class separability and speaker variability properties of the features using a Fisher linear discriminant measure and show that these features provide better class separability and better suppression of speaker-dependent information than the widely used mel frequency cepstral coefficient (MFCC) features.

Abstract

- We evaluate the technique on four different tasks: an in-car speech recognition task, the Aurora-2 matched task, the Wall Street Journal (WSJ) task, and the Switchboard task.
- The new feature extraction technique gives lower word-error-rates than the MFCC and perceptual linear prediction (PLP) feature extraction techniques in most cases.

Abstract

- Statistical significance tests reveal that the improvement is most significant in high noise conditions.
- The technique thus provides improved robustness to noise without sacrificing performance in clean conditions.
- Index Terms—Distortionless response, minimum variance, robust feature extraction for continuous speech recognition, spectral analysis, speech analysis.

Outline

- Introduction
- MVDR Spectral Envelope Estimation
- Feature Extraction for Speech Recognition
- Class Separability and Interspeaker Variability
- Speech Recognition Experiments
- Conclusion

Introduction

- 當消除其他不相干的特定語者資訊如pitch harmonics時，從語音訊號捕捉聲道轉換函數是一個對準確語音辨識的關鍵性需求。
- 在線性規劃（LP）技術中，頻譜的包絡線以一全極濾波器作藍本，此濾波器之係數以最小化頻譜與LP濾波器的頻率響應間的均方差估算出。
- 使用爲了計算上包絡線而採用的同步音調與自然頻率抓取技術的直接上包絡線估算演算法顯示好的結果，但計算成本很高且在噪音環境下易於產生非強健性行爲。

Introduction

- 本論文針對連續語音辨識提出一新的特徵萃取技術。這個技術的核心爲對頻譜估算的最小變異無失真響應方法（MVDR）。
- 在語音辨識中，除了頻譜包絡線的忠實再現，屬性統計如頻譜估計的偏差與變異也很受到關注。

MVDR Spectral Envelope Estimation

- 在非參數頻譜估計方法如FFT週期圖方法中，功率是用以interest的頻率爲中心的波段穿越濾波器輸出頻率單一採樣計算出。
- 兩個頻譜估算的統計特性爲interest，即偏差與變異數。

MVDR Methodology

- 在MVDR方法中，訊號功率（頻率爲ω_l）是以將訊號用一特殊FIR濾波器$h(n)$濾波，並計算輸出功率。

$$H(e^{j\omega_l}) = \sum_{k=0}^{M} h(k)e^{-j\omega_l k} = 1.$$

- 失真限制可寫爲 $\mathbf{v}^H(\omega_l)\mathbf{h} = 1$，而失真濾波器可由解決以下有限制式最佳化問題求得：

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{R}_{M+1} \mathbf{h} \text{ subject to } \mathbf{v}^H(\omega_l)\mathbf{h} = 1.$$

MVDR Spectral Envelope Estimation

- 其中R_{M+1}是一$(M+1)\times(M+1)$的常對角自相關矩陣，其解爲

$$\mathbf{h}_l = \frac{\mathbf{R}_{M+1}^{-1}\mathbf{v}(\omega_l)}{\mathbf{v}^H(\omega_l)\mathbf{R}_{M+1}^{-1}\mathbf{v}(\omega_l)}$$

- MVDR對頻率爲ω_l之訊號功率頻譜$S_{xx}(\omega)$的頻譜估算是爲了獲得最佳化受限濾波器$h_l(n)$輸出的功率，如

$$P_{MV}(\omega_l) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|H_l(e^{j\omega})\right|^2 S_{xx}(e^{j\omega})d\omega$$

$P_{MV}(\omega_l)$表MVDR功率頻譜估計值。

MVDR Spectrum Computation

- 在MVDR方法中對所有頻率的頻譜輸出功率可以用以下方法計算：

$$P_{MV}(\omega) = \frac{1}{\mathbf{v}^H(\omega)\mathbf{R}_{M+1}^{-1}\mathbf{v}(\omega)}$$

- 第M序MVDR頻譜將之參數化寫成

$$P_{MV}(\omega) = \frac{1}{\sum_{k=-M}^{M} \mu(k)e^{-j\omega k}} = \frac{1}{|B(e^{j\omega})|^2}$$

MVDR Spectrum Computation

- 其中參數 $\mu(k)$ 可由一使用LP係數 a_k 與預測失誤變異數 P_e 之適度非迭代計算獲得：

$$\mu(k) = \begin{cases} \frac{1}{P_e} \sum_{i=0}^{M-k} (M+1-k-2i) \\ \quad \times a_i a_{i+k}^*, & \text{for } k = 0, \ldots, M \\ \mu^*(-k), & \text{for } k = -M, \ldots, -1 \end{cases}$$

MVDR Properties

- 用MVDR頻譜的特性來解釋對語音模型的估算技術的優勢有：
  - 濾波器特性
  - 頻譜估算與包絡線估算
  - 對線性預測的連結
  - 偏壓與變異數屬性
  - 實證觀察

Empirical Observations

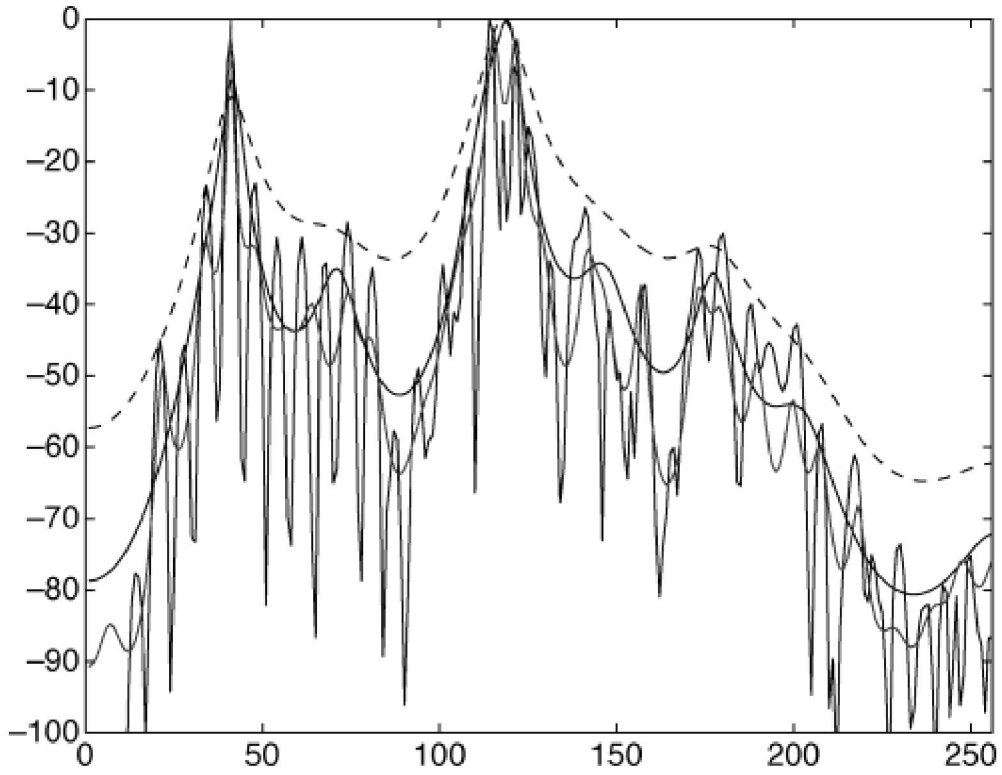

Fig. 1. Short-term FFT spectrum and the LP (solid) and MVDR (dashed) spectral envelopes.

Empirical Observations

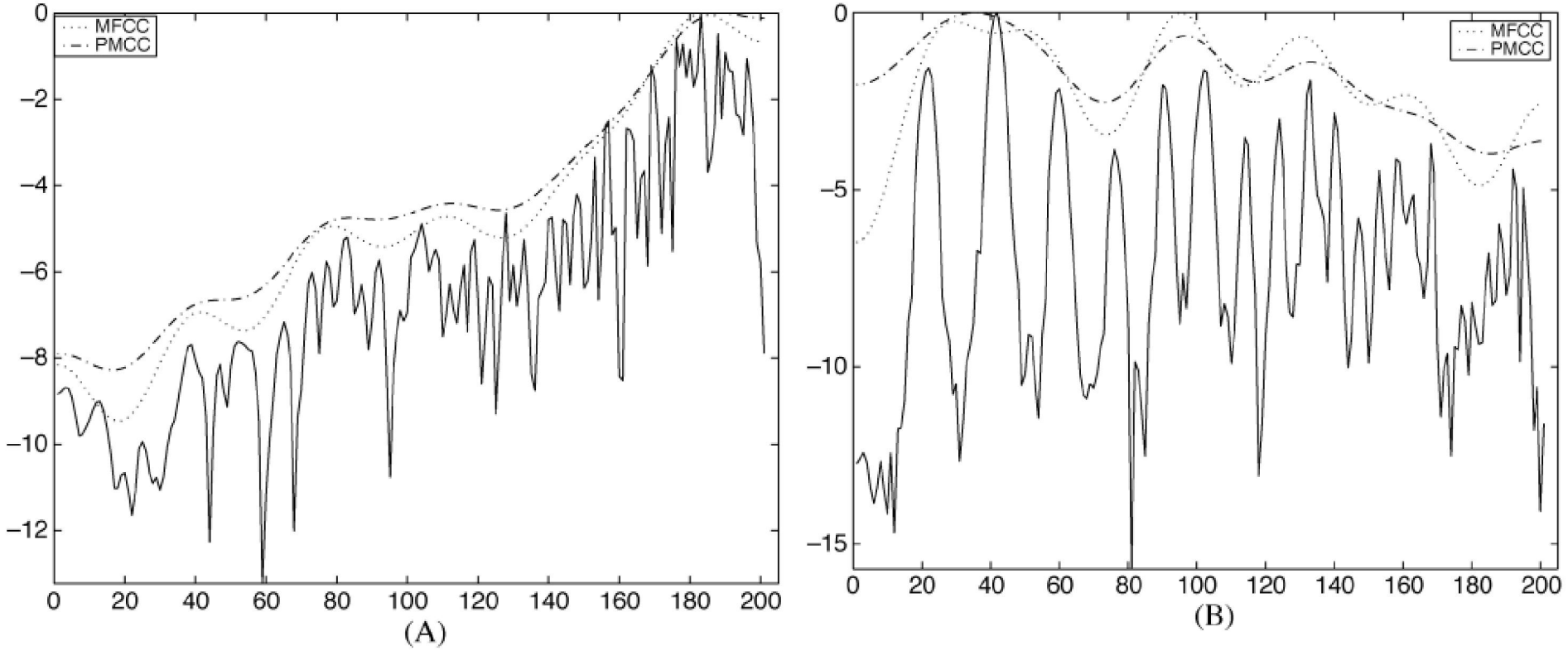

Fig. 2. Spectral envelopes for MFCC (dotted) and PMCC (dash-dotted) superimposed onto mel-warped FFT power spectrum (solid) for (A) unvoiced and (B) voiced sounds of a female speaker from WSJ database.

Feature Extraction for Speech Recognition

- 納入知覺資訊的MFCC與PLP兩種方法可用於與MVDR頻譜估算技術來獲得兩種不同的前端處理，我們稱爲MVDR-MFCC與PMCC。
- 從自相關估算取得MVDR係數$\mu(k)$有兩個優點：一爲由於將頻譜感知平滑化使自相關估算更加可信；二爲因MVDR估算的維數複雜度由於相對較小的mel濾波器輸出的維度而減少。

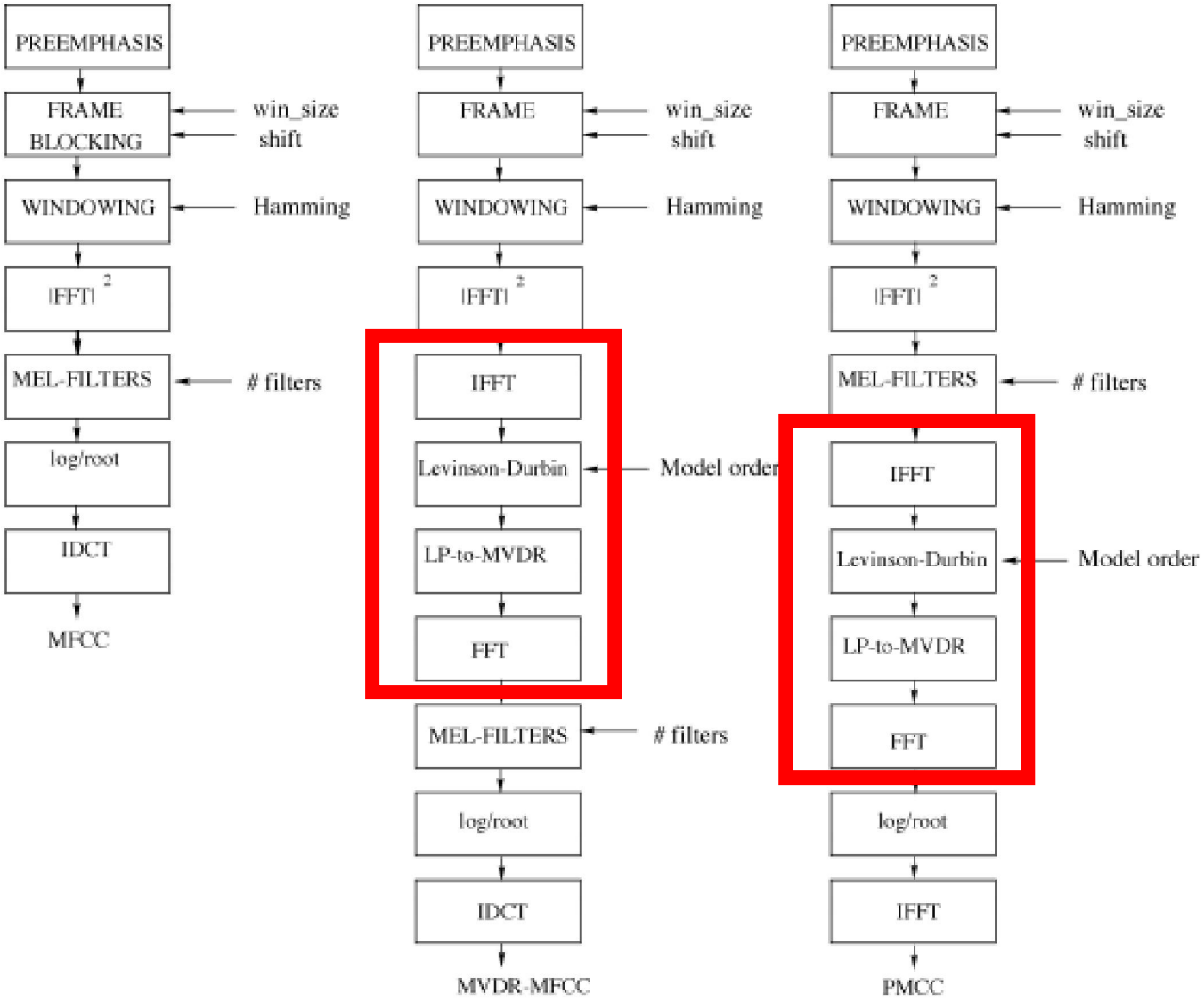

Fig. 3. Schematic diagrams of the MFCC, MVDR-MFCC, and PMCC front-ends.

Feature Extraction for Speech Recognition

- FFT方法有三個主要的步驟：
  - 從MVDR多項式用FFT計算功率頻譜
  - 取log值
  - 計算反FFT或DCT

Class Separability and Interspeaker Variability

- 我們設計一建構於線性判別分析（LDA）技術的方法來分析不同的特徵萃取方法中類別分離性與語者間的變異性。

Analysis of Class Separability

- 用Fisher's LDA計算類別間分離性如下：取得類別平均與整體平均

$$x_{c,i}, c = 1, 2, \ldots, C \qquad m_c = \frac{1}{N_c} \sum_{i=1}^{N_c} x_{c,i}$$

$$i = 1, 2, \ldots, N_c \qquad m = \frac{1}{N} \sum_{c=1}^{C} \sum_{i=1}^{N_c} x_{c,i}.$$

- 可得到類別內與類別間分散矩陣爲

$$S_W = \frac{1}{N} \sum_{c=1}^{C} \sum_{i=1}^{N_c} (x_{c,i} - m_c)(x_{c,i} - m_c)^T$$

$$S_B = \frac{1}{N} \sum_{c=1}^{C} N_c (m_c - m)(m_c - m)^T.$$

Analysis of Class Separability

- 當$d \ll D$，我們對一D次子空間之特徵向量進行線性預測，取得預測矩陣A

$$A = \underset{L}{\operatorname{argmax}} \frac{|LS_B L^T|}{|LS_W L^T|}$$

- 行列式量DM求法爲

$$\mathrm{DM} = \log\left(\prod_{i=1}^{d} \lambda_i\right)$$

Analysis of Class Separability

- 三種特徵萃取方法的DM爲

TABLE I
CLASS-SEPARABILITY MEASURE FOR MFCC, MVDR-MFCC, AND PMCC

	MFCC	MVDR-MFCC	PMCC
DM	-75.98	-75.60	-74.69

可看出PMCC的類別分離性最佳

- 使用隱藏Markov模型（HMM）作爲聲音類別進行分析

Analysis of Interspeaker Variability

- 以不同特徵萃取方法的強健性評估語者變異性，其中特徵向量 $x_{c,s,i}, c = 1, 2, \ldots, C$

$$s = 1, 2, \ldots, S, i = 1, 2, \ldots, N_{c,s}$$

- 語者s在類別c中的平均值爲

$$m_{c,s} = \frac{1}{N_{c,s}} \sum_{i=1}^{N_{c,s}} x_{c,s,i}$$

而類別c中所有語者的平均則爲

$$m_c = \frac{1}{M_c} \sum_{s:N_{c,s} \neq 0} m_{c,s}$$

Analysis of Interspeaker Variability

- 同樣取得整體平均

$$m = \frac{1}{M} \sum_{c=1}^{C} M_c m_c$$

- 與類別中和類別間平均分散矩陣

$$S_W = \frac{1}{M} \sum_{c=1}^{C} \sum_{s} (m_{c,s} - m)(m_{c,s} - m)^T$$

$$S_B = \frac{1}{M} \sum_{c=1}^{C} M_c (m_c - m)(m_c - m)^T.$$

Analysis of Interspeaker Variability

- 計算三種方法的DM值，同樣PMCC擁有最大的DM。因此，PMCC的語者間變異性爲三者最小。

TABLE II
INTERSPEAKER VARIABILITY MEASURES FOR MFCC, MVDR-MFCC, AND PMCC

Measure/Systems	MFCC	MVDR-MFCC	PMCC
DM	-74.57	-73.12	-70.44

Speech Recognition Experiments

- 比較兩種特徵萃取方法（MFCC、PLP）與加入MVDR後的方法（MVDR-MFCC、PMCC）在四種不同的資料庫：汽車內、AURORA2、WSJ與總機。

With an Automotive Database

TABLE III
WERs[%] FOR IN-CAR DATA WITH DIFFERENT FRONT-ENDS

Speed/Systems	MFCC	PLP	MVDR-MFCC	PMCC
00mph	1.18	1.14	1.14	1.13
30mph	2.19	1.93	2.16	1.97
60mph	6.65	5.93	6.22	4.92
all	3.34	3.01	3.18	2.68

TABLE IV
RELATIVE IMPROVEMENTS [%] OF PMCC WITH RESPECT TO MFCC, PLP, AND MVDR FEATURES ON THE IN-CAR TEST SET

Speed/Systems	MFCC	PLP	MVDR-MFCC
00mph	4.23	0.87	0.87
30mph	10.04	-2.07	8.79
60mph	26.01	17.03	20.90
all	19.76	10.96	15.72

With an Automotive Database

- 零假設（null hypothesis）機率

TABLE V
SIGNIFICANCE TESTS: $P(H_0)$ FOR THE PMCC-MFCC, PMCC-PLP, AND PMCC-MVDR PAIRS ON THE IN CAR TEST SET

Speed/Pair	PMCC-MFCC	PMCC-PLP	PMCC-MVDR-MFCC
00mph	0.61	0.92	0.92
30mph	0.09	0.75	0.14
60mph	2.2E-6	7.38E-7	3.09E-10
all	1.96E-13	1.38E-4	1.24E-8

- 可用於推導用MVDR方法的準確包絡線估算

With the Aurora Database

TABLE VI
WERs[%] FOR AURORA 2 (SET A) AND RELATIVE IMPROVEMENT OF PMCC WITH RESPECT TO THE PLP BASELINE

SNR	MFCC	PLP	MVDR-MFCC	PMCC	Rel. Imp.[%]
-5dB	65.60	65.56	65.28	61.26	6.6
0dB	28.51	28.20	28.36	25.23	10.5
5dB	9.27	8.73	9.65	8.50	2.7
10dB	3.23	3.28	3.23	3.26	0.6
15dB	1.65	1.86	1.77	1.50	19.4
20dB	1.29	1.32	1.22	1.04	21.2
Clean	0.91	0.78	0.89	0.78	0.0
0–20	8.79	8.68	8.84	7.89	9.1

TABLE VII
SIGNIFICANCE TESTS: $P(H_0)$ FOR PMCC-MFCC, PMCC-PLP, AND PMCC-MVDR PAIRS ON THE AURORA 2 (SET A) SET

SNR/Pair	PMCC–MFCC	PMCC–PLP	PMCC–MVDR-MFCC
-5dB	2.68E-13	4.4E-13	1.34E-11
Clean	0.249	1.0	0.3268
0-20	3.54E-9	2.01E-7	4.85E-10

LVCSR Experiments: WSJ

- MVDR在中高音語音有特別好的效果

TABLE VIII
WERs(%) AND SIGNIFICANCE TESTS FOR WSJ DEV/*eval* TEST SETS

Gender/Systems	MFCCs	PMCCs	Avg. Rel. Imp.	$P(H_0)$
Female	3.9/*4.5*	3.1/*3.9*	14.6	0.06
Male	5.5/*4.4*	4.9/*4.0*	10.0	0.13
Overall	4.9/*4.5*	4.2/*3.9*	12.8	0.01

LVCSR Experiments: Switchboard

TABLE IX
WER(%)S FOR SWITCHBOARD TASK WITH DIFFERENT FRONT-ENDS

	MFCC	PLP	PMCC model order=20	PMCC model order=30
Eval'98	41.9	38.3	38.2	37.5
Eval'00	26.5	24.5	24.2	24.3

TABLE X
SIGNIFICANCE TEST: PMCC WRT PLP AND MFCC

	PMCC20-MFCC	PMCC20-PLP	PMCC30-MFCC	PMCC30-PLP
Eval'98	2.1E-13	0.84	2.2e-308	0.09
Eval'00	4.3E-8	0.47	1.6E-7	0.63
Overall	2.2e-308	0.53	2.2e-308	0.12

Computational Considerations

TABLE XI
APPROXIMATE COMPUTATIONAL COMPLEXITY OF DIFFERENT FRONT-ENDS

Step / #Operations	MFCC	MVDR-MFCC	PMCC
Windowing	400	400	400
$\|\mathrm{FFT}\|^2$, N = 512	$512 + 512 \times \log_2(512)$	$512 + 512 \times \log_2(512)$	$512 + 512 \times \log_2(512)$
IFFT, $N = 512$	N/A	$512 \times \log_2(512)$	N/A
MVDR (M=80)	N/A	2×80^2	N/A
Filterbank (24)	2×257	2×257	2×257
MVDR (M=22)	N/A	N/A	2×22^2
FFT, N=128	N/A	N/A	$128 \times \log_2(128)$
log	Ignored	Ignored	Ignored
IDCT	24×13	24×13	N/A
IFFT, N=128	N/A	N/A	$128 \times \log_2(128)$
Total	6346	25855	8794

- WSJ實驗中前端處理需要的即時因子

TABLE XII
REAL-TIME FACTORS FOR WSJ TASK WITH MFCC AND PMCC

Systems	MFCC	PMCC	Rel. Imp.(%)
RTF	2.32	1.80	22.4

Conclusion

- PMCC比MFCC有更高的計算複雜度，但辨識失真卻減少。
- 整體而言在噪音干擾越強的環境底下所使用MVDR方法的效果就越好。