Combining Gaussian Mixture Model Front End with MFCC Parameters

Author : M.N Stuttle M.J.F Gales

Professor : 陳嘉平

Reporter : 許峰閤

ABSTRACT

Fitting a Gaussian mixture model (GMM) to the smoothed speech spectrum allows an alternative set of features to be extracted from the speech signal. These features have been shown to possess information complementary to the standard MFCC parameterisation. This paper further investigates the use of these GMM features in combination with MFCCs. The extraction and use of a confidence metric to combine GMM features with MFCCs is described. Results using the confidence metric on the WSJ task are presented. Also, GMM features for speech corrupted with additive noise are extracted from data corrupted with coloured addititve noise. Techniques for noise robustness and compensation are investigated for GMM features and the performance is examined on the RM task with additive noise.

Using Confidence Metrics for Feature Combination

- GMM可以用來調適語音頻譜使其平滑化

- 首先針對一段語音訊號加上window將其切爲許多寬度爲25.6ms的frame,並且window的移動頻率爲10ms

- 接著對每個frame做DFT,並且利用pitch-based raised cosine window使其平滑化

Using Confidence Metrics for Feature Combination

- 取一個N-component的高斯混合模型, 首先用EM演算法對時間t時各點的log likelihood作最佳化

$$l(\boldsymbol{x}(t)|\boldsymbol{\theta}(t)) = \sum_{k=1}^{K}\left(\ln \sum_{n=1}^{N} e_n(t)\mathcal{N}\left(x_k(t);\mu_n(t),\sigma_n^2(t)\right)\right)$$

$$\boldsymbol{\theta}(t) = \{\mu_1(t),\ldots,\mu_N(t),\sigma_1^2(t),\ldots,\sigma_N^2(t),e_1(t),\ldots,e_N(t)\}$$

Using Confidence Metrics for Feature Combination

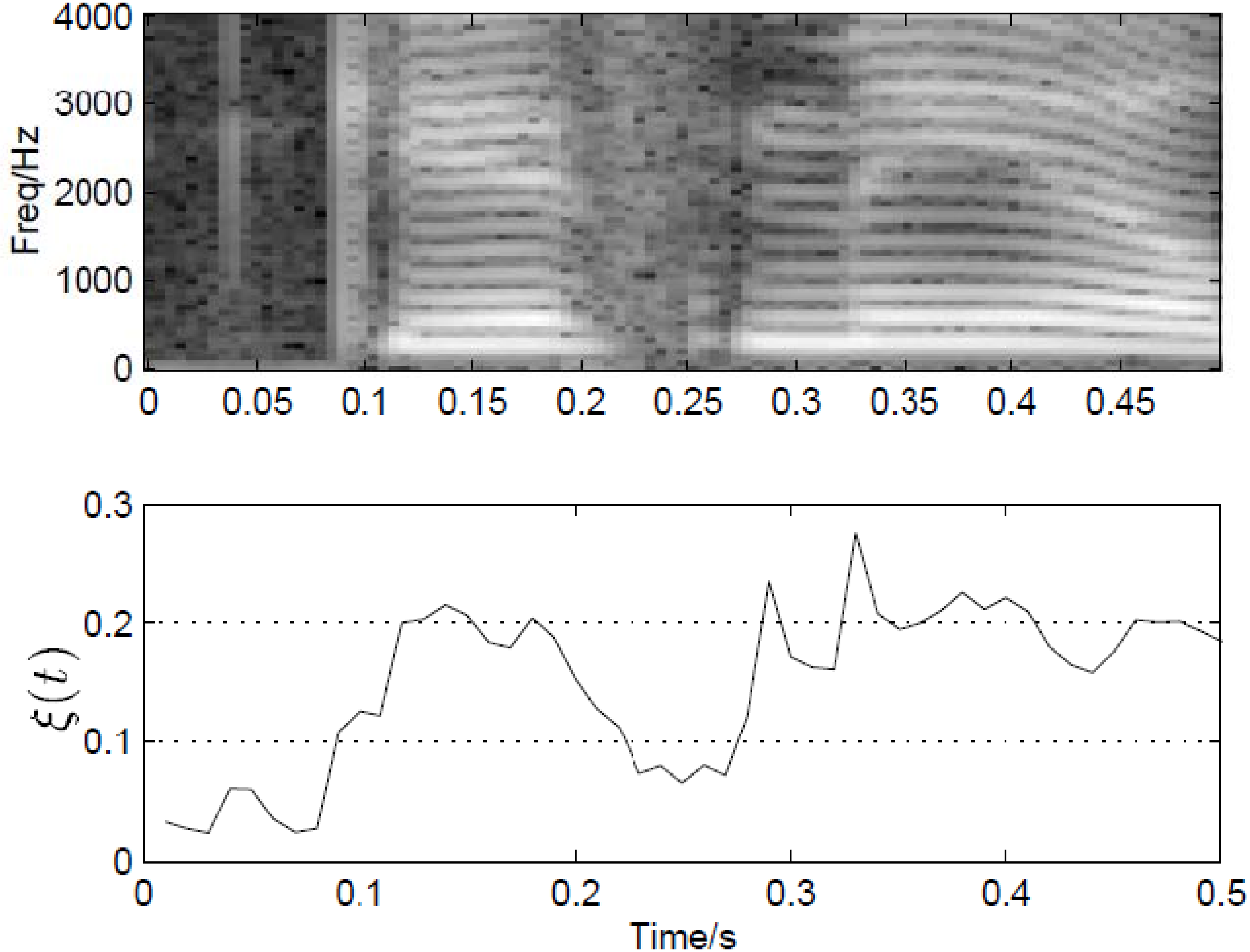
- 在此論文中的語音訊號都先通過4kHz的低通濾波器,並且採用6-component的高斯混合模型

- 頻譜中的峰值有時並沒有辦法明確的將其分類在某一個音素的類別中,所以我們需要給其一個信心值

Using Confidence Metrics for Feature Combination

$$\xi(t) = \beta\left[\prod_{n=1}^{N}\frac{\tilde{e}_n(t)+10.53}{\sigma_n(t)}\right]^{\frac{1}{N}}$$

由於 $\tilde{e}_n(t)$的值介於1~-10.53之間所以需要加10.53使其恆正


4000
3000
2000
1000
0
Freq/Hz
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.3
0.2
0.1
0
$\xi(t)$
0
0.1
0.2
0.3
0.4
0.5
Time/s

Using Confidence Metrics for Feature Combination

- 建立一個多串流系統來將GMM的參數與MFCC的資訊做結合, 得到一個機率分布函數

$$b_j(\boldsymbol{y}(t)) = \prod_{r=1}^{R} \Big[\sum_{m=1}^{M} c_{jrm} \mathcal{N}(\boldsymbol{y}_r(t); \boldsymbol{\mu}_{jrm}, \boldsymbol{\Sigma}_{jrm}) \Big]^{\gamma_r(t)}$$

c_{jrm} 爲 component m , stream r , state j 的 weight

$\boldsymbol{\mu}_{jrm}$ 爲 component m , stream r , state j 的 weight

$\boldsymbol{\Sigma}_{jrm}$ 爲 component m , stream r , state j 的 weight

Using Confidence Metrics for Feature Combination

$$\boldsymbol{y}_1(t) = [o_1(t), \ldots, o_{12}(t), l(t), \Delta, \Delta^2]^T$$
$$\boldsymbol{y}_2(t) = [\mu_1(t), \ldots, \mu_6(t), \Delta, \Delta^2]^T$$

$$\gamma_1(t) = 1 - \xi(t)$$
$$\gamma_2(t) = \left[\frac{\alpha_1}{\alpha_2}\right]\xi(t)$$

Compensating GMM Parameters in Noise

- 我們可以利用GMM來對noise speech做補償,使其可以逼近clean speech

- 首先對noise做一個Q-component GMM

$$\hat{\boldsymbol{\theta}}^{(n)} = \{\hat{\mu}_1^{(n)}, \ldots, \hat{\mu}_Q^{(n)}, \hat{\sigma}_1^{(n)2}, \ldots, \hat{\sigma}_Q^{(n)2}, \hat{e}_1^{(n)}, \ldots, \hat{e}_Q^{(n)}\}$$

Compensating GMM Parameters in Noise

$$l(\boldsymbol{x}(t)|\boldsymbol{\theta}(t),\hat{\boldsymbol{\theta}}^{(n)}) = \sum_{k=1}^{K} \ln \left(\sum_{q=1}^{Q} \hat{e}_q^{(n)} \mathcal{N}\left(x_k(t); \hat{\mu}_q^{(n)}, \hat{\sigma}_q^{(n)2}\right) + \sum_{n=1}^{N} e_n(t) \mathcal{N}\left(x_k(t); \mu_n(t), \sigma_n^2(t)\right) \right)$$

$\boldsymbol{\theta}(t)$ 爲clean speech parameters

$\hat{\boldsymbol{\theta}}^{(n)}$ 爲 noise speech average parameters

Compensating GMM Parameters in Noise

- 經由上式的補償過後,還必須解決noise masking的問題,所以後端的HMM也必須由noise model來調適

- 首先從HMM的state j component m的機率分佈函式（PDF）取得靜態平均,用此來估算平均的GMM參數 $\hat{\boldsymbol{\theta}}_{jm}$

Compensating GMM Parameters in Noise

$$l(\boldsymbol{x}_{jm} + \boldsymbol{q} | \hat{\boldsymbol{\theta}}_{jm}) =$$

$$\sum_{k=1}^{K} \left(\ln \sum_{n=1}^{N} \hat{e}_{jmn} \mathcal{N} \left(x_{jmk} + q_k ; \hat{\mu}_{jmn} , \hat{\sigma}_{jmn}^{2} \right) \right)$$

實驗

- 使用Resource Management 及 WSJ 語料庫
- RM語料庫含有1000個字彙
- WSJ語料庫含有65000個字彙
- 另外從Noisex corpus加入噪音

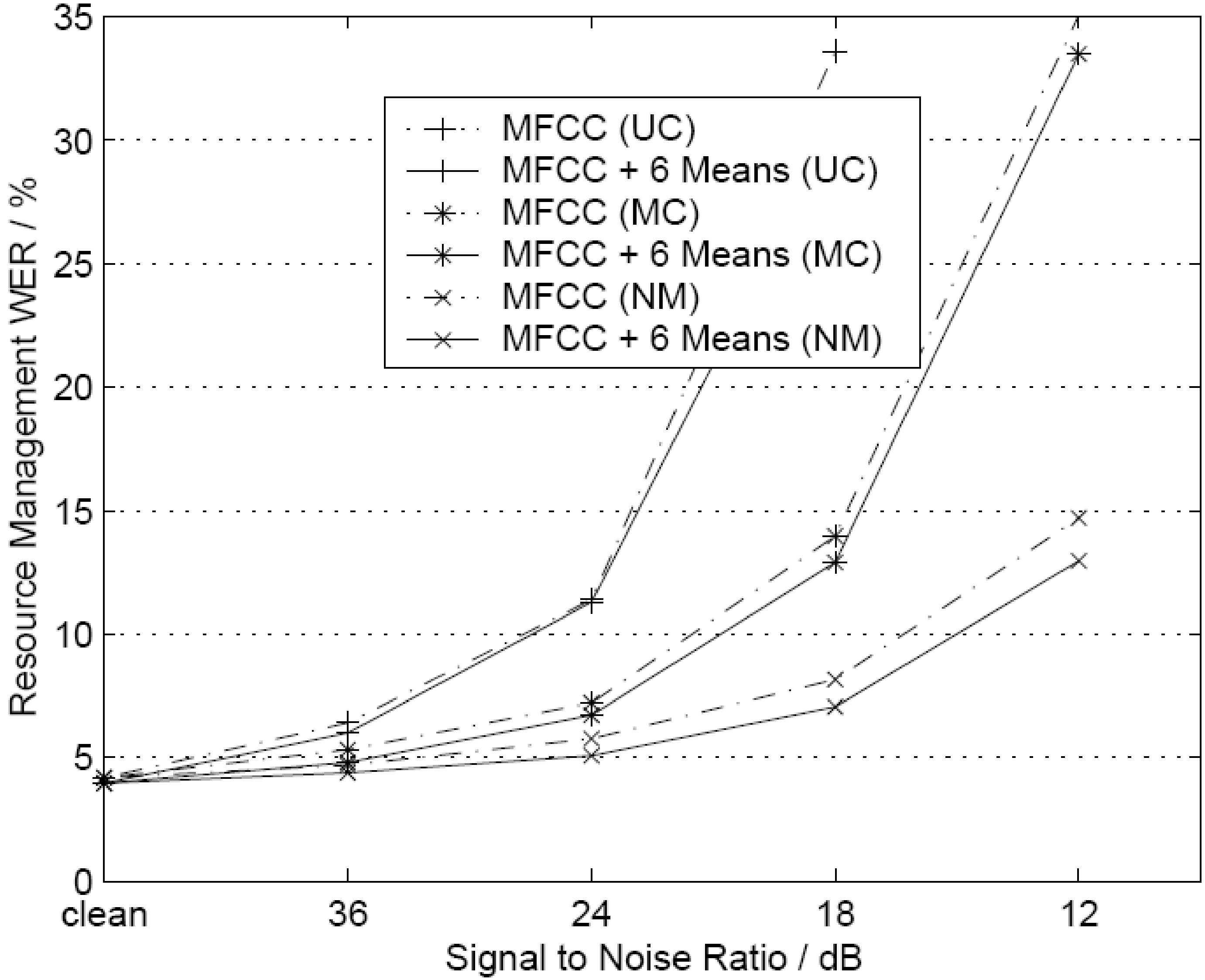
實驗

Confidence scale β	Clean	18dB SNR
0.0 (MFCC only)	4.19	32.3
0.1	3.95	29.6
0.2	3.94	29.6
0.3	4.12	30.4
0.4	4.23	31.9
0.5	4.32	33.5
$\xi(t) = 1$ (GMM only)	9.23	58.3

實驗

Description	WER /%
MFCC (UC)	32.3
MFCC (UC) + GMM (UC)	30.6
+ Confidence	29.6
MFCC (UC) + GMM (FC)	28.3
MFCC (UC) + GMM (MC)	22.1
MFCC (MC)	14.0
MFCC (MC) + GMM (MC)	13.1
+ Confidence	12.6

實驗



35
30
25
20
15
10
5
0
Resource Management WER / %
MFCC (UC)
MFCC + 6 Means (UC)
MFCC (MC)
MFCC + 6 Means (MC)
MFCC (NM)
MFCC + 6 Means (NM)
clean
36
24
18
12
Signal to Noise Ratio / dB