Temporal Modulation Processing of Speech signals for Noise Robust ASR

Author: Hong You , Abeer Alwan

Professor: 陳嘉平

Reporter :許峰閤

Outline

- Introduction
- Analysis
- Frequency adaptive modulatoin processing algorithm
- Experiment

Introduction

- Speech and noise is more accurately classified by low-passed modulation frequency than band-passed.

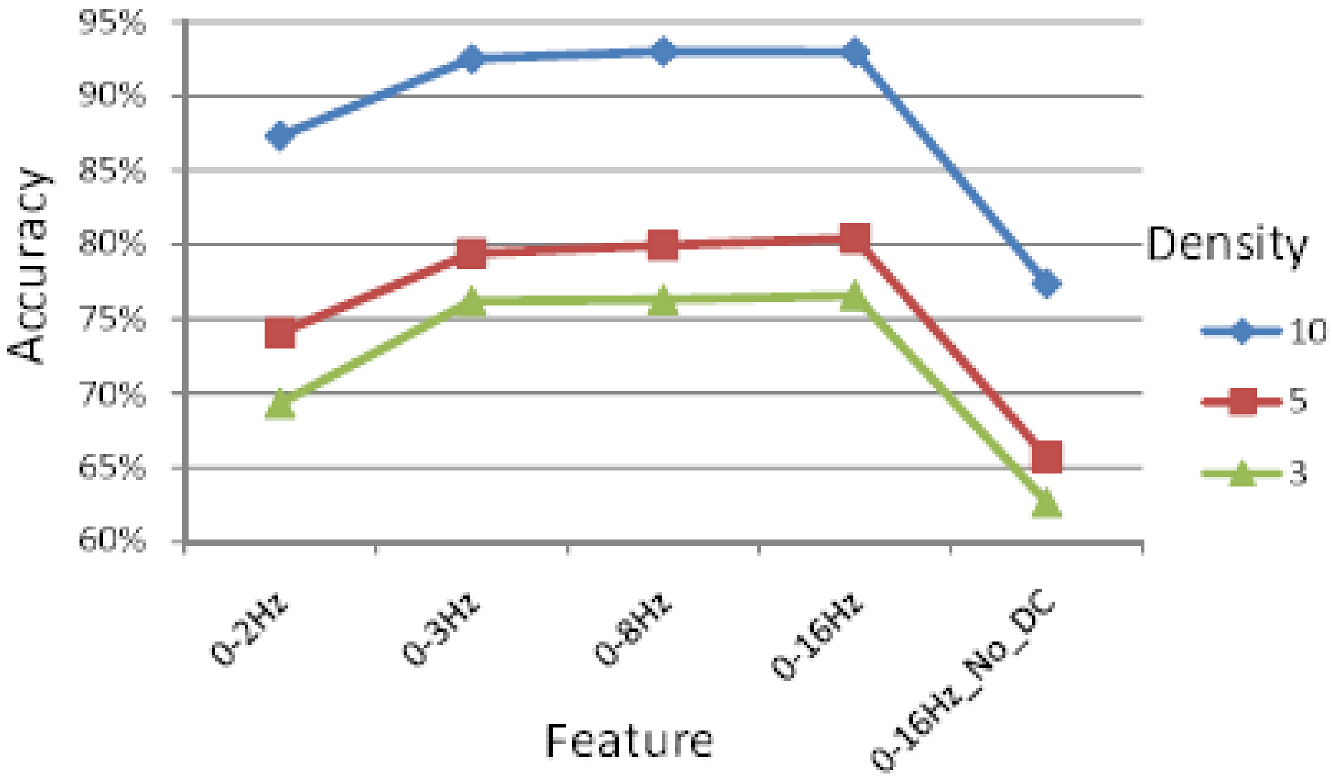
- The frequency adaptive modulatoin processing algorithm can attenuate noise sensitive MFs based on estimation from noise robust MFs.

Analysis

- Speech and noise signal have distinct modulation frequency(MF) characteristics.

- Using SVM trained to classify speech/noise modulation features.

Analysis



95%
90%
85%
80%
75%
70%
65%
60%
Accuracy
0-2Hz
0-3Hz
0-8Hz
0-16Hz
0-16Hz_No_DC
Feature
Density
10
5
3

Analysis

- Band-passed MF performs consistently worse than low-passed ones.

- The result show the important of combining DC modulation along with low MF for the task of speech/non-speech feature classification.

Speech and Noise Modulation characteristics

- We observe that speech MFs have smoother energy transition from 0 Hz to low MFs.

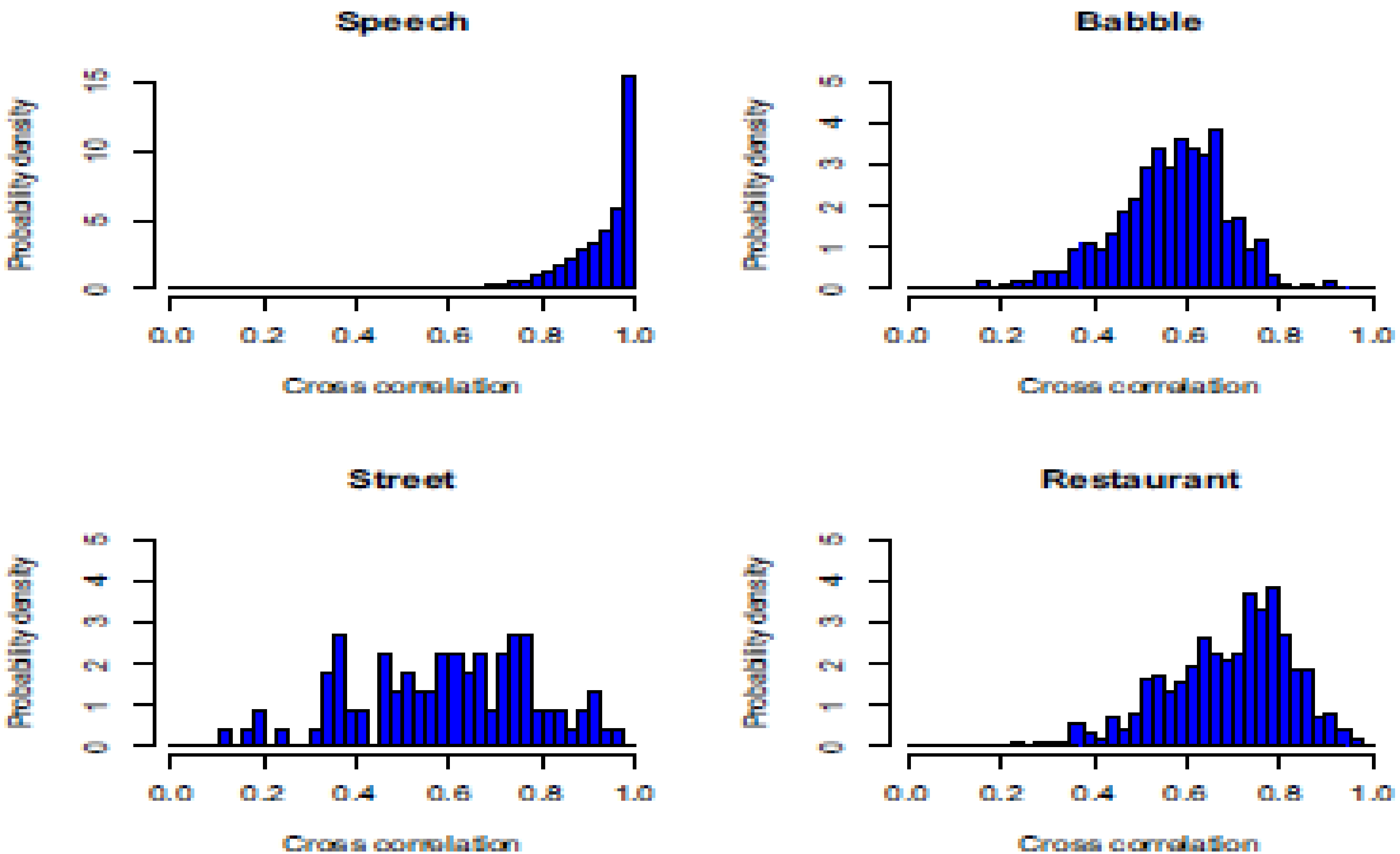
- For noise, sharp energy decrease occur from 0 Hz to low MFs.

Speech and Noise Modulation characteristics

- We analyze the cross-correlation between different MFs across acoustic frequencies.

- The hypothesis that the modulation pattern of speech signals over long segments is more consistent than that of noise.

Speech and Noise Modulation characteristics

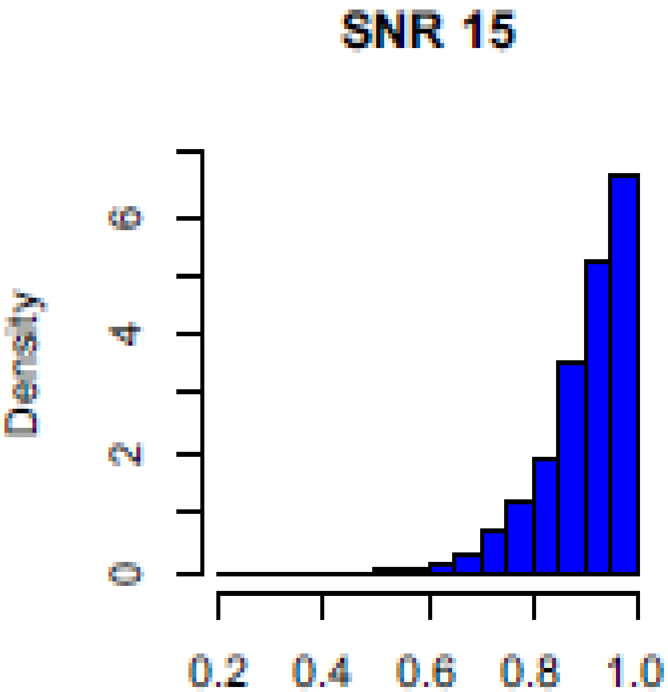


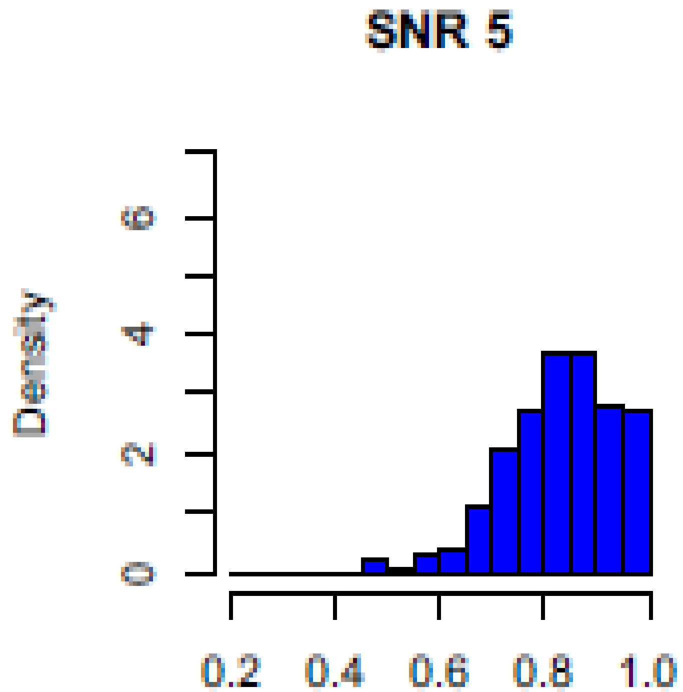
Speech and Noise Modulation characteristics

- And we will show the cross correlation of speech signal corrupted by subway noise at SNR 15db and 5db.

- It shows that additive noise considerably reduced cross-correlation between MFs.

Speech and Noise Modulation characteristics



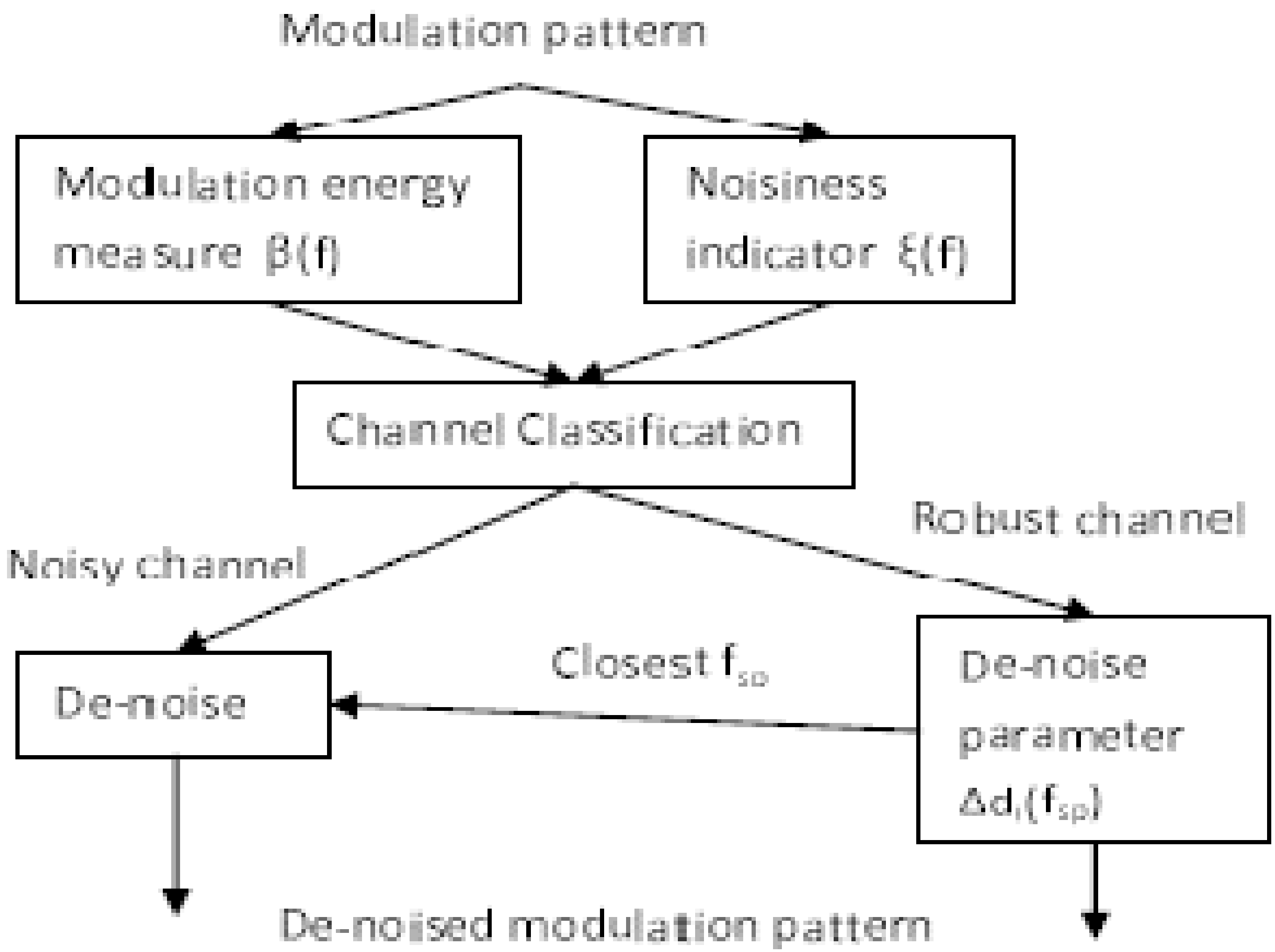


Frequency Adaptive Modulation processing Algorithm

- The basic idea is to denoised noise-sensitive MFs for noisy channels based on a linear combination of noise-sensitive and noise-robust MFs.

- We assume that noise-sensitive MFs range from 0 Hz to 2 Hz and noise-robust MFs range from 3 Hz to 8 Hz.

Frequency Adaptive Modulation processing Algorithm



Frequency Adaptive Modulation processing Algorithm

- Define:

$d_i(f)$ is the log magnitude at MF i and frequency f.

$$\beta(f)=\sum_{i=3}^{8} d_i(f)$$ Modulation energy measure

$$\varepsilon(f)=d_0(f)-d_{1.5}(f)$$ Noisiness indicator

Frequency Adaptive Modulation processing Algorithm

- To classify frequency channels, a recursive min-max scheme is used.

- There are 2 thresholds:

  T_β is an adaptive threshold adjusted ccording to local maximum of $\beta(f)$

  $T_\varepsilon = 5.5 + 0.33 \times$ mean of $\varepsilon(f)$ over 0< f <50

Frequency Adaptive Modulation processing Algorithm

- Speech modulation parameters, $\vartriangle d_i(f_{sp})$

  f_{sp} is the classified noise robust channel.

  For i = 0,1,2 Hz(MF)

$$\vartriangle d_i(f_{sp}) = d_i(f_{sp}) - \frac{1}{3}\sum_{j=3}^{5} d_j(f_{sp})$$

Frequency Adaptive Modulation processing Algorithm

- For noisy channel f, the closet f_{sp} is used to denoise noisy MFs. For i=0~2 Hz

$$d_i^{'}(f) =_{\triangle} d_i(f_{sp}) + \frac{1}{3}\sum_{j=3}^{5} d_j(f)$$

- A noisy MF is denoised if $|d_i^{'}(f) - d_i(f)|$ >1.5 db.

Experiments

- Using Aurora2 database.
- Using HTK tools.
- The MFCC features are used as our baseline system.

Experiments

	MFCC	RASTA-MFCC	ETSI-AFE	Adaptive
SNR20	97.05	95.00	98.4	95.60
SNR15	93.49	90.94	96.32	94.23
SNR10	78.72	80.2	91.53	86.34
SNR5	52.2	63.83	77.92	70.34
SNR0	26.01	37.98	50.91	43.75
SNR-5	11.18	16.79	20.4	18.67
Average	59.77	64.13	72.58	68.16