New Continuous Speech Feature Adjustment for a Noise-robust CSR System

Yiming SUN
Information Communication Networks Laboratory
Graduate School of Information Science and Technology
Hokkaido University, Sapporo, Japan
sunny@icn.ist.hokudai.ac.jp

Overview

- Introduction
- Conventional Methods
- Robust Continuous Speech Recognition (CSR) System
- Noise Disturbance
- Block Based DRA
- Results

Introduction

➢ Background

• The dynamic range adjustment (DRA) method has been developed as the compensation method for such difference in an isolated word and phrase.

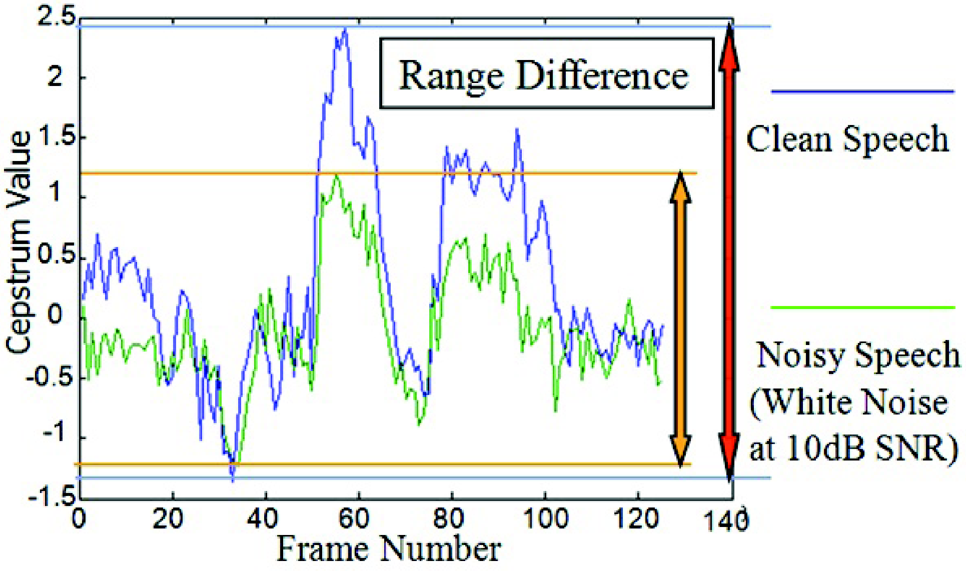


Fig.1 Noise influences in word feature vectors

➢ Summary

• The proposed method introduces a short time length block chosen stochastically from the feature sequence of continuous speech.

• The modified technique of a DRA is proposed to a CSR system.

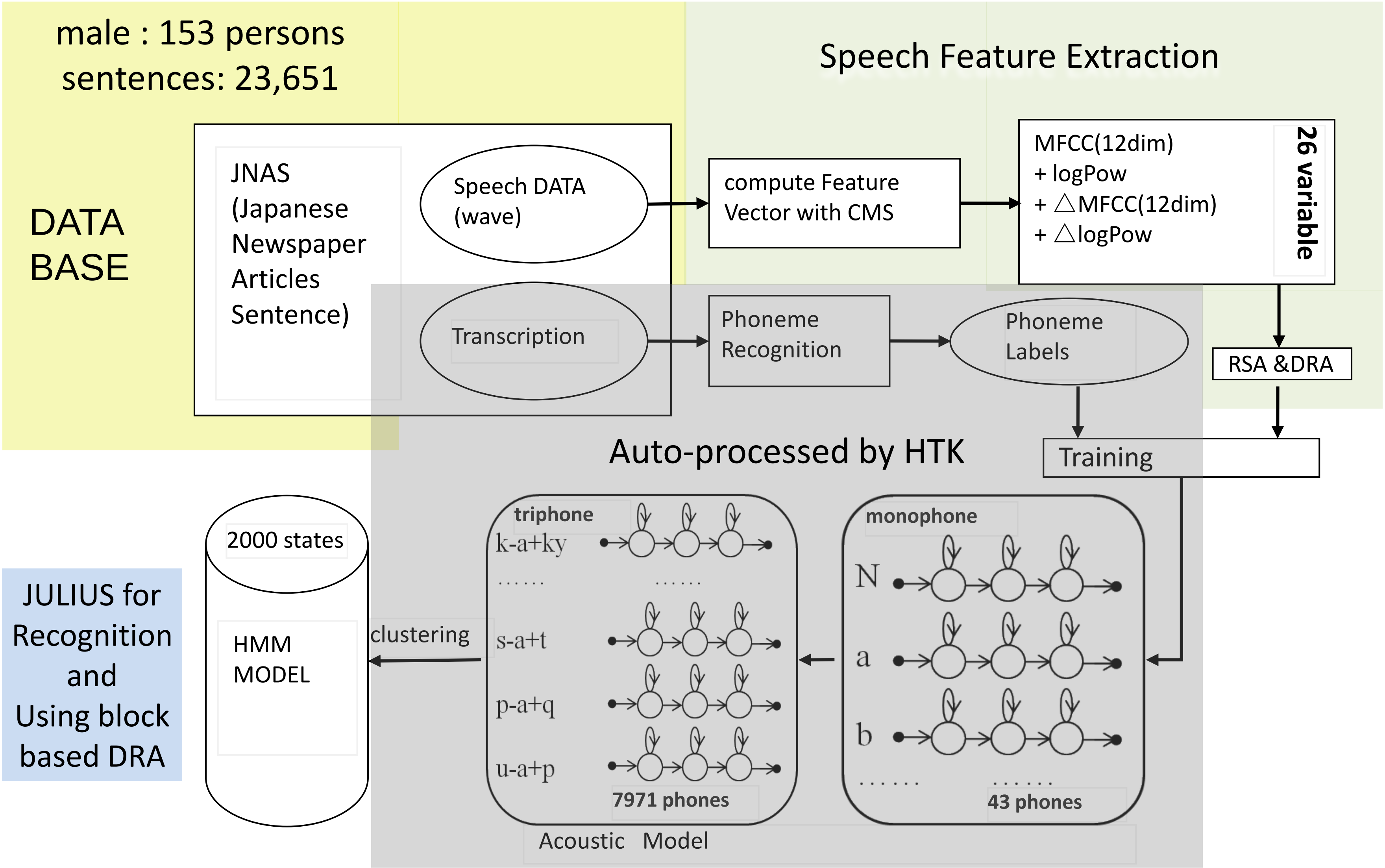
Conventional Methods

- CMS: cepstrum means subtraction
  - CMS is a channel normalization approach to compensate for the acoustic channel.

- RSA: running spectrum analysis
  - RSA is directly used in the modulation spectrum domain.
  - RSA can realize an ideal processing filter.
  - The components of low and high frequency are reduced by using RSA.

- DRA: dynamic range adjustment
  - Adjust the dynamic range of MFCC by normalizing the amplitude of each component.

Robust CSR System



male : 153 persons
sentences: 23,651

DATA BASE

JNAS (Japanese Newspaper Articles Sentence)

Speech DATA (wave)

Transcription

Speech Feature Extraction

compute Feature Vector with CMS

MFCC(12dim)
+ logPow
+ △MFCC(12dim)
+ △logPow

26 variable

RSA &DRA

Phoneme Recognition

Phoneme Labels

Training

Auto-processed by HTK

triphone

k-a+ky

……

s-a+t

p-a+q

u-a+p

7971 phones

monophone

N

a

b

……

43 phones

Acoustic Model

clustering

2000 states

HMM MODEL

JULIUS for Recognition and Using block based DRA

Fig.1 Structure of noise robust CSR system

Noise disturbance

- Sentence selection

  - A continuous speech has many non-speech parts and only noises. These parts effect DRA inappropriately.

  - The unbalance of several dynamic ranges existed in a continuous speech can be compensated.

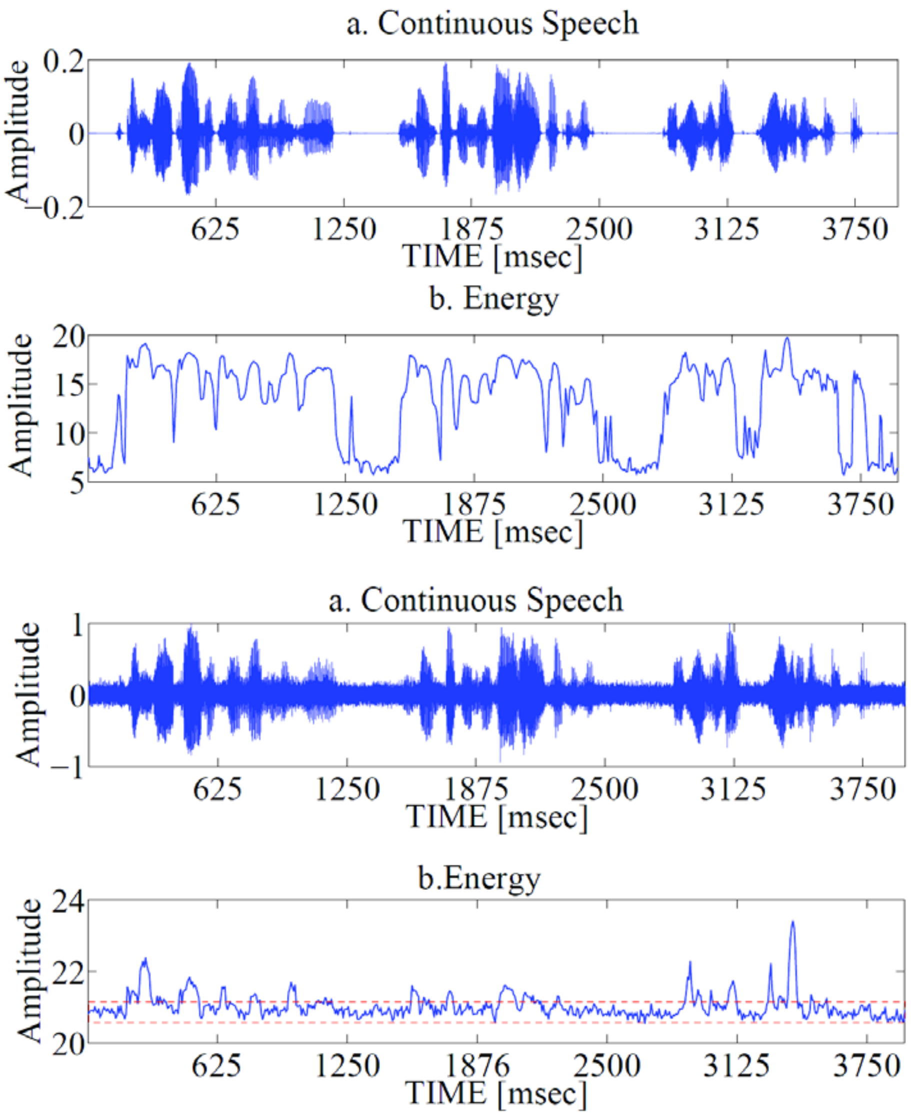


Fig.2 Noise disturbance in energy of continuous speech

Block Based DRA (1)

➢ A short sentence and blocks

- The algorithm finds out the maximum value in a given short sentence, i.e., “Peak Point” in Fig.3.

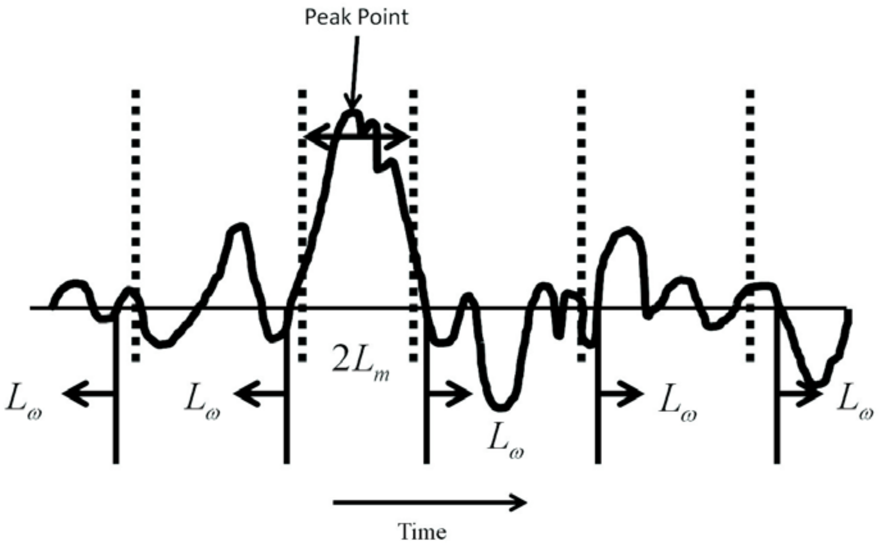


Fig.3 An example for a short sentence and blocks

➢ First step: block separation

- The main block is selected between the above two zero-crossing points nearby the “Peak Point” .

- The shortest length of other blocks is L_w.

Block Based DRA (2)

➢ Second step: determination of the maximum value

- $P_{\pm i,j}$ is defined as the maximum value within the ±i-th block
- Determine the maximum value among $P_{i,j}$ (i = 1, 2, ..., M) as $T_{1,j}$.
- If $P_{0,j} - T_{1,j} < \sigma_p$ and $T_{1,j} - P_{i,j} < \sigma_p$ then set $P_{i,j}$ as the adjustment value.
- If $P_{0,j} - T_{1,j} < \sigma_p$ or $T_{1,j} - P_{i,j} < \sigma_p$ then set $P_{i,j} = T_{1,j}$.
- If $P_{0,j} - T_{1,j} > \sigma_p$ and $T_{1,j} - P_{i,j} > \sigma_p$ then set $P_{i,j} = P_{0,j} - \sigma_p$.

➢ Third step: using block based DRA

- In each block, the following block-based DRA is applied:

$$p'_{k,i} = \frac{p_{k,i}}{P_{\pm i,j}},$$

Block Based DRA (3)

- An Example
  - The proposed algorithm uses the assumption in which there is not large difference between the adjustment values of neighbor-hood blocks.

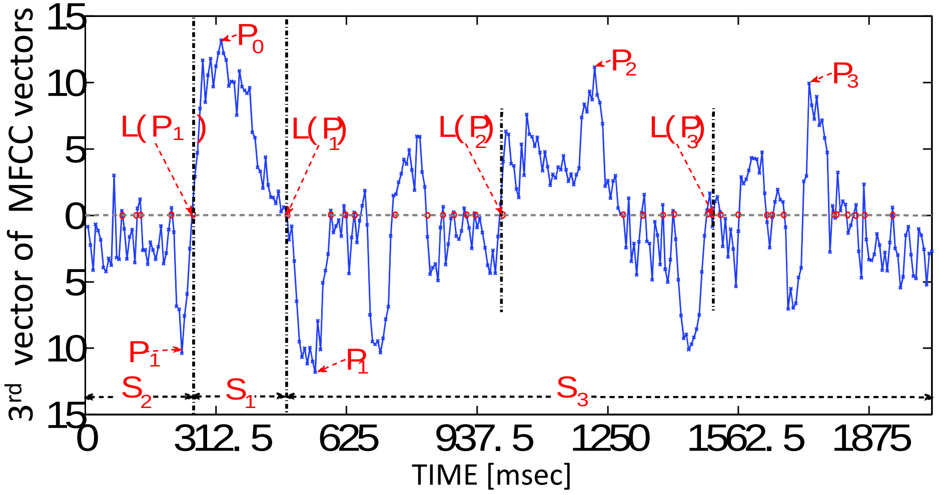


Fig.4 An example for separating blocks and determining maximum

- Parameter Setting (L_m)
  - The main block width includes at least a vowel.

Table 1: Long vowel frame average length

Phoneme	Means	Variance	Appear Times
a:	13.35	13.50	2054
e:	14.46	15.59	12688
i:	14.93	20.97	1724
o:	13.83	19.01	37657
u:	10.64	17.50	4831

Block Based DRA (4)

- Parameter Setting (L_w)

  - The recognition result becomes high when we set L_w = 80.

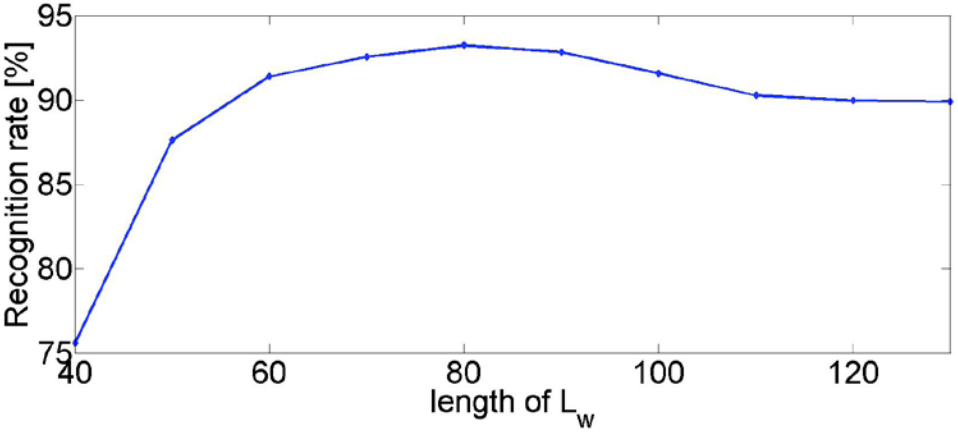


Fig.5 Recognition rate with different L_w

- Parameter Setting (σ_p)

  - The adjustment value focuses on preserving the continuity of the continuous speech features and keeping the relationship between the neighborhood blocks.

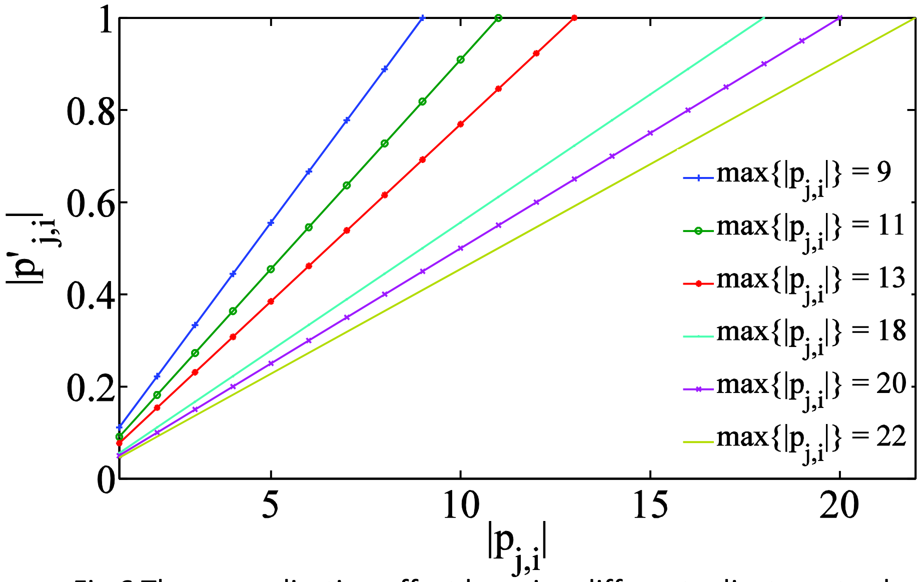


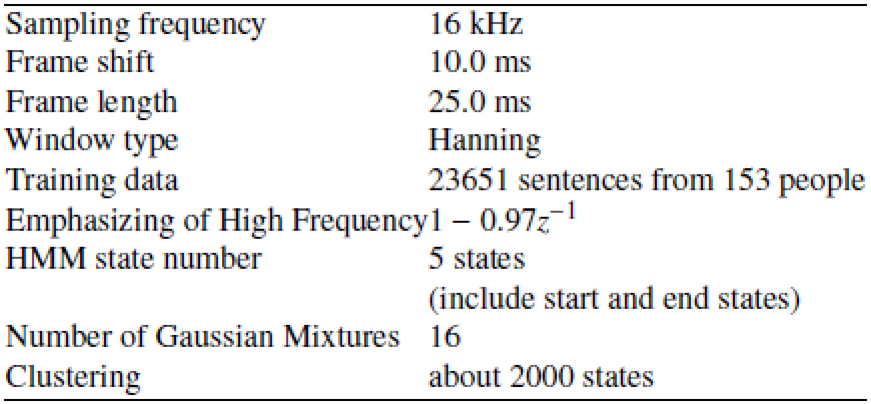
Fig.6 The normalization effect by using different adjustment values

Block Based DRA (5)

➢ Simulation

- The proposed method effectively increases the similarity between clean and noisy speech features, especially in the marked position from A to F.

Table 2: Acoustic analysis conditions



Sampling frequency	16 kHz
Frame shift	10.0 ms
Frame length	25.0 ms
Window type	Hanning
Training data	23651 sentences from 153 people
Emphasizing of High Frequency	$1 - 0.97z^{-1}$
HMM state number	5 states (include start and end states)
Number of Gaussian Mixtures	16
Clustering	about 2000 states

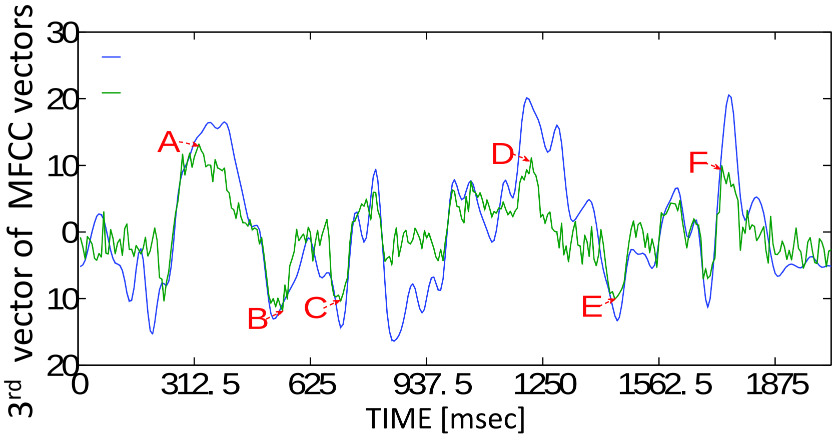


Fig. 7 Before DRA in CSR

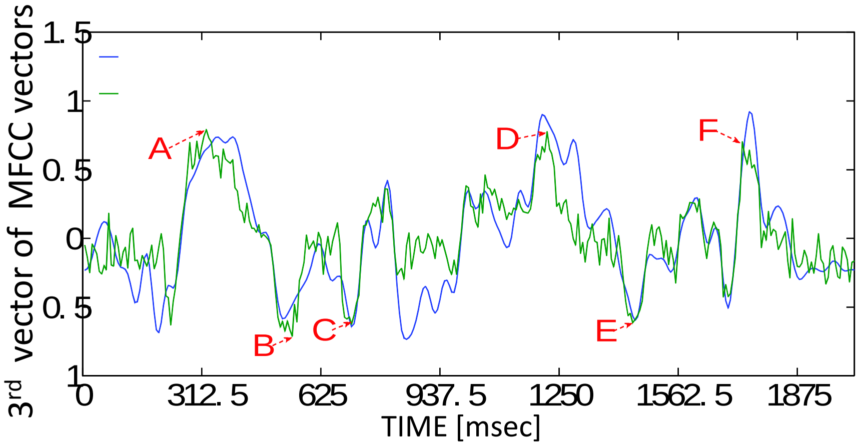


Fig. 8 Results for block based DRA in CSR

Results

Table 3: Noise types

Noise Type (15 kinds)	babble	buccaneer1	buccaneer2	destroyerengniner	destroyerops
	f16	factory1	factory2	hfchannel	leopard
	m109	machinegun	pink	volvo	white

$$\text{Percent Accuracy} = \frac{N-D-S-I}{N} \times 100\%$$

- Shows accurately the total performance

$$\text{Percent Correct} = \frac{N-D-S}{N} \times 100\%$$

- Shows the correct word recognition rate

N: Total number of words
D:Deletion errors
S:Substitution errors
I: Insertion errors

Table 4: Average recognition rates under clean and different SNR conditions

		Proposed		Original	
		Corr.	Acc.	Corr.	Acc.
known (clean)		93.22	92.29	92.69	91.49
unknown (clean)		83.90	82.43	82.77	81.52
known	SNR=20dB	80.08	77.72	77.80	75.82
	SNR=15dB	68.06	64.81	61.10	58.40
	SNR=10dB	49.93	46.25	39.23	36.04
unknown	SNR=20dB	73.76	71.31	72.46	70.23
	SNR=15dB	63.01	60.14	58.18	55.95
	SNR=10dB	47.91	44.75	37.06	35.19

Thank you!

Question?