Constructing Modulation Frequency Domain-Based Feature for Robust Speech Recognition

Author : Jeih-Weih Hung Wei-Yi Tsai

Professor:陳嘉平

Reporter:葉佳璋

Outline

- Introduction
- Temporal Filter Design In The Modulation Frequency Domain
- Constrained Optimization Problem
- Experimental Results and Discussion

Introduction

- Data-driven temporal filtering approaches based on a specific optimization technique have capable of enhancing the decimation and robustness of speech feature.
- The filter in these approaches are often obtained with statistics of the features in the temporal domain.

Introduction

- In this paper, we derive new data-driven temporal filters that employ the statistics of modulation spectra of the speech features.
- Three new temporal filtering approaches are proposed and based on three different constrained versions
  - Linear discriminant analysis (LDA)
  - Principal component analysis (PCA)
  - Minimum class distance(MCD)

Temporal Filter Design In The Modulation Frequency Domain

- An ordered sequence of M-dimensional feature vectors {x(n), n=1, 2, 3, …, N}, where n is the time index.
- Denote $x_m(n) = x(n, m)$, n=1,2, …,N, m=1,2,…, M noting that n is the time index and m is the feature index.

Temporal Filter Design In The Modulation Frequency Domain

$$
\begin{array}{cccccccc}
\begin{bmatrix} x(1,1) \\ x(1,2) \\ \vdots \\ x(1,m) \\ \vdots \\ x(1,M) \end{bmatrix}
\begin{bmatrix} x(2,1) \\ x(2,2) \\ \vdots \\ x(2,m) \\ \vdots \\ x(2,M) \end{bmatrix}
\begin{bmatrix} x(3,1) \\ x(3,2) \\ \vdots \\ x(3,m) \\ \vdots \\ x(3,M) \end{bmatrix}
& \cdots &
\begin{bmatrix} x(n,1) \\ x(n,2) \\ \vdots \\ x(n,m) \\ \vdots \\ x(n,M) \end{bmatrix}
& \cdots &
\begin{bmatrix} x(N,1) \\ x(N,2) \\ \vdots \\ x(N,m) \\ \vdots \\ x(N,M) \end{bmatrix}
&
\begin{array}{l} \rightarrow \{x_1(n)\} \\ \rightarrow \{x_2(n)\} \\ \quad\ \vdots \\ \rightarrow \{x_m(n)\} \\ \quad\ \vdots \\ \rightarrow \{x_M(n)\} \end{array} \\
\mathbf{x}(1) \quad \mathbf{x}(2) \quad \mathbf{x}(3) & \cdots & \mathbf{x}(n) & \cdots & \mathbf{x}(N) &
\end{array}
$$

Fig. 1. Representation of the time trajectories of feature parameters.

Temporal Filter Design In The Modulation Frequency Domain

- When an FIR filter $h_m(n)$ with length L is applied to $\{x_m(n)\}$, the output sample $\{y_m(n)\}$ are

$$y_m(n) = \sum_{u=0}^{L-1} h_u(u)\, x_m(n-u)$$

- $\{x_m(n)\}$ is processed by a running window of length L to obtain a set of L-length segments.

$$\tilde{x}_m(n) = x_m(n-L+1)\ldots\ \ x_m(n-1)\ \ x_m(n)$$

- Padding $h_m(n)$ and each of $\tilde{x}_m(n)$ with K-L zeros, $K \geq 2L$ where $H_m(k)$ and $X_m(n,k)$ are their K-point DFT.

By Plancherel theorem

$$y_m(n) = \frac{1}{K}\sum_{k=0}^{K-1} H_m(k) X_m^*(n,k) = \frac{1}{K}\sum_{k=0}^{K-1} H_m^*(k) X_m(n,k)$$

Temporal Filter Design In The Modulation Frequency Domain

- The instantaneous energy of the temporal filter output at time n is simply

$$\left|y(n)\right|^2 = \left|\frac{1}{K}\sum_{k=0}^{K-1} H_m(k) X_m^*(n,k)\right|^2$$

- It can be shown

$$\left|y(n)\right|^2 = \frac{2(k+2)}{K^2}\sum_{k=0}^{K/2}\left|H_m(k)\right|^2\left|X_m(n,k)\right|^2 = \frac{2(k+2)}{K^2} H^T X(n)$$

- If we define the instantaneous modulation spectral energy of the filter output as

$$\varepsilon_Y(n) = \sum_{k=0}^{K/2}\left|H_m(k)\right|^2\left|X_m(n,k)\right|^2 = H^T X(n)$$

$Let\ \{c_i, 1 \le i \le L\}\ be\ a\ set\ of\ real\ number.\ Then$

$$\left(\sum_{i=1}^{L} c_i\right)^2 = \sum_{i=1}^{L} c_i \sum_{j=1}^{L} c_j = \sum_{i=1}^{L}\sum_{j=1}^{L} c_i c_j$$

$$\le \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L}\left(c_i^2 + c_j^2\right) \quad \left(\because \left(c_i - c_j\right)^2 = c_i^2 + c_j^2 - 2c_i c_j \ge 0\right)$$

$$= \frac{1}{2}\sum_{i=1}^{L}\left(Lc_i^2 + \sum_{j=1}^{L} c_j^2\right) = \frac{1}{2}\left(L\sum_{i=1}^{L} c_i^2 + L\sum_{j=1}^{L} c_j^2\right) = L\sum_{i=1}^{L} c_i^2$$

$$Therefore, \left(\sum_{i=1}^{L} c_i\right)^2 \le L\sum_{i=1}^{L} c_i^2$$

both $\{\mathrm{H}(\mathrm{k})\}$ and $\{\mathrm{X}(\mathrm{n,k})\}$ are conjugate symmertic with respect to k=K/2. That is

$H(k) = \mathrm{H}^*(\mathrm{K\text{-}k})$,

$X(n, k/2) = X^*(n, K-k)\ for\ 1 \le k \le k/2 - 1$

and $\mathrm{H}(0)$,$\mathrm{H}(\mathrm{k}/2)$, $\mathrm{X}(\mathrm{n},0)$, and $\mathrm{X}(\mathrm{n,k}/2)$ are all real numbers.

Therefore, the instataneous energy of the temporal filter output is

$$\left|y(n)\right|^2 = \left|\frac{1}{K}\sum_{k=0}^{K-1} H(k)X^*(n,k)\right|^2$$

$$= \frac{1}{K^2}\left|H(0)\,X(n,0) + H(K/2)\,X(n,K/2) + \sum_{k=1}^{K/2-1}\left(H(k)\,X^*(n,k) + H^*(k)\,X(n,k)\right)\right|^2$$

$$\leq \frac{1}{K^2}\left|\left|H(0)\right|\left|X(n,0)\right| + \left|H(K/2)\right|\left|X(n,K/2)\right| + 2\sum_{k=1}^{K/2-1}\left|H(k)\right|\left|X(n,k)\right|\right|^2$$

$$\leq \frac{1}{K^2}\left|2\sum_{k=1}^{K/2}\left|H(k)\right|\left|X(n,k)\right|\right|^2$$

$$\leq \frac{4}{K^2}\left(\frac{K}{2}+1\right)\sum_{k=1}^{K/2}\left|H(k)\right|^2\left|X(n,k)\right|^2$$

$$= \frac{2(K+2)}{K^2}\sum_{k=1}^{K/2}\left|H(k)\right|^2\left|X(n,k)\right|^2$$

Therefore

$$\left|y(n)\right|^2 \leq \frac{2(K+2)}{K^2}\sum_{k=1}^{K/2}\left|H(k)\right|^2\left|X(n,k)\right|^2$$

Temporal Filter Design In The Modulation Frequency Domain

- Rewritten as $\left|y(n)\right|^2 \le \frac{2(K+2)}{K^2}\varepsilon_Y(n)$
- $\varepsilon_Y(n)$ can be approximately used to characterize the behavior of $\left|y(n)\right|^2$.
- Now the optimal vector H is found to maximize a specific objective function of $\varepsilon_Y(n)$, which is related to the statistics of X.
- We apply three optimization techniques: C-LDA, C-PCA, and C-MCD.

Constrained Optimization Problem

- Since each component of H is constrained to be real and nonnegative, it give rise to a constrained optimization problem.

$$H^{*} = \arg\max_{H} J\left(H\right),\ \text{subject to H} \geq 0$$

- In order to deal with the nonnegative for H, we introduce an intermediate variable vector

$$\bar{H} = \left[\bar{H}_0 \ \ \bar{H}_1 \ \ \cdots \bar{H}_{K/2}\right]$$

$$H_k = \left(\frac{\exp\left(\bar{H}_k\right)}{\sum_{m=0}^{K/2} \exp\left(\bar{H}_m\right)} \right)^{\frac{1}{p}},\ k = 0,1,2,\ldots K/2 \qquad \sum_{m=0}^{K/2} \bar{H}_k^{P} = 1$$

Constrained Optimization Problem

- Find the optimal H that maximizes *J*(H) through the intermediate vector $\bar{H}$.
- We use gradient decent algorithm to update $\bar{H}$

$$\bar{H}^{(\theta+1)} = \bar{H}^{(\theta)} + \varepsilon \left. \frac{\partial J}{\partial \bar{H}} \right|_{\bar{H} = \bar{H}^{(\theta)}}$$

- Where ε is the step size, and $\frac{\partial J}{\partial \bar{H}} = \frac{\partial H}{\partial \bar{H}} \frac{\partial J}{\partial H}$

$$\left(\frac{\partial H}{\partial \bar{H}} \right)_{ij} = \frac{1}{p} \left(\frac{\exp\left(\bar{H}_j\right)}{\sum_{m=0}^{K/2} \exp\left(\bar{H}_m\right)} \right)^{\frac{1}{p}-1} \times \left(\frac{\exp\left(\bar{H}_j\right) \delta_{ij} \sum_{m=0}^{K/2} \exp\left(\bar{H}_m\right) - \exp\left(\bar{H}_i + \bar{H}_j\right)}{\left(\sum_{m=0}^{K/2} \exp\left(\bar{H}_m\right) \right)^2} \right), 0 \le i, j \le K/2$$

- $\partial J / \partial \bar{H}$ is determined by the chosen objective function *J*(H).

Constrained Linear Discriminant Analysis

- LDA has been widely applied in pattern recognition. Its goal is to find the most discriminative representation of the data.
- To derive H is called constrained LDA(C-LDA).
- The squared magnitude spectrum X(n) is first labeled as one of the J classes or speech model.

Constrained Linear Discriminant Analysis

- The labeling process can be performed by means of the time alignment with pretrained models .
- Then the mean and covariance matrix for those X(n) labeled as belonging to each class j

$$\mu^{(j)} = \frac{1}{N_j}\sum_{n=1}^{N_j} X^{(j)}(n) \ and \ \Sigma^{(j)} = \frac{1}{N_j}\sum_{n=1}^{N_j}\left(X^{(j)}(n)-\mu^{(j)}\right)\left(X^{(j)}(n)-\mu^{(j)}\right)^T$$

Where $X^{(j)}(n)$ denote X(n) as belong to the jth class, N_j is the total number such $X^{(j)}(n)$.

Constrained Linear Discriminant Analysis

- The between-class and the within-class matrix of X can be defined

$$S_B = \sum_{j=1}^{J} N_j \left(\mu^{(j)} - \mu\right)\left(\mu^{(j)} - \mu\right)^T \; and \; S_w = \sum_{j=1}^{J} N_j \Sigma^{(j)} \; where \; \mu = \left(1/\Sigma_{j=1}^{J} N_j\right)\Sigma_{j=1}^{J} N_j \mu^{(j)}$$

- Denoting σ_B^2 and σ_W^2 as the between-class and within-class variance of ε_Y, the object function

$$J_{LDA} = \frac{\sigma_B^2}{\sigma_W^2} = \frac{H^T S_B H}{H^T S_W H}, \; subject \; to \; H \geq 0$$

$$\frac{\partial J_{LDA}}{\partial H} = \frac{2\left(H^T S_W H\right) S_B H - 2\left(H^T S_B H\right) S_W H}{\left(H^T S_W H\right)^2}$$

Constrained Principal Component Analysis

- C-PCA different from C-LDA, the X(n) do not need to be labeled.
- The mean and covariance of X can be estimate as

$$\mu = \frac{1}{N}\sum_{n=1}^{N} X(n) \ and \ \Sigma = \frac{1}{N}\sum_{n=1}^{N}\left(X(n)-\mu\right)\left(X(n)-\mu\right)^T$$

- σ^2 as the global variance of the ε_Y , the object function with C-PCA is

$$J_{PCA} = \sigma^2 = H^T \sum H, \ subject \ to \ H \geq 0 \qquad \frac{\partial J_{LDA}}{\partial H} = 2\sum H$$

Constrained Maximum Class Distance

- C-MCD is similar to C-LDA, X(n) is first labeled as one of the J classes or speech models.
- For simplicity, we assume $X^{(j)}$ is multivariate Gaussian distributed with mean $\mu^{(j)}$ and covariance $\Sigma^{(j)}$.
- The filter output ε_Y for the jth class denote as is a univariate Gaussian with mean $H^T\mu^{(j)}$ and variance $H^T\Sigma^{(j)}H$.
- The probability density function $g^{(j)}(x)$ of $\varepsilon_Y^{(j)}$ is

$$g^{(j)}(x) = N\left(x; H^T\mu^{(j)}, H^T\Sigma^{(j)}H\right) = \frac{1}{\sqrt{(2\pi)H^T\Sigma^{(j)}H}}\exp\left[-\frac{\left(x - H^T\mu^{(j)}\right)^2}{2H^T\Sigma^{(j)}H}\right]$$

Constrained Maximum Class Distance

- The distance between two different classes i and j of the filter output spectral energy ε_Y as

$$d_{ij} \triangleq \int_{-\infty}^{\infty} g^{(i)}(x)\log\frac{g^{(i)}(x)}{g^{(j)}(x)}dx = \log\frac{H^T\Sigma^{(j)}H}{H^T\Sigma^{(i)}H} + \frac{H^T\left(\mu^{(i)}-\mu^{(j)}\right)\left(\mu^{(i)}-\mu^{(j)}\right)H}{H^T\Sigma^{(j)}H} + \frac{H^T\Sigma^{(i)}H}{H^T\Sigma^{(j)}H} - 1$$

Which is Kullback-Leibler divergence between two Gaussian probability distribution.

- The objective function to be maximized is the sum of all class distances.

$$J_{MCD}(H) = \sum_i\sum_{i\neq j} d_{ij} \triangleq \sum_i\sum_{i\neq j}\left[\log\frac{H^T\Sigma^{(j)}H}{H^T\Sigma^{(i)}H} + \frac{H^T\left(\mu^{(i)}-\mu^{(j)}\right)\left(\mu^{(i)}-\mu^{(j)}\right)H}{H^T\Sigma^{(j)}H} + \frac{H^T\Sigma^{(i)}H}{H^T\Sigma^{(j)}H} - 1\right],\ \textit{subject to } H \geq 0 \quad A^{(i,j)} = \left(\mu^{(i)}-\mu^{(j)}\right)\left(\mu^{(i)}-\mu^{(j)}\right)^T$$

$$\frac{\partial J_{MCD}(H)}{\partial H} = \sum_i\sum_{i\neq j} d_{ij} \triangleq \sum_i\sum_{i\neq j}\left[2\left(\frac{H^T\Sigma^{(i)}H}{H^T\Sigma^{(j)}H}\right)\times\frac{\left(H^T\Sigma^{(i)}H\right)\Sigma^{(j)}H - \left(H^T\Sigma^{(j)}H\right)\Sigma^{(i)}H}{\left(H^T\Sigma^{(i)}H\right)^2} + \frac{2\left(H^T\Sigma^{(j)}H\right)A^{(i,j)}H - 2\left(H^TA^{(i,j)}H\right)\Sigma^{(j)}H}{\left(H^T\Sigma^{(j)}H\right)^2} + \frac{2\left(H^T\Sigma^{(j)}H\right)\Sigma^{(i)}H - 2\left(H^T\Sigma^{(i)}H\right)\Sigma^{(j)}H}{\left(H^T\Sigma^{(j)}H\right)^2}\right]$$

Experimental Results and Discussion

- AURORA Projection Database Version 2.0.
  - Each utterance in the clean training set is first converted into a sequence of 13-dimensional MFCCs.
  - The Filter length L, the DFT size K, and the exponent P are set to 101, 256, 4.
  - The proposed three temporal filter are then obtained.
  - Plus their delta and delta-delta features, then 39-dimensional feature are finally used.

Comparative Performance Analysis

- Some details of these techniques
  - RASTA
  - Linear-phase RASTA: a symmetric FIR filter is used to approximate the RASTA.
  - Spatial-temporal LDA: a supervector is constructed by concatenating nine neighboring 24-dimention(216) log spectral feature. These supervector are transformed by the projection matrix(whose obtained by LDA) to constitute the 39-dimension features.
  - Temporal LDA: obtained directly according to the characteristic of the features in the temporal domain
  - linear-phase temporal LDA: similar to LP RASTA.

Test	System	clean	20dB	15dB	10dB	5dB	0dB	-5dB	average (0~20dB)	Relative WER reduction
Test Set A	plain MFCC	98.91	94.99	86.93	67.28	39.36	17.07	8.40	61.13	
	RASTA	98.70	95.91	89.50	66.96	34.63	20.06	11.94	61.41	0.72
	LP RASTA	98.94	96.95	92.47	75.81	43.84	23.18	13.00	66.45	13.69
	ST_LDA	98.52	88.65	70.59	43.50	20.77	9.84	7.15	46.67	-37.20
	ST_LDA+D+A	98.78	95.45	88.05	70.05	43.07	18.00	7.09	62.92	4.61
	T_LDA	98.63	94.49	83.67	60.43	30.60	11.11	6.56	56.06	-13.04
	LP T_LDA	98.79	94.25	86.28	71.80	45.11	22.53	12.06	63.99	7.36
	MF_C-LDA	98.80	96.67	92.94	81.34	57.17	26.24	8.92	70.87	25.06
	MF_C-PCA	98.66	95.54	89.16	73.15	46.80	22.32	9.77	65.39	10.96
	MF_C-MCD	98.18	95.89	91.59	78.97	53.74	26.01	12.81	69.24	20.86

Test	System	clean	20dB	15dB	10dB	5dB	0dB	-5dB	average (0~20dB)	Relative WER reduction
Test Set B	plain MFCC	98.91	92.35	80.79	58.06	32.04	14.63	7.92	55.57	
	RASTA	98.70	96.74	91.93	74.90	44.30	23.66	13.02	66.31	24.17
	LP RASTA	98.94	97.38	93.80	81.52	53.04	27.99	14.28	70.74	34.14
	ST_LDA	98.52	83.90	64.89	37.92	16.19	6.66	5.63	41.91	-30.74
	ST_LDA+D+A	98.78	94.95	86.92	67.58	42.99	19.87	6.45	62.46	15.51
	T_LDA	98.63	93.72	87.06	66.18	36.54	17.85	9.70	60.27	10.58
	LP T_LDA	98.79	93.25	87.05	75.85	51.66	27.34	14.19	67.03	25.79
	MF_C-LDA	98.80	96.20	91.74	80.28	55.81	23.57	5.08	69.52	31.40
	MF_C-PCA	98.66	92.30	83.71	65.43	38.92	16.90	7.72	59.45	8.73
	MF_C-MCD	98.18	95.32	91.24	80.70	57.89	29.96	12.46	71.02	34.77

Test	System	clean	20dB	15dB	10dB	5dB	0dB	-5dB	average (0~20dB)	Relative WER reduction
Test Set C	plain MFCC	99.00	94.83	88.66	75.23	50.85	23.83	11.4	66.68	
	RASTA	98.69	95.40	87.70	63.31	34.94	21.12	12.72	60.49	-18.58
	LP RASTA	99.09	96.44	90.82	70.44	40.61	23.03	14.57	64.27	-7.23
	ST_LDA	98.07	86.06	72.19	52.51	31.65	17.20	9.55	51.92	-44.30
	ST_LDA+D+A	98.52	95.62	89.72	72.79	49.18	25.73	12.97	66.61	-0.21
	T_LDA	98.71	89.57	78.18	58.28	36.56	17.28	9.52	55.97	-32.14
	LP T_LDA	98.65	90.89	82.71	67.14	41.44	20.20	11.64	60.47	-18.64
	MF_C-LDA	98.87	96.00	91.70	80.66	59.34	32.28	15.52	71.99	15.94
	MF_C-PCA	98.77	94.41	88.22	74.13	52.50	26.86	12.97	67.22	1.62
	MF_C-MCD	98.36	94.85	88.88	73.48	46.13	24.81	16.25	65.63	-3.15

Comparative Performance Analysis

- We attempt to integrate the three proposed temporal filters with some other robustness technique.
  - CMVN:

$$y_{m,CMVN}(n) = \frac{\left[x_m(n) - \mu_m\right]}{\sigma^2}$$

  - CGN:

$$y_{m,CGN}(n) = \frac{\left[x_m(n) - \mu_m\right]}{\max\left[x_m(n)\right] - \min\left[x_m(n)\right]}$$

  - AFE

WORD RECOGNITION ACCURACIES (%) AND RELATIVE WER REDUCTION (%) AS COMPARED TO THE MFCC BASELINE FOR VARIOUS APPROACHES AT DIFFERENT SNR VALUES BUT AVERAGED OVER ALL THE NOISE TYPES IN TEST SET A OF THE AURORA-2 DATABASE

Test	System	clean	20dB	15dB	10dB	5dB	0dB	-5dB	average (0~20dB)	Relative WER reduction
Test Set A	Plain MFCC	98.91	94.99	86.93	67.28	39.36	17.07	8.40	61.13	
	CMVN	98.98	95.98	91.66	80.48	57.40	26.40	10.96	70.38	23.80
	CMVN+ C-LDA	98.85	97.02	94.15	87.34	71.81	41.97	16.13	78.46	44.58
	CMVN+ C-PCA	98.51	96.48	93.42	86.47	72.87	48.76	22.68	79.60	47.52
	CMVN+ C-MCD	98.28	96.15	93.11	86.30	72.25	47.81	21.85	79.12	46.28
	CGN	98.91	96.48	93.16	85.29	69.30	40.73	15.46	76.99	40.80
	CGN+ C-LDA	98.79	96.92	94.31	88.56	76.65	52.20	22.37	81.73	53.00
	CGN+ C-PCA	98.60	96.29	93.55	87.97	76.84	54.07	23.95	81.74	53.02
	CGN+ C-MCD	98.51	96.23	93.49	87.39	74.26	48.83	20.56	80.04	48.65
	AFE	99.10	98.14	96.75	92.99	84.06	60.72	28.86	86.53	65.35
	AFE+ C-LDA	98.85	97.77	96.52	93.17	85.35	63.65	29.14	87.29	67.30
	AFE+ C-PCA	98.95	97.94	96.54	92.94	83.63	60.91	28.88	86.39	64.99
	AFE+ C-MCD	98.85	97.79	96.58	93.30	85.40	63.95	29.75	87.40	67.58

WORD RECOGNITION ACCURACIES (%) AND RELATIVE WER REDUCTION (%) AS COMPARED TO THE MFCC BASELINE FOR VARIOUS APPROACHES AT DIFFERENT SNR VALUES BUT AVERAGED OVER ALL THE NOISE TYPES IN TEST SET B OF THE AURORA-2 DATABASE

Test	System	clean	20dB	15dB	10dB	5dB	0dB	-5dB	average (0~20dB)	Relative WER reduction
Test Set B	plain MFCC	98.91	92.35	80.79	58.06	32.04	14.63	7.92	55.57	
	CMVN	98.98	96.41	92.15	81.78	58.69	26.47	10.98	71.10	34.95
	CMVN+ C-LDA	98.85	97.23	94.75	88.41	72.18	42.33	15.28	78.98	52.69
	CMVN+ C-PCA	98.70	96.92	94.44	88.49	74.08	48.62	20.87	80.51	56.13
	CMVN+ C-MCD	98.12	95.99	93.10	86.62	73.11	50.1	22.85	79.78	54.49
	CGN	98.91	96.86	94.09	87.01	70.20	39.95	15.09	77.62	49.63
	CGN+ C-LDA	98.79	97.06	95.07	89.79	77.39	51.36	20.33	82.13	59.78
	CGN+ C-PCA	98.60	96.41	94.17	89.10	77.77	54.87	22.94	82.46	60.52
	CGN+ C-MCD	98.51	96.59	94.30	88.53	76.08	50.84	20.33	81.27	57.84
	AFE	99.10	98.01	96.29	92.26	81.27	57.59	26.32	85.09	66.44
	AFE+ C-LDA	98.85	97.22	95.43	91.54	82.12	60.16	28.40	85.29	66.89
	AFE+ C-PCA	98.95	97.62	95.70	91.20	80.10	56.76	25.43	84.27	64.60
	AFE+ C-MCD	98.95	97.32	95.47	91.73	82.13	60.41	28.19	85.41	67.16

Word Recognition Accuracies (%) and Relative WER Reduction (%) as Compared to the MFCC Baseline for Various Approaches at Different SNR Values but Averaged Over the Two Noise Types in Test Set C of the Aurora-2 Database

Test	System	clean	20dB	15dB	10dB	5dB	0dB	-5dB	average (0~20dB)	Relative WER reduction
Test Set C	plain MFCC	99.00	94.83	88.66	75.23	50.85	23.83	11.4	66.68	
	CMVN	99.12	95.51	88.71	74.21	51.25	24.30	10.49	66.80	0.36
	CMVN+C-LDA	98.83	96.80	93.30	84.63	66.66	37.82	14.90	75.84	27.49
	CMVN+C-PCA	98.66	96.10	92.43	84.67	69.79	48.04	23.04	78.20	34.57
	CMVN+C-MCD	98.25	95.62	92.37	84.29	70.02	48.02	23.81	78.06	34.15
	CGN	98.96	96.13	92.03	83.18	64.89	35.29	13.81	74.30	22.87
	CGN+C-LDA	98.91	96.94	93.95	87.80	74.11	47.11	19.55	79.98	39.92
	CGN+C-PCA	98.68	96.16	93.65	87.61	75.53	51.55	22.59	80.90	42.68
	CGN+C-MCD	98.63	96.40	93.11	86.44	72.29	46.81	19.68	79.01	37.00
	AFE	99.01	97.53	95.69	90.33	78.99	53.12	26.42	83.13	49.37
	AFE+C-LDA	98.86	97.52	95.74	91.67	81.29	56.47	25.44	84.54	53.57
	AFE+C-PCA	98.92	97.23	95.08	90.29	79.09	54.87	26.60	83.31	49.91
	AFE+C-MCD	98.87	97.53	95.83	91.81	82.07	58.10	26.63	85.07	55.19