Discriminative Speaker Adaptation with Eigenvoices

Jun Luo, Zhijian Ou, Zuoying Wang

Department of Electronic Engineering
Tsinghua University, Beijing, China
{luojun,ozj}@thsp.ee.tsinghua.edu.cn

Abstract

Eigenvoice is an effective speaker adaptation approach and capable of balancing the performance and the requirement for a large amount of adaptation data. However, the conventional Maximum Likelihood Eigen-Decomposition (MLED) method in eigenvoice adaptation is based on Maximum Likelihood (ML) criterion and suffers from the unrealistic assumption made by HMM on speech process, so alternative schemes may be more effective to improve the performance. In this paper, we propose a new discriminative adaptation algorithm called Maximum Mutual Information Eigen-Decomposition (MMIED) in which the mutual information between the training word sequences and the observation sequences is maximized. By the use of word lattice, the competing word hypotheses are taken into account to make the estimation more discriminative. MLED, MMIED and Maximum *a Posteriori* Eigen-Decomposition (MAPED) which is based on Maximum *a Posteriori* (MAP) criterion were all experimented to give a comprehensive comparison. Results showed that MMIED outperformed both MLED and MAPED.

1. Introduction

A speaker-dependent (SD) system performs better than a speaker-independent (SI) one, however, it requires a large amount of speaker-specific data, which may be impracticable in some applications. Speaker adaptation is to use limited amount of speaker-specific data to achieve performance approaching that of an SD system for a speaker [1]. Currently, dominant speaker adaptation techniques could be classified into three categories: the Maximum *a Posteriori* (MAP) [2] adaptation approaches, transformation-based adaptation approaches including Maximum Likelihood Linear Regression (MLLR) [3], and approaches related to speaker clustering such as the eigenvoice [4, 5] method.

Eigenvoice adaptation was shown to be an effective method for fast speaker adaptation with few data, and it has been applied to LVCSR system successfully. Experimental results showed that even with several seconds it could gain considerable improvements [5].

Most of these works were done based on Maximum Likelihood (ML) criterion. However, it is known that ML criterion would be optimal only if the following two conditions are satisfied [6]:

- Training data set is infinite;
- The true distribution of the speech process is an HMM.

Since the assumption made by HMM on speech process is in fact inaccurate, and training data is fairly limited in speaker adaptation, discriminative estimation such as Maximum Mutual Information (MMI) estimation are potentially more effective than ML estimation. Discriminative speaker adaptation with linear regression [7, 8] has been successfully applied with improved performance.

In this paper, we propose the MMI adaptation with eigenvoices. By the use of word lattice, possible competing word hypotheses are taken into account to give better discrimination. The re-estimation formula proposed by Gunawardana et al [7] for Extended Baum-Welch (EBW) algorithm is used.

The paper is organized as follows. In section 2 we review the conventional ML adaptation with eigenvoices. The MAP estimation is also introduced in section 3 to give a comprehensive comparison of existing algorithms. In section 4 the new estimation algorithm based on MMI criterion is discussed in detail. Section 5 gives the experimental results and section 6 concludes the paper.

2. ML Adaptation with Eigenvoices

The eigenvoice approach begins with a reference set of well-trained SD models. For each of the SD model, a supervector is formed by concatenating all of the mean vector parameters (supposed to be D-dimension). Then a dimensionality reduction technique principal component analysis (PCA) [9] is used to find the eigenvectors. The supervector for a new speaker is assumed to be a linear combination of selected eigenvectors.

Denote $\mathbf{C}$ as the sample covariance matrix of the supervectors, then by PCA method, $\mathbf{C}$ can be expressed as:

$$\mathbf{C} = [\mathbf{e}(1), \ldots, \mathbf{e}(D)] diag(\lambda_1, \ldots, \lambda_D) [\mathbf{e}(1), \ldots, \mathbf{e}(D)]^T \quad (1)$$

where $\mathbf{e}(i)$, $i = 1, 2, \ldots, D$, are eigenvectors and λ_i are the corresponding eigenvalues of $\mathbf{C}$ in descendent order ($\lambda_1 \geq \lambda_2 \ldots \geq \lambda_D$) which also represent their contributions to speaker variation. The top K eigenvectors, named as 'eigenvoices' are selected (where $K \ll D$). They account for most of the variation in the reference speakers. Let $\mathbf{e}(0)$ be the mean supervector, then for a new speaker, the supervector $\mathbf{S}$ could be represented as follows.

$$\mathbf{S} = \mathbf{e}(0) + \sum_{k=1}^{K} (x(k) \times \mathbf{e}(k)) \quad (2)$$

Given the adaptation data from a new speaker, only the coefficient vector $\mathbf{x} = (x_1, x_2, \ldots, x_K)^T$ of size K needs to be estimated.

Suppose that the single Gaussian state-output distribution is used. Given the adaptation data $\mathbf{O} = \mathbf{o}_1^T \triangleq \mathbf{o}_1, \cdots, \mathbf{o}_T$ with the corresponding transcription $\mathbf{W} = \mathbf{w}_1^N \triangleq w_1, \cdots, w_N$, let:

n: the dimension of the feature vector;
$\mathbf{o}_t$: the feature vector at time t;
μ_s: the mean vector for state s;
$\mathbf{C}_s$: the covariance for state s;
$\mathbf{e}_s(j)$: the subvector of eigenvoice j corresponding to state s;

Define $\gamma_s(t)$ as the occupation probability for state s at time t given sentence transcriptions and $\gamma_s^g(t)$ as the occupation probability for state s at time t without transcription[1], i.e.,

$$\gamma_s(t) = P(s_t = s|\mathbf{W}, \mathbf{O}) \tag{3}$$

$$\gamma_s^g(t) = P(s_t = s|\mathbf{O}) \tag{4}$$

Maximum Likelihood estimation method, called Maximum Likelihood Eigen-Decomposition (MLED) in [4] aims to maximize the likelihood $P(\mathbf{W}, \mathbf{O}|\mathbf{x})$ with respect to the unknown coefficient vector $\mathbf{x}$. This is done by iteratively maximizing an auxiliary function $Q(\mathbf{x}', \mathbf{x})$ as following:

$$Q(\mathbf{x}', \mathbf{x}) = P(\mathbf{W}, \mathbf{O}|\mathbf{x}') \times \sum_s \sum_t \gamma_s(t)[\log P(\mathbf{o}_t|s, \mathbf{x})] \tag{5}$$

where $\mathbf{x}'$ is current model, and $\mathbf{x}$ is the model to be estimated.

$$\begin{aligned}\log P(\mathbf{o}_t|s, \mathbf{x}) = &-\frac{n}{2}\log(2\pi) - \frac{1}{2}log|\mathbf{C}_s| \\ &-\frac{1}{2}(\mathbf{o}_t - \mu_s)^T\mathbf{C}_s^{-1}(\mathbf{o}_t - \mu_s)\end{aligned} \tag{6}$$

According to Equation 2, we get:

$$\mu_s = \mathbf{e}_s(0) + \sum_{k=1}^{K}(x(k) \times \mathbf{e}_s(k)) \tag{7}$$

Denote $\hat{\mathbf{x}}$ as the re-estimation of coefficient vector which maximizes $Q(\mathbf{x}', \mathbf{x})$ over $\mathbf{x}$. Let $\partial Q/\partial x(j) = 0, j = 1, \ldots, K$, we obtain the update equation for each $\hat{x}(j), j = 1, \ldots, K$:

$$\begin{aligned}&\sum_s \sum_t \gamma_s(t)(\mathbf{e}_s(j))^T\mathbf{C}_s^{-1}(\mathbf{o}_t - \mathbf{e}_s(0)) \\ &= \sum_s \sum_t \gamma_s(t) \times \sum_{k=1}^{K} \hat{x}(k)(\mathbf{e}_s(k))^T\mathbf{C}_s^{-1}\mathbf{e}_s(j)\end{aligned} \tag{8}$$

3. MAP Adapation with Eigenvoices

Note that the use of eigenvoices imposes a strong constraint on mean vectors, it is beneficial to explore the *prior* information in model estimations. Maximum *a Posteriori* criterion could be used for this purpose. The MAP estimation of acoustic model parameters θ is:

$$\hat{\theta} = \underset{\theta}{argmax}[P(\mathbf{W}, \mathbf{O}|\theta)P_0(\theta)] \tag{9}$$

where $P_0(\theta)$ denotes the prior probability of the known parameters θ. The estimation procedure under MAP criterion is called Maximum *a Posteriori* Eigen-Decomposition (MAPED) [10].

In [10], coefficients $x(k), k = 1, \ldots, K$ are modeled by a Gaussian distribution with mean $\mu_{x(k)}$ and variance $\sigma^2_{x(k)}$ respectively. Thus, the prior probability $P_0(\mathbf{x})$ is given by:

$$P_0(\mathbf{x}) = \sum_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma^2_{x(k)}}} \exp - \frac{(x(k) - \mu_{x(k)})^2}{2\sigma^2_{x(k)}} \tag{10}$$

[1] $\gamma_s^g(t)$ will only be used in MMIED, here we define it to keep consistency.

Since $x(k), k = 1, \ldots, K$ are the projections of the speaker supervector to the eigenvectors, the prior probability can be derived directly from the eigen-analysis of covariance matrix $\mathbf{C}$ [11].

Instead of modeling each individual coefficient $x(k)$, multi-dimensional Gaussian distribution is used to model supervector $\mathbf{S}$ with mean $\mathbf{e}(0)$ and covariance $\mathbf{C}$, then the prior probability is given by:

$$P_0(\mathbf{x}) = P_0(\mathbf{S}) = \mathcal{N}(\mathbf{S}|\mathbf{e}(0), \mathbf{C}) \tag{11}$$

Substitute the Eigenvoice expression of the speaker supervector from Equation 2 into Equation 11, and rewrite $\mathbf{C}$ as in Equation 1, then the log prior probability of coefficient vector $\mathbf{x}$ is given by:

$$\begin{aligned}\log P_0(\mathbf{x}) = &-\frac{1}{2}(\sum_{k=1}^{K} x(k)\mathbf{e}(k))^T(\mathbf{e}(1), \ldots, \mathbf{e}(D)) \\ &\times diag(\lambda_1^{-1}, \ldots, \lambda_D^{-1})((\mathbf{e}(1), \ldots, \mathbf{e}(D))^T(\sum_{k=1}^{K} x(k)\mathbf{e}(k)))\end{aligned} \tag{12}$$

Since $\mathbf{e}(i)^T\mathbf{e}(j) = 0, i \neq j$ and $\mathbf{e}(i)^T\mathbf{e}(i) = 1$[2], Equation 11 can be rewriten as[3]:

$$\log P_0(\mathbf{x}) = -\frac{1}{2}\sum_{k=1}^{K} \frac{x(k)^2}{\lambda_k} \tag{13}$$

Equation 13 gives that the mean of each coefficient $x(k)$ is equal to zero, and the variance is equal to the corresponding eigenvalue λ_k.

Redefine auxiliary function as:

$$\begin{aligned}Q'(\mathbf{x}', \mathbf{x}) =&P(\mathbf{W}, \mathbf{O}|\mathbf{x}') \\ \times\{&\sum_s \sum_t \gamma_s(t)[\log P(\mathbf{o}_t|s, \mathbf{x})] + \log P_0(\mathbf{x})\}\end{aligned} \tag{14}$$

Substitute $P_0(\mathbf{x})$ and let $\partial Q'/\partial x(j) = 0, j = 1, \ldots, K$, we get:

$$\begin{aligned}\sum_s \sum_t \gamma_s(t)(\mathbf{e}_s(j))^T\mathbf{C}_s^{-1}(\mathbf{o}_t - \mathbf{e}_s(0)) = &\sum_{k=1}^{K} \hat{x}(k) \\ \times[\sum_s \sum_t \gamma_s(t)(\mathbf{e}_s(k))^T\mathbf{C}_s^{-1}&\mathbf{e}_s(j) + \delta_{k,j} \cdot \frac{1}{\lambda_j}]\end{aligned} \tag{15}$$

for each $\hat{x}(j), j = 1, \ldots, K$. Here:

$$\delta_{k,j} = \begin{cases} 1, & \text{if } k = j \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

4. MMI Adaptation with Eigenvoices

The goal of MMI estimation is to maximize the mutual information [12]:

$$I(\mathbf{W}, \mathbf{O}|\theta) = log\frac{P(\mathbf{w}_1^N, \mathbf{o}_1^T|\theta)}{P(\mathbf{w}_1^N)P(\mathbf{o}_1^T|\theta)} \tag{17}$$

which is equivalent to maximize the conditional likelihood $P(\mathbf{w}_1^N|\mathbf{o}_1^T, \theta)$, namely Conditional Maximum Likelihood (CML) [13] estimation.

[2] It's the property of normalized eigenvectors.

[3] Though it's only derived for the case of full-ranked covariance C, it can also be adapted to other cases since the equation is only associated with the top K eigenvalues.

Gunawardana showed that for carefully chosen constant $d'(\mathbf{s}_1^T)$, iteratively updating the parameters as follows would increase the conditional likelihood:

$$\sum_{\mathbf{s}_1^T}[P(\mathbf{s}_1^T|\mathbf{w}_1^N,\mathbf{o}_1^T,\theta') - P(\mathbf{s}_1^T|\mathbf{o}_1^T,\theta')]\nabla_\theta \log P(\mathbf{o}_1^T|\mathbf{s}_1^T,\theta)|_{\hat{\theta}}$$
$$+\sum_{\mathbf{s}_1^T} d'(\mathbf{s}_1^T)\int P(\bar{\mathbf{o}}_1^T|\mathbf{s}_1^T,\theta')\nabla_\theta \log P(\bar{\mathbf{o}}_1^T|\mathbf{s}_1^T,\theta)|_{\hat{\theta}} d\bar{\mathbf{o}}_1^T = 0 \tag{18}$$

Here θ' and $\hat{\theta}$ means current estimation and re-estimation of model parameters respectively, $\mathbf{s}_1^T \triangleq s_1,\ldots,s_T$ denotes any of the T-length state sequences, and $\bar{\mathbf{o}}_1^T \triangleq \bar{\mathbf{o}}_1,\ldots,\bar{\mathbf{o}}_T$ represents any of the T-length observation sequences.

In eigenvoice-based adaptation, θ is the composition coefficient vector $\mathbf{x}$. Let $\mathbf{e}_s = (\mathbf{e}_s(1),\mathbf{e}_s(2),\ldots,\mathbf{e}_s(K))$, which is a matrix of $D \times K$, we can derive that

$$\nabla_{\mathbf{x}} \log P(\mathbf{o}_t|s,\mathbf{x}) = \mathbf{e}_s^T \mathbf{C}_s^{-1}(\mathbf{o}_t - \mathbf{e}_s(0) - \mathbf{e}_s \cdot \mathbf{x}) \tag{19}$$

Rearrange equation 18, and define:

$$\delta_s(q) = \begin{cases} 1, & \text{if } q = s \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

$$D'_s = \sum_{\mathbf{s}_1^T} d'(\mathbf{s}_1^T)\sum_{t=1}^{T}\delta_s(s_t) \tag{21}$$

we get:

$$(\sum_s(\sum_t(\gamma_s(t) - \gamma_s^g(t)) + D'_s)\mathbf{e}_s{}^T\mathbf{C}_s^{-1}\mathbf{e}_s)\hat{\mathbf{x}}$$
$$= \sum_s \mathbf{e}_s{}^T\mathbf{C}_s^{-1}(\sum_t(\gamma_s(t) - \gamma_s^g(t))(\mathbf{o}_t - \mathbf{e}_s(0)) + D'_s\mathbf{e}_s\mathbf{x}') \tag{22}$$

We call the resulting procedure Maximum Mutual Information Eigen-Decomposition (MMIED), since the target is to maximize the mutual information.

The $\gamma_s(t)$ is calculated by conventional forward-backward procedure, and $\gamma_s^g(t)$ can be efficiently computed by using word-lattice. The recognition result is organized as a word lattice as in Figure 1, where $w_{l,m}, l = 1,\ldots,L, m = 1,\ldots,M$ denotes the m^{th} candidate at the position l, L is the length of the normalized recognized word lattice[4], and M is the number of candidate words at each position. In addition, segment points of the states in every $w_{l,m}$ are stored in the word lattice. Given time t, word position $l(t)$ could be determined from this lattice. We assume a uniform unigram which is indeed the case in the *acoustic part* of the system. Thus $\gamma_s^g(t)$ can be expressed as:

$$\gamma_s^g(t) = \frac{\sum_{m:\Phi(w_{l(t),m},t)=s} P(\mathbf{O}_{s_{l(t)}}^{e_{l(t)}}|w_{l(t),m})}{\sum_m P(\mathbf{O}_{s_{l(t)}}^{e_{l(t)}}|w_{l(t),m})} \tag{23}$$

Here $\Phi(w_{l,m},t)$ means state in time t for candidate word $w_{l,m}$, s_l and e_l denote the start and end time of l^{th} position respectively.

The setting of const D'_s is a key issue in the update equation. Large value would cause slow convergence, while small value may cause unstable results. D'_s fixed to $E\sum_t \gamma_s^g(t)$ for const E is used in following experiments, as is suggested in [6].

[4] L is not necessarily equal to N, since it's depend on the recognition result.

$$\begin{array}{ccccc} w_{1,1}, & w_{2,1}, & \ldots, & w_{L-1,1}, & w_{L,1} \\ \vdots, & \vdots, & \ldots, & \vdots, & \vdots \\ w_{1,m}, & w_{2,m}, & \ldots, & w_{L-1,m}, & w_{L,m} \\ \vdots, & \vdots, & \ldots, & \vdots, & \vdots \\ w_{1,M}, & w_{2,M}, & \ldots, & w_{L-1,M}, & w_{L,M} \end{array}$$

Figure 1: *Illustration of a word lattice (Here a word is in fact a Chinese syllable and we use $M = 100$).*

5. Experimental Results

To evaluate the performance of MMIED, experiments were carried out on a Chinese LVCSR task using speech database for "China 863 Assessment". We tested our method for male and female speech respectively. Thus the database was split into two parts by gender information, which we would denote as 863-Male set and 863-Female set. For each set, the training data were the utterances from 83 speakers, each with 650 sentences. Thus there were 83 SD models to be constructed. Other 5 male and 6 female's data were used for evaluation respectively, each of them contributed 120 sentences.

All the Chinese characters are pronounced as one of the total 408 un-toned Chinese syllables in CV structure. A left-to-right HMM syllable model was used, with 2 states for consonant and 4 states for the vowel part. The total number of states is 857, and the acoustic features were 45-dimensional formed by 14 MFCCs along with normalized log-energy and their first and second order differentials, thus the dimension D of each supervector is $857 \times 45 = 38565$. Single Gaussian model was used for state-output probability density with full covariance matrix.

Here we focus on the *acoustic part*. The speech are decoded into free syllable strings without any grammar constraints, and the result was organized into syllable-lattices. No language model was used. Only the syllable Error Rate (SER) are reported for performance comparisons.

Adaptation were carried out in supervised mode, where the first 60 sentences were used to estimate the weight coefficients, and other 60 sentences reserved for evaluation. We chose $K = 60$ during the experiment. Baseline SER was 31.33% in 863-Male set and 45.16% in 863-Female set. The results of MLED, MAPED and MMIED with different const E settings ($E = 1.0$ and $E = 2.0$ respectively) are listed in Table 1. MMIED outperformed both MLED and MAPED according to this result.

Table 1: *Comparison of SER for MLED, MAPED and MMIED. Eigenvoice number K is chosen to be 60, and 60 sentences were used as adaptation data.*

	863-Male	863-Female
MLED	23.48%	24.92%
MAPED	23.46%	24.24%
MMIED(E=1.0)	22.75%	23.42%
MMIED(E=2.0)	23.22%	24.04%

We also compared the performance of MMIED, MLED and MAPED when the amount of adaptation data was increased gradually. In this procedure, the number of adaptation sentences was added gradually from 10 to 60. Results on 863-Male set are listed in Table 2, where the eigenvoices number is 60. When we increased the number of eigenvoices, minor improvement was

Table 2: *Performance of 863-Male with gradually added adaptation data, using $K = 60$ eigenvoices.*

nSent	MLED	MAPED	MMIED
10	28.31%	28.25%	28.20%
20	27.77%	27.76%	26.99%
30	26.42%	26.40%	25.72%
40	24.95%	24.95%	24.35%
50	24.43%	24.43%	23.82%
60	23.48%	23.46%	22.75%

Table 3: *Performance of 863-Male with gradually added adaptation data, using $K = 80$ eigenvoices.*

nSent	MLED	MAPED	MMIED
10	28.22%	28.24%	28.19%
20	27.79%	27.75%	26.58%
30	26.45%	26.43%	25.54%
40	25.25%	25.25%	24.38%
50	24.68%	24.67%	23.86%
60	23.54%	23.57%	22.89%

observed. The results with $K = 80$ are listed in Table 3. The value of constant E is fixed to be 1.0 during this procedure.

Experiments on 863-Female set gave similar result, in which MMIED also outperformed MLED and MAPED. Results are listed in Table 4 ($K = 60$) and Table 5 ($K = 80$).

6. Conclusions

Since Maximum Likelihood estimation is not optimal in practical situations, we applied Maximum Mutual Information estimation to eigenvoice adaptation, and it proves to be effective. Experimental results show there is observable gain over ML and MAP estimation.

Some questions are open for further investigation. One is the relationship between convergence and the setting of const D'_s. Since the choice of D'_s was always discussed for the unconstrained training cases, there may exist better choice in constrained cases. And the solution to MMIED is not unique. It was shown in [14] that EBW could be deducted from Quasi Newton Method (though trivial differences in expression was observed), which would give another MMIED approach. Finally, the generation of word lattice requires more detailed consideration.

7. References

[1] C. Lee, C. Lin, and B. Juang, "A study on speaker adaptation of parameters of continuous density hidden markov models," *IEEE Trans. Signal Proc.*, no. 4, pp. 806–814, 1991.

[2] J. Gauvain and C. Lee, "Maximum *a Posteriori* estimation for multivariate gaussian mixture observations of markov chains," *IEEE Trans. Speech Audio Proc.*, no. 2, pp. 291–298, April 1994.

[3] C.J.Leggetter and P.C.Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models," *Comput. Speech Lang.*, pp. 171–185, 1995.

[4] R. Kuhn, J. C. Junqua, P. Nguyen, and et al, "Rapid speaker adaptation in eigenvoice space," *IEEE Trans. Speech and Audio Proc.*, no. 6, pp. 695–707, 2000.

[5] H. Botterweck, "Very fast adaptation for large vocabulary continuous speech recognition using eigenvoices," in *Proc. ICSLP*, 2000, pp. 354–357.

[6] P.C.Woodland and D. Povey, "Large scale discriminative training for speech recognition," in *Proc. ITRW ASR*, 2000.

[7] A. Gunawardana and W. Byrne, "Discriminative speaker adaptation with conditional maximum likelihood linear regression," in *Proc. Eurospeech*, 2001, pp. 1203–1206.

[8] L.F.Uebel and P.C.Woodland, "Improvements in linear transforms based speaker adaptation," in *Proc. ICASSP*, 2001.

[9] I.T.Jolliffe, *Principal Component Analysis*. Springer, 1986.

[10] C. Huang, J. Chien, and H. Wang, "A new eigenvoice approach to speaker adaptation," in *Proc. ISCSLP*, 2004.

[11] J. Luo, Z. Ou, and Z. Wang, "Fast eigenspace-based map adaptation within correlation subspace," *Tsinghua Science and Technology (in chinese)*, pp. 829–832, 2004.

[12] L. Bahl, P. Brown, P. de Souza, and R. Mercer, "Maximum mutual information estimation of hidden markov model parameters for speech recognition," in *Proc. Eurospeech*, 1986, pp. 49–52.

[13] A. Nadas., "A decision theoretic formulation of a training problem in speech recognition and a comparison of training by unconditinoal versus conditional maximum likelihood," *IEEE Trans. Acoustics, Speech and Signal Proc.*, no. 4, pp. 814–817, 1983.

[14] Z. Ou, "Temporal dependent models for speech recognition," Ph.D. dissertation, Tsinghua University, 2003.

Table 4: *Performance of 863-Female with gradually added adaptation data, using $K = 60$ eigenvoices.*

nSent	MLED	MAPED	MMIED
10	32.70%	32.75%	31.16%
20	31.13%	30.84%	29.42%
30	29.45%	29.24%	28.05%
40	27.72%	27.16%	26.07%
50	25.99%	25.72%	24.67%
60	24.92%	24.24%	23.42%

Table 5: *Performance of 863-Female with gradually added adaptation data, using $K = 80$ eigenvoices.*

nSent	MLED	MAPED	MMIED
10	32.55%	32.39%	31.17%
20	30.92%	30.83%	28.89%
30	29.36%	29.12%	27.56%
40	27.48%	26.80%	25.76%
50	26.10%	25.43%	24.45%
60	24.87%	24.09%	23.13%