Acoustic-to-Articulatory Inversion Mapping with Gaussian Mixture Model

Tomoki Toda[†‡], *Alan W Black*[†], *Keiichi Tokuda*[‡]

[†]Language Technologies Institute, Carnegie Mellon University, USA
[‡]Graduate School of Engineering, Nagoya Institute of Technology, Japan
{tomoki,awb}@cs.cmu.edu, {tomoki,tokuda}@ics.nitech.ac.jp

Abstract

This paper describes the acoustic-to-articulatory inversion mapping using a Gaussian Mixture Model (GMM). Correspondence of an acoustic parameter and an articulatory parameter is modeled by the GMM trained using the parallel acoustic-articulatory data. We measure the performance of the GMM-based mapping and investigate the effectiveness of using multiple acoustic frames as an input feature and using multiple mixtures. As a result, it is shown that although increasing the number of mixtures is useful for reducing the estimation error, it causes many discontinuities in the estimated articulatory trajectories. In order to address this problem, we apply maximum likelihood estimation (MLE) considering articulatory dynamic features to the GMM-based mapping. Experimental results demonstrate that the MLE using dynamic features can estimate more appropriate articulatory movements compared with the GMM-based mapping applied smoothing by lowpass filter.

1. Introduction

An articulatory parameter, one of representations of speech, is useful for various applications such as speech coding [1], speech recognition [2][3], and speech synthesis [4]. Because it is much harder to record articulatory movements than the speech signal, many attempts at determining articulatory configurations from the speech signal have been studied. This process is called an acoustic-to-articulatory inversion mapping. It is well known that this mapping is one-to-many mapping.

Development of recording devices enables us to record the speech signal and movements of several articulators simultaneously. Available large enough quantities of parallel acoustic-articulatory data make it possible to apply a corpus-based approach to the inversion mapping. As one of the corpus-based methods, the inversion mapping with the acoustic-articulatory codebook has been proposed [5]. The optimum sequence of code vectors is selected under dynamic constraints on acoustic and articulatory parameters. Specifically, multiple acoustic frames are used as an input feature, and a measure capturing articulatory discontinuities is used for the selection. Richmond modeled the mapping with Neural Network based on mixture density estimation [6]. It has been reported that the multiple representation for articulatory probability density is effective for the inversion mapping. Hiroya et al. proposed the inversion mapping using speech production model based on the HMM [7][8]. In this model, the acoustic-articulatory correspondence is represented as a linear mapping in each state of the diphone HMMs. In this method, not only dynamic features of acoustic and articulatory parameters but also phonetic information that is needed for training HMMs are used as constraints for addressing the problem of the one-to-many mapping.

In this paper, we address the inversion mapping without constraints on phonetic information. As a mapping method, we employ the mapping method based on a Gaussian Mixture Model (GMM) that is often used for voice conversion [9]. We investigate the effectiveness of using multiple acoustic frames and multiple mixtures. Moreover, in order to improve the mapping accuracy, we apply maximum likelihood estimation (MLE) considering articulatory dynamic features to the GMM-based mapping. The MOCHA database [10] is used as acoustic-articulatory data in this paper.

The paper is organized as follows. In **Section 2**, the GMM-based mapping method is described. In **Section 3**, evaluations of the mapping are described. In **Section 4**, we apply MLE using dynamic features to the GMM-based mapping. In **Section 5**, the effectiveness of using articulatory dynamic features is described. Finally, we summarize this paper in **Section 6**.

2. GMM-Based Mapping

In the GMM-based mapping algorithm [9], a mapping function from an acoustic feature vector $\boldsymbol{x}_t$ to an articulatory feature vector $\boldsymbol{y}_t$ in frame t is defined as

$$\hat{\boldsymbol{y}}_t = \sum_{i=1}^{M} p(m_i|\boldsymbol{x}_t, \boldsymbol{\Theta}) \boldsymbol{E}(\boldsymbol{y}_t|\boldsymbol{x}_t, m_i, \boldsymbol{\Theta}), \quad (1)$$

$$\boldsymbol{E}(\boldsymbol{y}_t|\boldsymbol{x}_t, m_i, \boldsymbol{\Theta}) = \boldsymbol{\mu}_i^{(y)} + \boldsymbol{\Sigma}_i^{(yx)} \boldsymbol{\Sigma}_i^{(xx)^{-1}} (\boldsymbol{x}_t - \boldsymbol{\mu}_i^{(x)}), \quad (2)$$

$$p(m_i|\boldsymbol{x}_t, \boldsymbol{\Theta}) = \frac{w_i N(\boldsymbol{x}_t; \boldsymbol{\mu}_i^{(x)}, \boldsymbol{\Sigma}_i^{(xx)})}{\sum_{j=1}^{M} w_j N(\boldsymbol{x}_t; \boldsymbol{\mu}_j^{(x)}, \boldsymbol{\Sigma}_j^{(xx)})}, \quad (3)$$

where $\hat{\boldsymbol{y}}_t$ denotes an estimated articulatory feature vector. The total number of mixtures is M. A set of model parameters $\boldsymbol{\Theta}$ consists of weights, mean vectors and covariance matrices. The weight of the i-th mixture is w_i. The vectors $\boldsymbol{\mu}_i^{(x)}$ and $\boldsymbol{\mu}_i^{(y)}$ denote the mean vector of the i-th mixture for $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. The matrices $\boldsymbol{\Sigma}_i^{(xx)}$ and $\boldsymbol{\Sigma}_i^{(yx)}$ denote the covariance matrix of the i-th mixture for $\boldsymbol{x}$ and the cross-covariance matrix of the i-th mixture for $\boldsymbol{x}$ and $\boldsymbol{y}$. These covariance matrices are full. The normal distribution with $\boldsymbol{\mu}_i^{(x)}$ and $\boldsymbol{\Sigma}_i^{(xx)}$ is represented as $N(\boldsymbol{x}_t; \boldsymbol{\mu}_i^{(x)}, \boldsymbol{\Sigma}_i^{(xx)})$. As shown in these equations, the conditional mean vector in each mixture is calculated by a simple linear conversion taking account of the correlation between acoustic and articulatory features. The estimated articulatory feature is defined as the weighted sum of the product of each of the conditional mean vectors and the conditional probabilities that the input feature vector belongs to each one of the mixtures.

In order to estimate model parameters, a GMM on joint probability $p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\Theta})$ is trained [11].

3. Evaluation of GMM-Based Inversion Mapping

3.1. Experimental conditions

Acoustic-articulatory data of two speakers in MOCHA [10] was used. One was female (fsew0), and the other was male (msak0). The 460 British TIMIT sentences were uttered by each speaker.

We used electromagnetic articulograph (EMA) data, one of representations of articulatory data provided in MOCHA, as an articulatory parameter. The locations of seven articulators (top lip, bottom lip, bottom incisor, tongue tip, tongue body, tongue dorsum, and velum) are shown by x- and y-coordinates on the midsagittal plane. We performed a normalization process described in [6] for reducing the effect of noise resulting from measurement error. The 14-dimensional articulatory feature vector converted to Z-score was used. The frame shift was 10 ms.

As for an acoustic parameter, we used the 0-th through 24-th mel-cepstral coefficients extracted from 16 kHz sampling speech data. STRAIGHT analysis method was employed for this extraction [12]. The shift length was 10 ms. A feature vector was constructed by concatenating multiple acoustic frames.

The number of mixtures was varied from 1 to 64 (1, 2, 4, 8, 16, 32, 64). The number of input acoustic frames was varied from 1 to 21 (a current ± 0, 1, 3, 5, 7, 10 frames). When the number of frames was set to more than 5, the input vector dimension was reduced using PCA analysis technique with a loss of no more than 20% of the information.

A 1/5 cross validation test was conducted to measure the accuracy of the mapping under open conditions. The 460 sentences were divided into 5 partitions consisting of 92 sentences, and then one of the partitions was reserved for the testing by turns, while the other 4 partitions were used for training. The root mean square (RMS) error was calculated between the measured articulatory parameters and the estimated articulatory parameters. Finally, the average of the RMS error was calculated over the 5 combinations of the training and testing partitions.

3.2. Effectiveness of using multiple acoustic frames and multiple mixtures

The results are shown in **Fig. 1**. The RMS error shows the average of the errors for individual dimensions of the articulatory vector. Using multiple acoustic frames and multiple mixtures is obviously effective for improving the mapping accuracy. However, the degradation of accuracy is caused by excessively increasing the number of these parameters because a larger amount of training data is needed for estimating a larger number of parameters. Moreover, it can be observed that the optimum number of mixtures for each number of acoustic frames decreases as the number of acoustic frames increases. Consequently, the best mapping accuracy is achieved when the number of acoustic frames is set to 11 and the number of mixtures is set to 32 in this experiment. In that case, the RMS error is 1.61 mm (0.67 on z-score) and the correlation coefficient is 0.73 for the female speaker, and the RMS error is 1.53 mm (0.66 on z-score) and the correlation coefficient is 0.74 for the male speaker.

Figure 2 shows an example of the estimated articulatory movement using 32 mixtures and that using a single mixture. The measured articulatory movement is also shown in this figure. It can be seen that many discontinuous parts in the estimated articulatory movement caused by using the multiple mixtures because the correlation between frames is ignored in the GMM-based mapping.

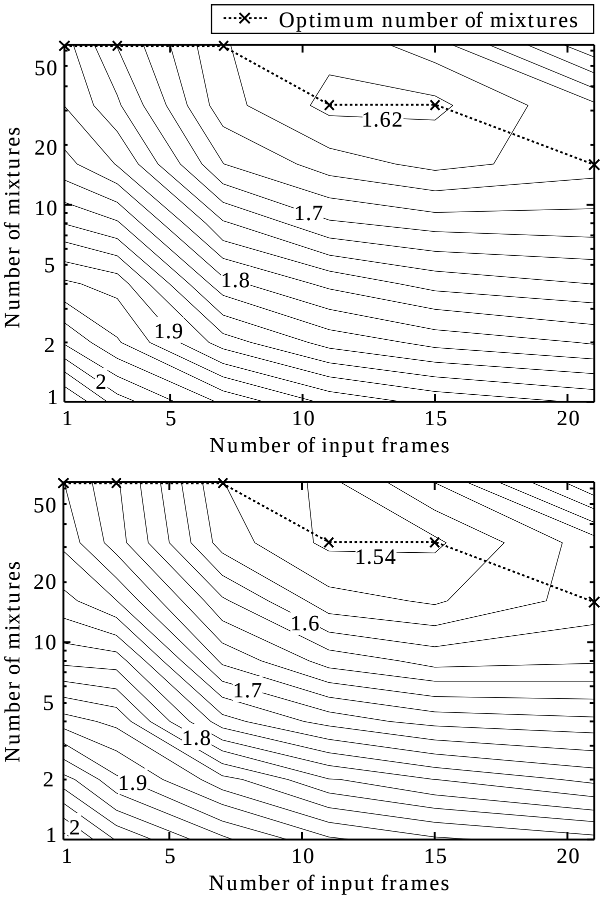

Figure 1: RMS error [mm] between measured and estimated articulatory movements. The upper figure shows the result for the female speaker, and the lower figure shows the result for the male speaker.

4. Maximum Likelihood Estimation Using Dynamic Features

In order to use a constraint on articulatory movements, we apply a parameter generation algorithm based on ML using dynamic features [13] to the GMM-based mapping. Hiroya et al. also use this technique in the inversion mapping with the HMM-based speech production model [8]. In the production model, a HMM state sequence is determined by the Viterbi search. Meanwhile, we use the EM algorithm described in [13] for maximizing likelihood in this paper.

Not only static but also dynamic features are used as the articulatory feature vector, which is given by

$$\boldsymbol{y}'_t = [\boldsymbol{y}_t^\top,\ \Delta \boldsymbol{y}_t^\top]^\top. \quad (4)$$

The conditional probability of the target feature vector $\boldsymbol{y}'_t$ for the given input feature vector $\boldsymbol{x}_t$ is written as

$$p(\boldsymbol{y}'_t|\boldsymbol{x}_t, \boldsymbol{\Theta}') = \sum_{i=1}^{M} p(m_i|\boldsymbol{x}_t, \boldsymbol{\Theta}')p(\boldsymbol{y}'_t|\boldsymbol{x}_t, m_i, \boldsymbol{\Theta}'), \quad (5)$$

$$p(\boldsymbol{y}'_t|\boldsymbol{x}_t, m_i, \boldsymbol{\Theta}') = N(\boldsymbol{y}'_t; \boldsymbol{E}'_t(i), \boldsymbol{D}'(i)), \quad (6)$$

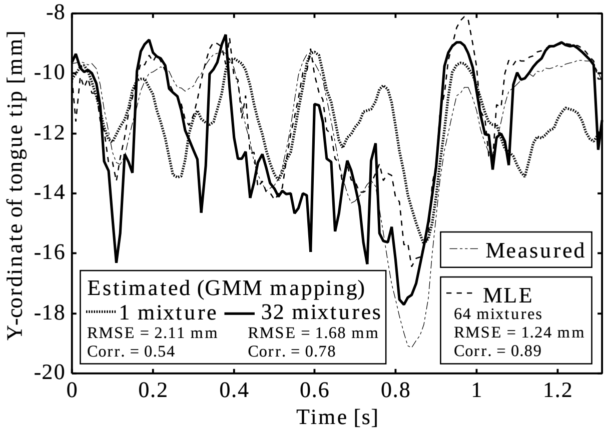

Figure 2: An example of estimated and measured articulatory trajectories.

where

$$\boldsymbol{E}'_t(i) = \boldsymbol{\mu}_i^{(y')} + \boldsymbol{\Sigma}_i^{(y'x)} \boldsymbol{\Sigma}_i^{(xx)^{-1}} (\boldsymbol{x}_t - \boldsymbol{\mu}_i^{(x)}), \quad (7)$$

$$\boldsymbol{D}'(i) = \boldsymbol{\Sigma}_i^{(y'y')} - \boldsymbol{\Sigma}_i^{(y'x)} \boldsymbol{\Sigma}_i^{(xx)^{-1}} \boldsymbol{\Sigma}_i^{(xy')}. \quad (8)$$

Model parameters are estimated by training a GMM on joint probability $p(\boldsymbol{x}, \boldsymbol{y}'|\boldsymbol{\Theta}')$.

Let $\boldsymbol{X} = [\boldsymbol{x}_1{}^\top, \boldsymbol{x}_2{}^\top, \cdots, \boldsymbol{x}_T{}^\top]^\top$ be a time sequence of the acoustic feature vector and $\boldsymbol{Y} = [\boldsymbol{y}_1{}^\top, \boldsymbol{y}_2{}^\top, \cdots, \boldsymbol{y}_T{}^\top]^\top$ be that of the articulatory feature vector. The relationship between a sequence of the static feature $\boldsymbol{Y}$ and a sequence of the static and dynamic features $\boldsymbol{Y}'$ can be represented as a linear conversion,

$$\boldsymbol{Y}' = \boldsymbol{W}\boldsymbol{Y} \quad (9)$$

where $\boldsymbol{W}$ is a transformation matrix consisting of coefficients of a delta window and 0 [13]. In order to maximize a likelihood function $p(\boldsymbol{Y}'|\boldsymbol{X}, \boldsymbol{\Theta}')$, we maximize an auxiliary function of a current feature vector sequence $\boldsymbol{Y}'$ and a new feature vector sequence $\hat{\boldsymbol{Y}}'$ defined by

$$\begin{aligned} Q(\boldsymbol{Y}', \hat{\boldsymbol{Y}}') &= \sum_{\text{all } \boldsymbol{m}} p(\boldsymbol{Y}', \boldsymbol{m}|\boldsymbol{X}, \boldsymbol{\Theta}') \log p(\hat{\boldsymbol{Y}}', \boldsymbol{m}|\boldsymbol{X}, \boldsymbol{\Theta}') \\ &= p(\boldsymbol{Y}'|\boldsymbol{X}, \boldsymbol{\Theta}') \Big\{ -\frac{1}{2} \hat{\boldsymbol{Y}}^\top \boldsymbol{W}^\top \overline{\boldsymbol{D}'^{-1}} \boldsymbol{W} \hat{\boldsymbol{Y}} \\ &\qquad + \hat{\boldsymbol{Y}}^\top \boldsymbol{W}^\top \overline{\boldsymbol{D}'^{-1}\boldsymbol{E}'} + \overline{K} \Big\}, \end{aligned} \quad (10)$$

where

$$\overline{\boldsymbol{D}'^{-1}} = \text{diag}\left[\overline{\boldsymbol{D}'^{-1}_1}, \overline{\boldsymbol{D}'^{-1}_2}, \cdots, \overline{\boldsymbol{D}'^{-1}_T}\right], \quad (11)$$

$$\overline{\boldsymbol{D}'^{-1}_t} = \sum_{i=1}^{M} \gamma_t(i) \boldsymbol{D}'(i)^{-1}, \quad (12)$$

$$\overline{\boldsymbol{D}'^{-1}\boldsymbol{E}'} = \left[\overline{\boldsymbol{D}'^{-1}_1 \boldsymbol{E}'_1}^\top, \overline{\boldsymbol{D}'^{-1}_2 \boldsymbol{E}'_2}^\top, \cdots, \overline{\boldsymbol{D}'^{-1}_T \boldsymbol{E}'_T}^\top\right]^\top, \quad (13)$$

$$\overline{\boldsymbol{D}'^{-1}_t \boldsymbol{E}'_t} = \sum_{i=1}^{M} \gamma_t(i) \boldsymbol{D}'(i)^{-1} \boldsymbol{E}'_t(i), \quad (14)$$

$$\gamma_t(i) = \frac{p(m_i|\boldsymbol{x}_t, \boldsymbol{\Theta}') N(\boldsymbol{y}'_t; \boldsymbol{E}'_t(i), \boldsymbol{D}'(i))}{\sum_{j=1}^{M} p(m_j|\boldsymbol{x}_t, \boldsymbol{\Theta}') N(\boldsymbol{y}'_t; \boldsymbol{E}'_t(j), \boldsymbol{D}'(j))}. \quad (15)$$

The constant $\overline{K}$ is independent of $\hat{\boldsymbol{Y}}'$. The sequence of the estimated articulatory static feature $\hat{\boldsymbol{Y}}$ that maximizes $Q(\boldsymbol{Y}', \hat{\boldsymbol{Y}}')$ is given by

$$\hat{\boldsymbol{Y}} = \left(\boldsymbol{W}^\top \overline{\boldsymbol{D}'^{-1}} \boldsymbol{W}\right)^{-1} \boldsymbol{W}^\top \overline{\boldsymbol{D}'^{-1}\boldsymbol{E}'}. \quad (16)$$

The target vector sequence given by the conventional mapping function is used as the initial vector sequence $\boldsymbol{Y}$. The new vector sequence $\hat{\boldsymbol{Y}}$ is calculated by the above equations, and then $\hat{\boldsymbol{Y}}$ is substituted for $\boldsymbol{Y}$. This procedure is iteratively performed until a certain convergence condition is satisfied.

There is little difference between the articulatory movements estimated with diagonal covariance matrices and that with full covariance matrices in our preliminary experiments. Therefore, we use only diagonal elements of $\boldsymbol{D}'(i)$.

5. Evaluation of MLE Considering Articulatory Dynamic Features

We investigate the effectiveness of MLE taking articulatory dynamic features into account by comparing with smoothing of the estimated articulatory trajectories by lowpass filter.

5.1. Experimental conditions

A 1/5 cross validation test was conducted in the same way as described in **Section 3.1**. The number of acoustic frames used as an input feature was set to 11. The number of mixtures was varied from 1 to 128. Cutoff frequency of lowpass filter was determined so that the RMS error was minimized in each dimension of the articulatory vector.

5.2. Results

Results of the MLE and the GMM-based mapping with the lowpass filtering are shown in **Fig. 3**. Results of the GMM-based mapping without the lowpass filtering are also shown in this figure. The RMS error can be reduced by performing smoothing of the estimated articulatory trajectories. Especially, the smoothing is effect when a number of mixtures are used because many articulatory discontinuities are caused by using multiple mixtures as mentioned in **Section 3.2**.

The mapping accuracy of the MLE and that of the GMM-based mapping with lowpass filtering is almost equal when the number of mixtures is small. However, when a large number of mixtures are used, the MLE can estimate more appropriate articulatory movements than the GMM-based mapping applied the lowpass smoothing. In the MLE, the smoothing of articulatory trajectories is performed considering the estimated static and dynamic probability density functions. In order to model joint probability density of both static and dynamic features, a larger number of mixtures are needed compared with the case of using only the static feature. When the number of mixtures is set to 64, the RMS error is 1.45 mm (0.61 on z-score) and the correlation coefficient is 0.79 for the female speaker, and the RMS error is 1.36 mm (0.59 on z-score) and the correlation coefficient is 0.80 for the male speaker. An example of the articulatory movement estimated by the MLE is also shown in **Fig. 2**.

We also measured the performance of the lowpass smoothing in the MLE with dynamic features. **Table 1** shows results for the female speaker when the number of mixtures is set to 64. The results of the GMM-based mapping with lowpass filtering are also shown in this table. It is shown that the improvements

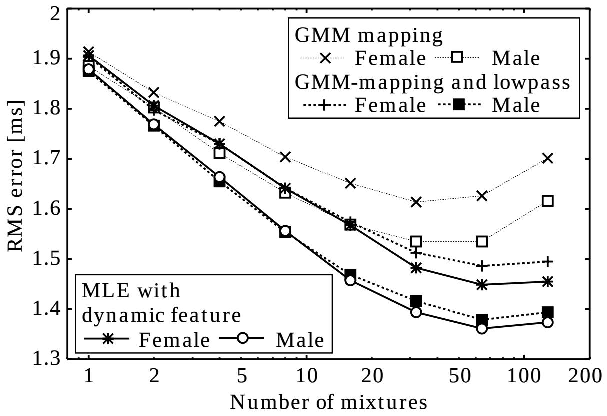

Figure 3: RMS error as a function of the number of mixtures.

by the lowpass smoothing in the MLE is much less than those in the GMM-based mapping. These results demonstrate that the smooth trajectories can be estimated by the MLE.

6. Conclusions

We performed the inversion mapping using a Gaussian Mixture Model (GMM). In order to address the problem of one-to-many mapping, we used multiple acoustic frames and multiple mixtures. From results of experimental evaluations, it was shown that the articulatory trajectories having many discontinuous parts are estimated by the GMM-based mapping with a large number of mixtures. In order to avoid the articulatory discontinuity, we applied the maximum likelihood estimation (MLE) considering articulatory dynamic features to the GMM-based mapping. Experimental results demonstrated that the MLE using dynamic features can estimate more appropriate articulatory movements compared with the GMM-based mapping applied the smoothing by lowpass filter.

Acknowledgment: This research was supported in part by JSPS (Japan Society for the Promotion of Science) Research Fellowships for Young Scientists. The authors are grateful to Dr. Hideki Kawahara of Wakayama University in Japan for permission to use the STRAIGHT analysis method.

Table 1: Effectiveness of articulatory smoothing by lowpass filter. The results for the female speaker are shown. “li”, “ul”, “ll”, “tt”, “tb”, “td” and “v” show lower incisor, upper lip, lower lip, tongue tip, tongue body, tongue dorsum and velum, respectively. “*_x” and “*_y” in each articulator show X- and Y-coordinates, respectively. In “RMSE” columns, numbers on the left side show results when not using lowpass filter and numbers on the right side show results when using lowpass filter. “Dec.” denotes a decrease rate of the RMS error. The number in each bracket shows the cutoff frequency of the lowpass filter

	GMM mapping		MLE with delta	
	RMSE[mm]	Dec.[%]	RMSE[mm]	Dec.[%]
li_x	0.97 → 0.89 (3.3)	8.18	0.89 → 0.88 (8.3)	0.32
ul_x	0.94 → 0.85 (2.8)	9.04	0.85 → 0.84 (5.0)	0.45
ll_x	1.28 → 1.16 (3.1)	9.28	1.16 → 1.16 (6.3)	0.45
tt_x	2.37 → 2.14 (5.0)	9.63	2.08 → 2.06 (8.3)	0.68
tb_x	2.22 → 2.04 (5.0)	7.91	2.00 → 1.99 (8.3)	0.61
td_x	2.10 → 1.94 (5.6)	7.53	1.90 → 1.89 (10.0)	0.58
v_x	0.45 → 0.41 (5.0)	8.16	0.41 → 0.41 (25.0)	0.46
li_y	1.29 → 1.19 (5.6)	7.21	1.17 → 1.17 (10.0)	0.50
ul_y	1.25 → 1.13 (4.5)	9.24	1.10 → 1.09 (7.1)	0.58
ll_y	2.49 → 2.30 (6.3)	7.74	2.25 → 2.23 (8.3)	0.70
tt_y	2.41 → 2.22 (7.1)	8.21	2.11 → 2.10 (12.5)	0.83
tb_y	2.17 → 1.97 (5.6)	9.40	1.88 → 1.87 (8.3)	0.73
td_y	2.33 → 2.09 (5.0)	10.54	2.01 → 2.00 (12.5)	0.59
v_y	0.51 → 0.48 (4.5)	6.89	0.47 → 0.47 (25.0)	0.39
Ave.	1.63 → 1.49	8.63	1.45 → 1.44	0.61

7. References

[1] J. Schroeter and M.M. Sondhi. Speech coding based on physiological models of speech production. *Advances in Speech Signal Processing*, S. Furui and M.M. Sondhi, Marcel Dekker New York, pp. 231–267, 1992.

[2] A.A. Wrench and K. Richmond. Continuous speech recognition using articulatory data. *Proc. ICSLP*, Beijing, China, pp. 145–148, Oct. 2000.

[3] J. Frankel, K. Richmond, S. King, and P. Taylor. An automatic speech recognition system using neural networks and linear dynamic models to recover and model articulatory traces. *Proc. ICSLP*, Beijing, China, Vol. 4, pp. 254–257, Oct. 2000.

[4] M.M. Sondhi. Articulatory modeling: a possible role in concatenative text-to-speech synthesis. *IEEE 2002 Workshop on Speech Synthesis*, Santa Monica, U.S.A., Sep. 2002.

[5] S. Suzuki, T. Okadome, and M. Honda. Determination of articulatory positions from speech acoustics by applying dynamic articulatory constraints. *Proc. ICSLP*, Sydney, Australia, pp. 2251–2254, Dec. 1998.

[6] K. Richmond. *Estimating articulatory parameters from the acoustic speech signal*. Ph.D. Thesis, The Centre for Speech Technology Research, University of Edinburgh, 2001.

[7] S. Hiroya and M. Honda. Determination of articulatory movements from speech acoustics using an HMM-based speech production model. *Proc. ICASSP*, Orlando, U.S.A, pp. 437–440, May. 2002.

[8] S. Hiroya and M. Honda. Acoustic-to-articulatory inverse mapping using an HMM-based speech production model. *Proc. ICSLP*, pp.2305–2308, Denver, U.S.A., Sep. 2002.

[9] Y. Stylianou. *Harmonic plus Noise Models for Speech, Combined with Statistical Methods, for Speech and Speaker Modification.* Ph.D. Thesis, Ecole Nationale Supérieure des Télécommunications, 1996.

[10] A. Wrench. The MOCHA-TIMIT articulatory database. *http://www.cstr.ed.ac.uk/artic/mocha.html*, Queen Margaret University College, 1999.

[11] A. Kain. *High Resolution Voice Transformation*. Ph.D. Thesis, OGI School of Science and Engineering, Oregon Health and Science University, 2001.

[12] H. Kawahara, I. Masuda-Katsuse, and A.de Cheveigné. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F_0 extraction: possible role of a repetitive structure in sounds. *Speech Communication*, Vol. 27, No. 3–4, pp. 187–207, 1999.

[13] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura. Speech parameter generation algorithms for HMM-based speech synthesis. *Proc. ICASSP*, pp. 1315–1318, Istanbul, Turkey, June 2000.