

# Integration of Metamodel and Acoustic Model for Dysarthric Speech Recognition

Hironori Matsumasa<sup>1</sup>, Tetsuya Takiguchi<sup>1</sup>, Yasuo Arika<sup>1</sup>, I-Chao LI<sup>2</sup>, Toshitaka Nakabayashi<sup>3</sup>

<sup>1</sup>)Graduate School of Engineering, Kobe University, JAPAN

<sup>2</sup>)Department of Economics, Otemon Gakuin University, JAPAN

<sup>3</sup>)Department of Clinical Developmental Disabilities, Kobe University, JAPAN

Email: mattu28@me.cs.scitec.kobe-u.ac.jp, {takigu, ariki}@kobe-u.ac.jp

**Abstract**— We investigated the speech recognition of a person with articulation disorders resulting from athetoid cerebral palsy. The articulation of the first words spoken tends to be unstable due to the strain placed on the speech-related muscles, and this causes degradation of speech recognition. Therefore, we proposed a robust feature extraction method based on PCA (Principal Component Analysis) instead of MFCC, where the main stable utterance element is projected onto low-order features and fluctuation elements of speech style are projected onto high-order features. Therefore, the PCA-based filter will be able to extract stable utterance features only.

The fluctuation of speaking style may invoke phone fluctuations, such as substitutions, deletions and insertions. In this paper, we discuss our effort to integrate a Metamodel and an Acoustic model approach. Metamodels have a technique for incorporating a model of a speaker's confusion matrix into the ASR process in such a way as to increase recognition accuracy. The integration of metamodels and acoustic models enables fluctuation suppression not only in feature extraction but also in recognition. The proposed method resulted in an improvement of 9.9% (from 79.1% to 89%) in the recognition rate compared to the conventional method.

**Index Terms**— dysarthric speech recognition, feature extraction, model integration

## I. INTRODUCTION

Recently, the importance of information technology in welfare-related fields has increased. For example, sign language recognition using image recognition technology [1] [2] [3], text reading systems from natural scene images [4] [5] [6], and the design of wearable speech synthesizers for those with voice disorders [7] [8] have been studied.

As for speech recognition technology, the opportunities in various environments and situations have increased (e.g., operation of a car navigation system [9], lecture transcription during meetings [10], etc.). However, degradation can be observed in the case of children [11], persons with speech impediments, and so on, and there has been very little research on people with speech impediments. There are 34,000 people with speech impediments

associated with articulation disorders in Japan alone, and it is hoped that speech recognition systems will one day be able to recognize their voices.

One of the causes of speech impediments is cerebral palsy. About 2 babies in 1,000 are born with cerebral palsy. Cerebral palsy results from damage to the central nervous system, and the damage causes movement disorders. Three general times are given for the onset of the disorder: before birth, at the time of delivery, and after birth. Cerebral palsy is classified as follows: 1) spastic type 2) athetoid type 3) ataxic type 4) atonic type 5) rigid type, and, a mixture of types [12].

In this paper, we focused on a person with an articulation disorder resulting from the athetoid type of cerebral palsy. Athetoid symptoms develop in about 10-15% of cerebral palsy sufferers. In the case of a person with this type of articulation disorder, the first movements are sometimes more unstable than usual. That means, the case of movements related to speaking, the first utterance is often unstable or unclear due to the athetoid symptoms. Therefore, we recorded speech data for a person with a speech impediment who uttered a given word several times, and we investigated the influence of the unstable speaking style caused by the athetoid symptoms.

In current speech recognition technology, the MFCC (Mel Frequency Cepstral Coefficient) has been widely used, where the feature is derived from the mel-scale filter bank output by DCT (Discrete Cosine Transform). In [13], PCA-based feature extraction has been studied. Also, we studied the application of kernel PCA to reverberant speech [14], and we proposed robust feature extraction based on PCA with more stable utterance data instead of DCT [15], where the main stable utterance element is projected onto low-order features while fluctuation elements of speech style are projected onto high-order features. As a result, the PCA-based filter will be able to extract stable utterance features only (Fig.1). Our proposed method improved the recognition accuracy, but the performance was not sufficient compared to that of persons with no disability.

The speech associated with articulation disorders may decrease intelligibility due to substitutions, deletions and insertions of phonemes. In [16], [17], a metamodel was

This paper is based on "Integration of Metamodel and Acoustic Model for Speech Recognition," by H. Matsumasa and T. Takiguchi and Y. Arika and I. Li and T. Nakabayashi, which appeared in the Proceedings of the Interspeech 2008, Brisbane, Australia, September 2008. © 2008 ISCA.

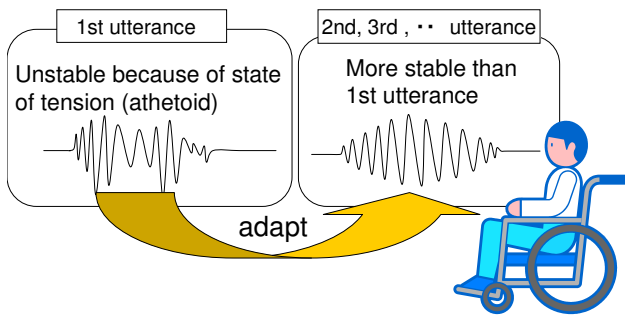


Figure 1. Corrective strategy for articulation disorders

introduced to increase recognition accuracy in speakers with low intelligibility by using the phoneme confusion matrix. However, the metamodel itself was effective only when there is limited adaptation data available for a speaker.

In this paper, we introduce a technique that combines acoustic models and metamodels that generate another word hypothesis for an utterance, where acoustic models are built by using a large amount of training data uttered by a person with an articulation disorder. Its effectiveness is confirmed by word-recognition experiments on speech data of a person with an articulation disorder, by comparing our previous method (PCA-based feature extraction) and metamodels only.

The paper is structured as follows. Section II introduces a Metamodel method. In Section III, we describe the robust feature extraction method using PCA. In Section IV, we describe the integration of a metamodel and acoustic model. In Section V, speaker-dependent and speaker-independent word-recognition experiments are shown. In the final section, Section VI, we summarize the ideas and main results in the paper and discuss necessary future work in this area.

## II. METAMODEL [16], [17]

In [16], [17], a metamodel was introduced to generate word hypotheses for an utterance using an unconstrained phoneme recognizer. An attractive alternative to correlating the phone hypotheses from the recognizers is to construct word hypotheses from the phone recognizer output.

For word recognition, we attempt to find the word  $\hat{w}$  for which the probability  $Pr(w|\mathbf{X})$  is the largest among all words  $w \in \mathcal{W}$ . Here  $\mathbf{X}$  is the acoustic signal input. Since the phoneme sequences form a proper subset  $p \in \mathcal{P}$ , we may rewrite  $Pr(w|\mathbf{X})$  as

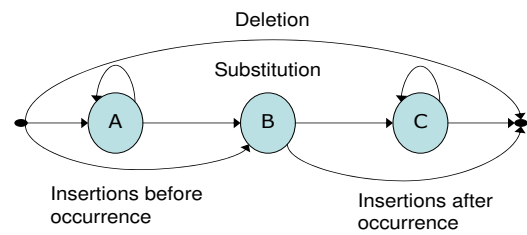
$$Pr(w|\mathbf{X}) = \sum_{p \in \mathcal{P}} Pr(w|p)Pr(p|\mathbf{X}). \quad (1)$$

Instead of choosing this subset of  $\mathcal{P}$ , we may choose the sequence  $\mathbf{p}^*$  obtained from a phoneme classification task:

$$\mathbf{p}^* = \arg \max_{p \in \mathcal{P}} Pr(p|\mathbf{X}). \quad (2)$$

Thus, a different approximation of  $Pr(w|\mathbf{X})$  gives

$$Pr(w|\mathbf{X}) \simeq Pr(w|\mathbf{p}^*)Pr(\mathbf{p}^*|\mathbf{X}). \quad (3)$$



Discrete probability distribution

Phoneme / State	A	B	C
a	0.1	0.7	0.2
i	0.4	0.1	0.3
u	0.3	0.05	0.3
e	0.1	0.05	0.1
o	0.1	0.1	0.1

Figure 2. Architecture of a phoneme metamodel (discrete HMM)

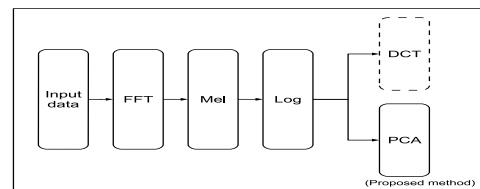


Figure 3. Feature extraction using PCA

The knowledge of  $\mathbf{p}^*$  from a phoneme-recognition experiment may be used to obtain an alternative decoding of the utterance by finding the word sequence that maximizes each side:

$$\hat{w} = \arg \max_{w \in \mathcal{W}} Pr(w|\mathbf{p}^*). \quad (4)$$

In [16], [17], a discrete HMM is used for a phoneme metamodel, as shown in Fig.2. Each state of a metamodel has a discrete probability distribution over the symbols for the set of phonemes. The central state  $B$  of a metamodel for a certain phoneme model takes account of correct decodings or substitutions. States  $A$  and  $C$  model (possibly multiple) insertions before and after the phoneme. The metamodels are trained in the same manner as acoustic HMMs using embedded Baum-Welch as follows:

- 1) Set the non-zero model transition probabilities to random values and seed the symbol distributions for each state equiprobably;
- 2) Use the reference transcriptions to concatenate the appropriate sequence of phoneme metamodels for each training set utterance; and
- 3) Use the decoded string  $\mathbf{p}^*$  for each utterance as training data and iterate re-estimation of the transition probabilities and distribution probabilities of the metamodels, using the Baum-Welch algorithm, until a suitable convergence criterion is reached.

## III. FEATURE EXTRACTION USING PCA

We proposed robust feature extraction using PCA with the more stable utterance data instead of DCT (Fig. 3),

where PCA is applied to the mel-scale filter bank output [15]. We computed the filter (eigenvector matrix) using the more stable utterance. Then we applied the filtering operation to the first utterance (unstably articulated utterance) in the log-spectral domain. Given the frame of short-time analysis  $n$  and frequency  $\omega$ , we represent the first utterance  $X_n(\omega)$  as the multiplication of the stable speech  $S_n(\omega)$  and the fluctuation element of speaking style  $H(\omega)$  in the linear-spectral domain:

$$X_n(\omega) = S_n(\omega) \cdot H(\omega). \quad (5)$$

The multiplication can be converted to addition in the log-spectral domain as follows:

$$\log X_n(\omega) = \log S_n(\omega) + \log H(\omega). \quad (6)$$

Next, we use the following filtering based on PCA in order to extract the feature of stable speech only,

$$\hat{S} = V^t X_{log}. \quad (7)$$

For the filter (eigenvector matrix),  $V$  is derived by the eigenvalue decomposition of the centered covariance matrix of a stable speech data set, in which the filter consists of the eigenvectors corresponding to the  $L$ -dominant eigenvalues.

#### IV. INTEGRATION OF METAMODEL AND ACOUSTIC MODEL

The articulation of speech uttered by persons with speech disorders tends to become unstable due to strain on their speech-related muscles. Therefore, the fluctuation in speaking style may invoke phone fluctuations such as substitutions, deletions and insertions. Metamodels are used for speaker adaptation of a person with articulation disorders in [16].

In this paper, we integrate metamodels and acoustic models for suppressing fluctuation. Figure 4 shows a schematic of the recognition system. The left half of the figure represents training flow and the right half represents test flow. First, the feature extraction of the test data is performed with PCA filter, where the filter is calculated using the stable speech. Secondly, we perform word-recognition and phoneme recognition using the acoustic model. Thirdly, we perform metamodel recognition, where the input feature is phoneme sequences obtained by phoneme recognition. Finally, we integrate acoustic likelihood and metamodel likelihood.

The metamodel-acoustic integration enables fluctuation suppression not only in feature extraction but also in recognition. The integration in recognition is represented as follows:

$$\begin{aligned} L_{Aco+Meta}^{\hat{w}_{N-best}} &= (1 - \alpha) \cdot L_{Aco}^{\hat{w}_{N-best}} + \alpha \cdot L_{Meta}^{\hat{w}_{N-best}} \\ &= (1 - \alpha) \cdot Pr(A|\hat{w}_{N-best})/N_{frame} \\ &\quad + \alpha \cdot Pr(p^*|\hat{w}_{N-best})/N_{phone} \end{aligned} \quad (8)$$

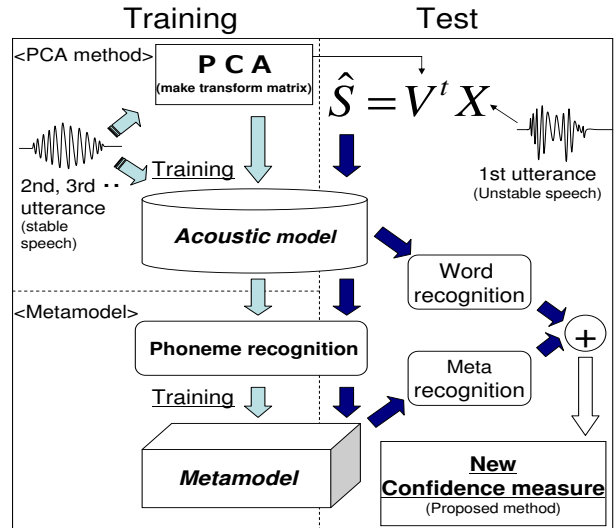


Figure 4. Integration of metamodel and acoustic model

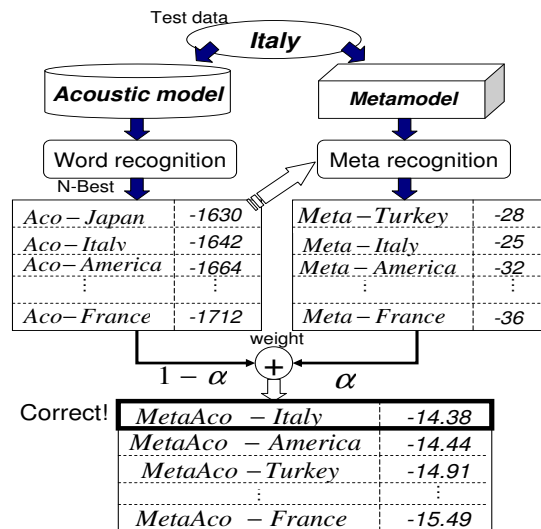


Figure 5. Example of integrated recognition

Here  $L_{Aco}$  and  $L_{Meta}$  represent acoustic likelihood and metamodel likelihood, respectively, and  $\alpha$  is the integration weight.  $N_{frame}$  and  $N_{phone}$  represent the number of acoustic frames and phoneme sequences outputted from the phoneme recognition, respectively. Figure 5 shows an example of integrated recognition. First we obtain N-best words ( $\hat{w}_{N-best}$ ) by word recognition. Then we perform metamodel recognition for only the N-best words and integrate the acoustic and metamodel likelihoods according to (8). In this figure,  $\alpha$  is set to 0.2.

#### V. RECOGNITION EXPERIMENT

##### A. Experimental conditions

The new model integration method was evaluated on word-recognition tasks for one person with an articulation disorder. We recorded 210 words included in the ATR Japanese speech database repeating each word five times (Fig. 6). The utterance signal was sampled at 16 kHz and windowed with a 25-msec Hamming window every 10

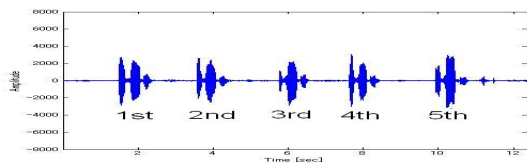


Figure 6. Example of recorded speech data

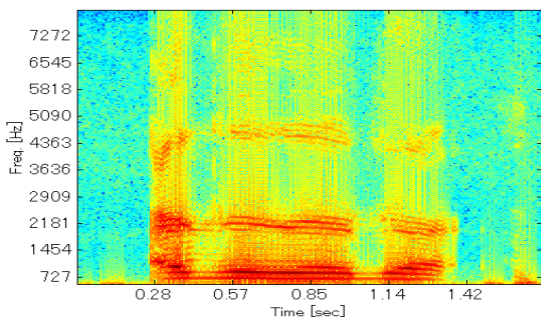


Figure 7. Example of a spectrogram from a person with an articulation disorder //a k e g a t a

msec. Then we clipped each utterance by hand. When we recognize the 1st utterance, the 2nd-5th utterances are used for training. We iterated this process for each utterance. Figure 7 shows an example of a spectrogram spoken by a person with an articulation disorder. Figure 8 shows a spectrogram spoken by a physically unimpaired person doing the same task. We used HTK [18] for all the experiments.

*B. Recognition using only acoustic models for articulation disorders*

Figure 9 shows the recognition rates using a speaker-independent model, where the speech data uttered by unimpaired persons are used. In a person with an articulation disorder, a recognition rate of only 3.5% was obtained, but in a physically unimpaired person, a recognition rate of 89.7% was obtained for the same task. It is clear that the speaking style of a person with an articulation disorder differs considerably from that of a physically unimpaired person.

We carried out speaker adaptation for improvement of recognition rates using a speaker-independent model, where MLLR (Maximum Likelihood Linear Regression) and MAP (Maximum A Posteriori) adaptation techniques [19] are used. The second utterance was used for adaptation, and the amount of adaptation data is 20, 40, 60, 100, and 210 words.

Figure 10 shows the recognition rates obtained using the speaker adaptation model. When the volume of adaptation data was small, little improvement in recognition rate was observed. When 210 words were used, however, the recognition rate improved significantly. Therefore, a large amount of adaptation data is necessary to realize an improvement in recognition rate when using an acoustic model of an unimpaired person.

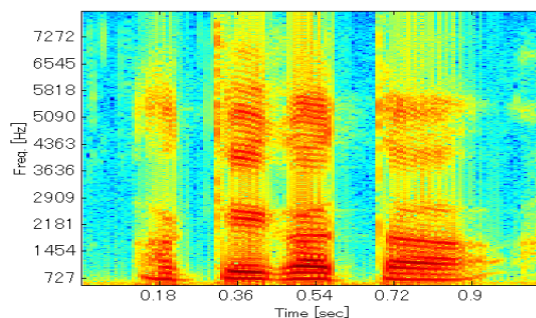


Figure 8. Example of a spectrogram from a physically unimpaired person //a k e g a t a

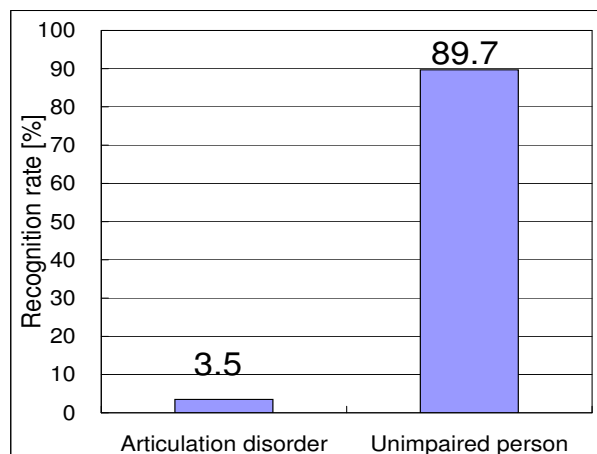


Figure 9. Recognition results for a speaker-independent model using training data uttered by unimpaired persons

It was difficult to recognize utterances using an acoustic model trained by utterances of a physically unimpaired person. Therefore, we trained the acoustic model using the utterances of a person with an articulation disorder. The acoustic model consists of an HMM set with 54 context-independent phonemes with 26 dimensional MFCC features (13-order MFCCs and their delta) and 6 mixture components for each state. Each HMM has three states and three self-loops.

TABLE I.  
RECOGNITION RATES FOR EACH UTTERANCE (ARTICULATION DISORDER)

1st	2nd	3rd	4th	5th
79.1	87.6	90.0	91.4	90.0

In a person with an articulation disorder, the recognition rate of the 1st utterance is 79.1%. As shown in Table I, it is lower than the others. The first utterance is the first intentional movement. It is conjectured that the speaker experienced a more strained state during the first utterance compared to subsequent utterances. So, athetoid symptoms occur and articulation becomes difficult. It is believed that this difficulty causes fluctuations in speaking style and degradation of the recognition rates.

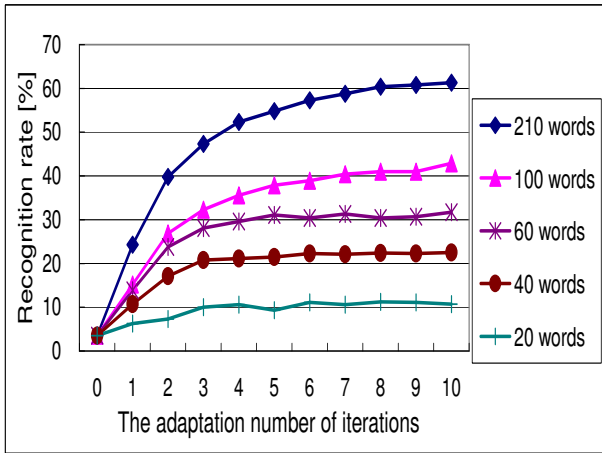


Figure 10. Adaptation results with MLLR and MAP estimation

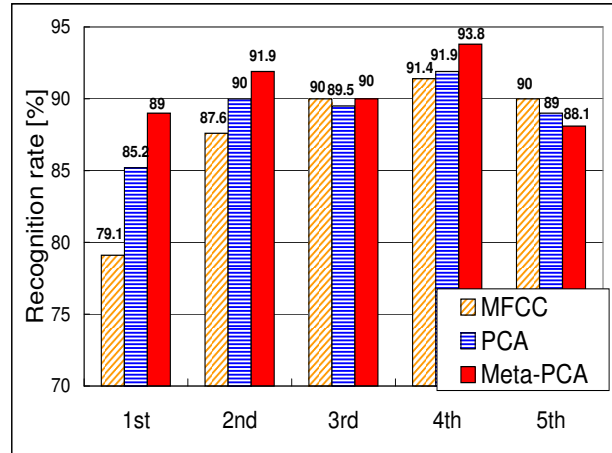


Figure 12. Recognition rate for each utterance for each method

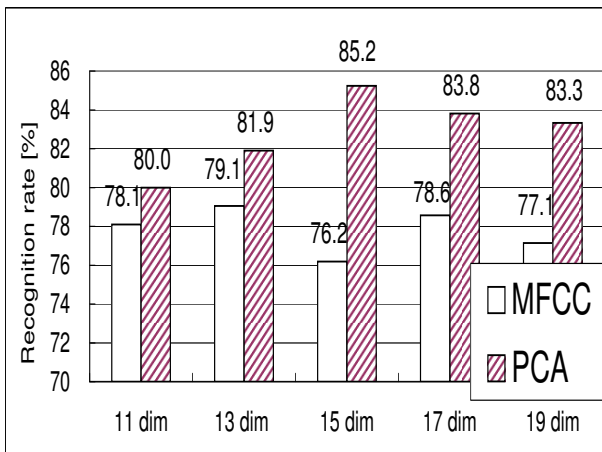


Figure 11. Recognition rates for the 1st utterance with PCA method

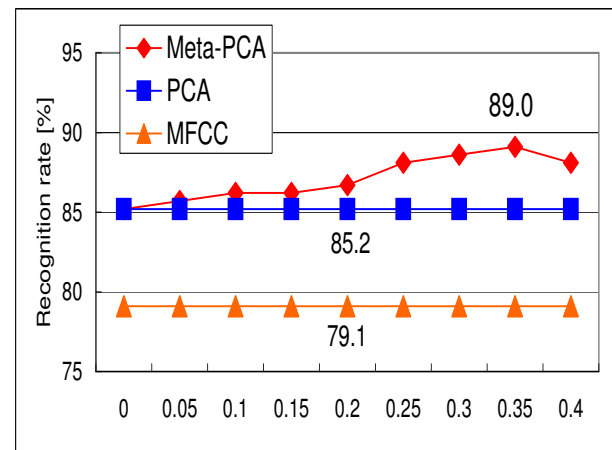


Figure 13. Recognition rates for the 1st utterance using the proposed method

C. Results using PCA-based feature extraction

For the feature extraction, PCA was applied to 24 mel-scale filter bank output, and then the delta coefficients were also computed. We experimented on the number of principal components, using 11, 13, 15, 17, and 19 dimensions. Figure 11 shows the recognition rates for the 1st utterance. Figure 12 shows the recognition rate for each utterance using MFCC (13-order and their delta) and PCA method (15-principal components and their delta).

As can be seen from Fig. 11, the use of PCA instead of DCT improves the recognition rates for the 1st utterance from 79.1% to 85.2% (13-order MFCC and 15-principal components). These results clearly show that the use of PCA achieves better performance than DCT when dealing with a 1st utterance. In addition, the recognition rates of the other utterances were equal to MFCC in recognition.

D. Results using integration of metamodels and acoustic models

We integrated metamodels and acoustic models by using three-best words. The weight was changed from 0 to 0.4. Figure 13 shows the recognition rates for the 1st utterance. Figure 12 shows the recognition rates for each

utterance using Metamodel-PCA, MFCC and PCA methods. As can be seen from Fig. 13, the use of metamodels and acoustic models improves the recognition rate for the 1st utterance from 79.1% to 89.0% with the weight 0.35. Figure 14, 15, 16 and 17 show the recognition rates for the 2nd, 3rd, 4th and 5th utterance, respectively. As can be seen from Fig. 13 to 17, all utterances except the 5th utterance obtained the best improvement with a weight of 0.35. So the difference of the best weight between each utterance was minimal. However, we may need to change the weight for each person with articulation disorders.

Figure 18 shows the average recognition rate for five utterances using Metamodel and the other methods. As can be seen from Fig. 18, the use of metamodels and acoustic models improves the average recognition rate for five utterances from 87.6% to 90.6% with a weight 0.35. These results clearly show that the use of integration achieves good performance. It can be expected that integration will decrease the influence of substitutions, deletions and insertions caused by phone fluctuations that are not taken into account in the feature extraction.

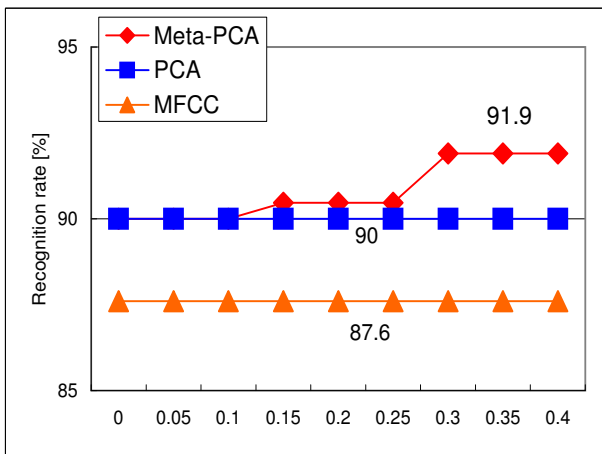


Figure 14. Recognition rates for the 2nd utterance using the proposed method

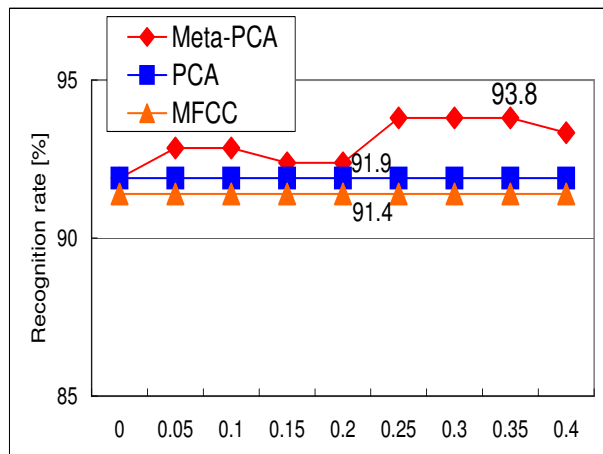


Figure 16. Recognition rates for the 4th utterance using the proposed method

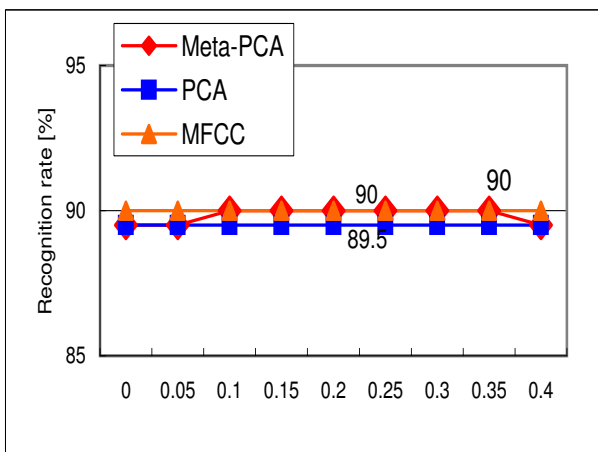


Figure 15. Recognition rates for the 3rd utterance using the proposed method

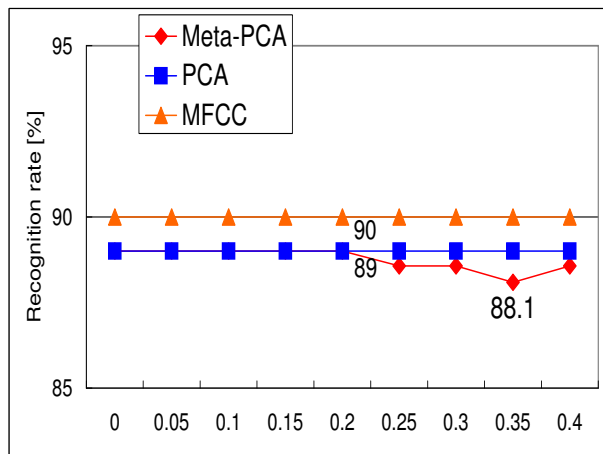


Figure 17. Recognition rates for the 5th utterance using the proposed method

*E. Results using integration of metamodels and acoustic models for another person with an articulation disorder*

We evaluated another person with an articulation disorder (Speaker B) to research the relationship between the weight and individuality. We recognized 585 utterances (195 words repeating three times) and used 2,820 utterances for training. The utterance signal was sampled at 16 kHz and windowed with a 25-msec Hamming window every 10 msec. The acoustic model consists of an HMM set with 54 context-independent phonemes with 24 dimensional MFCC features (12-order MFCCs and their delta) and 6 mixture components for each state. We integrated metamodels and acoustic models by using two-best words. The weight was changed from 0 to 0.5. Figure 19 shows the recognition rates for Speaker B.

As can be seen from Fig. 19, the use of metamodels and acoustic models improves the recognition rate from 56.6% to 58.5% with a weight 0.4. The experiment results show there was a variability of the best weight among persons due to differences in degree of dysarthria.

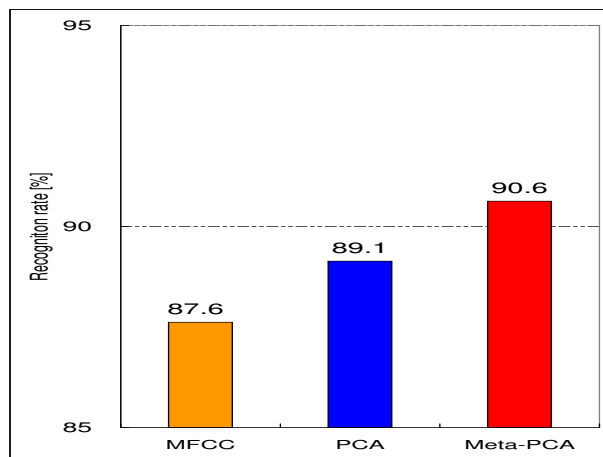


Figure 18. Average recognition rate for five utterances for each method

VI. SUMMARY

The articulation of speech uttered by a person with speech disorders tends to become unstable due to strain on their speech-related muscles. This paper has described a robust PCA-based feature extraction and integration

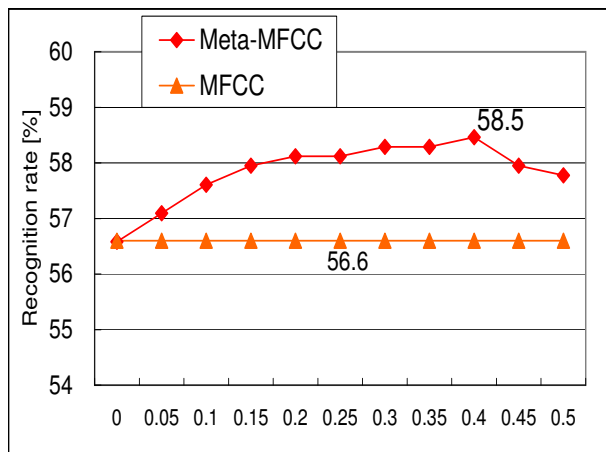


Figure 19. Average recognition rate for three utterances with the proposed method (Speaker B)

of metamodels and acoustic models. In the feature extraction, PCA is applied to the mel-scale filter bank output. It can be expected that PCA will project the main stable utterance elements onto low-order features, while elements associated with fluctuations in speaking style will be projected onto high-order features.

The fluctuation of speaking style may cause phone fluctuations such as substitutions, deletions and insertions. The integration of metamodels and acoustic models enables fluctuation suppression in the recognition process, where a phoneme metamodel handles correct decodings or substitutions. Because the degree of dysarthria differs from person to person, we need to change the integration weight for each person with articulation disorders.

The proposed method resulted in an improvement of 9.9% (from 79.1% to 89%) in the recognition rate compared to the conventional method (MFCC-based acoustic models only).

In this study, there were only two subjects tested, so in future experiments, we will increase the number of people tested and further examine the effectiveness of the proposed method.

#### REFERENCES

- [1] J. Lin, W. Ying and T. S. Huang, "Capturing human hand motion in image sequences," *IEEE Motion and Video Computing Workshop*, pp. 99–104, 2002.
- [2] T. Starner, J. Weaver and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12), pp. 1371–1375, 1998.
- [3] G. Fang, W. Gao and D. Zhao, "Large vocabulary sign language recognition based on hierarchical decision trees," *Proceedings of the 5th international conference on Multimodal interfaces*, pp. 125–131, 2003.
- [4] N. Ezaki, M. Bulacu and L. Schomaker, "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons," *ICPR 2004*, pp. 683–686, 2004.
- [5] M. K. Bashar, T. Matsumoto, Y. Takeuchi, H. Kudo and N. Ohnishi, "Unsupervised Texture Segmentation via Wavelet-based Locally Orderless Images (WLOIs)

and SOM," 6th IASTED International Conference COMPUTER GRAPHICS AND IMAGING, 2003.

- [6] V. Wu, R. Manmatha and E. M. Riseman, "Textfinder: an automatic system to detect and recognize text in images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11), pp. 1224–1229, 1999.
- [7] T. Ohsuga, Y. Horiuchi, A. Ichikawa, "Estimating Syntactic Structure from Prosody in Japanese Speech," *IEICE Transactions on Information and Systems*, 86(3), pp. 558–564, 2003.
- [8] K. Nakamura, T. Toda, H. Saruwatari and K. Shikano, "Speaking Aid System for Total Laryngectomees Using Voice Conversion of Body Transmitted Artificial Speech," *INTERSPEECH-2006*, pp. 1395–1398, 2006.
- [9] H. Chung, J. Park, Y. Lee and I. Chung, "Fast speech recognition to access a very large list of items on embedded devices," *IEEE Transactions on Consumer Electronics*, 54(2), pp. 803–807, 2008.
- [10] L. Lamel, E. Bilinski, J. L. Gauvain, G. Adda, C. Barras and X. Zhu, "The LIMSI RT07 Lecture Transcription System," *Springer Berlin/Heidelberg, Lecture Notes in Computer Science*, 4625, pp. 442–449, 2008.
- [11] D. Giuliani and M. Gerosa, "Investigating recognition of children's speech," *ICASSP2003*, pp. 137–140, 2003.
- [12] S. T. Canale and W. C. Campbell, "Campbell's Operative Orthopaedics," *Mosby-Year Book*, 2002.
- [13] S-M. Lee, S-H. Fang, J-W. Hung and L-S. Lee, "Improved MFCC Feature Extraction by PCA-Optimized Filter Bank for Speech Recognition," *Automatic Speech Recognition and Understanding, ASRU*, pp. 49–52, 2001.
- [14] T. Takiguchi and Y. Ariki, "Robust Feature Extraction Using Kernel PCA," *ICASSP2006*, pp.509–512, 2006.
- [15] H. Matsumasa, T. Takiguchi, Y. Ariki, I. LI and T. Nakabayashi, "PCA-Based Feature Extraction for Fluctuation in Speaking Style of Articulation Disorders," *INTERSPEECH-2007*, pp. 1150–1153, 2007.
- [16] O. C. Morales and S. Cox, "Modelling Confusion Matrices to Improve Speech Recognition Accuracy, with an Application to Dysarthric Speech," *INTERSPEECH-2007*, pp. 1565–1568, 2007.
- [17] S. Cox and S. Dasmahapatra, "High-Level Approaches to Confidence Estimation in Speech Recognition," *IEEE Trans. on SAP*, vol. 10, No. 7, pp. 460–471, 2002.
- [18] S. Young et. al., "The HTK Book," *Entropic Labs and Cambridge University*, 1995-2002.
- [19] V.V. Digalakis and L.G. Neumeyer, "Speaker Adaptation Using Combined Transformation and Bayesian Methods," *IEEE Trans. Speech and Audio Proc*, 4(4), pp. 294–300, 1996.

**Hironori Matsumasa** received a B.E. degree in computer science from Kobe University in 2007. His current research interests include speech recognition. He is a member of ISCA, ASJ.

**Tetsuya Takiguchi** received the B.S. degree in applied mathematics from Okayama University of Science, Okayama, Japan, in 1994, and the M.E. and Dr. Eng. degrees in information science from Nara Institute of Science and Technology, Nara, Japan, in 1996 and 1999, respectively. From 1999 to 2004, he was a researcher at IBM Research, Tokyo Research Laboratory, Kanagawa, Japan. He is currently a Lecturer with Kobe University. His research interests include robust speech recognition, signal processing, and microphone arrays. He received the Awaya Award from the Acoustical Society of Japan in 2002. He

is a member of the IEEE, the Information Processing Society of Japan, and the Acoustical Society of Japan.

**Yasuo Arika** received his B.E., M.E. and Ph.D. in information science from Kyoto University in 1974, 1976 and 1979, respectively. He was an assistant professor at Kyoto University from 1980 to 1990, and stayed at Edinburgh University as visiting academic from 1987 to 1990. From 1990 to 1992 he was an associate professor and from 1992 to 2003 a professor at Ryukoku University. Since 2003 he has been a professor at Kobe University. He is mainly engaged in speech and image recognition and interested in information retrieval and database. He is a member of IEEE, IPSJ, JSAI, ITE and IIEEJ.

**I-Chao Li** received his B.E., M.E. and Ph.D. in economics from Kobe University in 1995, 1997 and 2000, respectively. He was an assistant professor at Otemon Gakuin University from 2005 to 2008. Since 2008 he has been an associate professor at Otemon Gakuin University. He is mainly engaged in Social welfare and policies for disabled. He is a member of The Society of Economic Sociology, Japan Welfare Sociology Association and Japan Society for Disability Studies.

**Toshitaka Nakabayashi**

Department of Clinical Developmental Disabilities, Kobe University