Home Appliance Control Using Speech Recognition for a Person with an Articulation Disorder

Ryo AIHARA, Yuki TAKASHIMA, Tetsuya TAKIGUCHI, Yasuo ARIKI

Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada, Kobe, 657-8501, Japan

Abstract. In this paper we present a home appliance control system using automatic speech recognition for a person with an articulation disorder resulting from athetoid cerebral palsy. Such a people have a speech style that is unique to each person and the style changes depending on conditions which makes their speech difficult to recognize. In this paper, we build an acoustic model using dysarthric speech and adapt the model to their condition by using a model adaptation technique.

1 Introduction

In this paper, we focused on a person with a motor speech articulation disorder resulting from the athetoid type of cerebral palsy. In the case of a person with this type of articulation disorder, his/her movements are sometimes more unstable than usual. That means their utterances are often unstable or unclear due to the athetoid symptoms. Athetoid symptoms also restrict the movement of their arms and legs so there is great need for hands-free voice systems for them.

We introduce in this paper a home appliance control system using Automatic Speech Recognition (ASR) for a person with an articulation disorders. As reported in [1], the speech style of a person with an articulation disorder is unique to each person. Moreover, we investigated observed that their speech style can changes easily depending on their physical or mental conditions [1]. In this paper, in order to overcome such phenomena, we present an acoustic model using dysarthric speech and adapt the model to their speech conditions by using Maximum Likelihood Linear Regression (MLLR) and Maximum A Posteriori (MAP) [2]. Our method improved the recognition rate by 27.5% compared to an ASR system without adaptation.

2 Home Appliance Control System using ASR

Figure 1 shows the flow of our system. Our system can control 3 appliances. A user can input the commands by voice that are shown on the screen. Our system has 3 layers and recognition grammar is changed in each phase where the average number of commands is 8.

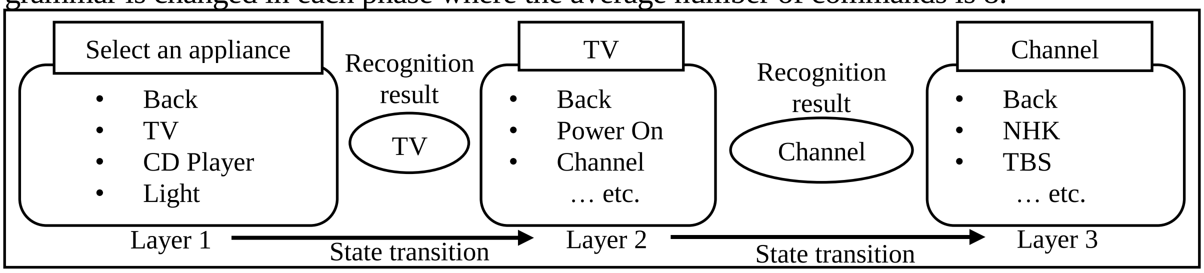

Fig. 1 Flow of Home Appliance Control System

3 Adaptation for Articulation Disordered Voice

MLLR uses linear transformation of Gaussian model parameters to adapt to a given speaker. The adapted n-dimensional mean vector $\hat{\boldsymbol{\mu}}$ is given as follows:

$$\hat{\boldsymbol{\mu}} = \mathbf{W}\boldsymbol{\xi} \qquad (1)$$

where $\mathbf{W} = [\mathbf{b} \quad \mathbf{A}]$ represents the $n \times (n+1)$ dimensional transformation matrix which consists of bias vector $\mathbf{b}$ and $n \times n$ dimensional transformation matrix $\mathbf{A}$. $\boldsymbol{\xi} = [\varpi \quad \mu_1 \quad \mu_2 \ldots \quad \mu_n]^{\mathrm{T}}$ represents the extended mean vector. ϖ represents the bias term. $\mathbf{W}$ is caliculated as follows by using adaptation data $\boldsymbol{\theta}$:

$$\mathbf{W} = \arg\max_{\overline{\mathbf{W}}} \{\boldsymbol{\theta} \mid \overline{\mathbf{W}}\} \quad (2)$$

MAP estimation maximizes posterior distribution by using adaptation data from a pre-trained acoustic model that is used as prior knowledge. The estimated n dimensional mean vector $\hat{\boldsymbol{\mu}}$ is given as follows:

$$\hat{\mu}_{jm} = \frac{N_{jm}}{N_{jm} + \tau} \bar{\mu}_{jm} + \frac{\tau}{N_{jm} + \tau} \mu_{jm} \quad (3)$$

where j , m , τ , N , $\bar{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ represent the number of the state of HMM, number of Gaussian mixtures in each state, weight of prior knowledge, likelihood of adaptation data, mean vector of adaptation data, and mean vector of pre-trained acoustic model, respectively. In this paper, we adapt a pre-trained acoustic model to their speech condition by using (1) and (3).

4 Experimental Results

Our task contains 5 scenes and 34 commands. Our subject is a Japanese male with athetoid-type cerebral palsy. Fig. 2 shows the word recognition accuracies comparing our method with the baseline. The acoustic model used in the baseline is trained with 1,050 articulation-disordered utterances that were recorded at his home. Test data is 125 articulation-disordered utterances that were recorded in our laboratory. The other baseline conditions are the same as [1]. The average baseline recognition rate is 56.4% because our subject got nervous due to the recording conditions and his speech style became different from the training data. After we adapted 10 utterances from the test data using the model adaptation technique, the average recognition rate improved to 83.9%.

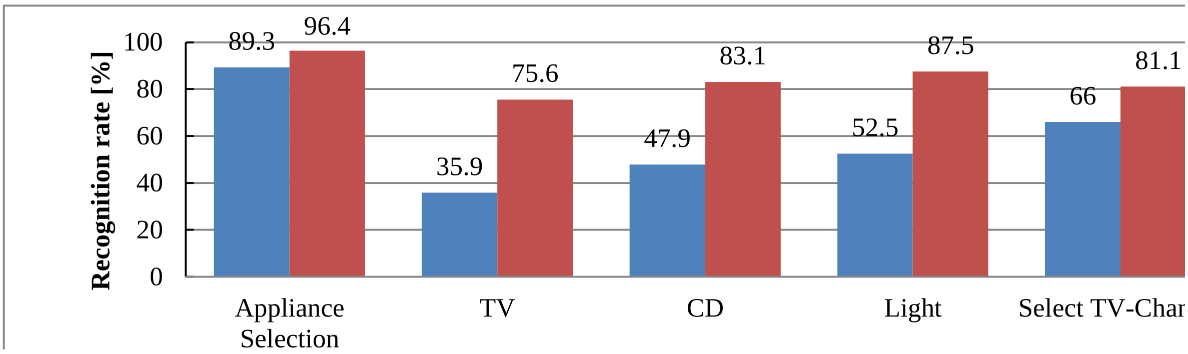

Fig. 2 Recognition results.

5 Conclusions

We introduced a hands-free home appliance control system using ASR for a person with athetoid cerebral palsy. The experimental results show that our acoustic model adaptation technique is able to improve the recognition rate of an articulation-disordered voice.

References

[1] H. Matsumasa *et. al.* Journal of Multimedia, Vol.4, Issue. 4, pp. 254-262, 2009.
[2] K. Shinoda, Electronics and Communications in Japan, Part 3, Vol. 88, No. 12, pp.25-42, 2005.