Title	ヤコビ適応法を用いた雑音・伝達特性・発声変形への同時適応
Author(s)	坂井，伸圭
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1559
Rights	
Description	Supervisor:嵯峨山 茂樹，情報科学研究科，修士

Jacobian Joint Adaptation to Noise, Channel and Lombard Effect

Nobuyoshi Sakai (010047)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 2002

Keywords: speech recognition, Jacobian adaptation, Lombard Effect.

1 Introduction

Because of the progress of speech recognition technology, an automatic speech recognition (ASR) system can show high recognition performance in an ideal environment without noise. However, the recognition rate decreases greatly in a real environment where noise and channel distortion exist. This is caused by the mismatch between the assumed acoustic model and the observed acoustic model. The factors that cause such mismatch include additive noise, channel distortion and a articulation variability which is called the Lombard effect caused by the presence of background noise.

In order to achieve robust ASR, it is necessary to adapt the acoustic models to those factors. A number of methods have been proposed so far for model adaptation to one of the factors. However, none has been proposed for the case that all of these factors are mixed simultaneously, because these factors affect the observed speech not linearly but non-linearly and hence aplying each adaptation technique for each factor does not solve the problem.

Then I discusses a method of adapting acoustic model to these factor simiultaneously, using Jacobian Adaptation which has a merit that it solves by linear approximation function using first order coefficient of the Taylor series. it has already been reported about Jacobian Joint Adaptation to Noise and Channel.

In this paper, the technique which models the movement of the formant frequency due to Lombard effect by using the frequency axis expansion and contraction in a spectrum area and apply Jacobian Adaptation it is proposed. In addition, it adapts acoustic model to noise, channel distortion, Lombard effect, and these change factors simiultaneously, and the effect is confirmed through the experiment.

2 Jacobian Adaptation to Lombard effect

The formant frequency makes about 1.5kHz a boundary and the one in the high frequency area is reported to a low region on the movement of the one in the low frequency number area to a high region by Lombard effect.

The movement of this formant frequency is modeled as follows by the use of the frequency axis expansion and contraction in a spectrum area.

Frequency f' after the frequency axis contracts to Spectrum of the ordinary voice, $S(f)$ is shown by the next expression by using frequency axis expansion and contraction coefficient λ and constant f_α

$$f' = \lambda f + (1 - \lambda) f_\alpha \tag{1}$$

Here, $\lambda > 1$ is meant shrinkage to face frequency f_α and extends oppositely for $\lambda < 1$. They are in the relation of the next expression in case of cepstrum of voice spectrum $S(f)$ is $\boldsymbol{C}$ and cepstrum of voice spectrum $S(f')$ is $\boldsymbol{C}_L$

$$\boldsymbol{C}_L = \boldsymbol{F}^{-1} \boldsymbol{F}_L \boldsymbol{C} \tag{2}$$

Here, $\boldsymbol{F}^{-1}$ is the reverse matrix of discrete cosine transform procession $(\boldsymbol{F})$, and $\boldsymbol{F}_L$ are the discrete cosine transform processions to with frequency axis expansion and contraction of the above-mentioned, and the (i, j) element is given by the next expression.

$$(\boldsymbol{F}_L)_{ij} = \cos \frac{i(\lambda(j + 0.5) + (1 - \lambda)\omega_\alpha)\pi}{N} \tag{3}$$

N is sampling point of and frequency ω_α is frequency axis constant correspond to f_α. If expansion and contraction coefficient λ changed only slight amount $\Delta\lambda$ due to Lombard effect, amount $\Delta \boldsymbol{C}_L$ of the change of cepstrum $\boldsymbol{C}_L$ is given by the next expression by a linear approximation by the Taylor development.

$$\Delta \boldsymbol{C}_L = \frac{\partial \boldsymbol{C}_L}{\partial \lambda} \Delta\lambda \tag{4}$$

Therefore, Lombard effect Jacobian matrix $\boldsymbol{J}_L \equiv \frac{\partial \boldsymbol{C}_L}{\partial \lambda}$ is At $\lambda \approx 1$

$$\boldsymbol{J}_L = \boldsymbol{F}^{-1} \frac{\partial \boldsymbol{F}_L}{\partial \lambda} \boldsymbol{C} \;\approx\; \boldsymbol{F}^{-1} \boldsymbol{G} \boldsymbol{C} \tag{5}$$

(i, j) element of $\boldsymbol{G}$ is

$$(\boldsymbol{G})_{ij} = \frac{-i(j + 0.5 - \omega_\alpha)\pi}{N} \sin\left(\frac{i(j+0.5)\pi}{N}\right) \tag{6}$$

This $\boldsymbol{J}_L$ is obtained in the assumed initial environment beforehand. In this case, change $\Delta\lambda$ of the frequency axis expansion and contraction coefficient are calculated from the expression 4 by the least square method using changes obtained by doing Viterbi Time Alignment to a vocal sample.

3 experiment result

The Jacobian Adaptation to Lombard effect from the acoustic model made by the ordinary voice to the model for Lombard voice was done, and it was experimented on the isolated-word speech recognition, and the effect was confirmed. Though the improvement of the recognition rate was about 2% in the Jacobian Adaptation which used the Lombard model even when best $\omega_\alpha$was used, the recognition rate improvement of about 15% was seen in the adaptation simultaneously at the time of having added the adaptation to multiplication and addition which is Jacobian Adaptation to noise and channel. It was able to be confirmed that the transformation element by addition and multiplication to Lombard Effect was large. Moreover, Jacobian Adaptation to Noise, Channel and Lombard Effect was done simultaneously to the Lombard voice to which noise and channel distortion weresuperimposed , and the improvement of the recognition rate of 16-19% was confirmed.