ECHO CANCELLATION AND MAP ADAPTATION FOR HANDS-FREE SPEECH RECOGNITION

6C－12

Sung-Ill Kim , Tetsuro Kitazoe
Department of Computer Science and Systems Engineering
Faculty of Engineering , Miyazaki University
1-1 , Gakuen Kibanadai Nishi , Miyazaki , 889-2192 Japan

Abstract

For some applications such as hands-free teleconferencing and telecommunication systems, the near-end speech signal to be transmitted is disturbed by ambient noise and by an echo due to the coupling between the microphone and the loudspeaker . In the present paper, we introduce a new approach using echo cancellation and Maximum A Posteriori (MAP) environmental adaptation technique to improve hands-free speech recognition accuracy . In this approach, it is shown that our proposed new system is effective for hands-free speech recognition in the echo and noise environments .

1 INTRODUCTION

In hands-free speech recognition, since the microphone and the loudspeaker are coupled, the sound from the loudspeaker is picked up by the microphone, both directly and indirectly and this is heard by the recognizer as echo, causing undesirable speech recognition results . In the past few years, many works have been performed in HMMs to improve the speech recognition accuracy with a close-talking microphone . But especially recently, the dissemination of hands-free communication systems requires to provide users with some comfort . Therefore, problems of reverberant speech recognition have to be solved to obtain a good distant-talking speech recognition accuracy . The problem of environmental noise including channel distortion or additive noise is another factor - apart from the acoustic echo which is fed back from the hands-free mode . For instance, the speech will be mixed with any noise present within the vehicle - echo cancellers[1,2] generally find it difficult to eliminate echo if there is a high level of background noise . Therefore we extend the MAP estimation to an environmental adaptation[3] for the recognition of hands-free speech signal obtained from different environments which have channel distortion or additive noise . In automobile application area, for instance, the hands-free microphone picks up only the desired speech and removes the undesired echo by using acoustic echo canceller and then the speech recognizer improves the echo-cancelled speech recognition accuracy by using environmental adaptaion technique in a channel distortion or additive noise environment of a car .

In this paper, we report an implémentation of new approach for hands-free speech recognition using echo canceller and MAP adaptation technique . We show that the proposed system is effective for hands-free speech recognition . We also report this hands-free speech recognition rates in comparision with close-talking ones .

2 ECHO CANCELLER AND ENVIRONMENTAL ADAPTATION

Acoustic echo was first encountered with the early video/audioconferencing studios and also occurs in a typical mobile situation, such as in a car . In this situation, sound from the loudspeaker is heard by the listener, as intended . However, this same sound is also picked up by the microphone, both directly and indirectly . The result of this reflection is the creation of multipath echo and multiple harmonics of echo encountered, which unless there are eliminated, are transmitted back to the distant end and is heard by the talker as echo . With echo cancellation, complex algorithmic procedures are used to compute speech models . This involves the system generating the sum from reflected echoes, of the original speech, then subtracting this from any microphone signal it picks up . The format of this "echo prediction" is learned by the echo canceller in a process called "adaptation" .

The MAP estimation is also called bayesian succesive estimation of the HMM parameters for the new speaker in a framework . The estimated mean vector value after given N samples is shown as

$$\hat{\mu}_N = \frac{\alpha\mu_o + \sum_{i=1}^{N} X_i}{\alpha + N} \tag{1}$$

where α is an adaptation parameter . The estimated covariance matrix by using N samples is

$$\begin{aligned} \hat{\Sigma}_N &= \frac{1}{\beta+N}\{X_N X_N^T - (\alpha+N)\mu_N\mu_N^T \\ &+ (\beta+N-1)\Sigma_{N-1} \\ &+ (\alpha+N-1)\mu_{N-1}\mu_{N-1}^T\} \end{aligned} \tag{2}$$

where β is a coefficient . In this paper, the values of α, β were set at 15 and 50 respectively, which were

determined experimentally . A block diagram of the overall hands-free speech recognition system based on acoustic echo canceller and MAP environmental adaptaion is shown in Figure 1 .

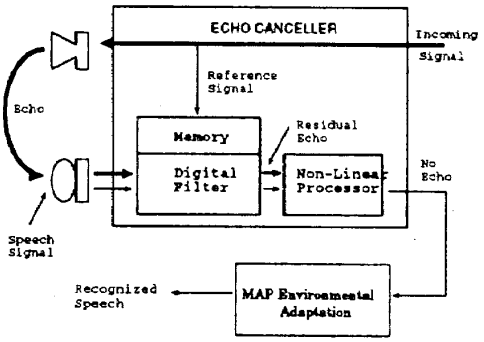

Figure 1: Overall block diagram of hands-free speech recognition system

3 EXPERIMENTAL CONDITIONS

We trained Japanese 40 phoneme HMM models using 5240 labeled word utterances of 10 male speakers and 503 sentences of 6 male speakers in the ATR Japanese speech database . In the test, we used ATR 100 phoneme balanced words . A set of 20 dimensional observation sequences including discrete duration information are obtained for recognition . Table 1 shows the preprocessing analysis condition of the speech data .

Table 1: Analysis of speech signal

sampling rate	16kHz,16bit
preemphasis	0.97
window function	16 msec Hamming window
frame period	5 ms
feature parameters	10-order MFCC +10-order delta MFCC +log power+delta log power +discrete duration information
model topology	3 state left-light phone model

4 EXPERIMENTS AND RESULTS

Figure 2 illustrates a plot of word recognition rates versus number of adaptation words for different speech signal modes . The close-talking and hands-free speech data were recorded using clean speech in a laboratory room environment, including ambient noise such as air conditioning systems, fans, etc . In hands-free mode experiments, we use a station noise data from the noise database of Japan Electronic Industry Development Association(JEIDA) as echo .

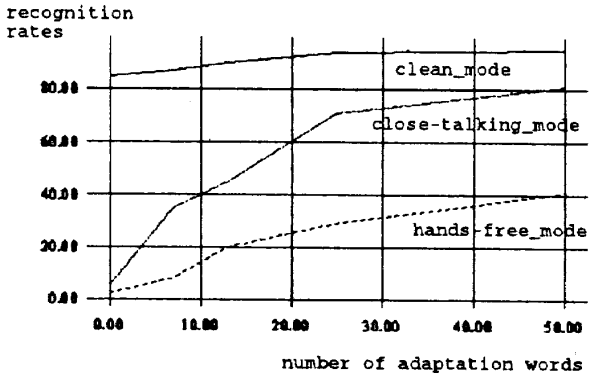

Figure 2: Comparision of three kinds of speech siganl modes

Though the recognition rates of the hands-free mode are lower than that of other modes owing to the over-cancelled speech signal and additive noise, it shows us that the combination of echo canceller and MAP adaptaion technique is effective for hands-free speech recognition .

5 CONCLUSIONS

This paper has described an efficient method of hands-free speech recognition based on the use of acoustic echo canceller and MAP environmental adaptation technique . The experimental results indicated that the hands-free mode using echo canceller and MAP adaptaion improved the recognition rate for the echo and additive noise environments . In the future, we will concentrate on the recognition rate improvement for hands-free speech recognition in task-specific applications .

References

[1] W. Kellermann : "Analysis and Design of Multirate Systems for Cancellation of Acoustical Echoes", Proc. ICASSP, pp. 2570-2573, 1988

[2] R. Martin and J. Altenhoner : "Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction", Proc. ICASSP, pp. 3043-3046, 1995

[3] Y. Tsurumi and S. Nakagawa : "An Unsupervised Speaker Adaptation Method for Continuous Parameter HMM by Maximum A Posteriori Probability Estimation", Proc. ICSLP, pp. 431-434, 1994