Topics

Improvement of Speech Recognition Performance in Car Environment

Hiroyuki Hoshino, Human Interface Lab.

1. Introduction

The degradation of recognition performance in a traveling car environment is one of the problems with the speech recognition function beginning to be adopted in navigation systems. A method for improving the recognition rate has been developed through this study by optimum filtering of the engine noise as one of the factors causing the degradation of the speech recognition performance in a car environment.

2. Result of recognition performance evaluation

Speech data (one hundred location names spoken by 5 men and 5 women) superposed with noise in 8 types of cars (during acceleration, deceleration, constant-speed driving and traveling on a coarse road surface) were prepared, and the speech recognition was tested using the HMM (Hidden Markov Model).

As a result of checking the relation between various acoustic characteristics of the noise and the recognition rate, it has been found that the speech articulation index (indicating ease of speech hearing in noisy environment) and the recognition rate have almost a positive correlation when driving at a constant speed and when driving during acceleration/deceleration (**Fig. 1**). Even when the articulation index is the same, the recognition rate is generally poorer and occasionally significantly drops during acceleration or deceleration compared to that during constant-speed driving.

As a result of checking the frequency spectrum of the noise, it has been found that the speech recognition rate significantly worsens when there are large engine noise components (periodic noise components) in the frequency range below 500 Hz (**Fig. 2**).

3. Method for improving the recognition performance

In this study, the cutoff frequency of the high-pass filter used for preprocessing of the recognition was varied according to the engine noise level. The characteristic representing the engine noise level can be calculated as the protrusion of the noise components from the frequency spectrum envelope or the difference between the overall frequency component level and the overall frequency spectrum envelope level. The engine noise band can be optimally cut off by varying the high-pass filter cutoff frequency in a range between 200 and 400 Hz according to the value of this difference (Fig. 2). **Fig. 3** shows the speech recognition improvement results during the second range acceleration of eight cars. The mean recognition rate of eight cars was improved from 46.0% to 57.9%. Especially for car E, the recognition rate was significantly improved from 28.2% to 66.8%.

Reference

1) Hoshino, H., et al. : "Improvement of Noise Robustness of Speech Recognition in Car Environments Using Acoustic Features of Car Interior Noise", Proc. Int. Workshop on Hands-Free Speech Commun., (2001), 151, ATR

(Report received on Sept. 25, 2001)

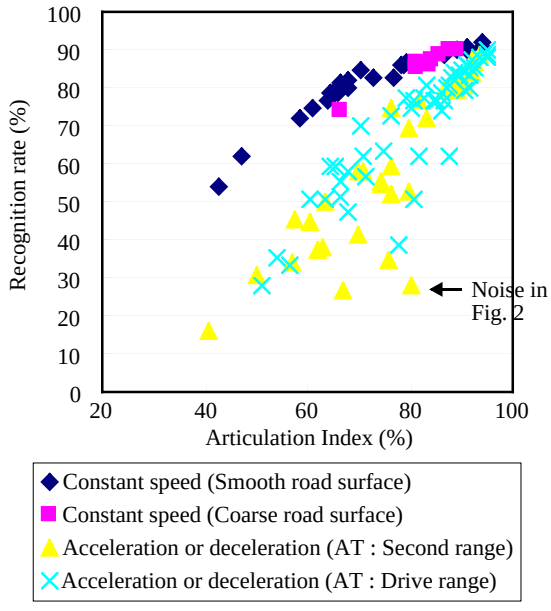

Fig. 1 Relation between the articulation index and the recognition rates under various driving conditions of eight cars.

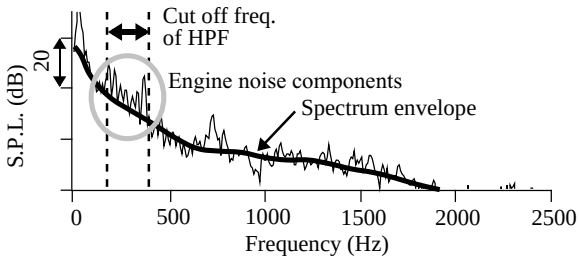

Fig. 2 Frequency spectrum and spectrum envelope of car interior noise, and cut off frequency area of high pass filtering.

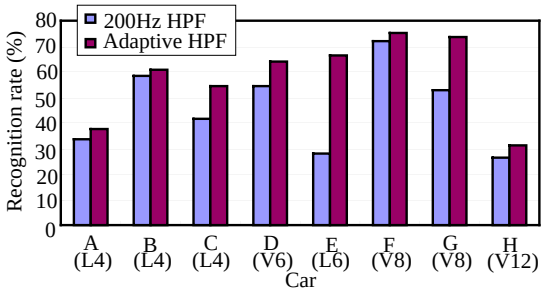

Fig. 3 Recognition results under the accelerating condition of eight cars.