Research Report

Noise-Robust Speech Recognition in a Car Environment Based on the Acoustic Features of Car Interior Noise

Hiroyuki Hoshino

車内音の音響的特徴を用いた自動車車室内におけるノイズロバスト音声認識

星野博之

Abstract

This paper describes an efficient method of improving the noise-robustness of speech recognition in a noisy car environment by considering the acoustic features of a car's interior noise. We analyzed the relationship between the Articulation Index values and the recognition rates in car environments under different driving conditions. We clarified that the recognition rate significantly worsens when the engine noise (periodic sound) components in the frequency range above 200 Hz were large. We developed a preprocessing method to improve the noise-robustness despite large amounts of engine noise. With this method, the cutoff frequency of the front-end high-pass filter is adaptively changed from 200 through 400 Hz according to the level of the engine noise components. The use of this method improved the average recognition rate for all eight cars under the second range acceleration condition by 11.9%, with the recognition rate for one of the cars being improved considerably by 38.6%.

Keywords: Robust speech recognition, Car interior noise, Noise suppression, High-pass filter, Articulation Index, Spectrum envelope

要　旨

本論文では，車室内騒音下における音声認識のノイズロバスト性を向上するための，騒音の音響的特徴を考慮した効率的な手法について述べる。我々は様々な走行状況における車室内環境下での会話明瞭度指標値と音声認識率との関係を解析した。その結果我々は，約200Hz以上の周波数帯域におけるエンジンノイズ（周期音）成分が大きい場合，認識率がかなり悪化することを明らかにした。そして，我々は大きなエンジンノイズ成分が存在する条件下においてノイズロバスト性を向上させる前処理法を開発した。これは，エンジンノイズ成分の大きさにより，前処理でのハイパスフィルタのカットオフ周波数を200Hzから400Hzの間で適応的に変化させるものである。この手法により，セカンドレンジ加速走行下における8車種の平均認識率は11.9%向上し，特にその中の1車種では認識率が38.6%向上した。

キーワード：ロバスト音声認識，自動車車室内騒音，ノイズ抑圧，ハイパスフィルタ，会話明瞭度指標，スペクトル包絡

1. Introduction

Recently, to improve driver safety and convenience, human-machine interfaces employing speech recognition have been widely adopted for in-vehicle information equipment such as navigation systems. One of the greatest problems facing speech recognition in car environments is the degradation of the recognition performance as a result of car interior noise while driving. Many research projects into noise-robust speech recognition in car environments have been performed and, as a result, robustness has improved remarkably in recent years.[1-4] For stationary or slowly varying additive noise, such as road and wind noise in car environments, spectral subtraction is a simple and efficient speech enhancement method.[5, 6] Non-linear spectral subtraction has also been proposed recently and has been shown to offer better performance.[7, 8] Both spectral subtraction and non-linear spectral subtraction use a time-averaged estimate of the noise spectra as their noise information. They do not, however, take the acoustic features of noise into account.

The acoustic features of car interior noise vary depending on the type of the car and the driving conditions, such as the vehicle speed, engine revs, road surface, and the direction and/or strength of the wind. This study considers only the car interior noise that is caused by the car itself and the driving conditions, such as engine noise, road noise and wind noise. Acoustically, engine noise is composed of periodic sound components at frequencies of less than 1000 Hz, road noise is composed of random noise components at frequencies of less than 1000 Hz, and wind noise is composed of random noise components at frequencies above 500 Hz inside the car. To efficiently improve the robustness of speech recognition in noisy car environments, we should consider the acoustic features of the car interior noise. Unfortunately, relatively few studies have specifically considered the acoustic features of noise in the development of noise-robust speech recognition.

The purpose of this study was to develop an efficient method of improving the noise-robustness of speech recognition in noisy car environments by considering the acoustic features of the car interior noise. We performed a speaker-independent isolated word recognition experiment in noisy environments under various driving conditions in eight types of car to clarify the acoustic features of the noise that affects the degradation of the recognition performance. We also developed an engine-noise-adaptive high-pass filtering method to improve the noise-robustness despite there being large amounts of engine noise.

2. Recognition experiment and analysis

2. 1 Speech and noise data

The recognition experiment used clean speech data and car interior noise data that had been recorded inside each of the eight cars. For the clean speech data, we recorded ten speakers' (five male and five female) voices, uttering 100 isolated words via a loudspeaker positioned to a point corresponding to a driver's mouth with a non-directional microphone mounted on the driver's sun visor. The output level of the loudspeaker was adjusted to that of one male speaker's average level. For the noise data, the interior noise data from eight cars was recorded under the driving conditions shown in **Fig. 1**.

The driving conditions included acceleration, deceleration, constant speed and coarse road surfaces. For the experiment, we made noise-overlapped speech data by adding the car interior noise data to the clean speech data.

2. 2 Recognition experiment

A speaker-independent isolated word recognition experiment was carried out using a speech recognition system that is based on the hidden Markov model (HMM). The noise-overlapped speech data was preprocessed by a 200-Hz high-pass filter, because this frequency range features an

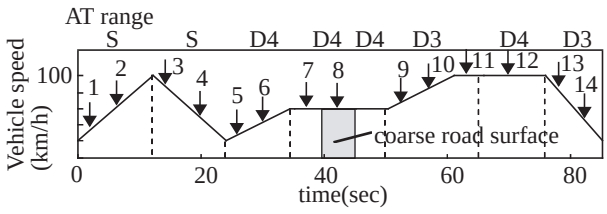

Fig. 1 The driving conditions for speech recognition experiments.

extremely large amount of noise components but very few speech components. Furthermore, this data was processed by non-linear spectral subtraction for speech enhancement.[3, 7, 8] The representation of speech was based on the LPC analysis. Speech was sampled at 12 kHz and features (16 cepstrum coefficients, 16 delta-cepstrum coefficients, log power, and delta log power) were extracted from Hamming-windowed 20-ms frames at a rate of 10 ms.

2. 3 Experimental results and analysis

Figure 2 shows the recognition results for noisy environments encountered under various driving conditions in eight different types of cars. We can see that the recognition rates differed between the eight cars and that they deteriorated significantly under acceleration conditions. However, they did not deteriorate so much in the case of coarse road surface conditions (driving condition No.8). The reason for this is that the main components of the road noise caused by the coarse road surface are of frequencies less than 200 Hz.

We examined the relationship between several acoustic features of the car interior noise after the 200-Hz HPF processing, and the recognition rates under constant speed conditions (driving conditions No.7, 8, 11, 12). The correlation coefficients of the typical acoustic features are listed below:

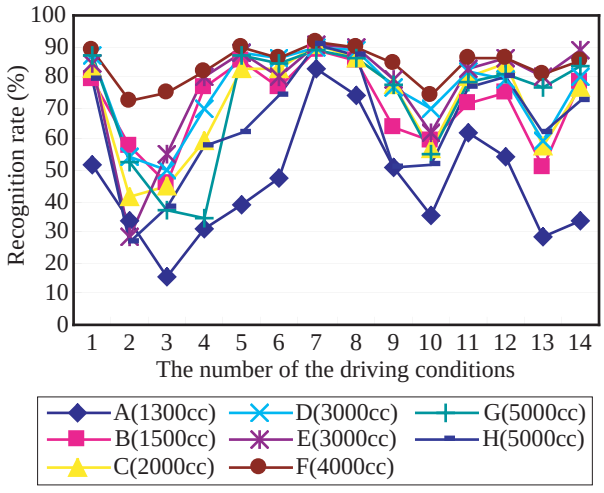

Fig. 2 Recognition results for noisy environments under various driving conditions in eight different types of cars.

- A-weighted overall level: r = 0.82
- Partial overall level above 450 Hz: r = 0.90
- Articulation Index: r = 0.96

There is a good correlation between the Articulation Index values for the noise and the recognition rates. The Articulation Index is a physical measure that is highly correlated with the intelligibility of speech. It is calculated by summing the evaluated values derived from the 1/3-octave band level of noise and the band-specific weighting factors of the 1/3-octave bands.[9, 10] Therefore, speech recognition performance in steady noise environments can be evaluated based on the Articulation Index. This means that the performance is degraded by the amount of noise in the main frequency bands of speech.

Figure 3 shows the relationship between the Articulation Index values and the recognition rates under the driving conditions encountered in the eight different cars. From this figure, we can determine the following:

(1) The recognition rates deteriorated under acceleration, relative to constant speed conditions, even though the Articulation Index values were the same.

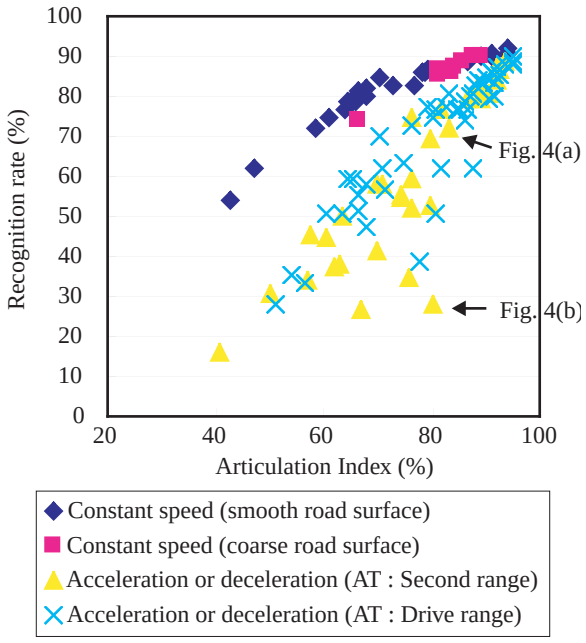

Fig. 3 Relationship between the Articulation Index values and the recognition rates under the driving conditions encountered in the eight different cars.

(2) The recognition rates under many of the acceleration or deceleration conditions deteriorated considerably in spite of the Articulation Index values being the same.

The first result indicates that the recognition performance was degraded in non-steady noise environments compared to steady noise environments, in spite of the Articulation Index values being the same. We verified this phenomenon by performing another experiment using simple non-steady noise that simulated car interior noise.

To investigate the cause of the second result, above, we analyzed the frequency spectrums of the two car interior noise samples shown in Fig. 3. **Figure 4** shows the analysis results before the 200-Hz HPF processing. Figure 4(a) shows an example in which the recognition rate is not low (driving condition No.2 with car F), and Fig. 4(b) shows an example in which the recognition rate is relatively low (driving condition No.2 with car E). From these figures, we found that the recognition rate deteriorated considerably when the large amounts of engine noise (periodic sound) components constituted a part of the frequency range above 200 Hz.

3. Adaptive high-pass filtering

3. 1 Method

From the analysis results described above, we developed a preprocessing method to improve the noise-robustness when there is a considerable amount of engine noise. With this method, the cutoff frequency of the front-end high-pass filter is adaptively changed from 200 through 400 Hz according to the level of the engine noise (periodic sound) components.

The level of the engine noise components (L_{en}) can be estimated from the difference between the power of the frequency spectrum P_S and the power of the spectrum envelope P_E of the noise, as follows.[11)]

$$L_{en} = 10 \log (P_S/P_E) \quad \cdots\cdots (1)$$

The frequency range was set to between 200 and 1000 Hz by considering the frequency area of the engine noise. The spectrum envelope was calculated from the liftered FFT cepstrum.

Figure 5 shows an analysis example of the frequency spectrum and spectrum envelope of the car interior noise under acceleration condition (driving condition No.2 with car E). It can be seen that the engine noise (periodic sound) components are a major constituent of the frequency range above 200 Hz. The L_{en} of this noise was 4.2 dB, and 400 Hz HPF processing produced the best recognition performance. On the other hand, L_{en} for the steady noise under all constant speed conditions did not exceed 1.6 dB, and 200-Hz HPF processing produced the best recognition performance.

From the examinations described above, we were able to determine a suitable cutoff frequency for HPF f_{cutoff} for L_{en}, as follows:

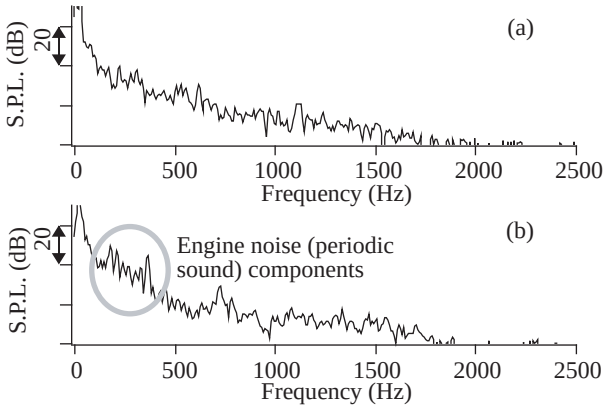

Fig. 4 Frequency spectrum analysis results of car interior noise. (a) An example in which the recognition rate is not low. (b) An example in which the recognition rate is relatively low.

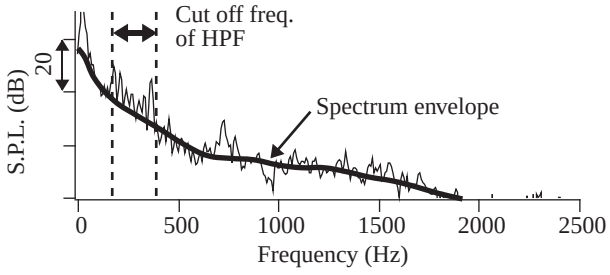

Fig. 5 An analysis example of the frequency spectrum and spectrum envelope of the car interior noise under acceleration condition (driving condition No.2 with car E).

特集

$$f_{cutoff} = \begin{cases} 200\text{Hz} & L_{en} < 1.8\text{dB} \\ 300\text{Hz} & 1.8 <= L_{en} < 2.3\text{dB} \\ 400\text{Hz} & L_{en} <= 2.3\text{dB} \end{cases} \quad \cdots\cdots\cdots (2)$$

A 500-Hz HPF degraded the recognition performance in all the environments evaluated in this study. This means that the speech frequency components above 400 Hz are necessary for speech recognition, even though the components were contaminated by noise.

Table 1 lists L_{en} and f_{cutoff} under acceleration conditions (driving condition: No.2) for each car. As shown here, the proposed method selects the HPF cutoff frequency according to the level of the engine noise components.

3. 2 Results

A speaker-independent isolated word recognition experiment was performed to evaluate the performance of the proposed method. The noise-overlapped speech data was preprocessed by engine-noise adaptive HPF instead of 200-Hz HPF. The other conditions were the same as in the first experiment.

Figure 6 shows the recognition results under the second range acceleration condition (driving condition: No.2). The average recognition rate for all eight cars under this condition was improved by 11.9 % through the use of this method, and the recognition rate for car E was improved by a considerable 38.6 %.

Figure 7 shows the relationship between the Articulation Index values and the recognition rates under a range of driving conditions for all eight cars when using the engine-noise-adaptive high-pass filtering. The relative deterioration under several different acceleration or deceleration conditions was improved, so the correlation coefficient between the Articulation Index values and the recognition rates under all acceleration and deceleration conditions rose from 0.88 to 0.94.

This result indicates that the correspondence between the machine performance and human performance of speech recognition was improved by the proposed method. However, the machine performance is still poor under conditions of non-steady noise.

Table 1 L_{en} and f_{cutoff} under the accelerating condition (driving condition: No.2).

Car	A	B	C	D	E	F	G	H
L_{en}(dB)	2.0	1.9	2.5	3.0	4.2	2.1	2.2	2.3
f_{cutoff}(Hz)	300	300	400	400	400	300	300	400

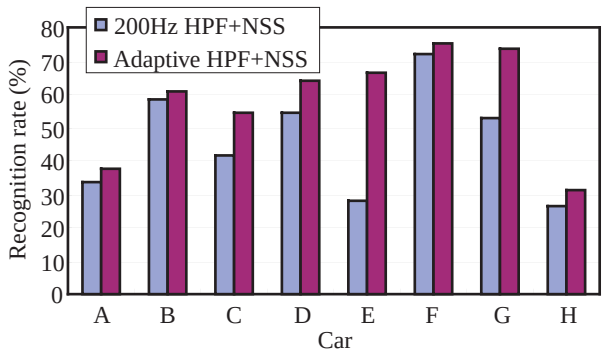

Fig. 6 Recognition results under the second range acceleration condition (driving condition : No.2).

4. Conclusions

We developed an engine-noise-adaptive high-pass

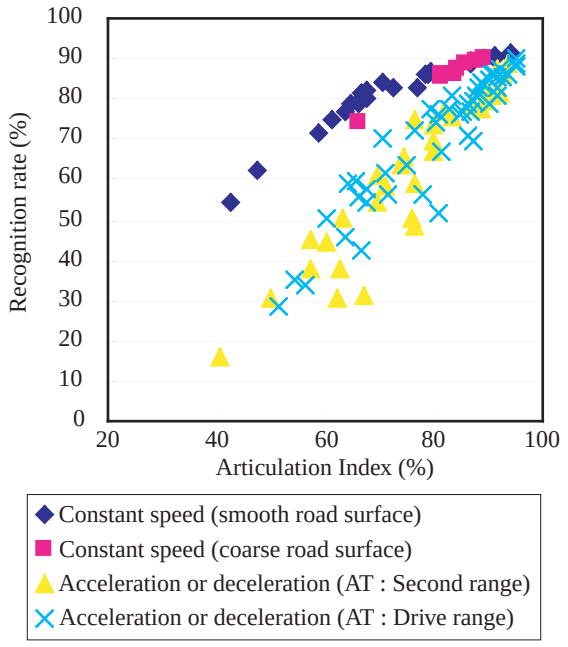

Fig. 7 Relationship between the Articulation Index values and the recognition rates under a range of driving conditions for all eight cars when using the engine-noise-adaptive high-pass filtering.

filtering method to improve the noise-robustness of speech recognition in noisy car environments. This was developed from our considering the relationship between the Articulation Index, which represents one of the acoustic features of noise, and the recognition rate. The method is relatively simple, and is also effective in improving the recognition performance despite its being degraded by large amounts of engine noise. Taking the acoustic features of noise into account as we did for this study appears to be an efficient approach to realizing noise-robust speech recognition. In the future, robust speech recognition under non-steady noise conditions should be examined.

References

1) Lecomte, I., Lever, M., Boudy, J. and Tassy, A. : "Car Noise Processing for Speech Input", Proc. ICASSP, **1989**(1989), 512, IEEE
2) Mokbel, C. E. and Chollet, G. F. A. : "Automatic Word Recognition in Cars", IEEE Trans. Speech and Audio Process., **3**-5(1995), 346
3) Shozakai, M., Nakamura, S. and Shikano, K. : "Robust Speech Recognition in Car Environments", Proc. ICASSP, **1998**-1(1998), 269, IEEE
4) Iwahashi, N., Pao, N. H., Honda, H., Minamino, K. and Omote, M. : "Stochastic Features for Noise Robust Speech Recognition", Proc. ICASSP, **1998**-2(1998), 633, IEEE
5) Boll, S. F. : "Suppression of Acoustic Noise in Speech using Spectral Subtraction", IEEE Trans. Acoust., Speech and Signal Process., **27**-2(1979), 113
6) Gong, Y. : "Speech Recognition in Noisy Environment: A Survey", Speech Commun., **16**(1995), 261
7) Lockwood, P. and Boudy, J. : "Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the Projection, for Robust Speech Recognition in Cars", Speech Commun., **11**(1992), 215
8) Vaseghi, S. V. and Milner, B. P. : "Noise Compensation Methods for Hidden Markov Model Speech Recognition in Adverse Environments", IEEE Trans. Speech and Audio Process., **5**-1(1997), 11
9) Beranek, L. L. : "The Design of Speech Communication System", Proc. Inst. Radio. Eng., **10**(1947), 880
10) ANSI S3.5-1969 : "American National Standard: Methods for the Calculation of the Articulation Index", American National Standards Institute, Inc., New York, (1969)
11) Hoshino, H. : "Balance of Car Interior Noise Components in Consideration of Masking Effect", Proc. inter-noise 2000, **1**(2000), 409

(Report received on Jan. 7, 2004)

Hiroyuki Hoshino
Year of birth : 1963
Division : ITS Lab. II
Research fields : Evaluation and analysis of car interior noise, Human machine auditory interface in car environment
Academic society : Soc. Autom. Eng. Jpn., Acoust. Soc. Jpn., Inst. Electron. Inf. Commun. Eng.