Voice Recognition Technology for Car-Mounted Devices

- Hitoshi Iwamida
- Toshitaka Yamato
- Kazuhiro Sakiyama
- Hideki Kitao
- Atsuyuki Takashima

Abstract

With car-mounted devices increasing in terms of their function and sophistication, voice recognition is attracting attention as a means to provide drivers a safe, easy-to-use interface with the devices in their vehicles.

In a bid to achieve sufficient voice recognition performance in various noise conditions, we have considered the issue in two separate research areas: voice-input and voice recognition processing.

This paper describes the way a multi-microphone system improves the SN quality of input signals, the way the rate of recognition is raised by identifying steady voice characteristics, and a voice recognition method that lessens the likelihood of noise-induced malfunction.

1. Introduction

The spread of highly functional and complex car-mounted devices such as navigation units and mobile phones is fueling concerns that road safety may deteriorate as drivers are distracted when they use or visually monitor such devices.

Accordingly, manufacturers are designing such devices so that only certain functions can be executed while the car is traveling, so that the driver does not have to use complicated functions. However, under such circumstances, it becomes impossible for the driver to obtain necessary information when required or to make full use of conventional devices as intended.

Human-machine interfacing based on voice recognition is attracting attention as a possible means of solving this problem.

Voice recognition is expected to allow drivers to minimize their eye and hand movement during operation of car-mounted devices, thereby permitting them to carry out with ease even those complicated operations that are normally prohibited during driving.

Accordingly, we have developed a noise filtering, voice recognition technology that can ensure high recognition performance even in noisy environments.

2. Purpose of Development

2.1 Technological Trends of Voice Recognition

The first voice recognition technology was applied to car-mounted devices to help speakers dependent carry out basic navigation operations or dial phone numbers on a hands-free car telephone simply by uttering a few words to activate a memory transmitter.

With existing technologies, it is now possible to set phone numbers and car navigation destinations classified by place names so that they can be called up by different users and with a vocabulary comprising thousands of words. To ensure that this technology can be used effectively, however, we need to achieve voice recognition performance that is adequate even in noisy environments. (Figure 1)

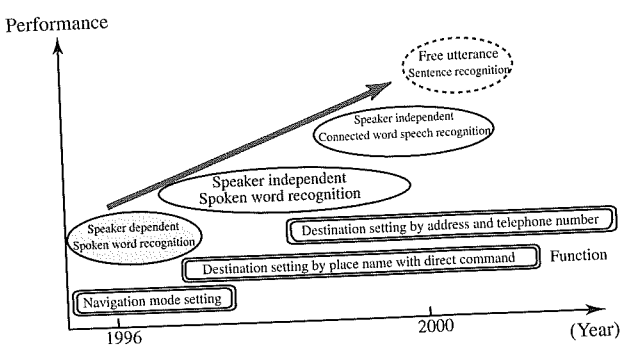


Figure 1 Trends in development of voice recognition technologies for vehicles

2.2 Fujitsu TEN's Approach

The vehicle cabin tends to be exposed to a number of noises. These noises include the sound of the vehicle traveling on the road, the sound of the engine running, and the sounds made by indicators, wipers, and other auxiliary devices (Figure 2). The loud noises generated when a vehicle is traveling at high speeds deteriorate the SN ratio of voice signal (S value) to noise signal (N value) and increase the voice recognition error rate. The voice recognition system may erroneously judge such noncontinuous noises as those generated by an indicator or wiper as being the sound of a voice, leading to recognition errors or other inconveniences.

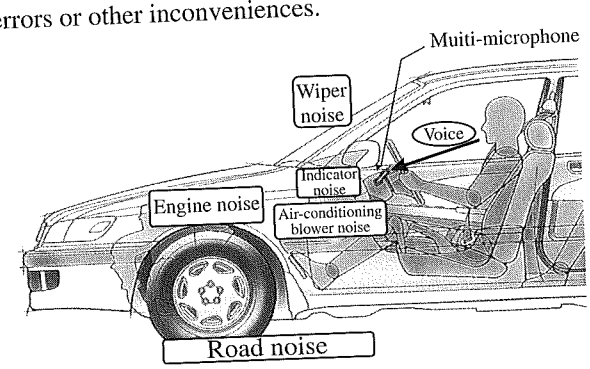


Figure 2 Noises in the vehicle cabin

We have developed a voice recognition system with a microphone of sharp directivity that can reduce the effects of noise on voice signals at the system's input section, thereby raising the voice recognition rate.

In a vehicle cabin, it may become necessary for the user to speak very loudly so as to be heard over the noise in the immediate environment. In such cases, the utterance made by the user will be different from the kind of utterance the user would normally make in a relatively quiet environment, such as an office environment. To boost the performance of the voice recognition system's basic processing section (the core of the voice recognition process), we developed an algorithm for an enhanced voice characteristic extraction function based on a large volume of voice data recorded in cars.

By analyzing the characteristics of device noises and voices, we also developed an algorithm to prevent indicator,

wiper, and other auxiliary device noises from causing voice recognition errors.

2.3 Outline of Voice Recognition

The voice recognition system has a voice input section at its first stage.

This section outputs in voice signal form a speaker's voice recorded through the microphone to the voice recognition processing section.

The voice recognition processing section roughly consists of an acoustic processing section and a word collation section. The acoustic processing section can further be divided into an analyzer that detects voice blocks in input signals and a phoneme collator that analyzes phonemes contained in the voice-block signals. The phoneme collator has phoneme dictionaries derived by statistically processing phonation patterns. (Figure 3)

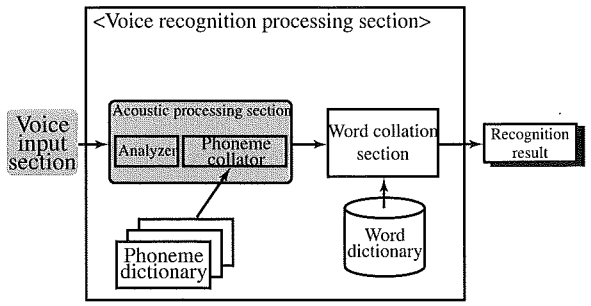


Figure 3 Outline of voice recognition processing

The word collation section compares and collates phoneme information output from the acoustic processing section with words registered in the word dictionary to extract the most probable words.

The word dictionary is written in text form for easy correction in accordance with the application.

3. Development of Noise Filtering Voice Recognition Technology

This chapter describes approaches for reducing noises at the voice input section to raise the recognition rate, for extracting voice characteristics in a stable manner at the acoustic processor, and preventing noises from causing malfunctioning.

3.1 Noise Reduction at the Voice Input Section

3.1.1 Purpose

To enhance the recognition performance of a car-mounted voice recognition system, it is essential to ensure that the SN ratio is high in noisy environments when road noises are generated while a vehicle is traveling on a road.

To improve the SN ratio, we set the input sensitivity to high for the voice incoming direction, and set it to low for

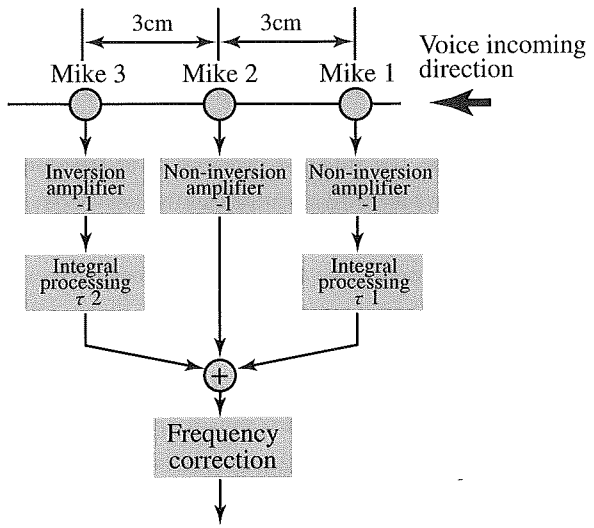


Figure 4 Configuration of multi-microphone system

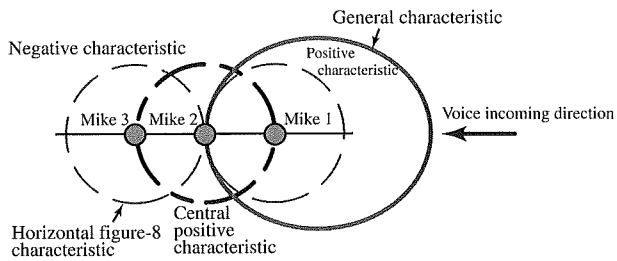


Figure 5 Simulation drawing of multi-microphone directivity pattern

the noise incoming direction by giving the voice input section sharp directivity.

We also proceeded with the development of a compact and low-cost microphone unit of the analog circuit type on the basis of the following prerequisites:

- Completing directivity control at the input section for voice recognition (microphone unit)
- Obtaining a flat frequency characteristic in a wide frequency band

3.1.2 Principle of the multi-microphone system

Figure 4 shows the configuration of the multi-microphone system and Figure 5 shows a simulated multi-microphone directivity pattern.

The desired directivity is realized through the following processing:

1) Arranging three non-directive microphones straight at regular intervals
2) Reversing the phase differences at the side microphones (Mike 1 and 3) and adding them by integral processing to generate a horizontal figure-8 characteristic with the middle microphone at the center
3) Adding the positive phase of the middle microphone (Mike 2) to generate directivity toward Mike 1

By using the low-frequency amplification characteristic of an integrator, the integral processing at the output sections of Mike 1 and Mike 3 improves the low-frequency characteristic where the horizontal 8-shaped characteristic is difficult to obtain. The frequency correction section compensates for the deterioration of the directivity in the 4 kHz band or a band of higher frequency by using the resonance characteristic of a low-pass filter.

This processing reduced the sensitivity of Mike 3 and strengthened that of Mike 1, thereby enabling directivity in the voice incoming direction.

3.1.3 Effects of directivity

Figure 6 shows a directivity pattern of the multi-microphone system that we have manufactured.

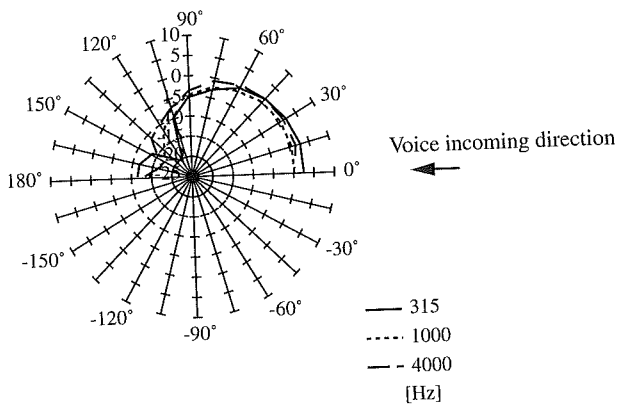


Figure 6 Directivity pattern of multi-microphone system

In the wide frequency band ranging from about 300 Hz to about 5 kHz, the multi-microphone system realized directivity with a sensitivity difference of 15 dB or more between the voice incoming direction (0 degree) and another direction (135 to 180 degrees).

Figure 7 shows an example of on-vehicle noise reduction with the multi-microphone system having the above sensitivity. The data compares the voice waveform for the word "Sapporo," uttered when the vehicle is traveling at 100 km/h, between the conventional single-microphone system and the multi-microphone system. From the noise waveforms before and after the voice section, we can see that the directivity of the multi-microphone system reduced the noise levels incoming from various directions while the vehicle is traveling.

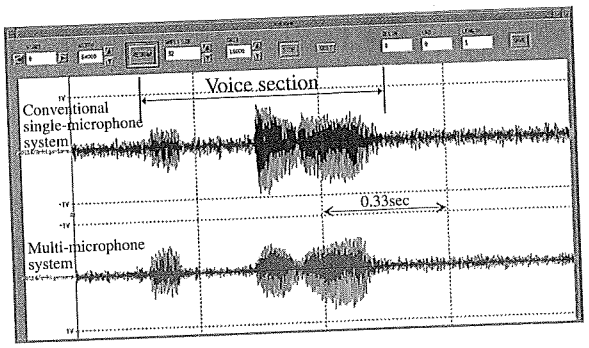


Figure 7 Comparison of waveforms during high-speed running

The optimum microphone mounting positions were also studied for the multi-microphone system. Compared with the conventional mounting at the center of the A pillar or on a sun visor, mounting on the steering column cover appears to be the optimum location for attaining a higher directivity effect.

The photo below shows the multi-microphone system and Table 1 lists its characteristics.

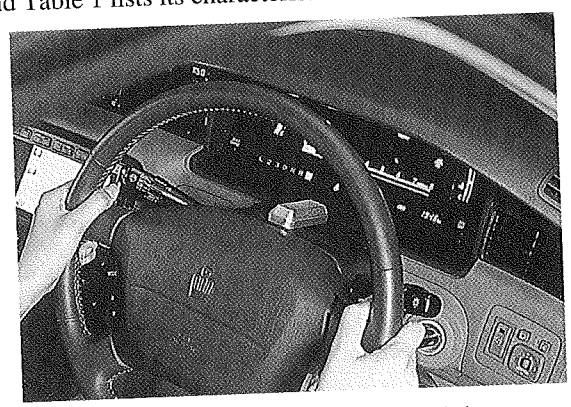


Table 1 Multi-microphone characteristics

Method	Analog processing using three-microphone array
Configuration	Compact unit with built-in directivity control circuit
Size	95 X 30 X 15 mm
Directivity (At 1 kHz)	-18 dB in the direction of 135 degrees -15 dB in the direction of 180 degrees Standard: 0 degree (voice incoming direction)
Frequency characteristic	300 Hz to 5 kHz (0±3dB in the direction of 0 degree)

3.2 Stable Voice Characteristic Extraction

A car-mounted voice recognition system should extract voice characteristics in a stable manner among various noises in a vehicle cabin. With the voice data of 62 people

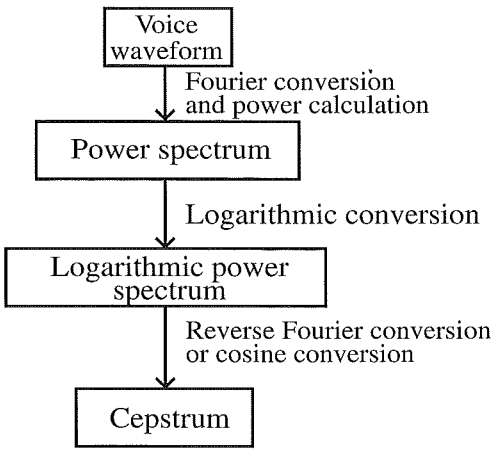


Figure 8 Relationship between voice waveform and characteristic extraction processing

collected during idling and high-speed running, we developed a voice recognition processing section as explained below.

3.2.1 Adopting Cepstrum

The conventional means of extracting voice characteristics is power spectrum. Power spectrum is produced from a voice waveform via Fourier conversion for development in a frequency area, after which it undergoes power calculation.

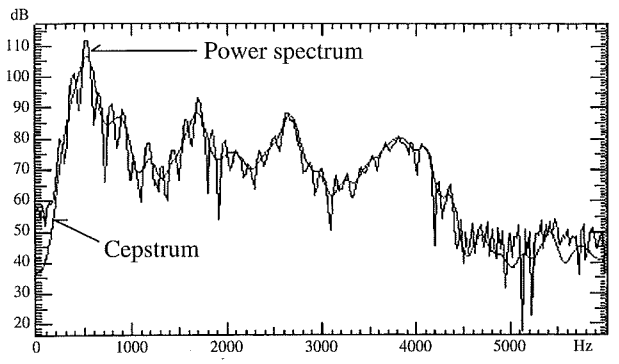


Figure 9 Comparison between power spectrum and spectral envelope of cepstrum

We adopted cepstrum produced by logarithmic processing and reverse Fourier conversion on a power spectrum. (Figure 8)

As Figure 9 shows, the greatest difference between power spectrum and cepstrum is in the fine structure. Power spectrum gives an adverse effect on pattern collation with a phoneme dictionary because the fine structure reflects a fine difference in voice input.

Since only a spectrum envelope (profile) can be extracted, the cepstrum allows comparatively stable phonemic collation even when the phonation is different or noises slightly change voice characteristics.

3.2.2 Extending the analytical window length

Voice characteristics are extracted from each frame of a fixed width at regular intervals. This fixed-width frame is called an analytical window. In general, extending the analytical window length raises the resolution in the frequency area but makes it difficult to observe transitional changes.

By experimental adjustments, we found that priority to the extraction of voice frequency characteristics would enhance the recognition performance. Therefore, the analytical window was set so that it would be longer than before.

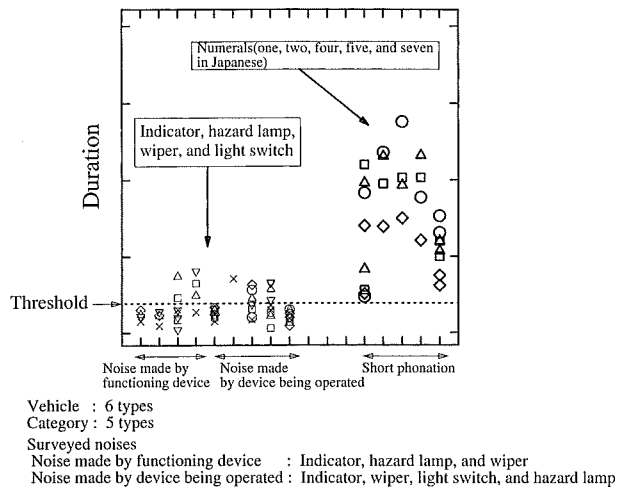


Figure 10 Duration of device noise and voice

3.2.3 Weighing on cepstrum

The voice characteristics in the cepstrum area show the tendency that voices are susceptible to being affected by noises in the low-frequency area and contain only a few effective components for voice recognition in the high-frequency area.

Therefore, we performed weighing on the cepstrum area to emphasize the middle-frequency area not susceptible to noises and containing effective components for voice recognition.

3.3 Preventing Noises from Causing Malfunction

As mentioned earlier, the noises that the vehicle cabin tends to be exposed to include the sound of the vehicle traveling on the road and the sounds made by indicators, wipers, and other auxiliary devices. These noises may

easily cause the voice detection section to malfunction. We therefore decided to carry out a survey focused on the duration of noises generated when devices are operated or devices are functioning and numeric phonation (one, two, four, five, and seven in Japanese) on six kinds of vehicles. (Figure 10)

The results of the survey revealed that noises from car-mounted devices and comparatively short phonation can be separated near the threshold shown in Figure 10.

Some words contain a soundless space for a stop or devocalization. Since the single sounds before and after this space tend to be shorter than regular sounds, it is risky to handle signals below the threshold as noises.

To solve this problem, we adopted a processing rule: "A short spontaneous signal is basically handled as a noise, but as a voice if a signal of the prescribed level or of a greater power is received within a specified period before or after the signal." In the preceding and succeeding periods, the threshold is set lower than the noise level of indicators and wipers to distinguish between noises and voices.

Some application software may not deal with single-sound words at all. If users utter single-sound words comparatively slowly, the noise and signal width can be further extended to realize a system that is more capable of filtering out noises.

4. Summary

By applying the three noise filtering approaches reported in the previous chapter, we succeeded in improving voice recognition performance and reducing malfunctioning as summarized here.

We combined a multi-microphone system capable of reducing noises at the voice input section with an acoustic processor for extracting voice characteristics in a stable manner. This combined system enabled the available vocabulary of 1,000 words uttered by random speakers and achieved 90% or higher recognition rate when a vehicle is traveling at 100 km/h. We also reduced the malfunction rate to about 10% by developing an acoustic processor that prevents noises from causing malfunctioning.

Compared with the products of our competitors that are already out on the market, the voice recognition system that we have developed is performing better. We plan to put a practical model on the market in the future.

5. Conclusion

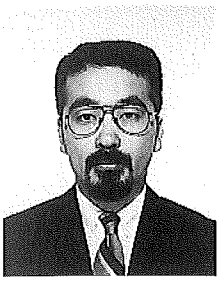
This paper has introduced Fujitsu TEN's approaches to noise filtering performance that poses the greatest problem when voice recognition is applied to vehicle use.

Voice recognition represents a prospective means of enabling users to operate with ease car-mounted devices that are becoming more functional and complicated. However, truly good operability cannot be realized by simply replacing manual switches with a voice recognition system.

In the field of personal computers, we have seen the debut of oral input software for word processing as well as the increasing activity involving the development of other connected word speech recognition techniques.

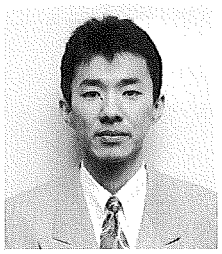
While enhancing noise filtering performance, we will keep developing connected word speech recognition technology. By realizing an efficient voice input system for different ways of verbal communication, we aim to construct a new type of human-machine interface so that a user can operate car-mounted devices comfortably and without any accompanying stress.

Authors



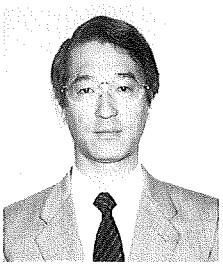
Hitoshi Iwamida

Joined Fujitsu Laboratories Ltd. in 1983. Engaged in voice recognition studies since then. Transferred to the Audiovisual System Laboratory of ATR from 1988 to 1991.



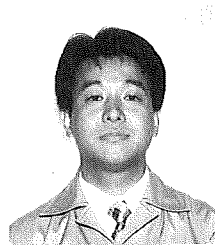
Kazuhiro Sakiyama

Employed by Fujitsu TEN since 1991. Engaged in developing digital signal processing systems. Currently assigned to Research & Development Department.



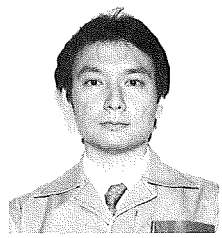
Atsuyuki Takashima

Employed by Fujitsu TEN since 1984. Engaged in developing acoustic and digital signal processing systems. Currently assigned as Assistant Department General Manager of Research & Development Department.



Toshitaka Yamato

Employed by Fujitsu TEN since 1985. Engaged in developing acoustic technology, then digital signal processing technology. Currently assigned to Research & Development Department.



Hideki Kitao

Employed by Fujitsu TEN since 1992. Engaged in developing digital signal processing technology, followed by development of voice processing technology. Currently assigned to Research & Development Department.