

In-vehicle Super-directive Microphone System for Speech Recognition Using Nonlinear Signal Processing with Two Microphones

Toshiya SHIKANO*

Junichi WATANABE**

Naoshi MATSUO***

ABSTRACT

A super-directive microphone system using digital signal processing has been developed for in-vehicle speech recognition systems for the North American market, in order to increase the recognition accuracy of these systems in high noise environments.

The developed microphone system uses a new super-directive algorithm with a high directional noise reduction rate that functions to reduce noise, such as passenger utterances. To enable adaptation of the algorithm to different in-vehicle noise environments, such as while operating the air conditioner and opening/closing the windows, the system also incorporates an omni-directional microphone and an analog high pass filter. The use of the new system increases recognition accuracy in a variety of noise environments.

1. Introduction

Honda has been gradually using North American navigation systems with speech recognition since the 2003 model year ACCORD. Questionnaire surveys conducted by Honda indicate that customers use speech recognition systems frequently and give them high marks for ease of operation. On the other hand, although tests conducted by Honda showed that these systems yield adequate recognition accuracy under usual road noise conditions, many requests have also been received indicating a preference for the further enhancement of recognition accuracy as sensed by the user. The way such systems have actually been used also indicated a need for modifications. Among the factors interfering with recognition, therefore, attention was focused on noises assumed to be particularly loud in the cabin, and thus the super-directive microphone using digital signal processing was developed for in-vehicle speech recognitions.

* Automobile R&D Center

** FUJITSU LIMITED

*** FUJITSU LABORATORIES LTD.

2. In-vehicle Noise and Its Countermeasures

2.1. In-vehicle Noise

The sources of cabin noise include the following:

- Road noise
- Wind noise that occurs when windows or sunroof are open
- Air conditioner noise and noise from air conditioner blower air hitting the microphone
- Sound of the person in the front passenger seat or other occupants talking

Classified by their physical properties, road noise, wind noise from open windows or sunroof, air conditioning noise, and the like, belong to the category of stationary noise. The sound of the person in the front passenger seat or other occupants talking belongs to the category of non-stationary (sporadic) noise. The sound of air conditioner blower air hitting the microphone is the noise of a clipping signal resulting when wind pressure

makes the diaphragm stick to the wall. This is a structural characteristic of the directional microphone. To summarize, these noises can be categorized as:

- Stationary noise
- Non-stationary noise
- Clipping signal that occurs under wind pressure

Microphone systems that would be effective against these noises were studied.

2.2. Non-stationary Noise and Super-directive Digital Signal Processing

A typical non-stationary noise in the cabin is the sound of passengers talking. Occasions when multiple passengers are in the car could include carrying children to and from school, going shopping at a mall, going to a restaurant for a meal, going on a family trip, and the like. On occasions like these, car navigation systems and other information equipment experience a higher frequency of operation.

Another factor is that the computations in speech recognition processing are performed on the basis of voice characteristics. Therefore, the voices of passengers have characteristics that seriously affect recognition when there is noise. A suppression ratio of 20–30 dB is needed for passenger voices in order to counter interference with recognition of the driver's voice. Since a signal-to-noise ratio of slightly less than 10 dB is the limit for recognition in white noise or similar noise, this is a challenging issue.

One method for suppressing non-stationary noise is to

use multiple microphones and distinguish the needed sound from the incoming direction of the sound source. This includes delay and sum beam forming, adaptive beam forming, and other such methods of directional suppression⁽¹⁾. Another approach is to distinguish the needed sound using independent component analysis, which assumes independent output from multiple noise sources⁽²⁾. With directional suppression, however, a high suppression ratio often cannot be achieved unless there is a sufficient number of microphones. The independent component analysis imposes a heavy computational load that makes it impractical for on-board applications. Therefore, a new super-directive digital signal processing method was selected⁽³⁾. This yields the targeted high noise suppression ratio for using two microphones.

Figure 1 shows the principle of the new super-directive algorithm. As indicated in Fig. 1 (a), when the sound waves of the driver's voice reach the two microphones, there will be a certain time difference between them according to the directions they come from. When this time difference is subjected to frequency axis conversion in a graph, the driver's utterance will be plotted as a nearly straight line sloping upward to the right, as shown in Fig. 1 (b), and it will be distinctly separate from the utterance coming from the passenger seat.

The effect of the new super-directive algorithm is shown in Fig. 2. Figure 2 (a) shows the sound pressure level of passenger utterances in a time series where the driver and passenger utterances occur alternately. When this is processed using the new super-directive algorithm, there is powerful selective suppression of just the

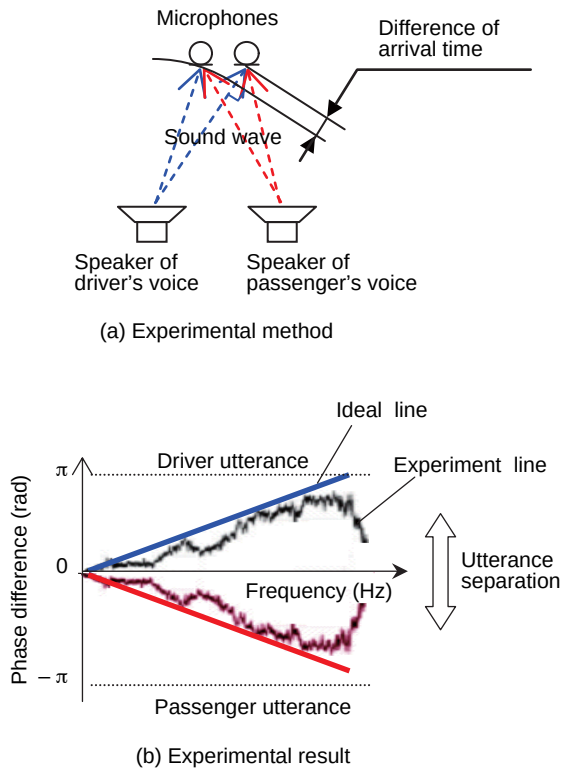


Fig. 1 Principle of channel separation between driver utterance and passenger utterance

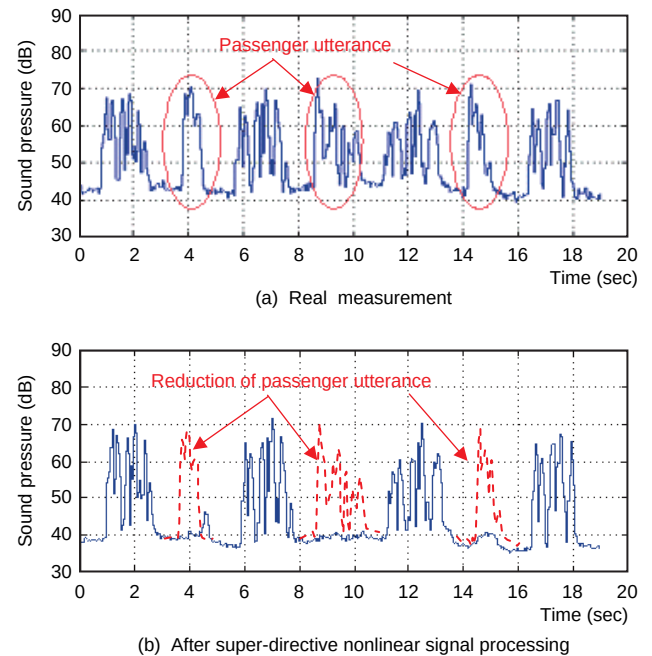


Fig. 2 Effect of channel separation

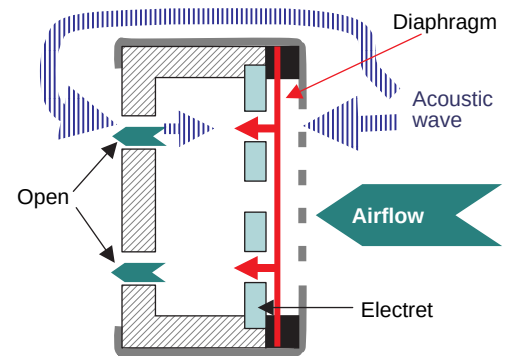
passenger utterance, as shown in Fig. 2 (b). The system is capable of picking out just the driver's utterance. Although done by directional suppression, this enables utterances by the driver to be distinguished from the crowded environment as though by the cocktail party effect in human hearing⁽⁴⁾.

2.3. Clipping Signals from Wind Pressure and the Omni-directional Microphone Element

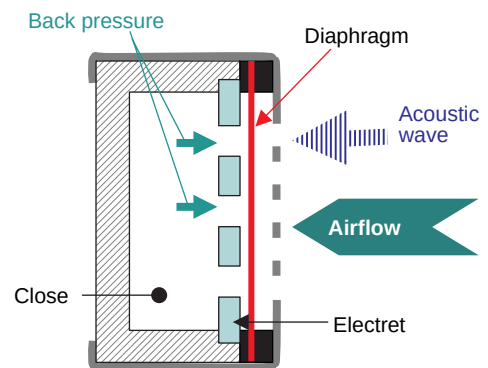
Directional microphone elements are ordinarily used as the microphone elements for on-board speech recognition and for hands-free telephones in order to selectively pick up the driver's voice. As shown in Fig. 3 (a), however, when the wind from an air conditioner is directed toward the microphone, the microphone will output a clipping signal with the voice signal in the case of a directional microphone element^{(5), (6)}. As shown in Fig. 4 (a), this happens because of the microphone structure, which is designed for directionality. The sound pressure against the front surface of the diaphragm and the sound pressure that leaks in from behind, against the back surface of the diaphragm, interfere with each other. In other words, the diaphragm in a directional microphone element is free of back pressure, and the diaphragm is readily displaced by wind pressure from in front, and by vibration or other factors apart from sound pressure that impinge on the microphone element. As a result, the diaphragm sticks to the front or back wall, generating a clipping noise. In contrast to this, as shown in Fig. 4 (b), is the omni-directional microphone⁽⁷⁾. This has back pressure to limit large displacements of the diaphragm, so that wind pressure and other such factors will not readily make the diaphragm stick to the wall. Therefore, as shown in Fig. 3 (b), no clipping signal is output.

The microphone in the cabin is attached at the front part of the roof, as shown in Fig. 5. This layout places

the microphone where the airflow from the dashboard vent will blow on it. In this situation, an omni-directional microphone of the closed cavity type that is not influenced by wind pressure (strictly speaking, a microphone with vents to bleed off pressure differentials so as to cancel out atmospheric pressure) is very resistant to wind pressure. Digital signal processing using two microphone elements provides directionality toward the driver and provides sound quality as good as or better than a directional microphone element. This made it possible to use omni-directional microphones, which perform well under wind pressure.



(a) Directional microphone



(b) Omni-directional microphone

Fig. 4 Comparison of microphone structures

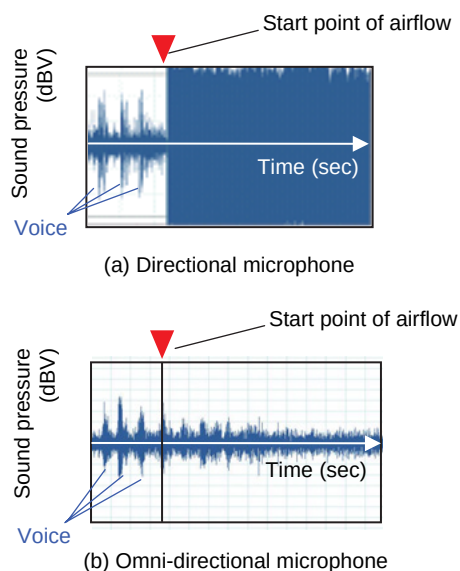


Fig. 3 Comparison of acoustic waves



Fig. 5 Direct airflow from air conditioner

2.4. Stationary Noise and Analog High Pass Filters

Various methods have been devised for signal processing to suppress stationary noise, such as spectral subtraction. When the noise reduction rate is raised, however, the processed result tends to have greater sound distortion relative to the original sound. When the sound distortion increases, unmatching occurs with the acoustic model that is designed for phoneme selection in the speech recognition computation described in a later section. This involves an issue of diminished recognition accuracy.

The omni-directional microphone with the frequency response characteristics shown in Fig. 6 (a) cannot attenuate the distinctive stationary noise in the cabin, which is concentrated at the lower frequency range. In this it differs from the directional microphone element with the low frequency attenuation response shown in Fig. 6 (b).

Therefore, a method of suppressing stationary noise was sought that would not necessitate the strong suppression performed by digital signal processing while also addressing the issues of the omni-directional microphone. The method adopted was to add an analog high pass filter with low frequency attenuation response that has an even steeper slope than the directional microphone element, as shown in Fig. 6 (c), and place it

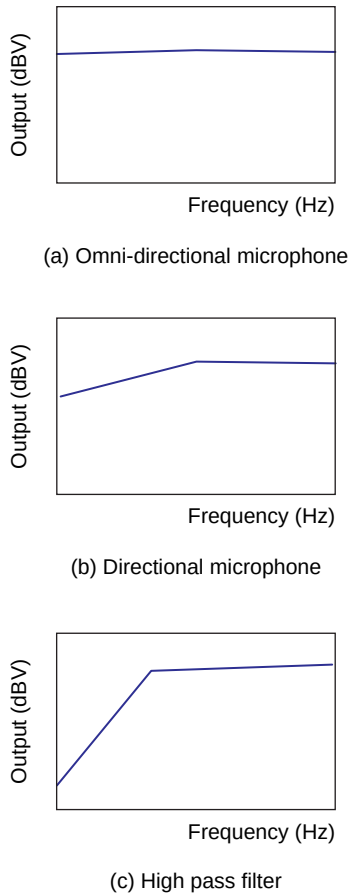


Fig. 6 Frequency response of microphone and high pass filter

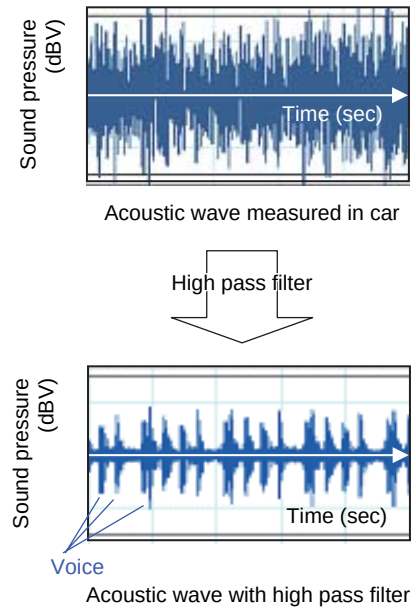


Fig. 7 Effect of high pass filter with omni-directional microphone

before the analog-digital conversion process. The crossover frequency of the high pass filter was set at or below the frequency needed by the speech recognition application in use.

Filtering is done at several tens of frequency bands during speech recognition computation. Since this is done before conversion from analog to digital, however, the information entropy (self-information) of the digital signal can be increased and the quality of the signal after processing enhanced. The effect is shown in Fig. 7. The background noise component at low frequencies is suppressed and the generated signal has a better signal-to-noise ratio.

3. In-vehicle Noise Suppression System

Sections 2.2 to 2.4 raise various issues with the in-vehicle noise environment. In order to address these issues, a method was developed to utilize the new super-directive algorithm in combination with the omni-directional microphone and the analog high pass filter to provide higher background noise suppression performance.

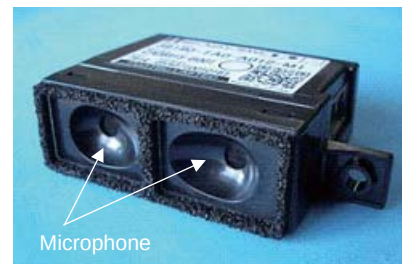


Fig. 8 Digital signal processing microphone unit

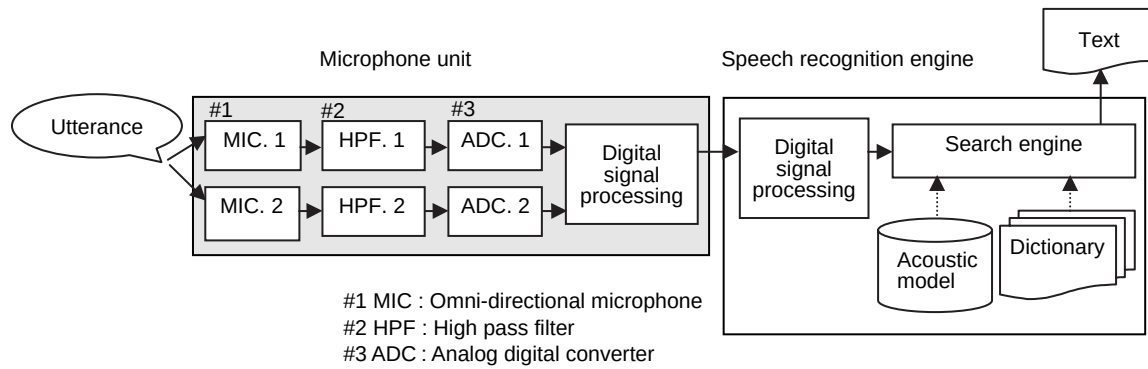


Fig. 9 Block diagram of voice recognition system

A summary view of the developed background noise suppression microphone system is shown in Fig. 8.

A flow chart of the process is shown in Fig. 9. Driver utterances are converted into electrical signals by the two microphone elements. The two signals go through their high pass filters and then undergo analog-to-digital conversion. These two signals are input to digital signal processing and subjected to noise suppression processing using the super-directive algorithm. The above describes the processing performed within the microphone unit. The flow calls for the signals output by this unit to be handed over to the "Speech recognition engine" in the car navigation unit. This enables current car navigation units to be incorporated into the system and used without modification.

4. Results

4.1. Non-stationary Noise Suppression Effect

Tests were performed while driving in an actual vehicle to compare speech recognition using the new and old microphones with utterances by the driver alone, and with simultaneous utterances by the driver and passenger. The sound pressure of the passenger utterances was equal to that of the driver utterances. Also, the periods of passenger utterance were made to overlap with and last longer than the driver utterances.

The results are shown in Fig. 10. The system provided a high level of recognition accuracy even when there were passenger utterances.

4.2. Air Conditioner Airflow Clipping Signal Suppression Effect

Tests were conducted to compare speech recognition using the new and old microphones while idling with the air conditioner turned off, with the fan turned to the highest setting and the airflow directed entirely to about shoulder level, and likewise with the airflow directed overhead (toward the microphones). Figure 11 shows the results.

The use of closed-back-type omni-directional microphones yielded high levels of recognition accuracy even when the air conditioner airflow was directly hitting the microphone elements. Having the air conditioner fan

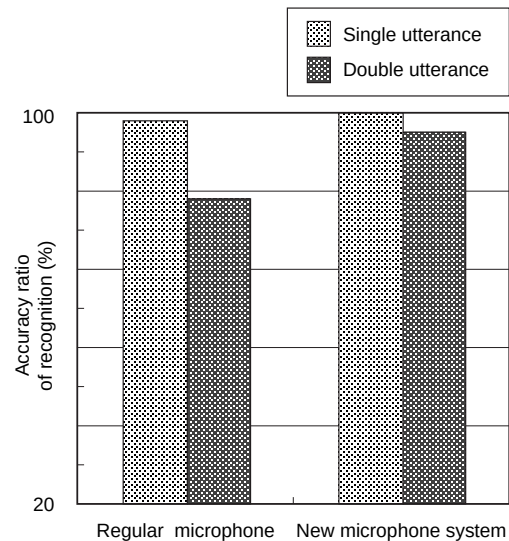


Fig. 10 Recognition accuracy for single utterance and double utterance

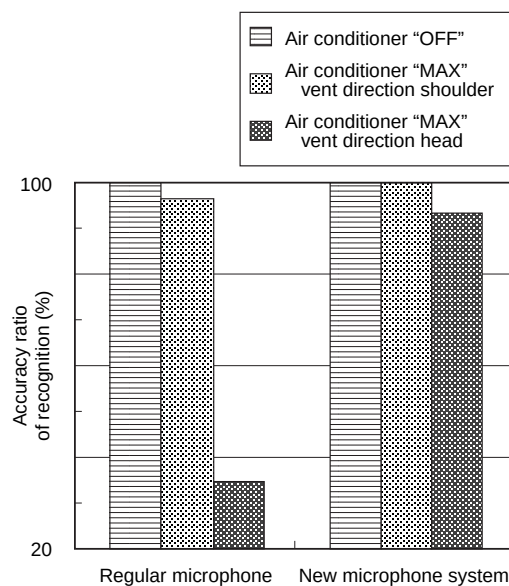


Fig. 11 Recognition accuracy with airflow from air conditioner

turned to the highest setting generated a large amount of noise in itself. This happens because the sound of turbulent flow from the air conditioner blower vent generates a resonance sound determined by the cabin volume. As will be described later, this is a type of wind-throb phenomenon⁽⁸⁾. The fact that the system provided a high level of recognition accuracy, even though the noise itself was very loud, indicates that the analog high pass filter described in section 2.4 was operating effectively, as well.

4.3. Stationary Noise Suppression Effect

Speech recognition tests were conducted with windows closed and with windows open. When cruising with open windows, the cabin noise reached 70 dBA or more. Utterances with ordinary sound pressure are covered over by the noise at that level, and a usable signal-to-noise ratio cannot be obtained. Therefore, this testing was conducted with utterances at a higher sound pressure in order to obtain a signal-to-noise ratio of several decibels.

The noise was very loud and conversation was complex when the windows were open, but, as shown in Fig. 12, the system provided a level of recognition accuracy more or less equal to when the windows were closed. The wind noise that occurs when windows or sunroof are open is a sound produced by a wind-throb phenomenon. It is generated by the sound of turbulent flow coupled with the Helmholtz resonance with the cabin volume. Consequently the central frequencies of the sound are constant regardless of the vehicle speed, and the on-board microphones are influenced at approximately 300 Hz or lower, including the harmonic overtones of the resonance frequencies. The noise resulting from this resonance sound has been suppressed by steep analog high pass filters, making it less likely that recognition accuracy would be decreased.

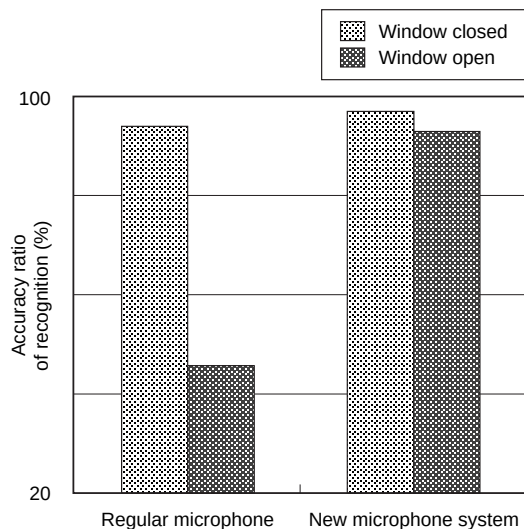


Fig. 12 Recognition accuracy with window closed and window open

5. Conclusion

In order to enhance the speech recognition accuracy of the system under the conditions of actual vehicle use, a microphone system was developed that combines a new super-directive algorithm with a high directional noise suppression ratio, omni-directional microphones, and analog high pass filters. These elements in combination complement each other for enhanced system feasibility and produce high levels of speech recognition accuracy, not only with ordinary road noise, but also with the wind throbb phenomenon that accompanies wind noise when windows or sunroof are open, with air conditioner noise, with air conditioner airflow hitting the microphones directly, and even with the sound of the person in the front passenger seat or other occupants talking.

References

- (1) Kaneda, Y.: Directivity Control using Microphone Array, The Journal of the Acoustical Society of Japan, Acoustical Science and Technology, Vol. 51, No. 5, p. 390-394 (1995)
- (2) Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis, Tokyo Denki University Press, 532 p. (2005)
- (3) Furui, S.: Onkyo Onsei Kougakaku, Kindaikagakusya, p. 28 (1992) (in Japanese)
- (4) JP, 19-318528, A (2007)
- (5) Nakamura, J.: Tanitsushikousei Kondensagata Maikurohon, Housougijutsusya notameno Maikurohon Kouza, Housougijutsu, Vol. 57, No. 5, p. 464-468 (1982) (in Japanese)
- (6) Nakamura, J.: Elektoretto, Housougijutsusya notameno Maikurohon Kouza, Housougijutsu, Vol. 57, No. 11, p. 1101-1108 (1982) (in Japanese)
- (7) Nakamura, J.: Mushikousei Kondensagata Maikurohon, Housougijutsusya notameno Maikurohon Kouza, Housougijutsu, Vol. 57, No. 4, p. 364-372 (1982) (in Japanese)
- (8) Sumitani, K., Maeda, K., Ichinose, K.: Development of Vehicle Fluid Dynamics : Flow around the Vehicle and Aerodynamic Characteristic, Nagare, Vol. 23, No. 6, p. 445-454 (2004)

■ Author ■



Toshiya SHIKANO



Junichi WATANABE



Naoshi MATSUO