Review

Speech-Based Interfaces in Vehicles

Toshihiro Wakita

Abstract

Recently, speech-based interfaces in vehicles have become a popular means of improving the accessibility of in-vehicle information equipment. In this edition, we present the results of our research into the "noise-robustness of speech recognition" and "driver distraction."

Keywords Speech recognition, Noise, Signal processing, Workload, Human factor

1. Introduction

Recently, speech-based interfaces in vehicles have become a popular means of improving the accessibility of in-vehicle information equipment. Although the driver must use his or her eyes, hands, and feet for driving, their speech and hearing are free to be used for other purposes. Speech-based interfaces are expected to become the primary means of interacting with in-vehicle information equipment in that everyone can use them easily and safely. Therefore, for researchers involved with speech-based interfaces, those interfaces used in vehicles are a very important research target. Over the last few years, sessions related to these speech-based vehicle interfaces have been held both as part of both auto- and speech-related conferences ("Speech Recognition in Vehicles" at the International Conference on Spoken Language Processing 2002,[1] "Speech Interface" at the Annual Congress & Exposition by the Society of Automotive Engineers of Japan). A large speech-related corpus has evolved in many countries,[2, 3] and standard data for speech recognition rate benchmarking has been defined.[4]

The development of speech-based interfaces for use in vehicles presents many technical problems. In addition to the common problems related to speech-based technology, vehicle applications incur two special challenges. The first is ensuring that the speech recognition is robust to vehicle noise, and the second is the problem of the driver being distracted by the use of the speech interface while driving. In this article, we present the results of our research into these two topics.

2. Speech-recognition in noisy vehicle environments

Because a moving vehicle's passenger compartment is subject to many types of noise, such as engine noise, wind noise, and road noise, the speech recognition rate tends to be poorer than in a quiet environment. In addition, the in-vehicle noise varies depending on the type of the car and the driving conditions, causing the degree of degradation to vary accordingly.[5] Thus, to improve the robustness of speech recognition, three techniques are used:

(1) Speech enhancement using multiple microphones

One method involves enhancing the driver's speech by using multiple, evenly spaced microphones combined with spatial signal processing. Typical algorithms include the delayed sum microphone array and the adaptive microphone array.[6] Another method is independent component analysis. This involves separating multiple signals based on the assumption that these signals are statistically independent.[7]

(2) Noise suppression using signal processing

To suppress noise, a spectrum subtraction method is usually used. This involves subtracting the estimated noise spectrum from the combined speech and noise signals.[8] Many variations of this method have been proposed.

(3) Pattern recognition with noise model

An acoustic model for speech recognition is usually developed using speech combined with vehicle noise. To improve the noise-robustness, a method using multiple acoustic models, each for different noise environments, has been proposed. Other methods include the synthesis of a noise acoustic model and a speech acoustic model.[9]

Using these techniques, speech recognition rates under steady driving conditions have been improved considerably but are still not perfect. In addition, there are now demands for noise-robustness capable of handling unsteady noise (ex. engine car noise during acceleration, passengers' speech, noise from an uneven road).

3. Safety evaluation of speech-based interfaces used in vehicles

Speech-based interfaces are expected to become the primary means of interacting with in-vehicle information equipment, given that everyone can use such interfaces safely. But, many accidents have occurred as a result of drivers using cellular phones while driving.[10] Therefore, we must clarify the types of speech-based interface that adversely affect driving performance. Recently, design guidelines for a vehicle's visual interfaces have been proposed.[11-13] For speech-based interfaces, we discuss a method for measuring the degree of distraction. A typical method is described below.

(1) Subjective evaluation:

This method involves the driver subjectively evaluating his or her workload when using a speech-based interface. NASA-TLX is a well-known example of this approach.[14]

(2) Physiological measurement:

This method involves measuring the driver's physiological characteristics while he or she is using a speech-based interface. Usually, the driver's heartbeat, amount of perspiration, and degree of eye movement are used. Recently, fMRI (functional Magnetic Resonance Imaging) has also been used to measure the amount of brain activity.

(3) Performance evaluation:

This method involves measuring driving performance while using a speech-based interface. Lane changing,[15] distance from the vehicle in front, and the time taken to react to changes in the road environment[16] have been proposed as performance factors to be measured.

In the future, design guidelines for speech-based interfaces for vehicles will be standardized based on the results of these measurement methods. There is currently a demand for a novel interface that will reduce the amount of driver distraction.

4. Contents of this special issue

In this special issue, we present four papers on "speech interfaces in vehicles". Two of the papers examine speech recognition while the other two consider interface design.

In the first paper, "Noise-robust Speech Recognition in a Car Environment Based on the Acoustic Features of Car Interior Noise", the improvement of speech recognition in the face of vehicle noise is discussed. The second paper, "Estimating Speech-Recognizer Performance Based on Log-Likelihood Difference Distribution of Word-Pairs", concerns recognition performance evaluation as it relates to product quality assurance. The third paper, "Voice Information System that Adapts to Driver's Mental Workload", is a proposal for a novel interface that adapts itself to the driver's workload. In the final paper, "Evaluating the Safety of Verbal Interface Use while Driving", a method of measuring the degree of driver distraction is discussed.

5. Conclusion

Speech-based interfaces present us with many technical problems. Nevertheless, a speech-based interface could be the ideal way of interacting with our vehicles, as it uses a technique that is very familiar to, and comfortable for, humans. Here at the automobile research laboratory, we will continue our research into speech-based interfaces.

References

1) Proc. Int. Conf. Spoken Lang. Process. (ICSLP 2002), (2002)
2) Kawaguchi, K., et al. : "Construction of Speech Corpus in Moving Car Environment", ICSLP 2000, (2000)
3) Hansen, J. H. L., et al. : "CU-Move : Analysis & Corpus Development for Interactive In-Vehicle Speech Systems," Eurospeech-2001, (2001), 2023
4) Hirsh, H. G. and Pearce, D. : "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions", ISCA ITRW ASR2000, (2000)
5) Hoshino, H., Terashima, R., Shimizu, T. and Wakita, T. : "Noise-Robust Speech Recognition in a Car Environment Based on the Acoustic Features of Car Interior Noise", Proc. Int. Workshop on Hands-Free Speech Commun., (2001), 151, ATR
6) Griffiths, L. J. and Jim, C. W. : "An Alternative Approach to Linearly Constrained Adaptive Beamforming", IEEE Trans. Antennas and Propagation, **30**(1982), 27
7) Asano, F., et al. : "Blind Source Separation in Reflective Sound Fields", Proc. of Int. Conf. on Acoust., Speech, and Signal Processing 2001, (2001), IEEE
8) Boll, S. F. : "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans. Acoust. Speech Signal Process., **27**(1979), 113
9) Gales, M. J. F. and Young, S. J. : "Robust Continuous Speech Recognition Using Parallel Model Combination", IEEE Trans. Speech Audio Process, **4**(1996), 352
10) Japan Safe Driving Center: Technical Report (online), available from <http://www.jsdc.or.jp/kenkyu/kenkyu. htm>, (accessed 2003-12-26)
11) Wakita, T., et al. : "Visual Behavior Model for Navigation System Operation while Driving", Proc. of 6th World Congr. ITS, (1999)
12) Ito, T., et al. : "Japan's Safety Guideline on In-Vehicle Display Systems", Proc. of 4th World Congr. ITS, (1997)
13) Zwahlen, H. T., et al. : "Safety Aspects of CRT Touch Panel Controls in Automobiles", Vision in Vehicles II, Ed. by Gale A. G., et al., (1988), 335, North Holland Press
14) Hart, S. G. and Staveland, L. E. : Hum. Ment. Workload, Ed. by Hancock, P. A., Meshkati, N., (1988), 139, Elsevier Sci. Publ. B. V., Amst.
15) Kimura, K., et al. : "In Vehicle Navigation System Operability while Driving", Proc. of 6th World Congr. ITS, (1999)
16) Kojima, S., et al. : "Evaluating the Safety of Verbal Interface Use while Driving", SAE Tech. Pap. Ser. No. 2001-01-0356(2001)
17) Uchiyama, Y., et al. : "Voice Information System that Adapts to Driver's Mental Workload", Proc. of the Hum. Factors and Ergon. Soc. 46th Annual Meet., (2002), 1871-1874, Hum. Factors and Ergon. Soc.

(Report received on Jan. 9, 2004)

Toshihiro Wakita
Year of birth : 1960
Division : ITS Lab. II
Research fields : Human interface in vehicles
Academic society : Acoust. Soc. Jpn.