INTRODUCTION OF SPEECH INTERFACE FOR MOBILE INFORMATIN SERVICES

Hirotaka Nakano

NTT DoCoMo Multimedia Laboratories
3-5 Hikarinooka, Yokosuka, Kanagawa, 239-8536 Japan
E-mail:nakano@mml.yrp.nttdocomo.co.jp

1. MOBILE INFORMATION SERVICE IN JAPAN

Popular Japanese mobile web-phones (for example, i-mode) are being widely used to connect to Internet providers (IP). The number of mobile subscribers reached 65,355,700 in September 2001, and 68.8% of them use mobile web-phones (44,936,800) [1].

The most popular service on mobile web-phone is e-mail. Currently, users type the messages using the ten standard keys on the phone. Several letters and Kana (Japanese phonetic characters) are assigned to each key, and the user steps through them by tapping the key repeatedly. After inputting several words, the user converts them into Kanji (Chinese character). Kana-Kanji conversion is still improving, and recently fast text input methods have been introduced, but these key input methods are still troublesome.

2. PROBLEMS OF ASR IN MOBILE ENVIRONMENT

A speech interface is expected to overcome this input difficulty. However, speech interfaces suffer several problems. The problems are both technical and social.

2.1. Technical problems

Since speech is a natural interface, users expect high communication performance. However, the recognition rates of current speech interfaces are insufficient to the point that they irritate the user. If only a speech interface is used, its poor browsing and editing performance are seen as problems. Current speech dialog systems are too rigid and formal; the service takes an initiative and forces the user to expend too much attention to follow it.

2.2. Social problems

Speech interfaces are effective only when used in a quite private area. In such areas, the voice commands are well captured by the system. However, most mobiles are used in quite noisy environments such as streets, stations, and on a train. In the densely populated cities of Japan, we hesitate to make noise in a public space and to disclose private information. Cellular phone use is prohibited on some trains, in theaters and other public spaces so as not to annoy other people.

3. SOLUTIONS

To use a speech interface in mobile information services, we hope to research and develop techniques that solve these problems. The goal is to make “my” cellular phone a really friendly tool. Three advances are needed now.

3.1. Enhanced recognition performance in noisy environments

Stable recognition against various background noises will allow speech interfaces to be used more widely.

3.2. Improved handling performance

The speech interface must cooperate with other interfaces, for example, after filling in a form with ASR, if the user wants to browse the form, it should be shown on a display. User condition (personal and environmental) changes frequently, so it is also important to switch to the best interface, even when accessing an information service.

3.3. Dialogue assistant processor

This processor is speaker adaptive; it compensates the current background noise to unify the ASR input. It assists the user in creating material. For example, the user can send a formal e-mail response simply by inputting the essential points into the processor in a casual style. This personal support should support other information services; since it will hold personal information, the processor should be placed on the user side, not a central system [2].

4. CONCLUTION

We should continue to investigate applicability of ASR. With solving these problems, ASR will be able to become one of the most powerful tool for text input.

[1] http://www.tca.or.jp/eng/daisu/index.html

[2] L. Comerford, D. Frank, P. Gopalakrishnan, R. Gopinath, J. Sedivy, The IBM Personal Speech Assistant. In *Proceedings of ICASSP-2001*, (May 2001), 1-4