Japanese Text-to-Speech system with new Auditory User Interface for Microsoft(R) Windows(R)

Takayuki Watanabe[1] and Tuneyoshi Kamae

Department of Physics, School of Science, University of Tokyo

Toru Kurihara

Department of Information Sciences (for the blind), Tsukuba College of Technology

1 Introduction

Computers became friendly with the advent of Graphical User Interface (GUI); visually impaired users, however, can not benefit from GUI. Traditional screen-readers are ineffective for complicated GUI because they do not care about the contextual structure of the visual output but speak the displayed characters as a plain text.

In 1994, Dr. Raman built up a new system, AsTER (Audio System for TEchnical Readings)[1]. AsTER is a computing system that renders the original content of an application into auditory space with various audio-formatting styles. Raman extended this approach to more general computing tasks and released a new speech interface, Emacspeak[2, 3], as an Emacs subsystem. In Emacspeak, user interfaces are separated from the contents and they arrange the contents so that the information is effectively presented for the sense of hearing. In other words, Emacspeak acts as an effective Auditory User Interface (AUI). These systems, however, can not treat Japanese and can run only on Emacs.

Thus, a new auditory rendering system based on Raman's work but can treat Japanese applications on Microsoft(R) Windows(R) is strongly required by Japanese visually impaired users.

2 Development of a new system

2.1 Voice-Meadow

We are developing a new multilingual Text-to-Speech (TTS) system with a 3-D AUI, Voice-Meadow (Fig. 1). Voice-Meadow is a natural

[1] Contact to: Takayuki Watanabe
address: 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033
phone: (03)3812-2111 ext.7622
Fax: (03)5802-4362
email: watanabe@phys.s.u-tokyo.ac.jp

extension of Emacspeak to Microsoft Windows. Meadow[4] is a multilingual Emacs for 32-bit Windows and is based on Emacs 20.

Voice-Meadow does not use speech recognition as an input method because a computer expert can use a keyboard regardless of one's visual ability and a keyboard is the most efficient input method for an expert.

The way of rendering the contextual information of the contents into auditory output is taken after that of Emacspeak. So, for example, Voice-Meadow offers an effective AUI for browsing and searching a calendar. It also can place outputs of Meadow's three frames (windows) in three different auditory-space, which enables the user to switch from one frame to another frame immediately and to be able to handle three different tasks. To cite another example, the text under edited can be assigned to a male voice with a bold voice for bold fonts, while other information such as the current line number can be assigned to a female voice.

Additional functions of Voice-Meadow are treatment of multilingual inputs and multilingual audio outputs. Since Japanese input is treated by IME independent of the application, auditory rendering of IME is treated by a dedicated separate application, Voice-IME (Fig. 1). Audio formatting is used to notice the status of IME to a user, i.e. whether he is inputting Japanese characters or not.

TTS server, a speech server of the current system as shown in Fig. 1, uses DirectSound(R) as a multimedia device and is capable of handling simultaneous inputs and 3-D output. It uses Microsoft's TTS engine when speaking English and Toshiba's engine when speaking Japanese. TTS functions are realized with use of Microsoft Speech API.

TTS client, an interface between Meadow and TTS server, can be used outside Meadow at DOS prompt. When standard outputs of "dir" or
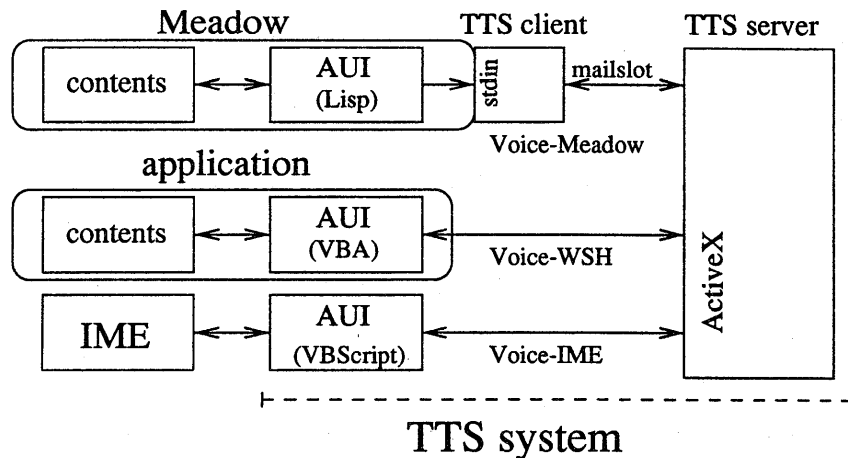
Figure 1: Configuration of the current TTS system.

"type" commands are put into the standard input of TTS client through a pipe, TTS server speaks the directory information or a content of the file.

2.2 Voice-WSH

Voice-Meadow is intended for experts because Meadow is not an easy application. There are, however, a lot of visually impaired persons who are not good at computers. Thus, a new system that will be a natural extension of Voice-Meadow to universal Windows applications is required.

Voice-WSH (Fig. 1) is designed to fulfill such demands. As clearly shown in Emacspeak, an intelligent AUI is indispensable to this system. We use Microsoft Visual Basic(R) for Applications (VBA) and Microsoft Visual Basic(R) Scripting Edition (VBScript) with Windows Scripting Host (WSH) as a rendering language of the AUI. VBA is not supported by all applications but major applications such as Microsoft Office. Since VBA and VBScript is a subset of Visual Basic, they are easy and users can customize AUI for their own purposes.

2.3 Application to other fields

The current TTS system can apply to other welfare such as helping people who have difficulty in reading. It also can apply to mobile computing such as medical rounds by doctors, wearable computers, and in-car use (e.g. navigation system). The application of this system to these markets will include a speech recognition system or a dedicated speech device.

3 Concluding remarks

The current suit of Voice system (Voice-IME, Voice-Meadow, and Voice-WSH) is still in developing stage but essential parts have been implemented. The current system, based on Emacspeak, is a TTS system much wider than the currently available commercial screen-readers such as ProTALKER(TM)97 and 95Reader. The heart of the current system is a context sensitive interactive AUI, which was implemented only by Emacspeak other than the current system.

Aural output is like an auto-scrolling display because it is temporal. At the same time, its eyesfree interaction is suitable for mobile computing. Thus, if a computer incorporates independent aural and visual interfaces, it will break a new world.

References

[1] T.V. Raman: *Audio System for Technical Readings*, Cornell University (1994).

[2] Emacspeak package is found at http://simon.cs.cornell.edu/Info/ People/raman/emacspeak/.

[3] T.V. Raman: *Auditory User Interfaces – Toward the Speaking Computer-*, Kluwer Academic Publishers (1997).

[4] Meadow, Multilingual enhancement to gnu Emacs with ADvantages Over Windows, was written by Miyashita Hisashi and edited by Takeyori Hara.