Multimodal Interfaces for Internet

Luc Julia
STAR Laboratory - SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025, USA
julia@speech.sri.com

Adam Cheyer
Artificial Intelligence Center - SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025, USA
cheyer@ai.sri.com

ABSTRACT

The World Wide Web frightened us: during the first two years of its popularity, we felt that human-computer interaction had been set back 30 years. Fortunately, Java expanded the possibilities for user interface design, providing a way to run complex programs over the net. In this paper, we present a Java-enabled application with a multimodal (pen and voice) interface over the web.

Our implementation approach was to add Java to the set of languages accepted by the Open Agent Architecture (OAA), a framework for rapidly prototyping complex applications, and particularly suited to those with multimodal interfaces [1]. Given the OAA's distributed nature, an OAA-based application can be run from a lightweight computer by downloading only the small user interface component, while the core of the application is implemented on a server composed of larger agents (e.g. speech recognition (SR), natural language, database) which cooperate and compete in parallel.

Despite the current lack of APIs for media input in Java, we chose voice entry as the primary input modality for the user [2], first by using a telephone to access a remote Nuance Speech Recognition server [3], and secondly by designing our own Java API for Speech Recognition using external native methods on the client.

ATIS [4], our first prototype application using SR over the telephone and the Java implementation of the OAA, has been publicly available on our web site for more than a year. Given the success of this first experiment (more than 4000 users), our next task was to bring multimodal concepts developed for several map-based applications [5] to the Web.

In order to achieve our multimodal objectives, it was necessary to adapt the pen modalities (gestures and handwriting) to an Internet and multiplatform context, i.e. the pen had to be interchangeable with any pointing device. For portability, the gesture recognition algorithms [6] were recoded directly in Java. A trade-off was made for handwriting, integrating a Java-enabled version of JOT, a character by character recognizer developed by CIC [7].

REFERENCES

1. Moran Douglas, Cheyer Adam, Julia Luc, Martin David. and Park Sangkyu. Multimodal User Interfaces in the Open Agent Architecture. In *Proceedings of IUI'97* (Orlando, FL), 1997, pp. 61-68.

2. Julia Luc and Cheyer Adam. Speech: A Privileged Modality. In *Proceedings of Eurospeech'97* (Rhodes, Greece), 1997, to appear.

3. http://www.nuancecom.com

4. Julia Luc, Cheyer Adam, Neumeyer Leo, Dowding John and Charafeddine Mehdi. HTTP://WWW.SPEECH.SRI.COM/DEMOS/ATIS.HTML. In *Proceedings of AAAI'97 Spring Symposium* (Stanford, CA), 1997, pp. 72-76.

5. Julia Luc and Cheyer Adam. A Multimodal Computer-augmented Interface for Distributed Applications. In *Proceedings of HCII'95* (Tokyo, Japan), 1995, pp. 237-240.

6. Julia Luc and Faure Claudie. Pattern Recognition and Beautification for a Pen based Interface. In *Proceedings of ICDAR'95* (Montreal, Canada), 1995, pp. 58-63.

7. http://www.cic.com