SALT-Based Multimodal Interactive Messaging
WA3-03

Supervisor: Dr. Albert Wong

Lam Yu Kwan, Ivan
Chan Tat Lun, Glen
Chan Kin Chung, Ben
A multimodal application on the Internet is one that allows the user to use a wide range of input devices such as speech, keyboard, and mouse concurrently and independently to interact with other users or programs. At the same time, the output of the multimodal application can be speech, sound effect, text and graphics for the users to access the Internet.

In our project, we are going to develop multimodal web application based on SALT (Speech Application Language Tags). With speech, the user can have an interactive communication as well as entertainment. We will implement some voice-activated functions with the use of speech in the webpages in order to support a variety of input devices for our multimodal application. Our webpages can be used by a game company to provide some games, chat rooms and customer services to Internet game players.

**SALT (Speech Application Language Tags)**

Speech Application Language Tags (SALT), founded in 2001 and representing over 70 technology leaders included Microsoft and Cisco Systems, are a lightweight set of extensions to existing markup languages, in particular HTML and XHTML that enable multimodal and telephony access to information, applications and Web services from PCs, telephones, tablet PCs and wireless personal digital assistants (PDAs).
Microsoft Speech add-in beta 3.0, which includes Speech Recognition Engine and Synthesis Engine, act as a bridge to communicate between Server and Input/Output device. The website contained seven components: website interface, two games, online chat room, system integration and windows applications control.
The powerful feature of SALT is demonstrated in the game website. Users can use speech without the aid of mouse and keyboard to obtain the company information, play games, chat with friends and even control the windows applications in their desktop computer. The speech synthesis function also provided a user-friendly interface and thus generated a more human-oriented environment.

After surveying 119 users, more than 60% of the respondents think that the overall speech usage performance was at the moderate level.