

Features

Multimodal interaction: Humanizing the human-computer interface

In everyday life humans use speech, gestures, facial expressions, touch to communicate. And, over long distances we resort to text messages and other such modern technology. Notably, when we interact with computers we rely exclusively on text and touch in the form of the keyboard/mouse and touch screens.

Kouichi Katsurada is an associate professor at Toyohashi Tech's Graduate School of Engineering with a mission to 'humanize' the computer interface. Katsurada's research centers on the expansion of human-computer communication by means of a web-based multimodal interactive (MMI) approach employing speech, gesture and facial expressions, as well as the traditional keyboard and mouse.

"Although many MMI systems have been tried, few are widely used," says Katsurada. "Some reasons for this lack of use are their complexity of installation and compilation, and their general inaccessibility for ordinary computer users. To resolve these issues we have designed a web browser-based MMI system that only uses open source software and de facto standards."

This openness has the advantage that it can be executed on any web browser, handle JavaScript, Java applets and Flash, and can be used not only on a PC but also on mobile devices like smart phones and tablet computers.

The user can interact with the system by speaking directly with an anthropomorphic agent that employs speech recognition, speech synthesis and facial image synthesis.

For example, a user can recite a telephone number, which is recorded by the computer and the data sent via the browser to a session manager on the server housing the MMI system. The data is processed by the speech recognition software and sent to a scenario interpreter, which uses XISL (extensible Interaction Scenario Language) to manage the human-computer dialogue.

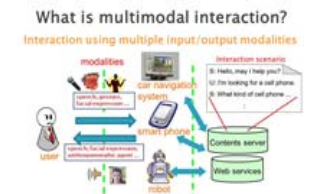
"XISL is a multimodal interaction description language based on the XML markup language," says Katsurada. "Its advantage over other MMI description languages is that it has sufficient modal extensibility to deal with various modes of communication without having to change its specifications. Another advantage is that it inherits features from VoiceXML, as well as SMIL used for authoring interactive audio-video presentations."

On the downside, XISL requires authors to use a large number of parameters for describing individual input and output tags, making it a cumbersome language to use. "In order to solve this problem, we will provide a GUI-prototyping tool that will make it easier to write XISL documents," says Katsurada.

"Currently, we can use some voice commands and the keyboard with the system, and in the future we will add both touch and gestures for devices equipped with touch displays and cameras," says Katsurada. "In other words, it is our aim is to make interaction with the computer as natural as possible."



Kouichi Katsurada



Description of multimodal interaction