Multimodal Interfaces and Accessibility

Tony Scarlatos
Stony Brook University

Lori Scarlatos
Brooklyn College

ABSTRACT

Digital multimedia provides a common platform for all content types, allowing for easy translation of one type to another. Translation of one type of information to another, such as auditory to visual, is key to making information accessible to individuals with impairments that limit their ability to comprehend certain media representations.

Recent advances in the processing power of desktop computer systems, and the decreasing cost of sensors and transducers, have also contributed to the accessibility of digital information, by making a diverse range of user inputs economical and feasible. Whereas the traditional keyboard and computer mouse interface excluded many users with visual or physical impairments, new technologies such as speech recognition have made interaction with digital content easy and intuitive.

In this paper, we describe three projects we have developed that facilitate communication between the hearing and hearing-impaired communities, and provide navigation of web-based content for visually impaired users.

The first project, iSign, translates spoken English into video clips of American Sign Language (ASL). The second project, FingerSpell, translates the hand gestures of the ASL alphabet into synthesized speech. And the third project, BookTalk, is a speech-enabled E-commerce site dealing in large print books and books on tape.

iSign

iSign was developed to address the slow rate of deaf students' vocabulary acquisition, an impediment to "mainstreaming" the students into public middle schools. In schools for the deaf instruction follows a "bilingual" strategy, where most subjects are taught in American Sign Language (ASL), and English is taught as a second language, using traditional techniques such as flash cards. But 95% of parents with deaf children do not know ASL, and cannot review vocabulary lessons with their children at home. By enabling parents to drill vocabulary lessons with their child it is anticipated that the student will acquire the vocabulary more quickly, recognized both as word shapes and lip movements.

iSign is a discrete speech recognition application, and therefore is speaker-independent and requires no training of the software. Based on groups of semantically related words we call albums, each vocabulary word is associated with a word shape, an illustration of the word, and a video clip of the ASL gesture for the word. The parent reads from the list of words, and the student watches their lips form the word. Then the student sees an ASL sign they recognize, the word shape, and an illustration of the meaning of the word. iSign is programmed so that additional albums of content can be added without recoding - the new albums are recognized instantly.

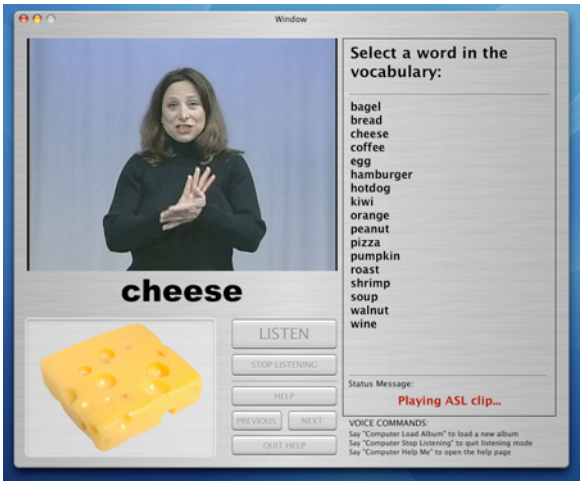

Figure 1. Result of saying "cheese" with the Food album open.

Figure 2. Linguist and author Noam Chomsky (center) testing iSign.

iSign can also be used by the hearing as an ASL learning tool; or as a speech training tool for advanced deaf students. If the student pronounces the vocabulary word correctly they get a visual confirmation of their success.

FingerSpell

Many words in English do not have a corresponding ASL gesture, and so these words are spelled out by a series of hand gestures, called finger spelling. Although "signing" a long speech would be time consuming and tedious this way, it is an adequate method for short messages. From the software developer's point of view it is considerably easier to translate a set of 26 simple hand gestures into strings than it would be to capture thousands of complex gestures and translate them into words. Since there are only minor differences in the hand signs for each letter, the gestures are actually parsed, which allows us a higher success rate of data capture and a lower error tolerance.

Figure 3. Data glove providing input to FingerSpell.

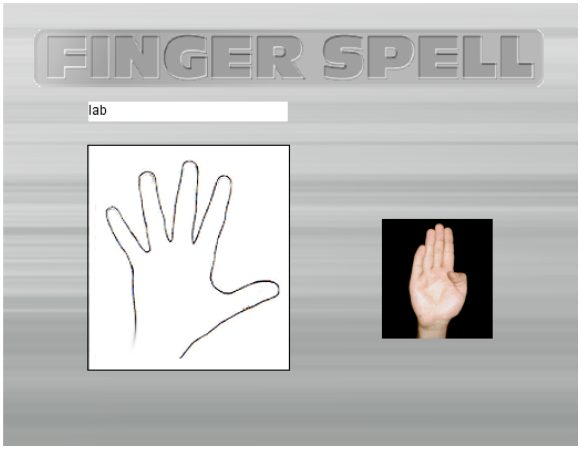

Figure 4. FingerSpell interface shows the word that was just spelled. Open-hand gesture indicates end of a word.

We couple the parsing of the hand signs with a fast dictionary search of 10,000 common words in English to allow the signer to spell words without having to sign each letter. This approach is similar to field completion in web browsers and forms. With this approach it is possible to spell most words in 4 gestures or less.

To capture the hand gestures we developed a custom data glove that is light and comfortable. 12 sensors, of both the touch and flex variety, are used to record finger positions. Recognized gestures add a letter to the current string, displayed on-screen. When the signer completes a word they make an "end word" gesture and the string is translated into synthesized speech. The state of the glove is also displayed on-screen providing feedback to users trying to learn finger spelling.

BookTalk

Visually impaired readers, such as the elderly, are able to satisfy their literary appetites with large print books or books on audio tape. But purchasing these books conveniently through the world wide web is challenging. Simply making the text size of the web page larger is not reliable, because of differences of text representation in various browsers, differing screen resolutions, etc.

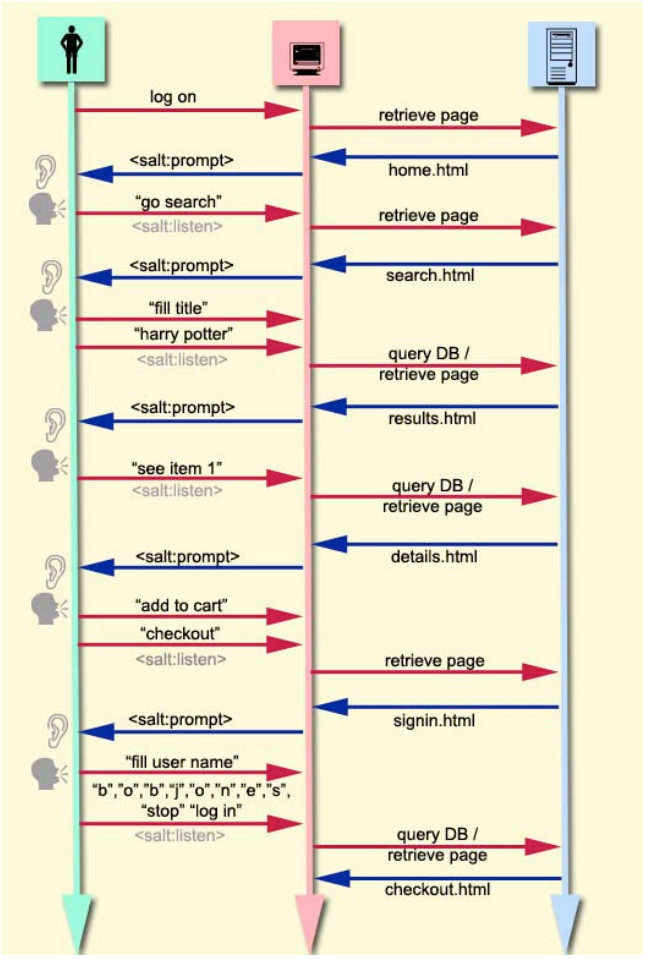

Figure 5. Flow of interactions during a simplified purchase session.

Using Speech Application Language Tags (SALT) we have developed a database driven web site that allows visually impaired users to verbally query the web site catalog and

have the query results read to them in a synthesized voice. Users can have excerpts from the book, liner notes, and reviews read to them, and are able to complete the order form verbally.