J. EDUCATIONAL TECHNOLOGY SYSTEMS, Vol. 33(2) 165-172, 2004-2005

FINGERSPELL: LET YOUR FINGERS DO THE TALKING*

TONY SCARLATOS

DMITRI NESTERENKO

State University of New York at Stony Brook

ABSTRACT

In this article we discuss an application that translates hand gestures of the American Sign Language (ASL) alphabet and converts them to text. The FingerSpell application addresses the communication barrier of the deaf and the hearing-impaired by eliminating the need for a third party with knowledge of the American Sign Language, allowing a user to use hand signs that will be translated to letters and words. Through the use of a data glove embedded with multiple input sensors, the application parses each hand and finger gesture signed by the user and detects unique cases which trigger application responses. Application responses come in both visual and auditory forms. The letter corresponding to the unique gesture will appear on the screen and play aloud using computer speech synthesis. The user may also sign an end of word gesture which cues the application to read the previous string of letters together, as one word. This allows a deaf person to communicate with another person without the need for the other person to understand sign language. In addition, this application can serve as a teaching tool for those learning how to sign. FingerSpell is a stepping stone to opening up lines of communication for the Deaf and hearing-impaired communities. Future research and advancements can extend the FingerSpell application to cover the full sign language (word signs) and alphabets of other sign languages.

*This work was supported by NSF CRCD grant #0203333.

INTRODUCTION

The language that is in primary use by the hearing impaired community in America is American Sign Language (ASL). ASL is composed of hand gestures that are combined with facial expressions to sign words. The vocabulary is extensive: a few thousand words can be signed with gestures [1]. Another approach to signing, commonly used to sign proper names and words that do not have a corresponding gesture, is to fingerspell them. Fingerspelling assigns each letter of the alphabet one of 26 unique hand gestures. Words are signed by concatenating these gestures together. This difference is similar to the distinction between pronouncing a word CAT, for example, and pronouncing C-A-T letter by letter [2] (see Figure 1).

This article discusses possibilities for filling the communication gap between the hearing-impaired and the larger community without the requirement for ASL translators.

HISTORY

In recent history the sociological perspective on Deaf[1] signers has shifted to view the group as a linguistic minority rather than individuals with a disability [2]. The American Sign Language has had a rough history of development, it was even banned throughout 1890-1935 at the suggestion that oral English should be the primary language to educate deaf children [1]. However, recent educational developments have focused on acknowledging signed language as a primary language, for education and communication, in children and adults. The majority of deaf children are born to hearing parents and it is important to give the parents as well as hearing-impaired individuals the tools to learn and communicate with ASL [1]. Studies have shown that individuals with solid skills in ASL and English have the higher scholastic performance noted in bilingual children (Vandenberg).

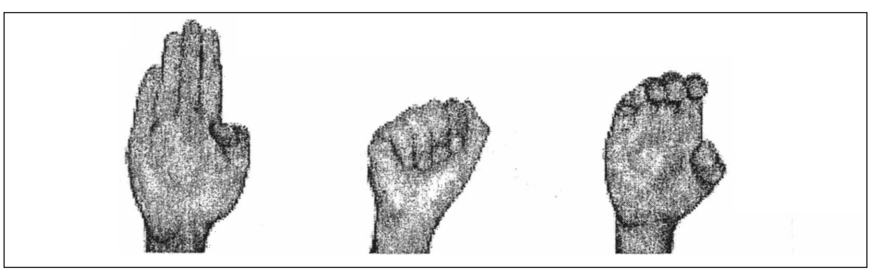

Figure 1. Finger spelling: hand signs for “b,” “a,” and “c” from left to right.

[1] We use the capitalized form “Deaf” to refer to the group of people with hearing disabilities. The lowercase “deaf” is used to refer to the audiological condition.

The rise of processing power, advancements in sensor technology, and multimedia proliferation on the PC market have allowed researchers to begin developing systems that attempt the translation of ASL into textual and spoken mediums. There are several different veins of research into the translation of gestures into text or speech. The first of these attempts to recognize the hand gestures of the signer. With thousands of unique gestures that correspond to words combined with facial expressions to indicate emotion, this approach involves complex sensor systems, visual information processing, and sophisticated algorithms to process and translate the data [1, 3]. Another complication is that in ASL the signer typically makes gestures for the word and then places that word in its appropriate place in the sentence [1]. In order to understand the meaning of a gesture, researchers use artificial intelligence and trainable systems.

The second approach is to process the individual hand signs and translate them into letters. This can be done either through visual processing or data gloves [4, 5]. The finger spelling approach is much easier computationally because the number of sensors or visual reference points is significantly reduced. The drawback on the user side is that it is tedious to maintain a lengthy conversation if each word has to be spelled out letter by letter.

Our solution attempts to use the computational and algorithmic ease of processing hand signs as compared to hand gestures, yet allow the signer to maintain a conversational speed. We chose this approach because we wanted to create an interface that is easy to learn, does not require the training time necessary for whole gesture processing, and is comfortable for the signer. We also considered the ease with which our system could be replicated. To solve the latter problem, all of the parts we use are readily available and are easy to combine. This also makes our solution very cost-efficient.

RELATED WORK

The classroom environment is far different from that of a laboratory and that is something that we have kept in mind. The goal of this project was to develop a tool for speech and hearing-impaired students which does not require significant learning time or complicated setups. Gesture recognition systems developed at computer science departments around the world require multiple input devices and we believe that their presence inside a classroom would be bulky and uncomfortable for the user.

One approach by Sidney Fels and Geoffrey Hinton that translates hand gestures to speech requires four input devices: a flex glove to process finger bends, a touch glove to process finger and wrist contact, a motion tracking device to track the hand, and a foot pedal to control volume [6]. Fluid gesture processing in this and most other cases requires the use of Neural Networks to train the system to distinguish among the individual variations of hand gestures that naturally arise in users. The training is necessary for the proper operation of the system and the time it takes is measured by tens of hours. Furthermore, because this device translates gesture directly to speech (similarly to text-to-speech software), during training users must be capable of hearing, reducing the user group to those who are speech-deficient but capable of hearing. Other gesture recognition systems such as the one developed by Waleed Kadous at the University of New South Wales, are cost efficient but suffer from their small dictionary size typically reduced to only commonly used words [3].

Visual recognition systems like the one developed at MIT suffer from occlusion and requirements for ambient light [7, 8]. It is a major problem of computer imaging and the solution is a balance of image extraction algorithms and requirements for the testing rooms. We believe that it is unwise to setup such requirements on the classrooms, opting instead to develop a system that will work under any condition.

Once we came to the conclusion that we would develop a finger spelling system rather than a gesture based system like other researchers, we were confronted with a choice of data acquisition gloves [4, 5]. Data gloves belong to a group of haptic data acquisition devices. They vary greatly in their purpose, function, and ergonomics. Some are cheap and simple such as the P5 glove [9] primarily used for gaming and others are sensor packed devices that are often used in virtual reality environments [10] whose prices are in the range of $10,000-$20,000.

Their effectiveness is also offset by their weight and bulkiness. Although there are no readily available data gloves that would be ideal for our purpose, we were opposed to creating one from scratch because that would significantly complicate the application of the device in a real environment. We also wanted the input from the glove to be compatible with any modern PC system, ideally through the use of USB, since that would remove the rigidity of a custom made processing device. We chose to focus our research on allowing the resulting text to be output to any application. We chose to harness inter-operability and flexibility of multimedia applications.

CONSTRUCTION

The construction of our data glove was completed with sensors purchased from InfusionSystems based in Montreal [11]. We combined a touch glove with six bend sensors: one for each finger and one for the wrist, for a total of twelve sensors. The sensor data is digitized with the iCubeX device supplied by InfusionSystems and acquired by a PC equipped with a USB port. The setup requires attaching the sensors to the glove in an intuitive way; the overall cost is under $600. The material of the glove supplied by InfusionSystems is similar to a golf or a baseball glove so the device is comfortable, allows the hand to breathe, does not wear the user out, and is very ergonomically-conscious (see Figure 2).

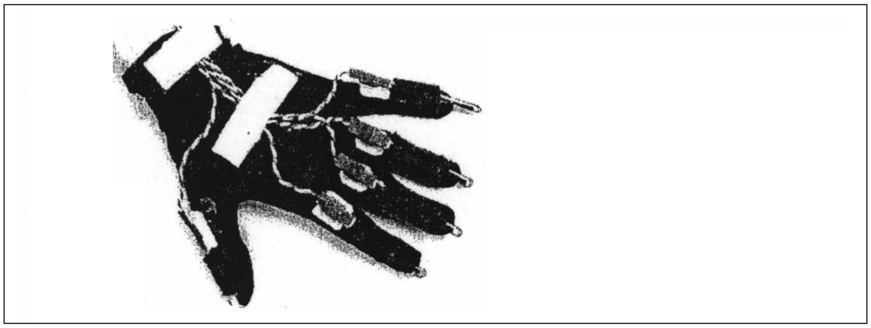

Figure 2. Our data glove for capturing hand signs. It is designed mostly from off the shelf parts. Touch sensors are embedded inside the globe, while the bend sensors are on the outside.

For the prototype, we processed the data with DirectorMX, a powerful multimedia tool that easily allowed us to bring various inputs together. We studied the finger spelling hand gestures and parsed them into groups of bend and touch signals. The result was a 12-digit code that is best suited to represent a sign of the ASL alphabet [12]. We have allowed for certain ambiguities to place the least number of restrictions on the signer by having two or even three codes correspond to the same letter, assuring that even if certain fingertip touches were not completed, the software would still "understand" the sign if the finger bend information was correct. The current state of the glove is continuously updated and compared with the letter codes. The amount of time between updates of the state is configurable by the signer, depending on their skill level. Those new to finger signing may want more time to adjust their hand gestures, while a signer who is familiar with the gestures may want to speed up gesture acquisition. To signal end-of-word we adapted the open palm hand sign, as it is not used to sign any letter and is a natural choice for a space sign. Once a word has been entered and the user signs end-of-word the text-to-speech module pronounces it.

Signing with finger spelling instead of hand gestures, while efficient from the computational point of view, has one obvious drawback of slowing down the communication. In standard practice, finger spelling is used to sign about 7% of the words, most of them are proper nouns and foreign words. In fact, the system was developed to allow the signing of foreign words in deaf communication [2]. To work around that we have built in an ability to search through a dictionary by narrowing down the word possibilities. It is often possible to sign a word very quickly with just three or four letters. By using a standard-sized dictionary with about 10,000 words and a fast search algorithm, we can find the desired word in real-time. Because finger spelling is less computationally and sensor intensive, we can provide a greater success rate of recognizing the sign the user had intended than if we had attempted full hand gesture recognition. Our dictionary search combines the extended communication that is possible through hand gestures with the simplicity of finger spelling. To cut down on mistakes, we eliminate excessive repetition of letters which would occur if the user keeps the same gesture while the current state is updated.

ACCESSIBILITY FAMILY

Using DirectorMX and the OSX operating system for the prototype allowed us to harness the accessibility features native to that platform. That places this device in a chain of features that make the computer friendly to users with disabilities. The combination of applications such as iSign [13], features built into the operating system such as speech synthesis, and this multimode interface device broadens the communication experience for people with disabilities.

The visual interface of our software reflects the notion that the user must have immediate feedback with the device, especially due to its primary purpose as a learning application (see Figure 3).

We focus on visually displaying the current state of the data glove as well as displaying a photo of the hand sign that corresponds to the letter that has been signed by the user. The combination is intended to help new users learn how to sign by providing them with visual feedback. The "current state" visually reflects

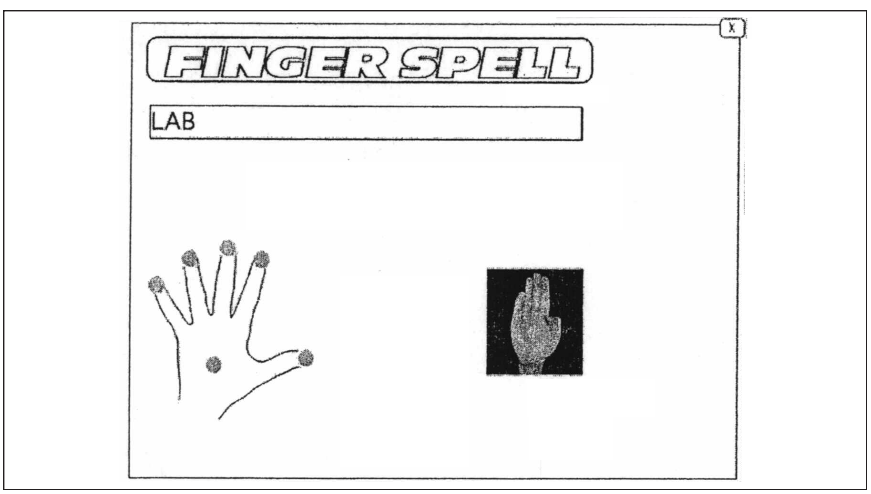

Figure 3. Our interface allows for maximum visual feedback: the "state" of the hand is displayed on the bottom left, the letter currently signed on the bottom right, and the word that is being spelled is shown on the top (in this case it's "lab").

the state of each of the touch sensors which can be either on or off and the state—extended, half bend, or a full bend—of the six bend sensors. The text field displays the current string. A letter is appended to the string whenever the user makes a sign that corresponds to a letter from the ASL alphabet.

CONCLUSION

The advances made in educational technique and multimedia proliferation have made a number of crossover applications possible. These applications should function as unobtrusive bridges between information and learning. Our current application strives to function as a multifunctional translator and tutor for people who are using finger spelling and ASL to communicate. We hope that ultimately devices like this will become powerful tools in real life enabling new modes of discourse.

FUTURE RESEARCH

Our future research includes field testing, making the glove capable of wireless communication, and including the alphabets of other languages to expand the freedom of use.

REFERENCES

1. H. Bornstein, *Manual Communication,* Gallaudet University Press, Washington, D.C., 1990.
2. I. Parasnis, *Cultural and Language Diversity,* Cambridge University Press, Massachusetts, 1996.
3. W. Kadous, *Grasph: Recognition of Australian Sign Language Using Instrumented Glove,* New South Wales University, October 1995. Available online at: http://www.ese.unsw.edu.au/~waleed/thesis.html.
4. J. Wang, Jennifer Shim, and L. Chen, *FingerPilot ASL,* Robotics Lab at Wellesley College, 2002. Available online at: http://cs.wellesley.edu/~jwang1/FingerPilot/project.html.
5. *A Glove That Speaks Volumes,* Wired News Article. Available online at: http://www.wired.com/news/gizmos/0,1452,49716,00.html.
6. S. Fels and G. Hinton, *Glove-TalkII: An Adaptive Gesture-to-Formant Interface,* Chi '95 Proceedings. Available online at: http://www.acm.org/sigchi/chi95/Electronic/documnts/papers/ssf_bdy.htm.
7. T. Starner, *Visual Recognition of American Sign Language Using Hidden Markov Models,* Master's thesis, MIT Media Laboratory, February 1995.
8. T. Starner and A. Pentland, *Real-Time American Sign Language Recognition from Video Using Hidden Markov Models,* IEEE Transactions on Pattern Analysis and Machine Intelligence, Los Alamitos, California, July 1997.

9. *Essential Reality P5 Data Glove.* Available online at: http://computer-shop.yellowpages.pl/12232R8P1828_Essential-Reality-P5-Gaming-Glove.html.
10. *Multi Purpose Data Gloves.* Available online at: http://www.doc.ic/ac.uk/~nd/surprise_97/journal/vol1/ncp/.
11. *Infusion Systems, Inc.* Available online at: http://www.infusionsystems.com/index.shtml.
12. *A Fingerspelling Reference Table.* Available online at: http://where.com/scott.net/asl/.
13. T. Scarlatos, L. Scarlatos, and F. Gallarotti, *ISign: Making the Benefits of Reading Aloud Accessible to Families with Deaf Children,* IASTED International Conference on Computer Graphics and Imaging, 2003.

Direct reprint requests to:

Tony Scarlatos
Computer Science
Stony Brook University
2204 Computer Science Bldg.
Stony Brook, NY 11794-4400
e-mail: tony@cs.sunysb.edu