An Intelligent Computer-Based System for Sign Language Tutoring

Tim Ritchings, PhD

Ahmed Khadragi, PhD\textsuperscript{1,2}

Magdy Saeb, PhD\textsuperscript{2}

\textsuperscript{1} School of Computing, Science and Engineering, The University of Salford, Greater Manchester, UK

\textsuperscript{2} Computer Engineering Department, College of Engineering and Technology, The Arab Academy for Science Technology and Maritime Transport, Alexandria, Egypt
An Intelligent Computer-Based System for Sign Language Tutoring

ABSTRACT A computer-based system for sign language tutoring has been developed using a low-cost data glove and a software application that processes the movement signals for signs in real-time and uses Pattern Matching techniques to decide if a trainee has closely replicated a teacher’s recorded movements. The data glove provides 17 movement signals from Bend sensors and push down switches and is used initially by the teacher to record selected signs. A trainee can study an animated 3D model of a hand and arm showing the recorded sign movements at leisure, before using the data glove to try and replicate the sign. Four Arabic Sign Language (ArSL) teachers provided 65 common posture and gesture signs, and then 10 trainees from the local community were asked to evaluate the system. They achieved an overall average accuracy of 93.8% in replicating the signs, despite finding the gestures harder than the postures to perform, and found approximately 18% of the signs difficult because of particular thumb/finger and wrist bending angles. Both the teachers and the trainees were familiar with these signs, and a usability questionnaire revealed that they preferred this approach to sign-language tutoring than the traditional human based method they had already experienced.

KEYWORDS Arabic Sign Language; data-glove, graphical interface; sign language tutoring.

Introduction

According to the World Federation of the Deaf in their Position Paper regarding the United Nations Convention on the Rights of People with Disabilities, there are 70 million Deaf people worldwide, and 80% of these lack education and are illiterate or semi-illiterate (WFD, 2003). They also commented that qualified sign language interpreters and technological assistance, which enhance accessibility, are woefully
inadequate or absent, and without access to language and information, full involvement and participation in life and life functions are seriously curtailed. The impact of these issues has been noted in a number of studies in the past 20 years which have revealed that while deaf and hard-of-hearing children in classrooms struggle with reading, writing and communication (Long et al., 2007; Antia et al., 2005; Long & Beil, 2005), there seemed to be no direct connection between deafness and mathematical difficulties (Nunes & Moreno, 1998), implying that their capacity for learning was no different to hearing children. This suggests that with support to help get over the communication barrier, children and adults can fully utilize their learning capabilities.

While there are an increasing number of people, particularly children, receiving cochlear implants in the United States, and ongoing debate about the quality of the benefits, in relation to the risks involved, sign language remains the main communication mechanism for Deaf people. Sign languages vary from country to country, the most common ones being the British Sign Language (BSL), the American Sign Language (ASL) and the Australian Sign Language (Auslan) (Nicola, 2009). The developing countries, where most of the Deaf people are illiterate, have traditionally lagged behind, but are now developing their sign languages. For example, ArSL is the official sign language in the Arab world (Abdel-Fattah, 2005). However, only a few individual attempts have been made to develop computer systems to teach ArSL, and the descriptions were short and lacking sufficient detail to provide much insight about the approaches taken to teach ArSL (Mohandes, 2001; Jarrah & Halwani, 2001).
Despite the variations in sign-language, the basic approach is to convey meaning through hand gestures, which is a composite action constructed by a series of hand postures (static hand configuration and hand location without any movements involved) that act as transition states (Chen et al. 2007). Many computer-based approaches to recognising signs have been described in the literature. For example, in 2000, a system with 10 hand-sensors that translated a signer’s finger movements into easy-to-read words via a handheld wirelessly connected display was described (Patterson, 2010). However, this system could only spell the words using finger spelling instead of recognizing full words. Another system was concerned with the translation of the ASL into speech (Rebollar, 2003), and used accelerometers to capture hand shape, hand orientation and hand location, which induced some accumulative errors, but also could not detect horizontal movements. It also included a sign classification algorithm, which involved a considerable processing and training overhead. Later, a sign recognition system was designed that recognised isolated signs from the ArSL Unified Dictionary using an instrumented Power Glove and a Support Vector Machines algorithm (SVM) (Mohandes & Buraiky, 2007). This Power Glove system used 4 finger sensors, and ultrasound tracking for the roll movement and 3D-positioning. The machine learning process introduced processing overheads and gave inaccurate results because of considerable data averaging. More recently, a group of engineering students developed the HandTalk Interactive Communication Glove (HandTalk, 2010), which again had single sensors on only four fingers, affecting the degree of precision in recognizing signs. The system experienced significant delays after each sign was performed while it was being recognised. Another system investigated was the Zerkin’s glove (Zerkin, 2010) that was
developed for capturing hand movements for integration in Maya and game engines, although there are known factors that can affect the tracking accuracy.

A key component of these computer-based systems for teaching sign language is the capture of the hand movements, and several approaches to tracking the movements have been considered, including vision-based techniques, motion capture data gloves, body suits and recent hand sign recognition systems. The vision-based tracking techniques used in Virtual Reality systems (Pavlovic et al., 1996; Azarbayejani & Pentland, 1996; Crowley et al., 1995; Coutaz. & Crowley, 1995) were inappropriate here because of occlusion and the constraints imposed by cameras being mounted in fixed pre-set locations. Electromagnetic tracking devices were also investigated (Polhemus, 2008), but tend to be large, heavy and expensive. Ultrasound tracking employed in other research was considered (Mohandes, & Buraiky, 2007). This technique requires a line of sight between the tracker and the sensors on the moving object which constrains user movement, and so was also ruled out.

A more practical approach to hand and finger tracking is through the use of data capturing hand gloves and full body tracking skeletons. CyberGloves were investigated by (Chao, 2001) and were found to have mean measurement errors of 1.7° across nine participants, although other researchers have recorded measurement errors as high as 6° (Kessler et al., 1995). It was observed that the accuracy was limited by cross-coupling effects between the sensors (Kahlesz et al., 2004). Other commercially available sign capturing gloves suffer from high price, low number of sensors or inaccurate tracking
means and motion confining tracking techniques (Dipietro, 2008). Full body tracking suits are heavy and expensive, and also lack finger tracking such as (Animazoo, 2009).

As sign language tutoring is time-consuming and usually occurs in formal one-to-many sessions involving a teacher who shows the trainees how to perform signs, this work explores the potential for a computer-based system that can support sign-language tutoring, that enables users to learn a sign language unsupervised and at their own rate in their home environment. Although this system provides a generic solution to sign language tutoring, for evaluation in this study, the Arabic Sign Language (ArSL) was used because of the expertise of the local sign-language teachers and the interests of the authors.

Methodology

The approach taken has been the development of a low-cost data glove and a software application that processes and displays the movement signals for a specific sign in real-time and uses Pattern Matching techniques to decide if a trainee has closely replicated a teacher’s recorded movements for that sign. The system was evaluated objectively by measuring the ability of a group of 10 trainees to accurately replicate 64 common ArSL signs recorded by 4 local teachers. The trainees were also asked to give a subjective indication of the difficulty they experience in trying to replicate each sign. As the trainees were already familiar with the majority of these signs, they were completed a subjective questionnaire related to the usability of the system and included their views on this
approach to sign language tutoring compared with the traditional approach that they has already experienced.

The design and testing of the data-glove, the description of the tutoring application, and the system evaluation strategy are discussed below.

**The Data-glove**

In order to capture the finger flexion, adductions, abductions (fingers moving closer and away respectively) and wrist movements, a variety of sensors were tried out, including small joystick sensor (CTS, 2010), a Panasonic slider potentiometer, and a rotary potentiometers (Digikei, 2010). They were all found to be unsatisfactory because of factors such as shape, size, inertial errors and unstable outputs. The most successful approach was to use flex sensors and the Flexpoint Bend Sensor (Flexpoint, 2010). It was found to be the best in terms of variable length, as they could be shortened, and could be fitted on individual joints. They also produced more accurate and stable readings, and found to be more durable, withstanding all trials and testing without breaking (Khadragi, 2011).

Various sensor holder modules were then connected to find the best way to mount these sensors in a stable manner on the fingers, including open chassis holders, wool gloves, metal and jointed skeletons. There were problems with some sensors getting loose and touching each other. The most resilient and stable glove was found to be the one shown in Figure 1, which has 17 mounted sensors. There are 2 Bend sensors mounted over each finger and thumb joints to measure the finger movements, a Bend sensor on the
top and on the bottom of the wrist (occluded in the Figure) to measure its pitch up and
down respectively, and a Bend sensor on the right side of the wrist to measure its yaw
(left/right movement). It should be noted that the Bend sensors only give a variable
output when bent in one direction, which is suitable for the fingers, but two are required
on opposite sides of the wrist to sense pitch up/down movements. Finally, in order to
detect finger abduction and adduction, 4 Push down switches are mounted to the side of
each finger starting with the thumb.

The 17 outputs from the data-glove were analogue voltages in the range 0 to 5 volts,
with an ATMEL AVR ATMEGA 16 (Atmel, 2007) microcontroller was responsible for
the collection of the raw signals and their conversion to digital form at a clock frequency
of 1MHz with 10 bits/sample/sensor. These digital signals transferred to the PC running
the tutoring application via the serial port at a Baud rate of 9600.

*The tutoring application*

Once the digital signals derived from the 13 Bend sensors and 4 switches on the data-
glove were transferred to the PC, the values were each converted to angles. This involved
a calibration process, where the voltages produced by specific bend angles were
measured. Since these sensors are linear in nature, all the various angles were calculated
linearly from any digital voltage value. This was confirmed for all the Bend sensors from
the calibration measurements. The voltage outputs from the push down switches were
either high (closed) or low (open).
As the signals from the two wrist Bend sensors were required to measure pitch up and down, 16 angles were derived from the 17 sensors and used to provide real-time control of a 3D animated model of the hand and arm as shown in Figure 2. The model replicates all the thumb/finger flexions and hand pitches up/down and yaws that can be made by the teachers and trainees. The model can be zoomed in/out, and studies from various viewing positions to help the trainee understand the teacher’s sign movements, before using the data-glove to try and replicate the sign. The trainee is able to select any sign to train for from the database of recorded signs from a list of pictures showing the object corresponding to each sign, as seen in the screen shot in Figure 3, and can practice this sign at their own rate until there was a good match with the recorded sign. The system produces continuous text feedback on the screen regarding the training status of each trainee.

The recorded signs were obtained from ArSL using the same data-glove in the same way as the as the trainee, and the full movements and timings are stored in a system database in order to replay the animated 3D movement for the trainees. However, in order to reduce the amount of data to be stored and used for the comparison of the trainee and teacher’s sign movement, the sign movements were also processed and represented in binary form. In the case of the Bend sensor angles, a straight and bent thumb/finger corresponded to 0° and 90° respectively. However, the teachers set tolerences of ±15° for finger flexes, and wrist movements, and so 0°-15° was coded as 0 and 75°-90° coded as 1. (The range of angles from 15°-75° were ignored and left for future work). The switch signals were coded as 0 (open) and 1 (closed). The final positions of the
thumb/fingers and wrists in a posture were thus encoded as a 17-bit binary pattern. (It should be noted that the two wrist sensors detecting pitch up and down were not independent as both could not be 1 at the same time.) In the case of a gesture, this binary pattern was recorded for each posture movement that formed the gesture. Although this approach loses information about the absolute speed of a sign, and the precise sequence of the thumb/finger and wrist movements, but does permit comparisons between the teacher’s movements and the slower and more variable speeds of a trainee. The application decided that the trainee had replicated the recorded posture when the two binary patterns matched exactly. While the system could give an indication of which angles were incorrect, the users found this confusing and this facility was eventually disabled.

The evaluation strategy

In order to evaluate the system, 4 ArSL teachers helped to create a database of 65 signs. Since signs generally are either posture based or gesture based so it was necessary to include both types of signs, and there were 33 posture based signs and 32 gesture based ones selected. Although the gesture signs only usually involved two postures in this study, it was considered interesting to see if the trainees experienced greater difficulty with the gestures. All the signs were chosen to be able to reflect all the movements that could be performed by a trainee within the limits of the 17 sensor configuration. They were chosen from various environments and fields such as animals, birds, alphabets, numbers, verbs, nouns and prepositions.
The evaluation of the system involved 10 trainees from the local deaf and hard of hearing community in Alexandria, and involved a set of controlled experiments, following a period of familiarisation with the data-glove and the tutoring application. These experiments allowed each trainee 5 attempts, without feedback from the application (which could have improved their performance), to reproduce each of the 65 signs. In each case the trainee’s posture was matched with the teacher’s recorded posture using the Pattern Recognition algorithm described earlier, and a mark of 1 was awarded for a successful sign posture. Each trainee could thus achieve a maximum score of 5 for each sign. In addition to monitoring the individual performance of the 10 trainees to be studied, the suitability of the data-glove suitability for capturing particular types of sign movements could also be investigated, and each trainee was asked to give a subjective indication of their difficulty in replicating the postures with the data-glove immediately after performing each sign.

In the subjective questionnaire related to the usability of the system the participants were asked the following four questions:

- How much experience have you had in the use of computers?
- How easy did you find it to use the tutoring application?
- How do you rate the stability of the tutoring application?
- How strongly would you recommend the use of the system in preference to traditional sign language teaching techniques.
They were also encouraged to prove any additional comments on their experience of using the system.

**Results**

In the objective tests, the number of successful thumb/finger and wrist movements each trainee achieved for each of the 65 signs was recorded. The overall accuracy of each of the 10 trainees is shown in Figure 4, and varied between 4.09 (82%) for Trainee 6 and 4.92 (98.4%) for trainee 7. The average for the whole group was 4.69 (93.8%). On inspection, Trainee 6 had small hands and that caused some degree of difficulty when bending the sensors exactly at their middle point, resulting in slightly out of tolerance range readings for some signs. In addition the data glove was loose, which caused the sensor not to stably reflect the exact finger joint angle while bending.

The accuracy achieved by the trainees on each of the 65 signs is shown in Figure 5a for the 33 postures and Figure 5b for the 32 gestures. It can be seen from the Figures that the ‘Up’ preposition (Sign56), and ‘Half’(60), proved to be with the least accurate at 4.4 (88%). This was because some of the sensors were unstuck and needed to be re-glued but the results were kept on record to judge the system fully with any glitches. Sign 60 also incorporated a left yaw bend angle that consequently produced faulty readings because of some wrist strains occurring to users. The easiest signs to produce were ‘Geem’(4), ‘400’(40), ‘2000’(42), ‘Dog’(50), ‘near’(55) and ‘TV’(65). The mean accuracies for the postures and gestures were 4.68±0.10(SD) and 4.69±0.14(SD) respectively, and an
unpaired t-test on the two populations gave a p-value of 0.43, suggesting that there was no statistically significant difference in the trainees’ performance on the postures and the gestures.

The scores for the subjective indication of the difficulty of replicating the each sign ranged between 0 (very easy) and 4 (very difficult). The subjective difficulty of each sign is shown in Figure 6a for the postures and Figure 6b for the gestures. (There are 13 signs with no error bars because all 10 trainees gave the same score) was computed. The mean scores for the postures and gestures were 1.03±0.99(SD) and 1.68±0.87(SD) respectively, and an unpaired t-test on the two populations gave a p-value of 0.0035, indicating that the trainees found the gestures more difficult to perform correctly than the postures. A total of 12 signs, representing approximately 18% of the database, were considered difficult (scores >2), because of certain joint bending angles. These signs, which included some postures and gestures, are listed in Table 1, together with the trainees’ remarks on which thumb/fingers and wrists movements caused the difficulty.

The scoring for the subjective assessment about the usability of the system varied between 1 (lowest) to 10 (highest), and the mean scores are given in Table 2. The results indicate that the software application was easy to use despite their limited experience with computers and stable (not crashing) and they found the system preferable for sign-language tutoring than the traditional human based approach that they had already experienced. The additional comments included a liking for the innovative way in teaching ArSL. They found that it was easier for them to learn the signs with the tutoring
system rather than dealing with a human teacher because it provided them with the time they required to practice with no limitation and with a high degree of freedom, eliminating the inconvenience of having to ask the teacher to move his hand to show it at different viewing angles during the class.

Discussion

The results presented in the previous section show that the tutoring system was able to record and accurately match the signals from the data-glove, and so could provide a useful tool for sign language tutoring. The subjective evaluation of the system by the trainees and the teachers, also supported this approach to sign language tutoring citing novelty, ease of use, flexibility and convenience to learn sign language at their own rate.

The trainees consistently achieved high accuracies in replicating the signs, despite finding some of these difficult because of particular thumb/finger and wrist bending angles. Erroneous signals, leading to mis-matches were investigated, and attributed to loose or poorly positioned sensors, this being the case for the trainee with small hands. Also, one sign incorporated a left yaw bend angle that consequently produced faulty readings because the trainee experienced some wrist strains. However, the data-glove proved to be sufficiently reliable for the operation of the system.

While there appeared to be no statistically significant difference in the trainees’ performance on the postures and the gestures, this could be deceptive as the movements for the gestures was not continuous, and there was a pause between each of the postures
in the gesture. Consequently the average score for the gestures might be expected to be similar to the average posture score. Further work is required to avoid this pause between the postures.

While it is difficult to compare these accuracies achieved here with those of other cited in the literature, the Power Glove system’s accuracy starts at about 48% for the Signs they used (Mohandes & Buraiky, 2007), while the minimum average accuracy achieved by the tutoring system was 82% by Trainee 6. Also, The Power Glove system’s maximum achieved accuracy was 91% which was comparable with the 98.4% achieved by Trainee 7. The Power Glove results also degraded dramatically at higher standard deviations because the averaging was not a representation of the actual features of the input signs. In addition, increasing the number of words had a bad impact on the accuracy of the Power Glove system, and with 120 words the accuracy of the system was found to be less than 5% depending on the classification algorithm parameters.

Conclusions

An interactive computer-based sign language tutoring system has been developed which incorporated push down switches to detect finger abduction and adduction. The Bend sensors were found to be a very effective low-cost light-weight movement sensors in this type of application. Sign movements were recorded for 33 posture signs and 32 gestures from 4 ArSL teachers. The tutoring system was evaluated by 10 trainees from the local deaf and hard-hearing community, who participated in objective tests of the system, which proved the ability of the application to record and recognise the signs
accurately using Pattern Matching techniques. They also gave subjective assessments of the degree of difficulty of each sign. All of the trainee’s were able to accurately replicate the 65 signs, although they found the gestures harder to perform that the postures. Approximately 18% of the signs were considered difficult because of specific hand and wrist movements.

In addition, both the teachers and the trainees gave a subjective assessment usability of the system, and the questionnaires revealed that even that the trainees who do not interact with computers in their everyday life they found the system to be easy to use, and in general terms they liked this approach to sign-language tutoring and found it preferable to the traditional human based approach that they had already experienced.

A valuable addition would be capturing data from the second signing hand too. This will allow more signs that their performance depends on two hands to be added to the prototype data glove to make it available to a wider range of sign language trainees.

References


Recognition Proceedings, IEEE, 3. (pp. 627-632). Vienna, Austria.


Figure 1 The data-glove with the finger and wrist mounted sensors

Figure 2 The 3-D graphical model of the hand and arm
**Figure 3** A screen shot of the tutoring system’s graphical interface

**Figure 4** The average accuracy for each trainee for the 65 signs
Figure 5a The average trainee accuracy for each of the posture signs

Figure 5b The average trainee accuracy for each of the gesture signs
Figure 6a The average trainee difficulty for the posture signs

Figure 6b The average trainee difficulty for the gesture signs
**Table 1** Trainees’ reason for difficulty with the movements. Starting by the Thumb, the fingers are called the Index, Middle, Ring and Pinkie.

<table>
<thead>
<tr>
<th>Sign No.</th>
<th>Name</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Letter ‘geem’</td>
<td>Thumb</td>
</tr>
<tr>
<td>7</td>
<td>Letter ‘heh’</td>
<td>Index to Pinkie finger</td>
</tr>
<tr>
<td>9</td>
<td>Letter ‘khaa’</td>
<td>Thumb up</td>
</tr>
<tr>
<td>12</td>
<td>Letter ‘meem’</td>
<td>Index finger</td>
</tr>
<tr>
<td>38</td>
<td>Number ‘200’</td>
<td>Right wrist Yaw</td>
</tr>
<tr>
<td>39</td>
<td>Number ‘300’</td>
<td>Wrist Yaw left + Thumb up</td>
</tr>
<tr>
<td>40</td>
<td>Number ‘400’</td>
<td>Wrist Yaw left + Thumb up</td>
</tr>
<tr>
<td>47</td>
<td>Word ‘talking’</td>
<td>Index + Middle fingers</td>
</tr>
<tr>
<td>48</td>
<td>Word ‘teaching’</td>
<td>Index to Pinkie finger</td>
</tr>
<tr>
<td>49</td>
<td>Word ‘welding’</td>
<td>Index finger</td>
</tr>
<tr>
<td>50</td>
<td>Word ‘dog’</td>
<td>Middle + Ring fingers</td>
</tr>
<tr>
<td>57</td>
<td>Word ‘arrow’</td>
<td>Wrist Yaw + Pinkie finger</td>
</tr>
</tbody>
</table>

**Table 2** Teacher and trainee responses to the system usability questionnaire.

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much experience have you had in the use of computers?</td>
<td>6.64±1.39</td>
</tr>
<tr>
<td>How easy did you find it to use the tutoring application?</td>
<td>8.42±0.94</td>
</tr>
<tr>
<td>How do you rate of the stability of the tutoring application?</td>
<td>9.57±0.65</td>
</tr>
<tr>
<td>How strongly would you recommend the use of the tutor system in preference to traditional sign language teaching techniques?</td>
<td>9.42±0.76</td>
</tr>
</tbody>
</table>