International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:8, 2008

3D Rendering of American Sign Language Finger-Spelling: A Comparative Study of Two Animation Techniques

Nicoletta Adamo-Villani

***Abstract*—**In this paper we report a study aimed at determining the most effective animation technique for representing ASL (American Sign Language) finger-spelling. Specifically, in the study we compare two commonly used 3D computer animation methods (keyframe animation and motion capture) in order to ascertain which technique produces the most 'accurate', 'readable', and 'close to actual signing' (i.e. realistic) rendering of ASL finger-spelling. To accomplish this goal we have developed 20 animated clips of finger-spelled words and we have designed an experiment consisting of a web survey with rating questions. 71 subjects ages 19-45 participated in the study. Results showed that recognition of the words was correlated with the method used to animate the signs. In particular, keyframe technique produced the most accurate representation of the signs (i.e., participants were more likely to identify the words correctly in keyframed sequences rather than in motion captured ones). Further, findings showed that the animation method had an effect on the reported scores for readability and closeness to actual signing; the estimated marginal mean readability and closeness was greater for keyframed signs than for motion captured signs. To our knowledge, this is the first study aimed at measuring and comparing accuracy, readability and realism of ASL animations produced with different techniques.

***Keywords*—**3D Animation, American Sign Language, Deaf Education, Motion Capture.

I. Introduction

COMPUTER animation of American Sign Language (ASL) has the potential to make digital content completely accessible to the Deaf because it provides a low-cost and effective means for adding sign language translation to any type of digital media. Like a video of a human ASL signer, computer animation technology allows for direct communication of ASL in a dynamic visual form that eliminates the need for closed captioning text, symbolic representations of the signs [1] or static sign images.

Manuscript received August 15th, 2008. This work was supported in part by NSF-RDE grant#0622900, by the College of Technology at Purdue University (grant#00006585), and by the Envision Center for Data Perceptualization at Purdue University.

Nicoletta Adamo-Villani is Associate Professor in the department of Computer Graphics Technology at Purdue University West Lafayette, IN (e-mail: nadamovi@purdue.edu).

The benefits of rendering sign language in the form of 3D animations have been investigated by several research groups [2-4] and commercial companies [5;6] during the past decade and the quality of 3D animation of ASL is slowly improving. However, its effectiveness and wide-spread use is still precluded by the low visual quality of the 3D virtual signers which result in low legibility of the signs and reduced appeal of the 3D avatars.

Three animation techniques are currently being used to render ASL signing: synthetic animation generated on-the-fly, keyframe animation, and motion capture. Although each method presents its own strengths and weaknesses, it is not clear which one produces the most accurate, intelligible, and realistic representation of the signs. The objective of the research work reported in the paper is to quantify the differences between ASL animation generated using keyframe technique and ASL animation produced using motion capture technology in order to determine the most effective method for rendering ASL signing. In particular, the study focuses on ASL finger-spelling which is a fundamental component of any sign language. Finger-spelling is essential for four reasons. It is used in combination with sign language for (1) names of people, (2) names of places, (3) words for which there are no signs and (4) as a substitute when the word has not yet been learned. Further, it is generally learned at the beginning of any course in sign language because the hand shapes formed in finger-spelling provide the basic hand shapes for most signs (see e.g. [7]).

To our knowledge, the work reported in the paper is the first study aimed at measuring and comparing realism and legibility of ASL 3D animations produced with different techniques. The higher goal of this research is to improve the realism and fluidity of ASL animation so that it can become an effective solution to the problem of deaf accessibility to digital media.

The paper is organized as follows. In section II we discuss the state-of-the-art in ASL animation and we give a brief description of keyframe animation and motion capture techniques. In section III we describe the user study and in section IV we report the findings. Discussion of results and conclusive remarks are included in section V.

II. Background

A. Animation of ASL

Unfortunately, current and past statistics show that many deaf students do not reach their potential [8]. Despite years of

trials with different teaching and communication methodologies, little progress has been made. There are several factors contributing to this problem: (1) A significant delay in deaf children's reading comprehension: 50% of students who are deaf leave high school with a reading level for English text that is below the fourth grade [9]. (2) The difficulty of (hearing) parents to convey in sign language basic concepts related to different school subjects (particularly math and science). (3) The inaccessibility to incidental learning (exposure to media in which concepts are practiced and reinforced). Deaf youngsters lack access to many sources of information (e.g., radio, conversations around the dinner table, educational software—the vast majority of educational software currently on the market requires the ability to speak or read/write English) and their incidental learning may suffer from this lack of opportunity [10]. Consequently, many concepts that hearing children learn incidentally in everyday life may have to be explicitly taught to deaf pupils in school, often through the use of an interpreter. Computer animation of American Sign Language (ASL) has the potential to remove many of the barriers to the education of deaf students because it provides a low-cost and effective means for adding sign language translation to any type of digital content.

Compared to video, animation technology has the following fundamental advantages. (a) 3-D animation offers great control over the visualization of the signs; the point of view of the virtual camera that renders the signing character and the location of the character in relation to the background can be optimized to enhance clarity. (b) The speed of the signing motion can be adjusted to the ASL proficiency of the user, of great importance, for instance, for children who are learning ASL. (c) Animated signs can be linked together smoothly, without abrupt jumps or collisions as would happen when concatenating video clips. (d) Animated signs can be decomposed into manual and non-manual components (i.e., hand motions and facial expressions) that serve as building blocks for creation of new sign animations. (e) Animations can be stored and transmitted remotely using only a small fraction of the storage and bandwidth costs of comparable video representations.

Moreover, some recent findings support the value of computer animation of ASL. The pioneering work in applying computer animation to ASL was carried out by Vcom3D [5]. Vcom3D products, now in use in over 30 school systems, have demonstrated the advantages of using three-dimensional animated characters that can communicate in ASL (and other variant sign languages) to provide multimedia access and increase English literacy for the Deaf. Data show improved reading comprehension as a result of using these products.

Three research groups in the USA are currently engaged in research and development of computer animation of ASL for improving deaf accessibility to educational content: Vcom3D, Stephen Austin State University, and Purdue University.

Vcom3D ASL animation is based on a system that translates high-level commands from an external application into character gestures and facial expressions which can be composed in real-time to form sequences of signs. Many signers argue that the main problem with fully synthetic animation of ASL is its inability to capture the nuances typically portrayed by skilled ASL users. While synthetic animation can approximate sentences produced by ASL signers, the individual handshapes and rhythm of signing are often unnatural, and the facial expressions do not convey meanings as clearly as a live signer [11].

Researchers at Stephen Austin State University, TX have recently begun investigating the advantages, if any, of ASL motion captured animations over videos of real signers. Specifically, the research group has created animations of science words and has evaluated them with a group of middle school and high school deaf students. Evaluation findings showed that students learned the ASL signs equally well using videos of signers and 3D motion captured animations, but had more difficulty re-producing the signs after using the 3D animations [11].

Purdue University Animated Sign Language Research Group [12] is focusing on research, development, and evaluation of innovative 3D animation-based interactive tools to improve K-6 math/science education for the Deaf. The research team is currently working on two projects: Mathsigner™ and SMILE™, funded by the National Science Foundation [13-14]. The animation of the ASL signs in Mathsigner™ and SMILE™ is based on the use of keyframe animation and motion capture technology, and on the development of a new blending algorithm which allows for creation of more natural transitions between individual signs [15]. In a survey of adult and children signer reactions to VcomD SigningAvatar® and a prototype of Mathsigner™, Mathsigner™ was rated significantly better on readability, fluidity, and timing, and equally good on realism [16].

Despite the substantial amount of ASL animation research, development and recent improvements, several limitations still exist that preclude animation of ASL from becoming an effective, general solution to deaf accessibility to digital media. One of the main problems is the low realism of the signing characters which is due primarily to non-natural handshapes, non-fluid motions across signs (i.e. movement epenthesis), lack of adherence to fundamental animation principles, and simplistic animation of facial expressions.The author believes that the first step toward improving the quality of ASL animation is to determine which animation technique (or combination of techniques) is the most efficient at producing realistic representations of the signing motions.

B. Keyframe Animation

Keyframe animation is a technique that was developed at the Walt Disney studios to produce animation frames more efficiently. "First an experienced animator would draw the most important, or "key", frames of an animation sequence. Then, the less experienced animators would draw the in-between frames that fell between the master animator's keyframes" [17].

Almost all 3D animation software is based on this keyframing approach. In a computer animation system the animator sets 'key' values to various objects' parameters (for instance, the rotations of the fingers or the position of the hand in space) and saves these values at particular points in the timeframe; this process is called 'setting keyframes'. After the animator has defined the keyframes, the 3-D software

interpolates the values of the object's parameters between the keyframes therefore generating the in-between frames, and, thus, the motion. To gain more control of the interpolation, a parameter curve editor is available in the majority of the 3D animation packages. The parameter curve editor shows a graphical representation of the variation of a parameter's value over time (the animation curve). Altering the shape of the curve results into a change in interpolation and, therefore, into a change in the speed of motion. For example, by changing the interpolation it is possible to avoid surface interpenetration (such as fingers intersecting each other) when transitioning from one hand shape to the next. The realism of keyframe animations depends largely on the animator's ability to set believable keyframes (for instance, realistic hand shapes) and on her ability to control the interpolation between the keyframes (i.e., the speed and fluidity of motion).

C. Motion Capture Animation

Motion Capture animation, also referred to as performance animation or digital puppetry, "..involves measuring an actor's (or object) position and orientation in physical space and recording that information in a computer-usable form" [18]. In general the position or orientation of the actor is measured by a collection of input devices (optical markers or sensors) attached to the actor's body. Each input device has 3 DOF and produces 3D rotational or translational data which are channeled to the joints of a virtual character. As the actor moves, the input devices send data to the computer model. These data are used to control the movements of the character in real time, or to generate the animation curves. Motion capture animation is often used when the animation of the 3D character needs to match the performance of the actor very precisely.

Different motion capture systems exist, such as mechanical, optical and electromagnetic. The motion capture devices used to generate the animations for this study included a Metamotion 19-marker optical system with a 6-camera set up and a pair of Immersion 18-sensor cybergloves. The optical system was used to record the signer's body motions (such as arm/shoulder/hand movements and waist/torso rotations); the cybergloves allowed for recording of finger motions and wrist rotations. Fig. 1 shows the system setup and the signer wearing the optical suit and cybergloves.

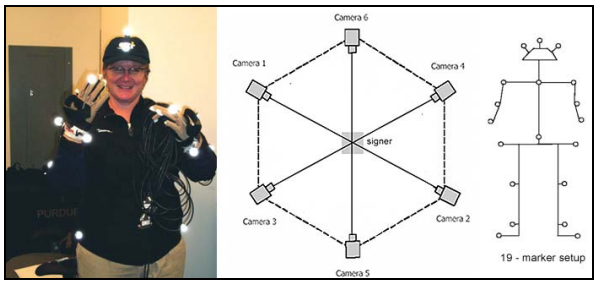

Fig. 1 From the left: signer with optical suit and cybergloves; diagram illustrating the 6-camera setup; diagram illustrating the 19-marker setup for the body

III. Description of the Study

The goal of the study was to measure and compare the *accuracy*, *readability* and *closeness to real finger-spelling* of 20 animation clips of finger-spelled words. 10 clips were produced using keyframe animation and 10 clips were produced with motion capture technology. Accuracy was measured by recognition (accuracy=1) or non-recognition of the word (accuracy=0). Readability and closeness were measured using a 10-point Likert Scale.

A. Materials

The Animation Clips

20 animation clips (10 keyframed + 10 motion captured) of 10 fingerspelled words. The words were: "apple", "cracker", "drain", "fruit", "heavy", "juice", "milk", "Nicoletta", "queen", and "zebra". The words were selected by a signer with experience in ASL. The choice was motivated by the following factors: (1) the words included almost all the letters of the manual alphabet (20/26); (2) two words included dynamic handshapes (i.e., "Z' as in "zebra" and "J" as in "juice"); (3) three words presented double letters (i.e., "apple", "nicoletta" and "queen"); and (4) several words presented challenging transitions between handshapes that could easily lead to surface interpenetration (i.e., the transition between "e" and "t" as in "Nicoletta" and between "v" and "y" as in "heavy").

The keyframed clips were produced using Autodesk Maya 8.5 software. The character (shown in Fig. 2) is a cartoon bunny designed, modeled and rigged by a senior student in the computer graphics technology department at Purdue University. The model of the virtual signer was built as a continuous polygonal mesh with a polygon count = 7020 polygons. The avatar was rigged with a body skeletal deformation system consisting of 35 joints attached to the surface with smooth skinning. The rig of the hands included 22 joints and 26 DOF (degrees of freedom) per hand. The character's body was animated using a combination of forward and inverse kinematics; the hands were animated using forward kinematics. Slow-in/slow-out was used as the initial interpolation between keyframes, however the animation curves were manually edited by the animator to represent subtle variations in the speed of motion, and to avoid intersection of surfaces. Videos of a signer finger-spelling each word were used as reference footage for the creation of the clips.

The motion captured clips were produced by recording the signing motion of each word directly from an experienced ASL signer; the motion was sampled 50 times per second. The collected motion data were edited using Autodesk Motion Builder 7.5 software and were applied to the same 3D avatar. The mocap rig of the character included the same number of skeletal joints as the one used for the keyframe animations. All animations were rendered to a resolution of 640x480 pixels using Maya software rendering algorithm; the lighting setup and camera position were maintained the same in all clips. The final animation sequences were output as Quick Time movies with Sorensen3 compression and a frame rate of 30 fps.

The Web Survey

The web survey consisted of 6 screens per animated clip (represented in Fig. 1), with a total of 120 screens (6x20). The animated sequences were presented in random order and each animation was assigned a random number. Data collection was embedded in the survey; in other words, a program running in the background recorded all subjects responses and stored them in an excel spreadsheet. The web survey can be accessed at: http://www.signedmotion.com/testing/

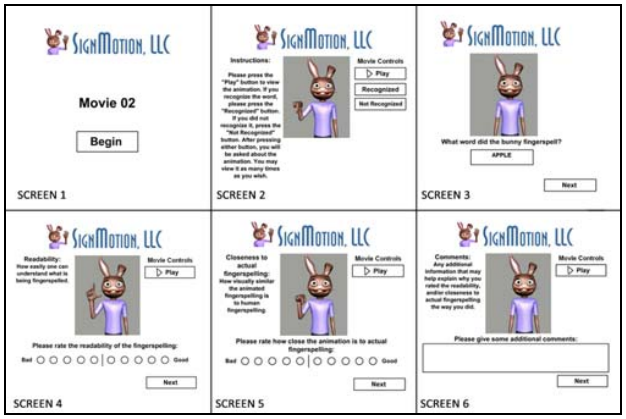

Fig. 2 The 6 screens of the web survey

B. Participants

71 subjects ages 19-45 with ASL skills. Subjects were recruited from the American Sign Language Club at Purdue University and from the department of Speech, Language and Hearing Sciences at Purdue University.

C. Procedure

Subjects were sent an email containing a brief summary of the research and its objectives, an invitation to participate in the study, and the http address of the web survey. Participants completed the on-line survey using their own computers and the survey remained active for 2 weeks. It was structured in the following way: For each clip, subjects were asked to (1) start the experiment by clicking on the "begin button" (screen 1, Fig. 2); (2) play the animation clip and input whether they recognized or did not recognize the word (screen 2, Fig. 2)- subjects were allowed to playback the clip as many times as needed; (3) enter the word in the text box, if recognized, or leave the text box blank, if not recognized (Screen 3, Fig. 2); (4) rate the readability of the animation using a 10-point Likert scale (Screen 4, Fig. 2); (5) rate the closeness to actual finger-spelling using a 10-point Likert scale (Screen 5, Fig. 2); and (6) enter additional comments (Screen 6, Fig. 2).

IV. Findings

All of the subjects' responses were collected in a Microsoft Excel spreadsheet and the statistical analyses were performed using SPSS v. 15 for Windows. The first factor that we evaluated was the number of correct and incorrect responses for keyframed signs as compared to the number of correct and incorrect responses for the motion captured signs. A simple cross-tabulation with all responses being either correct or incorrect revealed that the count for correct responses was 633 for the keyframed signs and 577 for the motion captured signs, while the count for incorrect responses was 87 for keyframed signs and 143 for motion captured signs. Fig. 3 shows a chart of the findings. A chi-squared test for independence yielded a very low p-value of 5.62E-05, confirming that the number of correct answers is correlated with the method used to animate the signs.

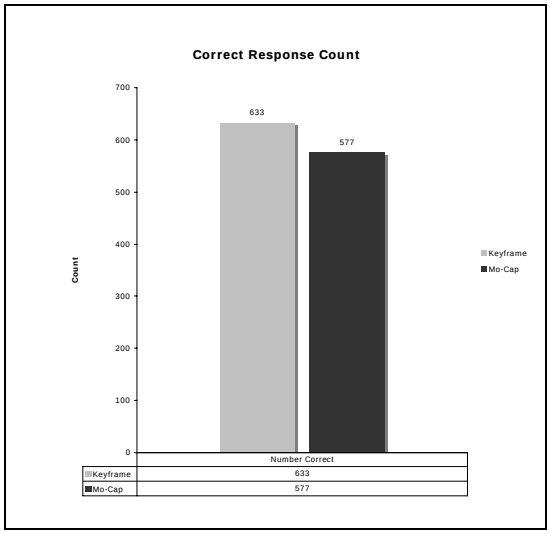

Fig. 3 Correct response count for keyframed (gray) and motion captured (black) animations

Chi-squared tests were used to determine the subject-dependence and word-dependence of the correct response count. The p-values for independence of the two tests were 4.91E-03 and 3.61E-97 respectively, confirming that the number of correct responses was also correlated with the subject and with the word presented to the subject. To account for the impact of subject and word as well as the method used, we ran a logistic regression test using all three variables. The logistic regression test revealed a p-value of 7.90E-08 for the method used, confirming correlation between the method and the number of correct responses. Furthermore, the coefficient values for the methods used to generate signs were 1.135 for keyframed signs with motion capture being the baseline value at 0, indicating that a correct response is more likely with keyframed signs than with motion captured signs.

The "readability" and "closeness" values reported by the subjects were analyzed using univariate ANOVAs. For each of the two variables, method of creation for the sign animation (keyframe or motion capture) was the factor taken into account, while word and subject were treated as block factors. The p-value for the method used was well below .05 in both tests, indicating that the method of creation did have an effect on the reported scores for closeness and readability. For each subject-word combination, we calculated the estimated marginal mean for the closeness score and the readability score for keyframed and motion captured signs. Without exception, the estimated marginal mean readability and

closeness was greater for keyframed signs than for motion captured signs in each subject-word combination. Figs. 4-7 show the resulting readability and closeness charts for the words "apple" and "zebra".

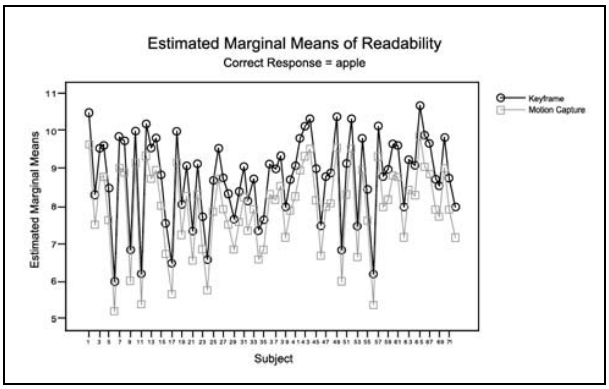

Fig. 4 Estimated Marginal Means of "Readability" for the animation of the word "apple". Values for keyframe technique are represented in black; values for motion capture technique are represented in gray

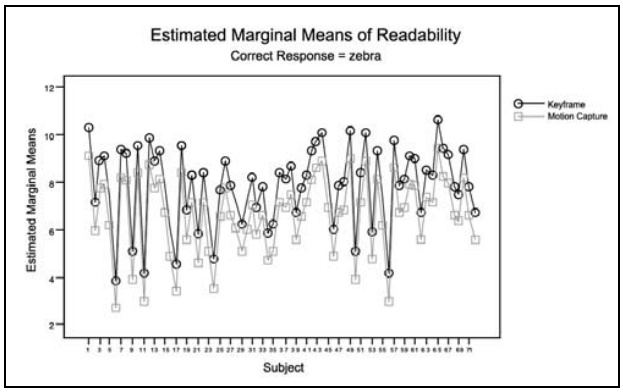

Fig. 5 Estimated Marginal Means of "Readability" for the animation of the word "zebra". Values for keyframe technique are represented in black; values for motion capture technique are represented in gray

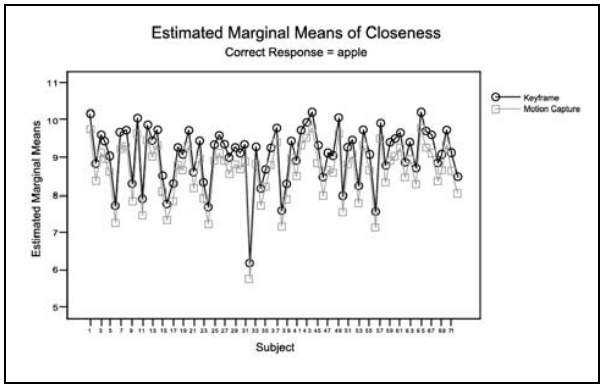

Fig. 6 Estimated Marginal Means of "Closeness to real fingerspelling" for the animation of the word "apple". Values for keyframe technique are represented in black; values for motion capture technique are represented in gray

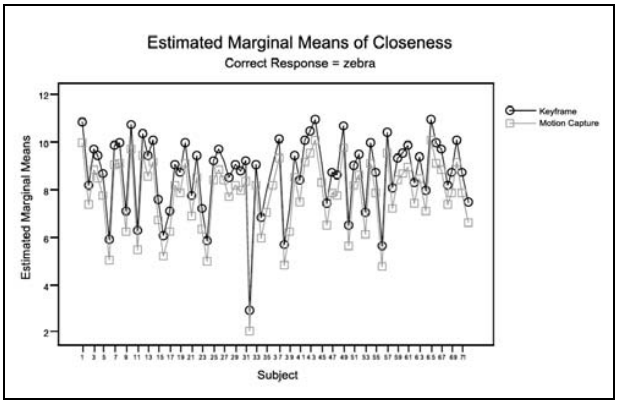

Fig. 7 Estimated Marginal Means of "Closeness to real fingerspelling" for the animation of the word "zebra". Values for keyframe technique are represented in black; values for motion capture technique are represented in gray

V. Discussion and Conclusion

In this paper we reported a user study aimed at measuring and comparing accuracy, readability and realism of ASL animations produced with two different methods (keyframe and motion capture). Findings showed that keyframe animation produces a more accurate, legible and close to real signing rendering of the signs than motion capture technique. These results might appear surprising if we consider that motion capture technology allows to measure and record motion data directly from an actor (a signer, in our case), and, therefore, to match the animation of the 3D character to the performance of the actor very precisely. Whereas keyframe animation does not involve any objective measure of the signer's motions and its life-like quality depends, almost entirely, on the animator's subjective ability to manually re-create the actor's gestures and movements.

The author believes that the lower ratings of motion capture animations can be explained by the following considerations:

1. Although the cybergloves allow for direct measurement of hand and finger movements (i.e., joint angles, wrist rotation and 3D spatial information), the recorded data can be imprecise due to system calibration inaccuracies, or to differences between the size of the signer's hands and the size of the gloves. In our case, there was a slight difference between the actor's hands' size and the gloves'.

2. In general, the recorded motions need to be edited when they are applied to a 3D avatar whose geometric proportions differ from the ones of the actor. In our case some editing was required in order to fit the signer's movements to the skeletal structure of the Bunny character.

3. The Motion Capture system captures all of the performer's motions, including secondary movements of the body such as shoulder and body jitters. While these subtle motions might make the animation more life-like and believable, they may take the viewers' attention away from the signing motion. Some of the participants' comments revealed that these secondary motions were often perceived as distracting; for instance, one of the subjects wrote "… there is

too much jitter in the avatar's body…it detracts from the clarity of the signs..".

4. Because the motion capture system captures the signer's motion precisely, it also captures the nuances of the signer's signing style (equivalent to a speaker's accent). The difference between keyframe animation and motion capture animation can be considered analogous to the difference between synthesized speech and recorded speech. While recorded speech sounds more natural, it is not necessarily more understandable, especially if the person speaking has an accent. On the contrary, in synthesized speech the words are pronounced correctly and therefore they can be easily understood, although they might sound unnatural and robotic.

While problem 3 could be easily solved by removing the actor's secondary motions, problems 1, 2 and 4 are likely to occur each time motion capture technology is used to create the signing animation. Therefore, currently, keyframe animation appears to be a more effective method for ASL rendering. However, the author believes that as mocap technology matures and gets perfected, many of the above mentioned problems will be overcome.

ACKNOWLEDGMENT

The author would like to thank Jesse Janowiak of Purdue Computer Graphics Department and Purdue Statistical Consulting Department for their help with the statistical analysis.

REFERENCES

[1] SignWriting. Available at: http://www.signwriting.org/
[2] N. Adamo-Villani, J. Doublestein, & Z. Martin, Sign Language for K-8 mathematics by 3D interactive animation. *Journal of Educational Technology Systems*, vol. 33, issue 3, 2005, pp. 241-257.
[3] S. Whitney, Adventure Games as Teaching Tools for Deaf and Hard of Hearing students. *Journal of Border Education Research* (In print)
[4] J. Vesel, Signing Science. *Learning & Leading with Technology*, vol. 32, issue 8, 2005, pp. 30-31.
[5] Vcom3D. Available at: http://www.vcom3d.com/
[6] eSIGN. Available at: http://www.visicast.cmp.uea.ac.uk/eSIGN/index.html
[7] M. Flodin, *Signing Illustrated.* Perigee Reference – The Berkley Publishing Group: New York, 1994.
[8] S. Burgstahler, Increasing the representation of people with disabilities in science. *Journal of Information Technology and Disability*, Vol. 24, No. 4, 1994.
[9] E. Sims, SigningAvatars. *Final Report for SBIR Phase II Project*, U.S. Department of Education, 2000.
[10] I. Rapin, Helping Deaf Children Acquire Language: Lessons from the Past. *International Journal of Pediatric Otorhinolaryngology*, vol. 11, 1986, pp. 213-223.
[11] S. Whitney, Mocap ASL for the Sciences. National Science Foundation RDE-DEI: award #HRD-0435679B, 2004.
[12] N. Adamo-Villani, & R. Wilbur, Novel approaches to deaf education. *Proc. of NSF First International Conference on Technology-Based Learning with Disability (LWD-07)*, OH, pp. 13-21, 2007.
[13] Mathsigner™. Available at: http://www2.tech.purdue.edu/cgt/i3/
[14] SMILE™. Available at: http://www2.tech.purdue.edu/cgt/i3/smile/
[15] N. Adamo-Villani, G. Beni & R. Wilbur, US patent application: "Interactive Animation System for Sign Language" 64145-00-US 12264-84 (filed on September 1st 2005).
[16] R. Wilbur, & N. Adamo-Villani, Software for Math Education for the Deaf. National Science Foundation RDE-FRI: award #0622900, 2006.
[17] M. O'Rourke, *Principles of Three Dimensional Computer Animation.* W. W. Norton & Company: New York, 2003.
[18] S. Dyer, J. Martin & J. Zulauf, Motion Capture White Paper. 12 Dec 1995. Available at: http://web.mit.edu/comm-forum/papers/furniss.html#5