The Effect of Rendering Style on Perception of Sign Language Animations

Tiffany Jen and Nicoletta Adamo-Villani$^{(\boxtimes)}$

Purdue University, West Lafayette, USA
tiffanyjenny66@gmail.com, nadamovi@purdue.edu

Abstract. The goal of the study reported in the paper was to determine whether rendering style (non-photorealistic versus realistic) has an effect on perception of American Sign Language (ASL) finger spelling animations. Sixty-nine (69) subjects participated in the experiment; all subjects were ASL users. The participants were asked to watch forty (40) sign language animation clips representing twenty (20) finger spelled words. Twenty (20) clips were rendered using photorealistic rendering style, whereas the other twenty (20) were rendered in a non-photorealistic rendering style (e.g. cel shading). After viewing each animation, subjects were asked to type the word being finger-spelled and rate its legibility. Findings show that rendering style has an effect on perception of the signed words. Subjects were able to recognize the animated words rendered with cel shading with higher level of accuracy, and the legibility ratings of the cel shaded animations were consistently higher across subjects.

Keywords: Sign language · Animation · Non photorealistic rendering · Cel shading · Deaf education

1 Introduction

Computer animation provides a low-cost and effective means for adding signed translation to any type of digital content. Despite the substantial amount of ASL animation research, development and recent improvements, several limitations still preclude animation of ASL from becoming an effective, general solution to deaf accessibility to digital media. One of the main challenges is low rendering quality of the signed animations, which results in limited legibility of the animated signs.

This paper investigates the problem of clearly communicating the ASL handshapes to viewers. With standard rendering methods, based on local lighting or global illumination, it may be difficult to clearly depict palm and fingers positions because of occlusion problems and lack of contour lines that can help clarify the palm/fingers configuration. "*Interactive systems that permit the lighting and/or view to be controlled by the user allow for better exploration, but non-photorealistic methods can be used to increase the amount of information conveyed by a single view*" [1]. In fact, it is common for artists of technical drawings

or medical illustrations to depict surfaces in a way that is inconsistent with any physically-realizable lighting model, but that is specifically intended to bring out surface shape and detail [1].

The specific objective of this experiment was to answer the research question of whether the implementation of a particular non-photorealistic rendering style (e.g., cel shading) in ASL fingerspelling animations can improve their legibility. The paper is organized as follows. In Sect. 2 (Background) we discuss computer animation of sign language, we define cel shading, and we explain the importance of ASL fingerspelling. In Sect. 3 (Study Design) we describe the user study, and in Sect. 3.5 (Findings) we report and discuss the results. Conclusion and future work are included in Sect. 4 (Conclusion).

2   Background

2.1   Computer Animation of Sign Language

Compared to video, animation technology has two fundamental advantages. The first one is scalability. Animated signs are powerful building blocks that can be concatenated seamlessly using automatically computed transitions to create new ASL discourse. By comparison, concatenating ASL video clips suffers from visual discontinuity. The second advantage is flexibility. Animation parameters can be adjusted to optimize ASL eloquence. For example, the speed of signing can be adjusted to the ASL proficiency of the user, which is of great importance for children who are learning ASL. The signing character can be easily changed by selecting a different avatar, hence the possibility of creating characters of different age and ethnicity, as well as cartoon characters appealing to young children.

Several groups have been focusing on research, development and application of computer animation technology for enhancing deaf accessibility to educational content. The ViSiCAST project [2], later continued as eSIGN project [3], aims to provide deaf citizens with improved access to services, facilities, and education through animated British Sign Language. The project is developing a method for automatic translation from natural-language to sign-language. The signs are rendered with the help of a signing avatar. A website is made accessible to a deaf user by enhancing the website's textual content with an animated signed translation encoded as a series of commands. Vcom3D commercializes software for creating and adding computer animated ASL translation to media [5,6]. The SigningAvatar®software system uses animated 3-D characters to communicate in sign language with facial expressions. It has a database of 3,500 English words/concepts and 24 facial configurations, and it can fingerspell words that are not in the database.

TERC [5,6] collaborated with Vcom3D and the National Technical Institute for the Deaf (NTID) on the use of SigningAvatar software to annotate the web activities and resources for two Kids Network units. Recently, TERC has developed a Signing Science Dictionary (SSD) [7,8]. Both the Kids Network units and the science dictionary benefit deaf children confirming again the

value of animated ASL. Purdue University Animated Sign Language Research Group, in collaboration with the Indiana School for the Deaf (ISD), is focusing on research, development, and evaluation of 3-D animation-based interactive tools for improving math and science education for the Deaf. The group developed Mathsigner, a collection of animated math activities for deaf children in grades K-4, and SMILE, an educational math and science immersive game featuring signing avatars [9,10].

Many research efforts target automated translation from text to sign language animation to give signers with low reading proficiency access to written information in contexts such as education and internet usage. In the U.S., English to ASL translation research systems include those developed by Zhao et al. [11], Grieve-Smith [12] and continued by Huenerfauth [13]. To improve the realism and intelligibility of ASL animation, Huenerfurth is using a data-driven approach based on corpora of ASL collected from native signers [14]. In France, Delorme et al. [15] are working on automatic generation of animated French Sign Language using two systems: one that allows pre-computed animations to be replayed, concatenated and co-articulated (OCTOPUS) and one (GeneALS) that builds isolated signs from symbolic descriptions. Gibet et al. [16] are using data-driven animation for communication between humans and avatars. The Signcom project incorporates an example of a fully data-driven virtual signer, aimed at improving the quality of real-time interaction between humans and avatars. In Germany, Kipp et al. [17] are working on intelligent embodied agents, multimodal corpora and sign language synthesis. Recently, they conducted a study with small groups of deaf participants to investigate how the deaf community sees the potential of signing avatars. Findings from their study showed generally positive feedback regarding acceptability of signing avatars; the main criticism on existing avatars primarily targeted their low visual quality, the lack of non-manual components (facial expression, full body motion) and emotional expression. In Italy, Lesmo et al. [18] and Lombardo et al. [19] are working on project ATLAS (Automatic Translation into the Language of Sign) whose goal is the translation from Italian into Italian Sign Language represented by an animated avatar. The avatar takes as input a symbolic representation of a sign language sentence and produces the corresponding animations; the project is currently limited to weather news.

Despite the substantial amount of ASL animation research, development and recent improvements, several limitations still preclude animation of ASL from becoming an effective, general solution to deaf accessibility to digital media. One of the main problems is low visual quality of the signing avatars due to unnatural motions and low rendering quality of the signed animations, which results in limited legibility of the animated signs.

2.2   Rendering of Sign Language Animations

The visual quality of the ASL visualization depends in part on the underlying rendering algorithm that takes digital representations of surface geometry, color, lights, and motions as input and computes the frames of the animation. With
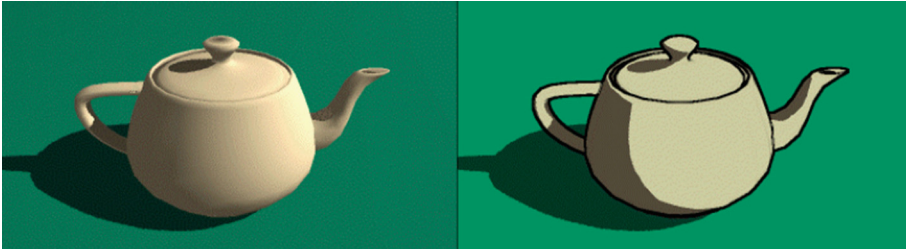
Fig. 1. Teapot model rendered by a photo-realistic rendering algorithm (left) and by a cel shading algorithm (right) [20]

photorealistic rendering methods, based on local lighting or global illumination, it may be difficult to clearly depict palm and fingers positions because of occlusion problems and lack of contour lines that can help clarify the palm/fingers configuration. Non-photorealistic methods, such as cel shading, could be used to increase the amount of information conveyed by a single view.

Cel shading is a type of non-photorealistic rendering in which an image is rendered by a computer to have a "toon" look that simulates a traditional hand-drawn cartoon cel. The toon appearance of a cel image is characterized by the use of areas selectively colored with a fill, a highlight, shading, and/or a shadow colors. Contour lines can be used to further define the shape of an object and color lines may be used to define the different color areas in the colored image. The contrast of the color lines and the thickness of the contour lines can be adjusted to improve clarity of communication. The type of cel shading used in this study produced images that have a stylized hand drawn look, with constant-size outlines and uniformly colored areas. Figure 1 shows a simple 3D model rendered by a photo-realistic rendering algorithm and by a cel shading algorithm with contour lines, one level of shading and shadows.

2.3   ASL Finger Spelling

Learning finger spelling is important, as it is very difficult to become fluent in ASL without mastering fingerspelling. 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. It is generally learned at the beginning of any course in sign language also because the hand shapes formed in finger spelling provide the basic hand shapes for most signs [21]. In spite of its importance and its apparent simplicity, high fluency in finger spelling is not easy to acquire, mainly for the reasons outlined. Achieving fingerspelling proficiency requires the visual comprehension of the manual representation of letters and one reason students experience difficulty in fingerspelling recognition is its high rate of handshape presentation. Most signs in ASL use no more than two hand shapes [22], but fingerspelling often uses as many handshapes as there are letters in a word.

3   Study Design

The objective of the study was to determine whether cel shading allowed the subjects to better recognize the word being signed to them. The independent variable for the experiment was the implementation of cel shading in ASL animations. The dependent variables were the ability of the participants to understand the signs, and their perception of the legibility of the finger-spelled words. The null hypothesis of the experiment was that the implementation of cel shading in ASL animations has no effect on the subjects' ability to understand the animations presented to them and on the perception of their legibility.

3.1   Subjects

Sixty-nine (69) subjects age 19–64, thirty-five (35) Deaf, thirteen (13) Hard-of-Hearing, and twenty-one (21) Hearing, participated in the study; all subjects were ASL users. Participants were recruited from the Purdue ASL club and through one of the subject's ASL blog (johnlestina.blogspot.com/). The original pool included 78 subjects, however 9 participants were excluded from the study because of their limited ASL experience (less than 2 years). None of the subjects had color blindness, blindness, or other visual impairments.

3.2   Stimuli Animations

Forty animation clips were used in this test. The animations had a resolution of 640×480 pixels and were output to Quick Time format with Sorensen 3 compression and a frame rate of 30 fps. Twenty animation clips were rendered with cel shading and twenty animation clips were rendered with photorealistic rendering with ambient occlusion. Both sets of animations represented the same 20 finger-spelled words. Camera angles and lighting conditions were kept identical for all animations. The animations were created and rendered in Maya 2014 using Mental Ray. Figure 2 shows a screenshot of one of the animations in Maya; Fig. 3 shows 4 frames extracted from the photorealistic rendering animation and 4 frames extracted from the cel shaded animation.

3.3   Web Survey

The web survey consisted of 1 screen per animated clip with a total of 40 screens (2×20). Each screen included the animated clip, a text box in which the participant entered the finger-spelled word, and a 5-point Likert scale rating question on perceived legibility (1 = high legibility; 5 = low legibility). 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 also included a demographics questionnaire with questions on subjects' age, gender, hearing status and experience in ASL.
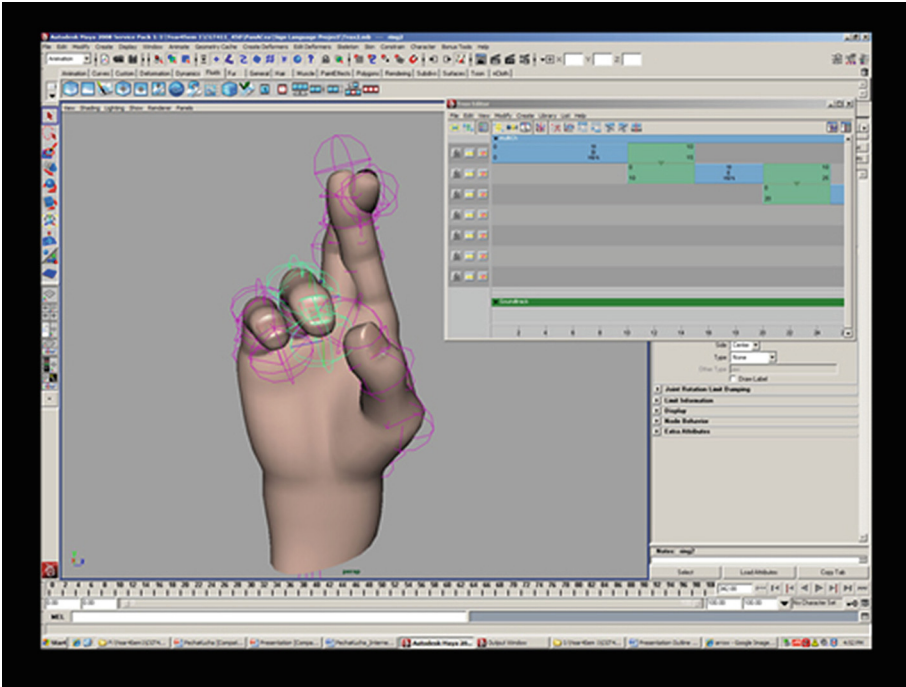
Fig. 2. The twenty words shown in the animations were: "cracker," "heavy," "can," "drain," "fruit," "milk," "Kyle," "child," "movie," "awesome," "axe," "bear," "voyage," "kiosk," "wild," "adult," "year," "duck," "love," and "color." The words were selected by a signer with experience in ASL. The choice was motivated by two factors. The words include almost all the letters of the manual alphabet (20/26), and the majority of these words present challenging transitions between hand-shapes. Since finger-spelling does not rely on facial expressions or body movements, the animations showed only the right hand.

3.4   Procedure

Subjects were sent an email containing a brief summary of the research and its objectives (as specified in the approved IRB documents), 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: the animation clips were presented in randomized order and for each clip, subjects were asked to (1) view the animation; (2) enter the word in the text box, if recognized, or leave the text box blank, if not recognized; (3) rate the legibility of the animation. At the end of the survey, participants were asked to fill out the demographics questionnaire.
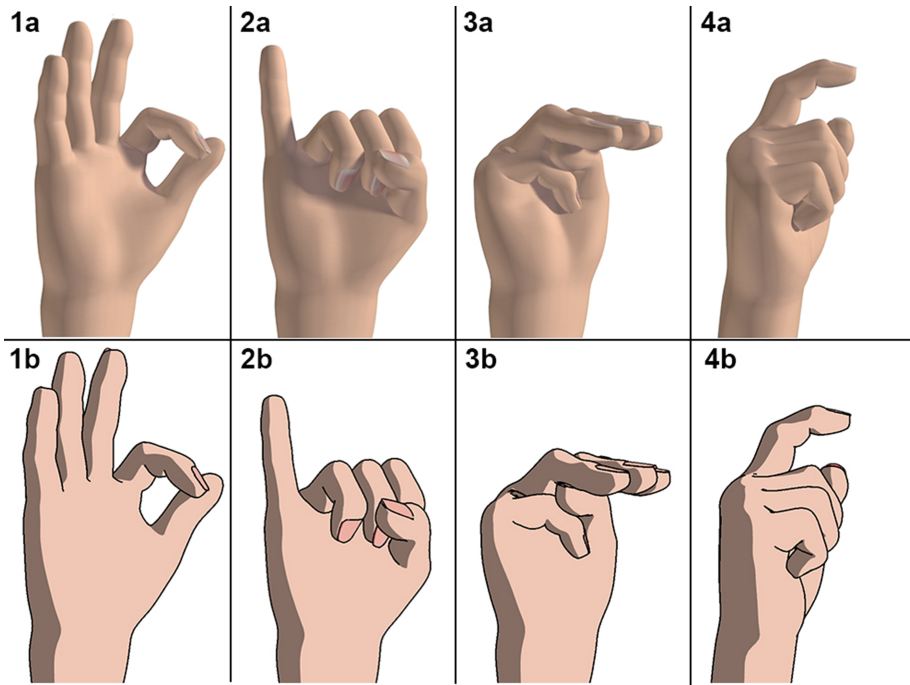
Fig. 3. Handshapes rendered with photorealistic rendering method (top, 1a–4a); handshapes rendered with cel shading (bottom, 1b–4b)

3.5   Findings

For the analysis of the subjects' legibility ratings a paired sample T test was used. With twenty pairs of words for each subject, there were a total of 1,380 rating pairs. The mean of the ratings for animations rendered with photorealistic rendering was 2.21, and the mean of the ratings for animations rendered with cel shading was 2.12. Using the statistical software SPSS, a probability value of .048 was calculated. At an alpha level of .05, the null hypothesis that cel shading had no effect on the user's perceived clarity of the animation was therefore rejected. Perceived legibility was significantly higher for the cel shaded animations than for the photorealistic rendered animations. Figure 4 shows the breakdown of the subjects' ratings of the animations.

For the analysis of the ability of the subjects to recognize the words, the McNemar test, a variation of the chi-square analysis, was used. Using SPSS once again, a probability value of .002 was calculated. At an alpha level of .05, a relationship between cel shading and photorealistic rendering and the subject's ability to identify the word being signed was determined. Word recognition was higher with cel shading across all subjects.

Two extraneous variables that were not considered during the design phase were revealed by the feedback provided by the subjects at the end of the survey:
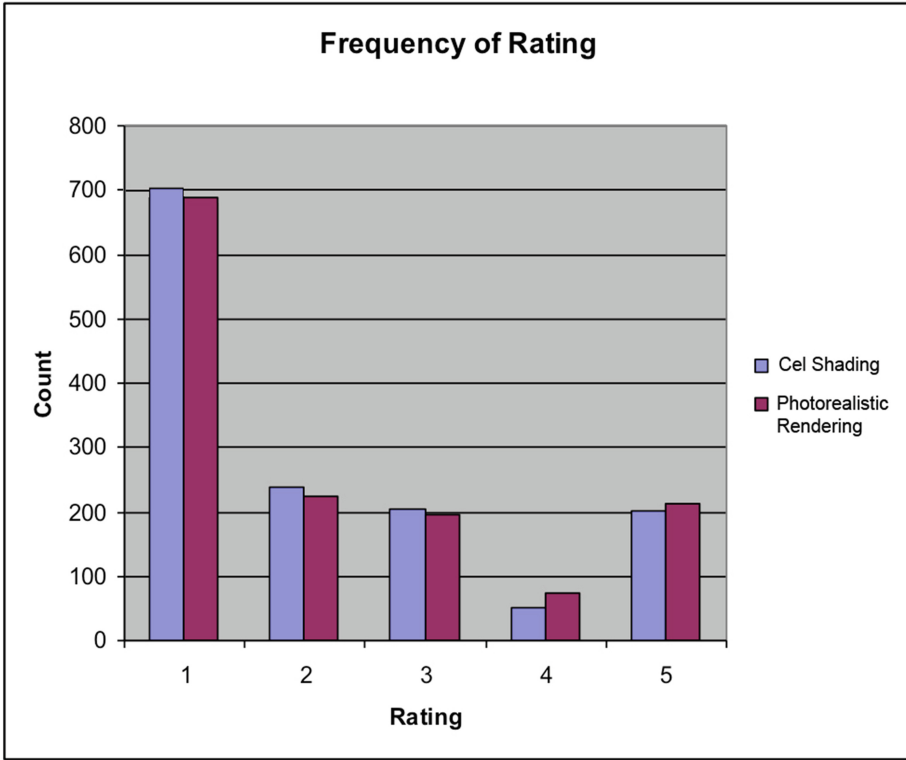
Fig. 4. A graph showing the breakdown of the subjects' ratings of the animations. Lower ratings, which indicate higher legibility, were more frequent with cel shaded animations, whereas higher ratings were more common for animations with photorealistic rendering.

(1) variation in subjects' computer screen resolution and (2) variation in subjects' internet connection speed. (1) Some subjects had a low screen resolution, which forced them to scroll down to see each animation. This might have caused the subjects to miss a part of the word being signed. (2) Since the survey was posted online, connection speed was also a problem. Several subjects mentioned that the animations were choppy and jumpy at times and caused them to miss some letters. In both cases, since the results from the survey were being compared within the subjects, that is, one subject's responses in one category were being compared to his/her responses in the other category, those extraneous variables did not have a substantial impact on the results.

4   Conclusion

In this paper we have reported a user study that aimed to determine whether rendering style has an effect on subjects' perception of ASL fingerspelling ani-

mations. Findings from the study confirmed our hypothesis: rendering style has an effect and non-photorealistic rendering (specifically, cel shading) improves subjects' recognition of the finger-spelled words and perceived legibility of the animated signs. Although the study produced significant results, it was limited to ASL finger spelling and the animations showed only the 3D model of the right hand. In future work we will extend the study to full-body avatars and complex 2-handed signs that involve body movements and facial expressions. As mentioned in the introduction, the authors believe that sign language animation has the potential to improve deaf accessibility to digital content significantly. The overall goal of this study, and other previous studies [23–25], is to advance the state-of-the-art in sign language animation by improving its visual quality, and hence its clarity, realism and appeal.

References

1. Rusinkiewiz, S., Burns, M., De Carlo, D.: Exaggerated shading for depicting shape and detail. ACM Trans. Graph. **25**(3), 1199–1205 (2006). (Proc. SIGGRAPH)
2. Elliott, R., Glauert, J.R.W., Kennaway, J.R., Marshall, I.: The development of language processing support for the ViSiCAST project. In: Proceedings of ASSETS 2000, pp. 101–108 (2000)
3. eSIGN (2003). http://www.visicast.cmp.uea.ac.uk/eSIGN/index.html
4. Vcom3D (2007). http://www.vcom3d.com
5. Sims, E.: SigningAvatars. In: Final Report for SBIR Phase II Project, U.S. Department of Education (2000)
6. TERC (2006). http://www.terc.edu/
7. Vesel, J.: Signing science. Learn. Lead. Technol. **32**(8), 30–31 (2005)
8. Signing Science (2007). http://signsci.terc.edu/
9. Adamo-Villani, N., Wilbur, R.: Two novel technologies for accessible math and science education. IEEE Multimedia **15**(4), 38–46 (2008). Special Issue on Accessibility
10. Adamo-Villani, N., Wilbur, R.: Software for math and science education for the deaf. Disabil. Rehabil. Assistive Technol. **5**(2), 115–124 (2010)
11. Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N.I., Palmer, M.: A machine translation system from English to American sign language. In: White, J.S. (ed.) AMTA 2000. LNCS (LNAI), vol. 1934, pp. 54–67. Springer, Heidelberg (2000)
12. Grieve-Smith, A.: SignSynth: a sign language synthesis application using web3D and perl. In: Wachsmuth, I., Sowa, T. (eds.) Gesture and Sign Language in Human-Computer Interaction. LNCS, vol. 2298, pp. 134–145. Springer-Verlag, Berlin (2002)
13. Huenerfauth, M.: A multi-path architecture for machine translation of English text into American sign language animation. In: Student Workshop at the Human Language Technology Conference/North American Chapter of the Association for Computational Linguistics (HLTNAACL) (2004)
14. Huenerfauth, M.: Cyclic data-driven research on American sign language animation. In: Proceedings of SLTAT 2011, University of Dundee, UK (2011)

15. Delorme, M. Braffort, A., Filhol, M.: Automatic generation of French sign language. In: Proceedings of SLTAT 2011, University of Dundee, UK (2011)
16. Gibet, S., Courty, N., Duarte, K.: Signing avatars: linguistic and computer animation challenges. In: Proceedings of SLTAT 2011, University of Dundee, UK (2011)
17. Kipp, M., Heloir, A., Nguyen, Q.: A feasibility study on signing avatars. In: Proceedings of SLTAT 2011, University of Dundee, UK (2011)
18. Lesmo, L., Mazzei, A., Radicioni, D.: Linguistic processing in the ATLAS project. In: Proceedings of SLTAT 2011, University of Dundee, UK (2011)
19. Lombardo, V., Nunnari, F., Damiano, R.: The ATLAS interpreter of the Italian sign language. In: Proceedings of SLTAT 2011, University of Dundee, UK (2011)
20. Decaudin, P.: Cartoon-looking rendering of 3D-scenes. In: Syntim Project INRIA, p. 6 (1996)
21. Flodin, M.: Signing Illustrated: The Complete Learning Guide. Berkley Publishing Group, New York (1994)
22. Battison, R.: Lexical Borrowing in American Sign Language. Linstok Press, Silver Spring (1978)
23. Adamo-Villani, N.: 3D rendering of American sign language finger-spelling: a comparative study of two animation techniques. In: Proceedings of ICCIT 2008 - 5th International Conference on Computer and Instructional Technologies, vol. 34, pp. 808–812 (2008)
24. Adamo-Villani, N., Wilbur, R., Eccarius, P., Abe-Harris, L.: Effects of character geometric model on the perception of sign language animation. In: IEEE Proceedings of IV 2009 - 13th International Conference on Information Visualization, Barcelona, pp. 72–75 (2009)
25. Adamo-Villani, N., Kasenga, J., Jen, T., Colbourn, B.: The effect of ambient occlusion shading on perception of sign language animations. In: Proceedings of ICCEIT 2011, Venice, Italy, pp. 1840–1844 (2011)