

Toward an evaluation model for signing avatars

Shada Bennbaia

Mada Center

Abstract –

Signing avatars can make a significant impact on the lives of deaf people by making information accessible anytime and anywhere. With technological development, sign language avatars can be the cost-effective communication solution that will remove the barriers between deaf people and the world. However, most researchers are not part of the deaf community. To this day, the deaf community has little to no knowledge about the signing avatar technology. Thus, researchers have created and used evaluation methods to involve deaf people and feedback to develop and improve sign language avatars based on their needs and requirements. In the article, evaluation methods and tools used to assess signing avatars' functionality, acceptability, and shortcomings were presented and discussed.

Keywords: Sign Language Processing, Evaluation Methodology, Signing Avatar

Introduction

Signing avatars are virtual signers that produce sign language to improve the communication and information accessibility for a deaf individual. They are not intended to replace human sign language interpreters but to co-exist and support them in different domains. For instance, interpreters are required when the sign language translation is highly sensitive and must be as accurate as possible, such as in doctor appointments. At the same time, the avatar can serve the purpose of translating standardized text or automatically translating dynamic content, such as announcements and websites [1]. To read and write without auditory cues is a challenging task for a deaf person which causes many of them to leave school low level of reading/writing abilities. It limits their access to all written content. One of the early adopted solutions was the use of recorded human signers. However, this solution requires high production costs, and the recorded videos cannot be modified or even anonymized after production. As opposed to using virtual characters to produce sign language where the appearance can be completely customized, the animations can be dynamic and easily adjusted when needed, can have an interactive behavior, and the production of new content is relatively cost-effective [2].

Any new technology often faces the challenge of acceptability within the targeted users. Moreover, most of the developers of signing avatars are hearing researchers, which can make deaf people skeptical about the technology for historical reasons [2]. Accepting the sign language avatar within the deaf community is crucial for the successful implementation of the technology. Previous works tried to involve deaf people in developing and evaluating signing avatars but an

efficient effort to clarify the overall acceptance is not made yet. It is essential to identify the issues that cause the rejection of the technology to solve them. Researchers used different approaches to assess the needs of the deaf individuals, incorporate them in the development process and use their feedback to evaluate the signing avatar's performance. Comprehensibility of signing avatars is a significant factor that affects acceptance and performance. Assessing comprehensibility is a challenging and not straightforward task. No unified methodology exists to test sign language avatar comprehensibility [3]. In the ViSiCAST/eSIGN projects, comprehensibility tests have been carried out. One limitation is that these tests primarily used few participants [4].

In the following sections, focus groups, as one of the most known methods to collect qualitative data, are presented with examples followed by performance metrics used to quantify sign language avatar performance. This task is part of the Jumla Sign Language project supported by the Mada Innovation Program [5].

Focus groups

Focus groups are well-known tools used to evaluate human-computer interaction and extract empirical data in research and analysis. The aim of focus groups in evaluating sign language avatars is to elicit deaf people's opinions and feedback toward the signing avatar in the research and development phase to improve the technology and better address the users' needs. Through interactions and comments, this method enables researchers and developers to get in-depth information from the focus group participants about their preferences and how to prioritize them and what issues they consider important. Usually, the focus group is controlled discussions led by an expert with 3-10 potential end-users, where the expert presents the product and guides the discussion.

Balch et al. [6] were the first to use the focus group method with deaf people where they found that this method is very productive in terms of extracting information and understanding the needs and requirements of participants. Kipp et al. used focus group methods to evaluate the acceptability and comprehensibility of sign language avatars [2]. They provided a few recommendations to ensure the quality of the focus group sessions: a) use visual materials such as images, icons, and videos; b) ensure the environment is sign language friendly; C) utilize interactive tools such as voting and open discussions, and the focus group session with online assessments. Kipp et al. [2] conducted a focus group study, complemented by an online questionnaire, to evaluate the German deaf community's opinion about sign language avatars.

Two focus groups with 5 and 3 deaf participants. As shown in Figure 1, the participants sat in a circle and the sessions were video recorded for each group for further detailed analysis. Different sign language avatars were presented to the focus group, and they were asked to rate them and give their feedback on the signing avatar's performance. The deaf participants mainly criticized the appearance and the animation of the signing avatars and described it as unnatural, robotic, and emotionless. These criticisms showed that the nonmanual features of the SL avatars such as the mouth patterns, face expressions, and natural body movements, are essential for deaf people to accept the signing avatar technology. To quantify the focus group results, the deaf participants

were asked to vote on the most critical avatar feature. Facial expressions got the most votes, followed by natural body movement, emotions, appearance, and comprehensibility.



Figure 1. Focus group from the Deaf community

Researchers around the world have used the focus group method to test signing avatars that produce different signing languages such as American Sign Language, Swiss-German Sign Language, Japanese Sign Language, Brazilian Sign Language, Turkish Sign Language, British Sign Language, and many more [7]–[14]. Evaluating signing avatars through a focus group from the deaf community proved a necessary step to incorporate the end-user in the development process and increase their understanding and acceptability of the technology. Moreover, it is also an essential step for researchers to understand the end-user since most of the researchers are not from the deaf community and provide them with a communication tool that satisfies their needs and requirements.

Performance metrics

Even though focus groups and questionnaires provide a good evaluation of the animated avatars, the development process requires detailed, quantified performance metrics to measure and benchmark the development and evolution of these signing avatars. One method to quantify focus group study is to let the participants objectively measure the accuracy of the sign language translation. For example, with one of the first known sign language avatars, TESSA [7]–[14], the

group was asked to indicate whether the sign language phrase produced by the avatar is accurate and easy to understand and, if not, what can be the reason. Through this simple method, they were able to identify the phrases identification average accuracy to be 61%, where 70% of the identification errors were identified to be due to unclear signs.

Using machine translation methods to translate and produce sign language requires performance metrics to precisely evaluate the translation process. San-Segundo et al. [15] used the scoring metrics of BiLingual Evaluation Understudy (BLEU), which calculates the statistical difference between the machine translation and original translation, and Sign Error Rate (SER), which is the percentage of wrong produces signs. However, in this study, the SER was reported to be 31.6% and the BLEU to be 0.5780. Similarly, Patel et al. [16] evaluated the performance of their machine translation avatar system by statistically measuring the accuracy of voice recognition, grammar search and sign synthesis which in total achieved an average translation accuracy of about 77%. Moreover, they used the processing time to benchmark their work faster at publishing the study at 0.85 seconds. In another work by Oh et al. [17], they used the word translation, the correctly machine-translated word ratio, to evaluate their signing avatar for a weather forecast system.

Furthermore, System Usability Scale (SUS) has been used to evaluate the usability of the signing avatar systems by the deaf user. El-Gayyar et al. [18] used this performance metric with a focus group of 5 deaf people to evaluate an Arabic signing avatar application within a limited domain. They achieved a 79.8% average SUS score, indicating that the developed application is acceptable.

It is important to note that to evaluate a sign language avatar system, comprehensive testing should be conducted considering all sign language production characteristics such as accurate translation, non-manual signals, spatial representation, and avatar appearance and naturality, using both qualitative studies and performance metrics.

Conclusion

Producing sign language through animated avatars is a challenging task due to the complex nature of sign language. Evaluating signing avatar systems is not only essential to measure the development progress but also to increase the engagement and acceptability of the deaf community as the signing avatar is a solution that can make their daily lives more accessible and more independent. For that reason, complimentary evaluation methods should be combined to efficiently test and evaluate the sign language avatar's performance, comprehensibility, and acceptability.

References

- [1] S. Ebling, "Evaluating a Swiss German Sign Language Avatar among the Deaf Community," *Proc. Third Int. Symp. Sign Lang. Transl. Avatar Technol.*, no. October, 2013.
- [2] M. Kipp, Q. Nguyen, A. Heloir, and S. Matthes, "Assessing the deaf user perspective on sign language avatars," in *ASSETS'11: Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, 2011, pp. 107–114. doi: 10.1145/2049536.2049557.

- [3] M. Huenerfauth, L. Zhao, E. Gu, and J. Allbeck, "Evaluation of American sign language generation by native ASL signers," in *ACM Transactions on Accessible Computing*, 2008, vol. 1, no. 1. doi: 10.1145/1361203.1361206.
- [4] J. R. Kennaway, J. R. W. Glauert, and I. Zwitserlood, "Providing signed content on the Internet by synthesized animation," *ACM Trans. Comput.-Hum. Interact.*, vol. 14, no. 3, 2007, doi: 10.1145/1279700.1279705.
- [5] D. Al Thani, A. Al Tamimi, A. Othman, A. Habib, A. Lahiri, and S. Ahmed, "Mada Innovation Program: A Go-to-Market ecosystem for Arabic Accessibility Solutions," in *2019 7th International conference on ICT & Accessibility (ICTA)*, 2019, pp. 1–3.
- [6] G. I. Balch and D. M. Mertens, "Focus group design and group dynamics: Lessons from deaf and hard of hearing participants," *Am. J. Eval.*, vol. 20, no. 2, 1999, doi: 10.1177/109821409902000208.
- [7] M. J. Davidson, "PAULA : A Computer-Based Sign Language Tutor for Hearing Adults," 2006.
- [8] S. Ebling and J. Glauert, "Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community," *Univers. Access Inf. Soc.*, vol. 15, no. 4, pp. 577–587, 2016, doi: 10.1007/s10209-015-0408-1.
- [9] I. Zwitserlood, M. Verlinden, J. Ros, and S. Van Der Schoot, "SYNTHETIC SIGNING FOR THE DEAF: eSIGN."
- [10] T. Uchida *et al.*, "Sign language support system for viewing sports programs," in *ASSETS 2017 - Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 2017, pp. 339–340. doi: 10.1145/3132525.3134768.
- [11] T. Uchida *et al.*, "Systems for supporting deaf people in viewing sports programs by using sign language animation synthesis," *ITE Trans. Media Technol. Appl.*, vol. 7, no. 3, pp. 126–133, 2019, doi: 10.3169/mta.7.126.
- [12] J. R. F. Brega, I. A. Rodello, D. R. C. Dias, V. F. Martins, and M. de P. Guimarães, "A virtual reality environment to support chat rooms for hearing impaired and to teach Brazilian Sign Language (LIBRAS)," 2014.
- [13] C. Eryiğit, H. Köse, M. Kelepir, and G. Eryiğit, "Building machine-readable knowledge representations for Turkish sign language generation," *Knowl.-Based Syst.*, vol. 108, 2016, doi: 10.1016/j.knosys.2016.04.014.
- [14] S. Cox *et al.*, "TESSA, a system to aid communication with deaf people," in *Annual ACM Conference on Assistive Technologies, Proceedings*, 2002, pp. 205–212. doi: 10.1145/638286.638287.
- [15] R. San-Segundo *et al.*, "Speech to sign language translation system for Spanish," *Speech Commun.*, vol. 50, no. 11–12, pp. 1009–1020, 2008, doi: 10.1016/j.specom.2008.02.001.
- [16] B. D. Patel, H. B. Patel, M. A. Khanvilkar, N. R. Patel, and T. Akilan, "ES2ISL: An Advancement in Speech to Sign Language Translation using 3D Avatar Animator," in *Canadian Conference on Electrical and Computer Engineering*, 2020, vol. 2020-Augus. doi: 10.1109/CCECE47787.2020.9255783.
- [17] J. Oh, S. Jeon, M. Kim, H. Kwon, and I. Kim, "An avatar-based weather forecast sign language system for the hearing-impaired," *IFIP Adv. Inf. Commun. Technol.*, vol. 436, 2014, doi: 10.1007/978-3-662-44654-6_51.
- [18] M. M. M. M. El-Gayyar, A. S. A. S. Ibrahim, and M. E. E. Wahed, "Translation from Arabic speech to Arabic Sign Language based on cloud computing," *Egypt. Inform. J.*, vol. 17, no. 3, pp. 295–303, Nov. 2016.