Are sign language avatars ready for the real world? Examining the evidence

R. J. Wolfe, PhD.
Professor, School of Computing
DePaul University
Chicago, Illinois USA

On a nearly daily basis, we see news stories enthusiastically describing avatar technologies that claim to "solve" deafness. At a first glance, this would appear to be an encouraging development to those in the Deaf[1] community who routinely face challenges when interacting with the hearing world. There are barriers to job opportunities, education and government services because of the lack of available certified sign language interpreters. Businesses, health care services, schools and governmental agencies are always looking for ways to avoid increased costs when providing Deaf accessibility and for these organizations, the abundance of happy news articles would seem to imply that automatic interpretation between signed and spoken languages is just around the corner and the new technology will keep accessibility costs to a minimum.

So, the question is, "Has avatar technology progressed to the point where it is ready for practical use?" A clear-eyed discussion of this question is essential for determining the best use of scarce resources. Attempting to deploy a technology before it has matured can lead to a waste of time and money. Even worse, it can take away from other options that have an established track record of success including interpreter services and interpreter training programs. The frustrating result can be a decrease, rather than an increase, in accessibility.

The rest of this discussion focuses on two topics. The first topic is a discussion of the question, "What is sign language avatar technology and what is its potential?" The second topic provides a checklist of questions to ask when encountering one of the happy news stories proclaiming a technological breakthrough in Deaf accessibility.

As commonly used, the term "sign language avatars," refers to three major research efforts. It's important to distinguish among these three areas in order to understand the likelihood of developing a technology that's ready for practical use. The three areas are

1. Sign language recognition. This is the conversion of signed language to written text in a spoken language. This component uses a camera and/or 3D sensing equipment to record the motions of a person who is signing. It converts the video or 3D data into a representation of sign language, and from there, into a written form of a spoken language.
2. Spoken language to sign language translation. This area of research applies natural language processing techniques to convert a written form of a spoken language into a text representation of a signed language. Unfortunately, there is no universally-accepted written form of signed languages, so it is not possible for this step to produce signing that is readable.
3. Avatar display of signed languages. The goal of this research is to take the unreadable text representation and display it as an animated video using an avatar. The goal of this component is to produce animations of signing that have a natural flow and are easy to read.

Let's look at each of these in turn, in order to analyze their readiness for practical applications.

[1] The term "big D" Deaf refer to people who share a sign language and a culture. The term "little d" deaf refers to the audiological condition of not hearing

The first area, sign language recognition, must accommodate the many variations in signing styles. Further, signs can change shape depending on how they're being used. Because of the fluid variability of sign production, this technology was originally limited to recognizing an extremely small number of words (Starner, Weaver, & Pentland, 1998). It often required a signer to wear specialized equipment such as data gloves or a motion capture suit. (Abhishek, Qubeley, & Ho, 2016). Other approaches place restrictions on the physical appearance of the signer and the environment surrounding the signer (Koller, Zargaran, Ney, & Bowden, 2016). At present, the accuracy of the best systems for recognizing continuous signing is less than 70%. Compare this to Google Voice, which has an accuracy rate of 95%. (Protalinski, 2017). From this information it is safe to conclude that sign language recognition is still a research work-in-progress and not ready for practical application.

The second area, spoken-to-signed translation, also involves a conversion, but in the opposite direction from sign language recognition. Efforts to convert written to signed language have been ongoing for 25 years. Early efforts used grammar rules to construct an ASL syntactic structure (Zhao, et al., 2000) and typically focused on highly predictable input that followed a script and used a limited and parameterized vocabulary. Examples include automated weather reports, and interactions with a postal clerk or airport security personnel (Grieve-Smith, 2001) (Cox, et al., 2002) (Lancaster, et al., 2003).

Since then, other efforts are looking to deeper constructs to represent language such as the use of an Interlingua between the spoken and signed languages. (Veale, Conway, & Collins, 1998)(Huenerfauth, Marcus, & Palmer, 2006). Recently this approach seems to have fallen from favor as the automatic translation field has embraced corpus-based techniques. A corpus is a collection of texts previously translated into multiple languages by expert human translators. The automatic translation then can then convert a new text from one language into another by searching the corpus. Examples of this approach include Google Translate (Johnson, et al., 2017) and DeepL translator (DeepL GmbH, 2019).

Corpus-based techniques can produce translations with accuracy in the 80-90% range (Popescu-Belis, 2019) , but their success relies on having huge amounts of text to analyze (Maucec, Brest, & Kacic, 2005). For example, Europarl, an early corpus of 11 spoken languages, has over 300 million words (Koehn, 2005). However, such large corpora do not yet exist for signed languages. The largest sign language corpus is currently less than 0.5% of Europarl (Konrad, 2018). It will still be many years before the size of sign language corpora rival the size of spoken-language corpora and signed/spoken translations yield the same accuracy rates as currently seen for automatic translations between two spoken languages.

However, this component has matured to the point of practicality for one limited area. It is practical in situations where the communication is in a single direction from spoken to sign language and the spoken text follows a script that is highly predictable. Examples of this include pre-recorded customer announcements in train station or hotels.

With this one exception, the second area, spoken-to-signed language translation, is also still a work-in-progress in need of more research. This is consistent with the views of the World Association of Sign Language Interpreters and the World Federation of the Deaf on the capabilities of avatar technology (World Federation of the Deaf, 2018).

A major contributing to the fact that spoken-to-sign translation technology lags behind spoken-to-spoken language translation is that is requires the display of signed language via avatar. This is the last of the three research areas. Its purpose is to display signed language that is natural and easy to read.

These two qualities (naturalness, ease of reading) are essential if an avatar display can leave the research lab for use in practical applications.

Avatar display has been an active area of research for nearly 40 years (Poizner, Bellugi, & Lutes-Driscoll, 1981). Researchers began with simple stick figures and have continued to improve the realism ever since. Although a single image of many of today's avatars can look quite appealing, the key to their effectiveness is the way that they move and how well they portray facial nonmanual signals. Because essential information in signed languages is conveyed via the face, a signing avatar should have a face as agile and expressive as a Deaf signer. Current avatars are only capable of producing some of the mouthings and mouth gestures of the signed languages of the world, (Brumm, Johnson, Hanke, Grigat, & Wolfe, 2019) so this lack of a capability puts a limit on the variety of sentences they can produce.

When an avatar's motion is natural and flowing, like the motion of a human signer, the avatar becomes easy to understand. The comprehension of current avatars varies widely, depending on what is being signed. For isolated words, studies have reported comprehension rates of over 90% (Ebling, et al., 2017), but for complete sentences, the comprehension rates are in the 60% range (Smith & Nolan, 2016).

Thus, the very best of today's avatars are capable of producing extremely short sentences that have a good chance to being understood. For limited applications, this technology may be ready for practical application involving pre-recorded customer announcements in train station or hotels. It can also produce isolated words in a dictionary format to support hearing students in interpreter training programs.

Given this background on the maturity of the three aspects of sign language avatar technology, here are important questions to pose whenever encountering a news article that proclaims a new breakthrough:

1. Where is the technology currently being used? Is the news article only describing a successful test in a laboratory or does it describe an application that's being used in a real-world situation?
2. What is the scope of the claim? There are many situations, such as classrooms, and doctor visits where automatic interpretation will simply never replace human interpreters. Their knowledge of culture, history and context all contribute to their effective interpretation, and this is beyond what automatic interpretation can do.
3. Does the headline of the news article accurately match the article's contents? A headline will be as exciting as possible to attract reader attention to the article.
4. What is the level of Deaf involvement in the project? Is the group lead by a Deaf researcher? Are there Deaf researchers on the team? The Deaf perspective is essential in any Deaf accessibility project.
5. Did the news article include a quote from a member or members of the Deaf community? The opinions of hearing people do not matter; If a technology is one that will be used by the Deaf community then the Deaf community should be consulted.

In conclusion, the answer to the question, "Are sign language avatars ready for the real world?" is "for the most part, not yet." However, with more Deaf researchers involved in this effort, the results will have a greater likelihood of making those short, everyday barriers of language a little easier.

References

Abhishek, K. S., Qubeley, L. C., & Ho, D. (2016). Glove-based hand gesture recognition sign language translator using capacitive touch sensor. *2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)*, (pp. 334-337).

Brumm, M., Johnson, R., Hanke, T., Grigat, R.-R., & Wolfe, R. (2019). Use of Avatar Technology for Automatic Mouth Gesture Recognition. *Poster presented at SignNonmanuals: Der Beitrag nicht-manueller Elemente zur Syntax der Österreichischen Geb_rdensprach (ÖGS), Graz, Austria.* Retrieved from http://signnonmanuals.aau.at/sites/default/files/Wolfe_Graz%20nonmanuals%20workshop%202019%20Poster.pdf

Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., & Abbott, S. (2002). Tessa, a system to aid communication with deaf people. *Proceedings of the fifth international ACM conference on Assistive technologies*, (pp. 205-212).

DeepL GmbH. (2019, August 08). *Press Information.* Retrieved from DeepL: https://www.deepl.com/press.html

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals, 51*, 779-795.

Ebling, S., Johnson, S., Wolfe, R., Moncrief, R., McDonald, J., Baowidan, S., . . . Tissi, K. (2017). Evaluation of Animated Swiss German Sign Language Fingerspelling Sequences and Signs. *International Conference on Universal Access in Human-Computer Interaction*, (pp. 3-13).

Fratarcangeli, M., & Schaerf, M. (2004). Realistic modeling of animatable faces in MPEG-4. *Computer Animation and Social Agents*, (pp. 285-297).

Grieve-Smith, A. B. (2001). SignSynth: A sign language synthesis application using Web3D and Perl. *International Gesture Workshop*, (pp. 134-145).

Huenerfauth, M., Marcus, M., & Palmer, M. (2006). *Generating American Sign Language classifier predicates for English-to-ASL machine translation.* Ph.D. dissertation, University of Pennsylvania.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., . . . others. (2017). Google_s multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics, 5*, 339-351.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. *MT summit*, *5*, pp. 79-86.

Koller, O., Zargaran, O., Ney, H., & Bowden, R. (2016). Deep sign: Hybrid CNN-HMM for continuous sign language recognition. *Proceedings of the British Machine Vision Conference 2016.*

Konrad, R. (2018). Statistik: Token-Tags.

Lancaster, G., Alkoby, K., Campen, J., Carter, R., Davidson, M. J., Ethridge, D., . . . others. (2003). Voice activated display of American Sign Language for airport security. *Technology and Persons with Dsabilities Conference.*

Loomis, J., Poizner, H., Bellugi, U., Blakemore, A., & Hollerbach, J. (1983). Computer graphic modeling of american sign language. *ACM SIGGRAPH Computer Graphics*, *17*, pp. 105-114.

Maucec, M. S., Brest, J., & Kacic, Z. (2005). Slovenian to English machine translation using corpora of different sizes and morpho-syntactic information. *Language Technologies Conference: proceedings of the 9th International Multiconference Information Society IS*, *2006*, pp. 222-225.

Othman, A., & Jemni, M. (2011). Statistical Sign Language Machine Translation: from English written text to American Sign Language Gloss. *International Journal of Computer Science Issues (IJCSI), 8*, 65.

Poizner, H., Bellugi, U., & Lutes-Driscoll, V. (1981). Perception of American sign language in dynamic point-light displays. *Journal of experimental psychology: Human perception and performance, 7*, 430.

Popescu-Belis, A. (2019). Context in Neural Machine Translation: A Review of Models and Evaluations. *arXiv preprint arXiv:1901.09115*.

Protalinski, E. (2017, May 17). *Google's speech recognition technology now has a 4.9% word error rate.* Retrieved from Venture Beat: https://venturebeat.com/2017/05/17/googles-speech-recognition-technology-now-has-a-4-9-word-error-rate/

Shantz, M., & Poizner, H. (1982). A computer program to synthesize American Sign Language. *Behavior Research Methods & Instrumentation, 14*, 467-474.

Smith, R. G., & Nolan, B. (2016). Emotional facial expressions in synthesised sign language avatars: a manual evaluation. *Universal Access in the Information Society, 15*, 567-576.

Starner, T., Weaver, J., & Pentland, A. (1998). Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on pattern analysis and machine intelligence, 20*, 1371-1375.

Tobin, A. (2015). Is Google Translate Good Enough for Commercial Websites?; A Machine Translation evaluation of text from English websites into four different languages. *Reitaku Review, 21*, 94-116.

Veale, T., Conway, A., & Collins, B. (1998). The challenges of cross-modal translation: English-to-Sign-Language translation in the Zardoz system. *Machine Translation, 13*, 81-106.

World Federation of the Deaf. (2018). WFD and WASLI Statement on Use of Signing Avatars. Retrieved from http://wfdeaf.org/news/resources/wfd-wasli-statement-use-signing-avatars

Zhao, L., Kipper, K., Schuler, W., Vogler, C., Badler, N., & Palmer, M. (2000). A machine translation system from English to American Sign Language. *Conference of the Association for Machine Translation in the Americas*, (pp. 54-67).