Automated Sign Language Translation: The Role of Artificial Intelligence Now and in the Future

Lea Baumgärtner, Stephanie Jauss, Johannes Maucher and Gottfried Zimmermann
Hochschule der Medien, Nobelstraße 10, 70569 Stuttgart, Germany

Keywords: Sign Language Translation, Sign Language Recognition, Sign Language Generation, Artificial Intelligence.

Abstract: Sign languages are the primary language of many people worldwide. To overcome communication barriers between the Deaf and the hearing community, artificial intelligence technologies have been employed, aiming to develop systems for automated sign language recognition and generation. Particularities of sign languages have to be considered - though sharing some characteristics of spoken languages - since they differ in others. After providing the linguistic foundations, this paper gives an overview of state-of-the-art machine learning approaches to develop sign language translation systems and outlines the challenges in this area that have yet to be overcome. Obstacles do not only include technological issues but also interdisciplinary considerations of development, evaluation and adoption of such technologies. Finally, opportunities and future visions of automated sign language translation systems are discussed.

1 INTRODUCTION

A basic understanding of the particularities and notation of sign language serves to better understand the importance and challenges of automated sign language translation. We use both the terms deaf and Deaf throughout this paper, the first referring solely to the physical condition of hearing loss, the latter emphasizing the Deaf culture, meaning people often prefer to use sign language and were often, but not obligatorily, born deaf (Matheson, 2017).

1.1 Importance and Particularities of Sign Languages

There are 70 million deaf people around the world (World Federation of the Deaf (WFD),), around 80,000 deaf people and 16 million with hearing loss living in Germany (Deutscher Gehörlosen-Bund e.V. (DGB),). Many of those deaf and hard-of-hearing (DHH) individuals use sign languages as their primary language. Linguistics have shown that sign languages have similar properties as spoken languages, such as phonology, morphology and syntax (Bavelier et al., 2003). Signs, like speech, are combinations of phonological units, especially hand shape, hand position and hand movement (Bavelier et al., 2003). Like syllables and words build spoken language, those units build signs.

There are, however, differences in how sign languages work in detail, some of which are mentioned using the example of American Sign Language (ASL) in the following. Since suffixes to verbs like in spoken languages are not possible, ASL tenses are built by adding words in the beginning or at the end of a sentence. Moreover, ASL verbs of motion, for example, include information about path, manner and orientation. According to the (World Federation of the Deaf (WFD),), there are over 300 different ones around the world (European Union of the Deaf (EUD), 2012).

Furthermore, countries with the same spoken language can have different sign languages, such as Germany and Austria (European Union of the Deaf (EUD), 2012). Globalization nurtured attempts to create an international sign language for deaf people to communicate cross-border. Today, the pidgin language International Sign (IS) is mostly used for this purpose (European Union of the Deaf (EUD), 2012).

1.2 Sign Language Notation

A fundamental problem regarding sign languages is the lack of a standardized transcription system. There have been several attempts to develop notations during the last decades. (McCarty, 2004) considers the Stokoe notation to be the most known of those systems that are actually being used. Both of them have been developed for ASL but have been transferred to

Baumgärtner, L., Jauss, S., Maucher, J. and Zimmermann, G.
Automated Sign Language Translation: The Role of Artificial Intelligence Now and in the Future.
DOI: 10.5220/0010143801700177
In *Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications (CHIRA 2020)*, pages 170-177
ISBN: 978-989-758-480-0

other sign languages and are used in a variety of countries worldwide (McCarty, 2004; Kato, 2008). The Stokoe notation describes signs through three symbols – “tab” for the location in which the sign is made, “dez” for the hand shape and “sig” for the hand movement – and uses 55 symbols (12 tab, 19 dez and 24 sig) (McCarty, 2004). Another notation system for sign languages that has its roots in the Stokoe system and is commonly used today is the HamNoSys (Hamburg Notation System for Sign Languages). It is applicable internationally since it does not refer to national finger alphabets and was improved after being originally mostly based on ASL (Hanke, 2004). HamNoSys describes signs on a mainly phonetic level, including “the initial posture (describing non-manual features, handshape, hand orientation and location) plus the actions changing this posture in sequence or in parallel” (Hanke, 2004, p. 1) for each sign. Another widely used way to describe signs are glosses. They represent signs on a higher level providing spoken language morphemes that convey the signs’ meaning (Duarte, 2019). This kind of notation is consequently dependent on the sign language and written language used. Currently, the most common way of gathering data of sign language is through video files, which compared to texts have drawbacks in storage, complexity and cost. Therefore, the need for a text-based notation system is still applicable. Even though some sign language notation systems are being used around the world, there is no standard notation system for sign language on an international scale, even though this would make gathering datasets for developing sign language translations systems much easier (Tkachman et al., 2016).

2 AI IN AUTOMATED SIGN LANGUAGE TRANSLATION

In this paper, sign language translation is understood to include the translation of sign language to text as well as of text to sign language. Thus, both areas of sign language recognition and sign language generation are presented in the following, comparing different Artificial Intelligence (AI) methods, especially their architecture, performance and challenges.

2.1 Sign-Language-to-Text Translation (Sign Language Recognition)

Over the past years, several approaches have been researched and analyzed in the area of automatic sign language translation. Although sign languages largely depend on phonological features like hand location, shape and body movement (Bragg et al., 2019), some recognition approaches are based on static hand shape while others focus on continuous dynamic sign gesture sequences. Feed-forward networks are mostly used for static hand shape recognition and Recurrent Neural Networks (RNN) are often employed for sequential recognition of body movements. This section will outline a few possibilities of sign-language-to-text translation based on a variety of AI methods and solutions. Those can be differentiated in static hand sign recognition approaches (separable into vision-based and sensor based-methods) compared in section 2.1.1 and dynamic sign language recognition approaches outlined in section 2.1.2.

2.1.1 Static Hand Sign Translation

One possibility of automatic sign language translation is based on static hand signs. The natural language alphabet can be mapped directly to a set of different hand shapes. The recognition of those static signs can be realized by an AI classification model. Most challenging is the feature extraction such as recognition of hand details like fingers, hand rotation and orientation. Numerous vision-based approaches for solving the hand sign recognition problem have been investigated over the past years. In 2019, Fayyaz and Ayaz analyzed the performance of different classifier architectures based on a static sign language image dataset (Fayyaz and Ayaz, 2019). For Support Vector Machine (SVM), the authors achieved an accuracy of approximately 83% with Speeded Up Robust Features (SURF) and 56% without SURF (Fayyaz and Ayaz, 2019). The Multilayer Perceptron (MLP) achieved an accuracy of approximately 86% with SURF features and 58% with manually extracted features (Fayyaz and Ayaz, 2019). (Pugeault and Bowden, 2011) (2011) researched static hand shape recognition in real-time, using depth information for hand detection in combination with hand tracking. (Pugeault and Bowden, 2011) used random forests as classifiers. The result of their work is a dataset consisting of 500 samples. The authors achieved the best performance by using “the combined vector (mean precision 75%), followed by appearance (mean precision 73%) and depth (mean precision 69%)” (Pugeault and Bowden, 2011, p. 5) and implemented a graphical user interface which was able to run on standard laptops.

Based on the numerous investigated experiments and papers, it can be concluded that translating static hand signs is fairly well researched. It is mostly an image recognition problem that can be solved by common state-of-the-art image processing solutions. Since static hand signs represent characters without context or creating sentences, static hand sign translation is not sufficient for daily life use cases.Therefore, dynamic sequence-to-sequence sign language translation seem to be more promising as a communication method for the daily life of signers.

2.1.2 Dynamic Sign Language Translation

The translation of dynamic sign language requires more complex network architectures such as RNNs because input data are time-based sequences. Since machine learning and natural language processing methods improved over the past years, the possibilities of sign language translation have been enhanced, too. The main challenge is to map the continuous sign language sequences to spoken language words and grammar. Particularly, separate movements and gestures cannot be mapped to spoken language directly. Therefore, a gloss notation might be used. In the following, we will outline some approaches to capture the hand and body movements for the purpose of continuous sign language recognition. In 2018, (Camgoz et al., 2017) created a sign-language-to-speech translation system based on sign videos using an attention-based encoder-decoder RNN and Convolutional Neural Network (CNN) architecture. The researchers produced the first publicly accessible dataset for continuous sign language translation, called "RWTH PHOENIX-weather 2014" (Camgoz et al., 2018). To overcome the problem of one-to-one mapping of words to signs, they integrated a CNN with attention mechanism before the RNN to model probabilities (Camgoz et al., 2018). They conclude that the networks performed quite well – except when mentioning numbers, dates or places (Camgoz et al., 2018).

DeepASL was published in 2017 by (Fang et al., 2017). Its architecture consists of hierarchical bidirectional recurrent neural networks (HB-RNN) in combination with a probabilistic framework based on Connectionist Temporal Classification (CTC) (Fang et al., 2017). DeepASL achieved an average translation accuracy of 94.5% on a word level, an average word error rate of 8.2% on a sentence level on unseen test sentences as well as 16.1% on sentences signed by unseen test users (Fang et al., 2017). The ASL translator can be integrated into wearable devices such as tablets, smartphones or Augmented Reality (AR) glasses and enable face-to-face communication between a deaf and a hearing person. An example has been shown by the authors in a system translating both performed signs by the deaf person into spoken English and spoken words by the hearing person into English text that will then be projected into an AR glasses of the deaf person. This makes DeepASL a useful and effective daily life sign language interpreter (Fang et al., 2017).

The described applications demonstrate the importance and possibilities of dynamic sign language recognition. While static hand sign recognition may be a first approach, dynamic sign language translation appears to be more useful for signing people in their daily life.

2.2 Text-to-Sign-Language Translation (Sign Language Generation)

2.2.1 Importance

While the importance of sign-language-to-text translation may seem more obvious because it enables deaf people to be understood by persons who do not understand sign language, the vice versa translation from text to sign language is sometimes seen as less important. This is reflected in less research existing in this field (Duarte, 2019).

Nevertheless, text-to-sign-language translation should be considered important. Sign language videos can make information more accessible to those who prefer sign language representation through videos or animations over rarely used and for many more difficult to understand text representation (Bragg et al., 2019; Elliott et al., 2008). Pre-recorded videos, however, face some problems: production costs are high, later modification of the content is not possible and signers cannot remain anonymous (Kipp et al., 2011a; Kipp et al., 2011b). That is why animated avatars, understood as being "computer-generated virtual human[s]" (Elliott et al., 2008, p. 1) are the most common way to present generated sign language. They provide "similar viewing experiences" as videos of human signers (Bragg et al., 2019, p. 5) while being much more suitable for automatic generation and solving the mentioned problems: their appearance can be adapted to suit the use case and audience and animations can be dynamically adjusted which allows real-time use cases (Kipp et al., 2011a; Kipp et al., 2011b).

2.2.2 Technological Approaches

Two kinds of approaches generating such avatar animations can be distinguished: motion-capturing (human movements are tracked and mapped to an avatar) and keyframe animations (the entire animation is computer-generated) (Bragg et al., 2019). They face various challenges and will be compared in the following.

(Elliott et al., 2008) (2007) describe a pipeline for sign language translation based on the projects ViSiCAST and eSIGN. In the first step, English text is translated into phonetic-level sign language notations of German Sign Language (DGS), Dutch Sign Language (NGT) or British Sign Language (BSL). In the second step, a real-time avatar animation is generated from the output of the second step. The researchers used the HamNoSys notation (and even improved it for their needs) as well as Gestural SiGML (Signing Gesture Markup Language) (Elliott et al., 2008; Kacorri et al., 2017). (De Martino et al., 2017) have taken a different approach while improving the Falibras system that automatically translates Brazilian Portuguese text to Brazilian Sign Language (Libras) animated by an avatar for their use case. Therefore, they combined Statistical Machine Translation (SMT) with Example-Based Machine Translation (EBMT) to enable translations for unseen texts as well as translations of ambiguous terms dependent on the context and frequency of the occurrence in previous translations. (Morrissey and Way, 2005) have used a similar approach before: Using the ECHO corpus they automatically translated English text into Gloss notation of three sign languages including non-manual features. Even if test sentences where combined of parts out of the corpus only 60% of the resulting translations were considered coherent by them.

The study (De Martino et al., 2017) conducted had the goal to develop a system that presents marked texts to students via an animated 3D avatar next to the text so they can experience written and signed content at the same time. They built a corpus with signers being tracked by motion capture technology and the videos were transcribed as Portuguese and English text and gloss sequences describing the recorded signs' hand and facial expressions. In the evaluation, the intelligibility score of the signing avatar was on average 0.926, which compared to the 0.939 score achieved by human signer videos can be considered very high. However, only comprehensibility of those signs was evaluated and not whether the animations feel natural etc. That is why the project is currently extended to not only enlarge the corpus, but also include non-manual features such as facial expressions in the Intermediary Language (De Martino et al., 2017).

In 2011, (Kipp et al., 2011a) extended their EMBR system presented one year before. This is a general-purpose real-time avatar animation engine that can be controlled via the EMBRScript language (Kipp et al., 2011a). (Kipp et al., 2011a) evaluated the comprehensibility of their system by comparing the avatar animations with human signers, combining objective count of the glosses understood and subjective specialist opinions. The avatar led to quite varied understandability and reached around 58% sentence-level comprehension in comparison to human signers on average, which the authors state is close to ViSiCAST results. The researchers think that 90% comprehensibility will be possible if more linguistic research in the area of non-manual features and prosody is conducted.

3 CONCLUSION

3.1 Challenges

Although numerous approaches have been researched in the areas of sign language recognition and generation, it is still a long way to achieve fully automated, real-time translation systems. These challenges are summarized in the following.

3.1.1 Use Cases

There are several application domains for sign language translation that pose questions of transferability but also ethical questions. First, various use cases involve various requirements from vocabulary to platform and interface (Bragg et al., 2019).

Furthermore, the tenability of use cases might be limited. In a statement, the WFD (as cited in (Al-khazraji et al., 2018)) expressed worries about using computed signing avatars in certain contexts in which information is very critical and suggest possible use cases only for static content that requires no interaction and can be pre-created. The Deaf community is also concerned that automated sign language translation will replace professional human translators which is why (Al-khazraji et al., 2018) demand researcher's responsibility to properly evaluate their systems and consider these concerns before deployment. (Kipp et al., 2011b) likewise identified mainly one-way communication domains with not too complex or emotional content after assessing deaf study participants' perspective. Dialogic interaction with avatars could not be pictured by them,only the translation of simple sentences, news, guides and texts. Nevertheless, the overall attitude towards signing avatars Kipp, Nguyen et al. (2011) experienced was positive and even increased during the study. (Bragg et al., 2019) on the other hand assess interactive use cases as compelling, for example, personal assistant technologies. They call for the development of real-world applications, i.e. considering real use cases and constraints. This includes focusing on the recognition of sign language sequences rather than of single signs to enable fluent daily life conversations between signers and non-signers (Bragg et al., 2019; Fang et al., 2017).

3.1.2 Sign Language Complexity, Internationality & Notation

The WFD reminds in a statement (as cited in (Al-khazraji et al., 2018)) that sign languages are full languages and cannot be translated word by word. As with text-to-text translation, human assessment, especially by members of the Deaf community, is vital for developing translations systems suitable for real-life scenarios. Sign languages additionally vary among themselves, too.

(Kipp et al., 2011a) found that the multimodality of sign language which requires synchronization of various body parts is a reason why state-of-the-art signing avatars reach a comprehensibility of at best only around 60% and 70%. Moreover, many methods in machine learning and natural language processing have been developed for spoken or written languages and cannot easily be transferred to sign languages that have various structural differences (Bragg et al., 2019). This affects the context changing a sign's meaning or non-manual features extending over multiple signs (Bragg et al., 2019). In addition, notations vary through studies and languages and there is no reliable standard written form for sign languages, even though such a default annotation system could largely advance training sign language recognition and generation systems (Bragg et al., 2019). (Bragg et al., 2019) see a big potential in sharing annotated datasets which would help to enlarge training data and reduce error. Thus, accuracy and reliability could be enhanced and costs reduced. Furthermore, a standard annotation system could also benefit sign language users in general, as it would enable the use of text editors or email systems. The right level of abstraction compared to a dynamic sign language (as is the case with written vs. spoken language), however, has yet to be defined (Bragg et al., 2019).

3.1.3 Full Automation

A big limitation most current sign language translation systems face is, as (Bragg et al., 2019) state, that user intervention is required and they are not fully automated. This is particularly true for avatar generation where different parameters are often defined by humans to make the avatar seem more natural, even though machine learning methods have been proposed, for instance, the approaches by (Adamo-Villani and Wilbur, 2015) and (Al-khazraji et al., 2018) described above.

In the area of sign language generation, especially motion-capturing technology involves a high effort and (Elliott et al., 2008) evaluate motion-capturing technologies as not feasible as they are also limited in not allowing to reuse individual components. Furthermore, the degree of automation is not described precisely in all studies, nor are machine learning architectures.

3.1.4 Datasets

One reason for the large extent of human involvement in translation systems is the generation of corpora. These exist commonly in the form of videos or motion-capture data of human signers which entail multiple problems (Bragg et al., 2019). Generating datasets in-lab, on the other hand, may generate higher quality but is commonly more expensive as well as less scalable, realistic and generalizing for real-life low-quality equipment scenarios, according to Bragg et al. (2019). They continue to explain that enlarging the corpora by collecting data on production is useful and scalable but an initial dataset is needed.

The diversity and size of those datasets are significant factors influencing the performance of automated sign language translation systems (De Martino et al., 2017). Content, size and format of training data depend on the use case (Bragg et al., 2019). Evaluating these datasets is another problem (Bragg et al., 2019) but applies to all steps in developing automated sign language translation systems. Furthermore, not all corpora are published open-source and accessible for public research which is criticized by (Bragg et al., 2019) who conclude that "few large-scale, publicly available sign language corpora exist" and even the largest of them are a lot smaller than corpora of other research areas such as speech recognition. In Table 1, (Bragg et al., 2019) compare public datasets most commonly used for sign language recognition. There are even more problems in dataset generation. One of them is that they have to be created for all languages. Many of the existing corpora are based on ASL, according to (Bragg et al., 2019). Additionally, annotations should be included. That correlates with the lack of a standard notation system discussed in section 3.1.2, since annotations vary in format, linguistic granularity, cost and software (Bragg et al., 2019). A written form for sign languages would also enable datasets to exist without video content, upon (Bragg et al., 2019). Moreover, the authors report, many existing corpora contain only individual signs (see Table 1), which they state is not sufficient for real-world use cases. Further problems are unknown proficiency and demographic data of signers and the lack of signer va-

Table 1: Comparison of Popular Public Sign Language Video Corpora Commonly Used for Sign Language Recognition. Reprinted from “Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective” by D. Bragg, T. Verhoef, C. Vogler, M. Ringel Morris, O. Koller, M. Bellard, L. Berke, P. Boudreault, A. Braffort, N. Caselli, M. Huenerfauth and H. Kacorri, 2019, p. 5.

Dataset	Vocabulary	Signers	Signer-independent	Videos	Continuous	Real-life
Purdue RVL-SLLL ASL [65]	104	14	no	2,576	yes	no
RWTH Boston 104 [124]	104	3	no	201	yes	no
Video-Based CSL [54]	178	50	no	25,000	yes	no
Signum [118]	465	(24 train, 1 test) -25	yes	15,075	yes	no
MS-ASL [62]	1,000	(165 train, 37 dev, 20 test) -222	yes	25,513	no	yes
RWTH Phoenix [43]	1,081	9	no	6,841	yes	yes
RWTH Phoenix SI5 [74]	1,081	(8 train, 1 test) -9	yes	4,667	yes	yes
Devisign [22]	2,000	8	no	24,000	no	no

riety (Bragg et al., 2019). Another issue in the development of sign language translation systems is that it “requires expertise in a wide range of fields, including computer vision, computer graphics, natural language processing, human-computer interaction, linguistics, and Deaf culture” (Bragg et al., 2019, p. 1). Hence, an interdisciplinary approach is essential achieving systems that meet the Deaf community’s needs and are technologically feasible (Bragg et al., 2019).

Currently, most researchers are not members of the Deaf community (Kipp et al., 2011b), which might lead to incorrect assumptions made about sign language translation systems, even if strong ties to the Deaf community exist (Bragg et al., 2019). On the other hand, Deaf people being the target user group have often little knowledge about signing avatars (Kipp et al., 2011b). Concluding, a strong information exchange between researchers and the Deaf community must be developed (Kipp et al., 2011b). To enable Deaf and Hard of Hearing (DHH) individuals to work in this area, animations tools, for instance, should be improved or developed to serve deaf people’s needs (Kipp et al., 2011a).

3.1.5 Interdisciplinary & Involvement of the Deaf Community

Adaptations are possibly not only one-sided: (Bragg et al., 2019) believe that the Deaf community will be open to adapt to technological change: Like written languages evolve due to technology, for example, character limits and mobile keyboards, they believe, sign languages could, for instance, become simpler to support written notation and automated recognition.

3.1.6 Evaluation

Comparability of automated sign language translation systems largely depends on the evaluation methods – however, these vary a lot from research to research (Kipp et al., 2011a). An objective assessment on which approaches are most promising is therefore not possible.

Standardized evaluation methods have been proposed, for instance, by (Kacorri et al., 2017). The authors found that especially four questions influence a participant’s rating on signing avatars: which school type they attend(ed), whether they use ASL at home, how often they use modern media and video files and their attitude towards the utility of computer animations of sign language. They stress that next to explaining the evaluation methods, collecting characteristics of the participants in such studies is essential. (Kacorri et al., 2017) have released their survey questions in English and ASL and hope for evaluation standards in the field of sign language generation.

3.1.7 Acceptance within the Deaf Community

Following the research on signing avatars by (Kacorri et al., 2017), acceptance within the Deaf community is vital for the adoption of sign language generation technologies. (Bragg et al., 2019) believe that the Deaf user perspective has to be properly analyzed and that enforcing technology on the Deaf community will not work. However, avatar generation “faces a number of technical challenges in creating avatars that are acceptable to deaf users (i.e., pleasing to view, easy to understand, representative of the Deaf community, etc.)” (Bragg et al., 2019, 7).

To begin with, avatar rendering itself faces challenges: A large range of hand shapes must be presentable, relative positions must be converted to absolute ones, body parts must not collide and all should seem natural and happen in real-time (Elliott et al., 2008). Small-scale systems exist that work even on mobile devices, too, for instance, an app that (Deb et al., 2017) describe (see section 3.2). But real-time rendering is not yet sufficient for avatars signing in a way that is perceived adequately natural. (Kipp et al., 2011b) have investigated the acceptance of signing avatars for DGS within the German Deaf community in 2011. They found that hand signs received positive critique but upper body movement was criticized as not being sufficient to appear natural (Kipp et al., 2011b). Another challenge are transactions that should look realistic which is especially difficult to achieve from motion-capture data (Bragg et al., 2019). It gets even more challenging because movements can convey endless nuances (Bragg et al., 2019). A similar issue arises due to the numerous possible combinations of signs, according to (Bragg et al., 2019).

One approach to improve prosody in sign language avatars by (Adamo-Villani and Wilbur, 2015) is described in section 2.2.2. So far, realism of signing avatars seemed to be the goal. However, avatars also face the uncanny valley problem. That means that strong but not fully realistic avatars are not received as pleasing, which could explain that the cartoonish avatar was ranked best in the study by (Kipp et al., 2011b). Lastly, the acceptance of signing avatars will likely rise with the involvement of the Deaf community in research and development of sign language systems, as described in section 3.1.5. (Kipp et al., 2011b) found that participation of Deaf persons in such studies significantly increases the positive opinion about signing avatars.

3.2 Opportunities and Visions

When the challenges outlined in section 3.1 can be resolved, opportunities for automated sign language translation systems are enormous. Especially a standard notation and bigger sign language datasets could significantly evolve training and performance of sign language recognition and generation technologies (Bragg et al., 2019; De Martino et al., 2017). They would also entail numerous advantages of their own – such as a written form of sign languages, accurate dictionaries and better resources for learning sign languages (Bragg et al., 2019). In the near future, static one-way information could be presented through signing avatars, for example next to text on websites. This is already the case in some small-scale projects, for instance a part of the website of the city of Hamburg (Elliott et al., 2008).

Various research has been conducted in an educational context, not only for making material more accessible to students using sign language, but to assist those who want to learn a sign language, too. One example is the system that (Deb et al., 2017) describe. An Augmented Reality (AR) application was developed that presented 3D animations of signs on mobile devices via AR as an overlay to the scanned Hindi letters. This involves various technologies of image capturing, processing, marker tracking, animation rendering and augmented display. If it could be extended to whole texts and transferred to different languages and sign languages, there would be numerous use cases for this system.

Involving the additional step of speech processing, students of the NYU have developed a proof of concept app that can translate American Sign Language to English speech as well as vice versa and display a signing avatar through augmented reality (Polunina, 2018). Unfortunately, technological details are not given. Taking this concept further, a daily life application based on smartphone technologies could be developed and automatically translate speech to sign language and vice versa. A range of (spoken and signed) languages could be supported and the signer might additionally be able to choose or individualize the signing avatar.

Concluding, the mentioned approaches are promising. In the future, they could enable sign language users to access personal assistants, to use text-based systems, to search sign language video content and to use automated real-time translation when human interpreters are not available (Bragg et al., 2019). With the help of AI, automated sign language translation systems could help break down communication barriers for DHH individuals.

REFERENCES

Adamo-Villani, N. and Wilbur, R. B. (2015). ASL-Pro: American Sign Language Animation with Prosodic Elements. In Antona, M. and Stephanidis, C., editors, *Universal Access in Human-Computer Interaction. Access to Interaction*, volume 9176, pages 307–318. Springer International Publishing, Cham.

Al-khazraji, S., Berke, L., Kafle, S., Yeung, P., and Huenerfauth, M. (2018). Modeling the Speed and Timing of American Sign Language to Generate Realistic Animations. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '18*, pages 259–270, Galway, Ireland. ACM Press.

Bavelier, D., Newport, E. L., and Supalla, T. (2003). Children Need Natural Languages, Signed or Spoken. Technical report, Dana Foundation.

Bragg, D., Verhoef, T., Vogler, C., Ringel Morris, M., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., and Kacorri, H. (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '19*, pages 16–31, Pittsburgh, PA, USA. ACM Press.

Camgoz, N. C., Hadfield, S., Koller, O., and Bowden, R. (2017). SubUNets: End-to-End Hand Shape and Continuous Sign Language Recognition. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3075–3084, Venice. IEEE.

Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., and Bowden, R. (2018). Neural Sign Language Translation. In

The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

De Martino, J. M., Silva, I. R., Bolognini, C. Z., Costa, P. D. P., Kumada, K. M. O., Coradine, L. C., Brito, P. H. d. S., do Amaral, W. M., Benetti, Â. B., Poeta, E. T., Angare, L. M. G., Ferreira, C. M., and De Conti, D. F. (2017). Signing avatars: Making education more inclusive. *Universal Access in the Information Society*, 16(3):793–808.

Deb, S., Suraksha, and Bhattacharya, P. (2017). Augmented Sign Language Modeling(ASLM) with interactiondesign on smartphone - an assistive learning and communicationtool for inclusive classroom.pdf. In *ScienceDirect - Procedia Computer Science 125 (2018) 492–500*, Kurukshetra, India. EvliseervBie.Vr .B.V.

Deutscher Gehörlosen-Bund e.V. (DGB). Gehörlosigkeit. http://www.gehoerlosen-bund.de/faq/geh%C3%B6rlosigkeit.

Duarte, A. C. (2019). Cross-modal Neural Sign Language Translation. In *Proceedings of the 27th ACM International Conference on Multimedia - MM '19*, pages 1650–1654, Nice, France. ACM Press.

Elliott, R., Glauert, J. R., Kennaway, J., Marshall, I., and Safar, E. (2008). Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4):375–391.

European Union of the Deaf (EUD) (2012). International Sign. https://www.eud.eu/about-us/eud-position-paper/international-sign-guidelines/.

Fang, B., Co, J., and Zhang, M. (2017). DeepASL: Enabling Ubiquitous and Non-Intrusive Word and Sentence-Level Sign Language Translation. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems - SenSys '17*, pages 1–13, Delft, Netherlands. ACM Press.

Fayyaz, S. and Ayaz, Y. (2019). CNN and Traditional Classifiers Performance for Sign Language Recognition. In *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing - ICMLSC 2019*, pages 192–196, Da Lat, Viet Nam. ACM Press.

Hanke, T. (2004). HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts. *LREC*, 4:1–6.

Kacorri, H., Huenerfauth, M., Ebling, S., Patel, K., Menzies, K., and Willard, M. (2017). Regression Analysis of Demographic and Technology-Experience Factors Influencing Acceptance of Sign Language Animation. *ACM Transactions on Accessible Computing*, 10(1):1–33.

Kato, M. (2008). A Study of Notation and Sign Writing Systems for the Deaf. page 18.

Kipp, M., Heloir, A., and Nguyen, Q. (2011a). Sign Language Avatars: Animation and Comprehensibility. In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Vilhjálmsson, H. H., Kopp, S., Marsella, S., and Thórisson, K. R., editors, *Intelligent Virtual Agents*, volume 6895, pages 113–126. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kipp, M., Nguyen, Q., Heloir, A., and Matthes, S. (2011b). Assessing the deaf user perspective on sign language avatars. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '11, pages 107–114, Dundee, Scotland, UK. Association for Computing Machinery.

Matheson, G. (2017). The Difference Between d/Deaf and Hard of Hearing. https://blog.ai-media.tv/blog/the-difference-between-deaf-and-hard-of-hearing.

McCarty, A. L. (2004). Notation systems for reading and writing sign language. *The Analysis of Verbal Behavior*, 20:129–134.

Morrissey, S. and Way, A. (2005). An Example-Based Approach to Translating Sign Language. In *Second Workshop on Example-Based Machine Translation*, page 8, Phuket, Thailand.

Polunina, T. (2018). An App That Translates Voice and Sign Language. https://wp.nyu.edu/connect/2018/05/29/sign-language-app/.

Pugeault, N. and Bowden, R. (2011). Spelling It Out: Real–Time ASL Fingerspelling Recognition. In *1st IEEE Work- Shop on Consumer Depth Cameras for Computer Vision, in Conjunction with ICCV'2011, Barcelona, Spain*, Barcelona, Spain.

Tkachman, O., Hall, K. C., Xavier, A., and Gick, B. (2016). Sign Language Phonetic Annotation meets Phonological CorpusTools: Towards a sign language toolset for phonetic notation and phonological analysis. *Proceedings of the Annual Meetings on Phonology*, 3(0).

World Federation of the Deaf (WFD). Our Work. http://wfdeaf.org/our-work/.