Supporting dialogue between healthcare professionals and Deaf patients through automated text-to-sign translation

Floris Roelofsen
Institute for Logic, Language and Computation
Amsterdam, the Netherlands
f.roelofsen@uva.nl

Lyke D. Esselink
Institute for Logic, Language and Computation
Amsterdam, the Netherlands
lyke.d.esselink@gmail.com

Shani E. Mende-Gillings
Institute for Logic, Language and Computation
Amsterdam, the Netherlands
semg98@gmail.com

Anika S. Smeijers
Amsterdam University Medical Centre
Amsterdam, the Netherlands
a.s.smeijers@amsterdamumc.nl

Abstract

It is challenging for healthcare professionals to communicate with Deaf patients,[1] even more so in times of COVID-19. Sign language interpreters can often not enter hospitals and face masks make lipreading impossible.

To address this urgent problem, we developed a system which allows healthcare professionals to translate sentences that are frequently used in the diagnosis and treatment of COVID-19 into Sign Language of the Netherlands (NGT). Translations are displayed by means of videos and avatar animations. The architecture of the system is such that it could be extended to other applications and other sign languages in a relatively straightforward way.

1 Background and motivation

Research has shown that it is difficult for healthcare professionals to communicate with Deaf patients (Fellinger et al., 2012), especially in times of COVID-19 (McKee et al., 2020). Sign language interpreters can often not enter hospitals and clinics, interpreting via video relay is not always viable, and face masks make lipreading impossible (Grote and Izagaren, 2020).

To understand how these difficulties are perceived by Deaf people, we conducted an online questionnaire in January and February 2021. 145 deaf sign language users from the Netherlands, aged 20 to 84, participated. The questionnaire was formulated both in NGT (through videos) and in written Dutch. The results (which will be presented in detail elsewhere) show that the inability of healthcare professionals to communicate in sign language is indeed perceived as a very significant threat. Specifically, 50% of all participants stated they were worried about getting COVID-19 because NGT interpreters are not always allowed into the hospital, while, for comparison, only 30% is worried about getting COVID-19 because friends and relatives would not be allowed to visit in the hospital.

To address this problem, we have developed a flexible, modular system which allows healthcare professionals to translate phrases that are frequently used in the diagnosis and treatment of hospital patients, in particular phrases related to COVID-19, from Dutch or English into Sign Language of the Netherlands. For a limited number of sentences, translations are displayed by means of pre-recorded videos. In addition, the system is able to generate translations that are displayed by means of an animated avatar. The present paper concentrates on describing the latter part of the system.

We have concentrated on Dutch and English as the source languages and Sign Language of the Netherlands (NGT) as the target sign language, mainly because of our familiarity with these languages, the Deaf community and the practice of healthcare professionals in the Netherlands. The general problem we aim to address, however, is not specific to the Netherlands but manifests itself worldwide.[2] Therefore, we have aimed to design

[1] It is common to refer to people who are part of the deaf community and primarily communicate using sign language as Deaf people, with a capital 'D'.

[2] The World Federation of the Deaf estimates that there are around 70 million Deaf people around the world: https://wfdeaf.org/.

the system in such a way that it could be extended to include other source and target languages in a relatively straightforward way. In this respect, our system diverges from some existing text-to-sign translation systems, which are tailor-made for a specific target sign language and not easily portable to other languages (see Section 3 below).

The system we have developed only translates text into sign language, not the other way around. This means, for instance, that if a doctor uses the system to ask a Deaf patient an open-ended question such as *How do you feel?*, and if the patient gives an elaborate answer in NGT, the doctor will most likely not be able to understand the answer and our system will not be of help in this case. Making this possible would require incorporating sign recognition technology (see, e.g., Zhou et al., 2020), which has been beyond the scope of our project so far. Note, however, that if a doctor uses our system to ask a more specific yes/no question such as *Do you feel dizzy?*, then the answer in NGT—involving a head nod in the case of *yes* and a head shake in the case of *no*—will most likely be perfectly clear for the doctor even without a general understanding of NGT. Thus, the current system is able to support relatively simple dialogues, but it is limited in scope and certainly does not (yet) offer a full-fledged dialogue system. We view it as a first, but critical step toward a more comprehensive solution.

A qualified human sign language interpreter should, whenever available, be preferred over a machine translation system. Still, it is worth investigating the extent to which a machine translation system can be of help in situations in which a human interpreter cannot be employed, especially in the medical setting where effective, instantaneous communication between healthcare professionals and patients can be of critical importance.

2 Brief background on sign language

Evidently, we cannot provide a comprehensive overview here of the linguistic properties of sign languages in general (see, e.g., Baker et al., 2016), nor of NGT in particular (see Klomp, 2021). We will, however, highlight some important features which any text-to-sign translation system needs to take into account.

First of all, sign languages have naturally evolved in Deaf communities around the world. This means that, contrary to a rather common misconception, there is not a single, universal sign language used by all Deaf people worldwide, but many different sign languages, just as there are many different spoken languages.

Second, Deaf people often have great difficulties processing spoken language even if presented in written form. The median reading level of deaf adolescents when finishing high-school is comparable to that of hearing children of around 8 years old (Wauters et al., 2006; Kelly and Barac-Cikoja, 2007). This may be surprising at first sight but not so much upon reflection. Imagine what it would be like as a native speaker of, say, English, to learn Hebrew or Greek just by looking at the characters, without being told how these characters are pronounced. Thus, for healthcare professionals to communicate with Deaf patients exclusively through written text would not be satisfactory, especially since in the medical setting it is critical to avoid miscommunication, to obtain reliable informed consent for interventions, and to foster an environment in which patients feel maximally safe.

Third, there is generally no direct correspondence between the sign language used in a given country and the spoken language used in that same country. For instance, while English is the mainstream spoken language both in the US and in the UK, American Sign Language (ASL) and British Sign Language (BSL) differ considerably from each other, as well as from spoken English. Such differences do not only pertain to the lexicon, but also to grammatical features such as word order. This means in particular that, to translate a sentence from English to ASL or BSL it does not suffice to translate every word in the sentence into the corresponding sign in ASL/BSL and then put these signs together in the same order as the words in the English sentence.

Fourth, signs are generally not just articulated with the hands, but often also involve facial expressions and/or movements of the head, mouth, shoulders, or upper body. These are referred to as the *non-manual* components of a sign. A text-to-sign translation system has to take both manual and non-manual components of signs into account.

Fifth, related to the previous point, non-manual elements are not only part of the *lexical* make-up of many signs, but are also often used to convey certain *grammatical* information (comparable to *intonation* in spoken languages). For instance, raised eyebrows may indicate that a given sentence is a question rather than a statement, and a head shake expresses negation. Such non-manual grammatical markers are typically 'supra-segmental', meaning that they do not co-occur with a single lexical sign but rather span across a sequence of signs in a sentence. Sign language linguists use so-called *glosses* to represent sign language utterances. For instance, the gloss in (1) represents the NGT translation of the question *Are you going on holiday?*.

(1) brow raise
YOU HOLIDAY GO

Lexical signs are written in small-caps. They always involve a manual component and sometimes non-manual components as well. The upper tier shows non-manual grammatical markers, and the horizontal line indicates the duration of these non-manual markers. In this case, 'brow raise' is used to indicate that the utterance is a question. A text-to-sign translation system should be able to integrate non-manual elements that convey grammatical information with manual and non-manual elements that belong to the lexical specification of the signs in a given sentence (Wolfe et al., 2011). This means that a system which translates sentences word by word, even if it re-orders the corresponding signs in accordance with the word order rules of the target sign language, cannot be fully satisfactory. More flexibility is needed: word by word translation can be a first step, but the corresponding signs as specified in the lexicon, must generally be adapted when forming part of a sentence to incorporate non-manual markers of grammatical information.

3 Relevant previous work

Sign synthesis A crucial prerequisite for text-to-sign translation is sign *synthesis*: the ability to create sign language avatar animations. Broadly speaking there are two ways to achieve this: key-frame animation (e.g., Delorme et al., 2009) and motion capture (e.g., Gibet et al., 2011).

While motion capture makes it possible to obtain a library of high-quality animations for lexical signs, a disadvantage of this technique is that animations for lexical sign obtained in this way are difficult to modify so as to incorporate non-manual grammatical markers (Courty and Gibet, 2010). In principle, the same problem also applies to libraries of lexical signs obtained by means of key-frame animation. However, in this case, there is a promising strategy to overcome the problem. Namely, rather than directly animating each lexical sign, it is possible to generate key-frame animations of lexical signs procedurally from structured specifications of the phonetic properties of these signs (Elliott et al., 2004). Such phonetic properties include (but are not limited to) the initial location, shape and orientation of the hands, possibly movements of the hands and other body parts, and facial expressions. Several formalisms have been developed to specify the phonetic properties of signs in a structured, computer-readable fashion (see Courty and Gibet, 2010, for an overview). Arguably the most extensively developed and most widely used formalism is the Sign Gesture Markup Language (SiGML) of Elliott et al. (2004), further extended in Glauert and Elliott (2011) and based on the HamNoSys notation originally developed for the annotation of sign language corpora (Prillwitz et al., 1989; Hanke, 2004). For illustration, our SiGML encoding of the NGT sign WHAT is given in Figure 1. As can be seen in the figure, both manual components (handshape, location, movement) and non-manual features (mouth, face, head) are encoded.

SiGML specifications can be converted into key-frame animations by the JASigning avatar engine (Elliott et al., 2004; Kennaway et al., 2007; Jennings et al., 2010). This approach makes it possible, in principle, to integrate non-manual grammatical markers with the lexical signs that make up a sentence, although such functionality has not yet been thoroughly implemented in systems based on SiGML and JASigning to our knowledge.

Given these considerations, we opted to use SiGML and JASigning as a basis for sign language synthesis, and to implement a new functionality to automate the integration of non-manual grammatical markers with lexical signs. A basic library of SiGML specifications of around 2000 lexical signs in NGT was already compiled in the course of previous projects (Zwitserlood et al. 2004; Kennaway et al. 2007; Prins and Janssen 2014; Esselink 2020). While we have had to extend this library with healthcare-related as well as some general-purpose signs (see Section 5.2 below), the availability of an initial repertoire of signs encoded in SiGML was essential for a timely development of the system.

Text-to-sign translation Turning now to previous work on text-to-sign *translation*, two general approaches can again be distinguished, differing mainly in the type of intermediate representation

```
<hamgestural_sign gloss="WAT">                                                          
    <sign_manual>                                                                       Manual
        <handconfig handshape="finger2" thumbpos="across"/>              ]
        <handconfig extfidir="u"/>                                       ] handshape
        <handconfig palmor="d"/>                                         ]
        <location_bodyarm location="shoulders" side="right_at"/>         ] location
        <wristmotion motion="swinging" size="small"/>                    ] movement
    </sign_manual>
    <sign_nonmanual>                                                                    Non-manual
        <mouthing_tier>                                                  ]
            <mouth_gesture movement="L30"/>                              ] mouth
        </mouthing_tier>                                                 ]
        <facialexpr_tier>                                                ]
            <eye_brows movement="FU" speed="0.8"/>                       ] face
            <eye_lids movement="SB" speed="0.8"/>                        ]
            <eye_gaze direction="AD" speed="0.8"/>                       ]
        </facialexpr_tier>                                               ]
        <head_tier>                                                      ]
            <head_movement movement='SL' />                              ] head
        </head_tier>                                                     ]
    </sign_nonmanual>
</hamgestural_sign>
```

Figure 1: SiGML encoding of the NGT sign WAT ('what').

that is employed in going from text to sign.

In the first approach, which we will refer to as the **gloss approach**, a given input sentence is transformed into a gloss of the corresponding sign language utterance. Next, based on this gloss representation, an avatar animation is generated.

(2) The gloss approach:
text $\Longrightarrow$ gloss $\Longrightarrow$ animation

This approach is taken, for instance, by HandTalk,[3] a Brazilian company that provides an automated text-to-sign translation service with Brazilian Portuguese / English as possible source languages and ASL / Brazilian Sign Language as possible target languages. HandTalk uses machine learning techniques to map input texts to the corresponding glosses, and a combination of key-frame animation and motion capture techniques to generate animations based on a given gloss.

In the second approach, which we refer to as the **phonetic approach**, the given input sentence is transformed into a sequence of phonetic representations of signs. Next, based on these phonetic representations, an avatar animation is generated.

(3) The phonetic approach:
text $\Longrightarrow$ phonetic rep. $\Longrightarrow$ animation

This approach has been taken in work based on SiGML and JASigning (Zwitserlood, 2005; Kennaway et al., 2007; Prins and Janssen, 2014; Battaglino et al., 2015; Ebling and Glauert, 2016; David and Bouillon, 2018, among others). Applying machine learning techniques to carry out the first step, from text to phonetic representations, is not feasible because it would require the availability of large parallel corpora of texts and the corresponding phonetic sign representations, which do not exist and would be very costly to create. The process of manually generating phonetic representations is highly time-consuming and requires expert knowledge of SiGML or a similar formalism. Rayner et al. (2016) have created a framework to ease this process, which is especially helpful if the sentences that need to be translated are all variations of a limited set of templates. For instance, the framework has been used successfully to develop an application for translating railway announcements (David and Bouillon, 2018). In less restricted domains, however, generating phonetic representations still requires expert knowledge of SiGML or similar formalisms and remains very time-intensive.

The gloss approach and the phonetic approach have complementary pros and cons. An advantage of the gloss approach is that it enables the use of machine learning technology to carry out the first part of the translation process. Disadvantages are that (i) the animation of each individual sign involves a lot of manual work, (ii) grammatical non-manual elements cannot be straightforwardly integrated with lexical signs, and (iii) all components of the system are tailor-made for a particular target sign language, i.e., no part of the system can be re-used when a new target language is considered. In particular, since no gloss-based system

[3] https://handtalk.me/en

currently exists for NGT, this approach was not viable for our purposes.

Advantages of the phonetic approach are that (i) grammatical non-manual features can in principle be integrated with lexical signs (though this possibility remains largely unexplored) and (ii) part of the system, namely the software that generates avatar animations based on phonetic representations (i.e., JASigning or a similar avatar engine) is not language-specific and can be used for any target sign language. The main disadvantage is that the initial step from text to phonetic representations involves a lot of manual work.

Based on these considerations, we opted for an approach that employs *both* a gloss representation *and* a phonetic representation in going from a given input text to an avatar animation of the corresponding sign language utterance. This *modular* approach is further described in Section 4.

4 A modular approach

As depicted in Figure 2, our modular approach breaks the translation process up into three steps: (i) a **gloss translation step**, which maps a given input sentence in Dutch or English to a gloss representation of the corresponding NGT sentence, (ii) a **phonetic encoding step**, which transforms the NGT gloss into a computer-readable phonetic representation, in our case in SiGML, and (iii) a **animation step**, which generates an avatar animation based on the phonetic representation.

Suppose, for instance, that we need to translate the Dutch/English sentence in (4).

(4) Waar doet het pijn?
Where does it hurt?

The first step is to convert this sentence into the corresponding NGT gloss in (5), where 'whq' stands for the non-manual marking that is characteristic for constituent questions in NGT. While empirical studies have found quite some variation in the actual realisation of 'whq' in NGT (Coerts, 1992; de Vos et al., 2009), furrowed eyebrows are seen as the most canonical realisation (Klomp, 2021).

(5) $\overline{\text{PAIN WHERE}}^{\text{whq}}$

The second step is to map this gloss representation to a phonetic representation in SiGML, a fragment of which is displayed in Figure 2. Finally, this SiGML representation is fed into the JASigning avatar engine, which generates an animation (as shown in Appendix A).

5 Implementation

In implementing the general approach outlined above, several choices still can/need to be made. Our choices in this regard were driven by the specific objective to address the urgent need for a translation tool to aid healthcare professionals in communicating with Deaf patients, ensuing from the current pandemic. Two requirements follow from this objective: (i) the system had to be developed within a short time-frame, and (ii) high accuracy of the delivered translations was more important than broad approximate coverage.

Our aim has therefore *not* been to automate the entire translation process. In particular, automating the process of mapping input sentences to the corresponding NGT glosses using machine learning techniques would not have been feasible within a short time-frame, and would, even in the somewhat longer term, most likely result in an unacceptably low accuracy rate for use in a healthcare setting.[4] We therefore mainly focused on the automation of the phonetic encoding step.

5.1 Collecting phrases for translation

We collected a set of phrases that are commonly used during the diagnosis and treatment of COVID-19, based on consultation with healthcare professionals at the Amsterdam University Medical Centre (AUMC) as well as direct experience (one of the authors is a medical doctor). We also consulted a list of phrases that was used in the SignTranslate system (Middleton et al., 2010).[5]

The resulting corpus was then divided into three categories: video-only, avatar-only, and hybrid. The first category, video-only, consisted mainly of sentences that could be divided into three further categories: emotional, complex, and informed consent. Sentences concerning the patient's emotional well-being require a high level of empathy to be

[4] Prins and Janssen (2014), who investigated the feasibility of automated sign language translation for children television programs, drew the same conclusion.

[5] The SignTranslate system was developed in the UK to translate 500 phrases that are common in the healthcare setting from English to British Sign Language. Translations were always displayed by means of pre-recorded videos, not by means of avatar animations. Since the system was developed more than 10 years ago, the phrases it could translate were evidently not specifically related to COVID-19. However, many general-purpose phrases are also relevant in the diagnosis and treatment of COVID-19.

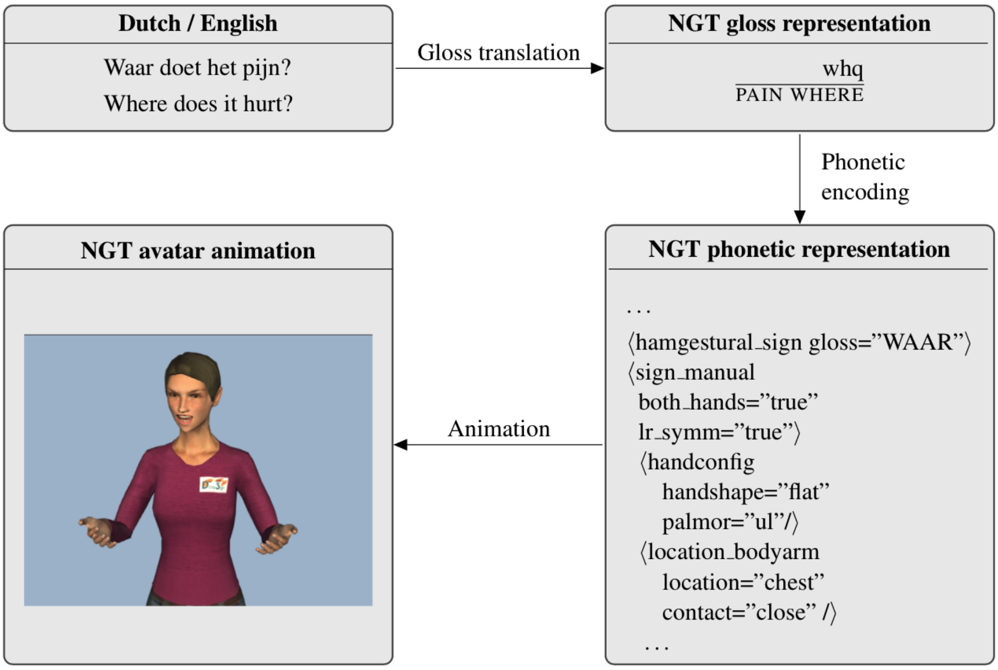


Figure 2: Overview of the translation process.

conveyed, which is difficult to achieve in a satisfactory way with an avatar given the current state of the art. We therefore deemed that video translations were necessary for these sentences. Sentences were classified as complex when they involved a combination of several statements and/or questions, or required a demonstration of pictures or diagrams along with an explanation (see Appendix B for an example). Finally, in the case of questions and statements concerning informed consent it is especially important to leave no room for potential misunderstandings. To ensure this, we chose to always offer video translations of these sentences.

The second category, avatar-only, consisted of sentences with many variations differing by only one word or phrase, indicating for instance the time of day or a number of weeks. It would not have been feasible to record a video translation for each version of these sentences.

The third category, hybrid, consisted of sentences that do not fall into one of the other two categories. For these, the system offers both a video translation and an avatar translation. In some cases, the avatar translation is slightly simplified compared to the video translation.

After categorising all of the sentences, those from the first and third category were translated into NGT and recorded by a team consisting of a sign language interpreter and a native Deaf signer. The Deaf signer that is visible in the videos was chosen for her clear signing style without a specific dialect, and her neutral reputation within the Deaf community. The level of complexity of the sentences involved was not deemed to require validation through backwards translation (Smeijers et al., 2014). Translations were checked by a sign linguist who is also a medical doctor. This resulted in a collection of 139 video translations. The sentences from the second and third category (including all variations) together comprised 7720 sentences for avatar translation.

5.2 Encoding lexical signs and non-manual grammatical markers in SiGML

Based on the obtained corpus, we determined for which lexical NGT signs and which non-manual grammatical markers we would need a SiGML representation in order for the system to be able to translate all the sentences. For some lexical signs, SiGML representations had already been created in previous projects (see references above). Based on our collection of video translations as well as the online video dictionary of the Dutch Sign Language Centre,[6] we created SiGML representations for 118 additional lexical signs, as well as three crucial non-manual grammatical elements: negation, yes/no question marking, and constituent question marking.

[6]https://www.gebarencentrum.nl

5.3 Constructing SiGML representations for full sentences

In order for the system to operate fast at run-time, we pre-processed all sentences and stored SiGML representations of their translations in a database. At run-time, the system only queries this database and does not compute any translations on the fly.

To construct the SiGML representations of full sentences, we implemented a program that, when given the gloss representation of a sentence in NGT, creates the SiGML code for that sentence. In this process, non-manual grammatical elements are integrated with the phonetic representations of lexical signs. For instance, in the case of yes/no questions, the program makes sure that the sentence ends with the general interrogative sign in NGT (palms up, raised eyebrows) and changes the non-manual component of the last sign before this general interrogative sign to include raised eyebrows, in line with what we observed in our collection of video translations. In the case of wh-questions, the general interrogative sign was also always appended at the end of the sentence. Although the use of this sign in questions is in fact *optional* in NGT (Coerts, 1992), we expect that it increases comprehension in the case of avatar translations.

5.4 User interface

We developed an online user interface. The user enters a sentence or a sequence of search terms. Based on their input they are presented with a list of available sentences from the database. These sentences differ depending on the translation mode chosen (video/avatar). After selecting a sentence the translation is offered in the chosen format.

As mentioned earlier, some of the possible input sentences differ only in one word or phrase. These sentences can be thought of as involving a general template with a variable that can take several values, such as a day of the week, a time of day, or a number of times / minutes / hours / days / weeks / months. When a user wants to translate such a sentence, they first select the template and then provide the intended value for the variable. For example, they may select the template "I am going to explain more at *time*", and then select a particular time (as illustrated in Appendix C).

While JASigning in principle offers a number of different avatars for sign language animation, there are differences in execution between these avatars. Our user interface therefore only makes use of one of the avatars, Francoise, and does not allow the user to choose between different options. We intend to further optimise the visualisation of the avatar in future work.

6 Discussion: video translations versus avatar animations

As a first step in evaluating the system we have consulted extensively with a prominent member of the Deaf community in the Netherlands who has years of experience in advising organisations (especially museums and hospitals) on how to make their services more accessible to Deaf people.

Based on these consultations and our own experiences in developing the system, we believe that the following considerations concerning the advantages and limitations of video and avatar translations in a healthcare setting will be helpful in guiding further work in this direction.

An important advantage of avatar technology is that it provides flexibility and scales up more easily than video translation. Once a library of animated signs has been created, and a procedure to integrate non-manual grammatical markers has been implemented, translations for many sentences can be generated. This makes it particularly straightforward to provide translations for sentences that differ only slightly from each other (e.g., in a phrase indicating the time of day).

A disadvantage, however, of avatar translations is that they can be less natural and more difficult to comprehend. While several empirical studies have reported promising comprehension rates for JASigning avatars (see, e.g., Kennaway et al., 2007; Smith and Nolan, 2016), our consultant indicates that certain avatar translations offered by our system may well be difficult to interpret for some users. Certain signs differ from each other only in rather subtle ways, and may be indistinguishable when produced by the JASigning avatar. Certain facial expressions and body movements of the avatar are quite unnatural, which can add to the difficulty of understanding translations. Certainly, the avatar's ability to display emotional empathy is very limited. This makes it undesirable to use avatar translations in situations where such empathy is required, as is often the case in medical settings.

Video translations, on the other hand, have their own benefits and drawbacks. They are better than avatar translations in terms of naturalness and comprehensibility, especially in the case of complex sentences. Moreover, our consultant indicates that patients are likely to feel more comfortable watching a video of a human interpreter rather than an animated avatar in a situation in which their physical well-being is at stake.

The main disadvantage of a video translation system is its inability to scale up efficiently. All translations have to be recorded separately, even ones that are almost identical. Cutting and pasting video fragments of individual signs to create new sentences does not yield satisfactory results.

Another disadvantage of a video translation system, though possibly less significant, is the difficulty of maintaining consistency among the translations that are offered. Every human interpreter has their own signing 'style'. Therefore, if the interpreter often changes from one video translation to the next, the patient will have to constantly readjust to a new signing style. Unfortunately, as the number of sentences in a corpus grows, it becomes less realistic to use the same interpreter for all translations.

A general advantage that a machine translation system (using either pre-recorded videos, or avatar technology, or both) may sometimes have over a human interpreter, especially in the healthcare domain, concerns privacy. A patient may receive sensitive information, and may not want this information to be known to anyone else. In this case, employing a human interpreter has a disadvantage (though this may of course be outweighed by the higher level of translation accuracy and empathy that can be provided by a human interpreter).

It is important to emphasise that constructing sign language translations in either format is a time-consuming affair, though for different reasons. Building a corpus of video translations is time intensive because every translation has to be recorded separately. For avatar translations, it takes time to encode individual signs. These are reusable, however, which becomes especially attractive as the number of required translations grows. However, the overall preference for one method over another is context-dependent: pros and cons should be carefully weighed in each specific context.

7 Conclusion and future work

We have investigated the potential of automated text-to-sign translation to address the challenges that the current pandemic implies for the communication between healthcare professionals and Deaf patients. We have motivated a modular approach to automated text-to-sign translation, and have built a first prototype system following this approach. We have discussed various prospects and limitations of the system.

In future work, we intend to extend the coverage of the present translation system and employ the general approach motivated here to develop text-to-sign translation systems for different domains, e.g., for announcements at airports or railway stations, an use case which has already been explored to some extent for other sign languages (Battaglino et al., 2015; Ebling and Glauert, 2016).

We also intend to employ the technology presented here to develop a system to support parents of deaf children in learning sign language. More than 95% of deaf children have hearing parents (Mitchell and Karchmer, 2004). These parents typically do not know any sign language when their child is born. Moreover, it is often difficult for them to learn a sign language well and fast enough to provide their child with sufficient accessible language input. Resources such as textbooks and dictionaries are scarce. Classes are expensive and often only subsidised to a very limited extent. In the Netherlands, less than 10% of hearing parents of deaf children communicate with their child in sign language (Knoors and Marschark, 2012). This percentage is presumably even much lower in countries where fewer resources are available. Consequently, many deaf children are at high risk of *language deprivation* during the first years of their lives—the critical period for language acquisition. A recent article in the bulletin of the World Health Organisation (Murray et al., 2019) draws attention to this alarming situation and its far-reaching repercussions. A late onset of language acquisition does not only dramatically lower the capacity to learn language throughout life (Mayberry et al., 2002), but also impacts a child's cognitive and socio-emotional development, and often leads to mental health problems across the lifespan (Humphries et al., 2016; Hall, 2017). We expect that a text-to-sign translation system especially designed to translate child-directed speech could be of great help to hearing parents in the first stages of learning to communicate with their children through sign language, and could therefore play an important role in reducing the risk of language deprivation and its detrimental repercussions.

Acknowledgments

We would like to express our gratitude to Bastien David, who provided essential help in the process of creating the SiGML representations for the additional lexical signs. ZonMW ERC Inge, John

References

Anne Baker, Beppie van den Bogaerde, Roland Pfau, and Trude Schermer. 2016. *The linguistics of sign languages: An introduction*. John Benjamins Publishing Company.

Cristina Battaglino, Carlo Geraci, Vincenzo Lombardo, and Alessandro Mazzei. 2015. Prototyping and preliminary evaluation of a sign language translation system in the railway domain. In *Universal Access in Human-Computer Interaction*, pages 339–350.

Jane Coerts. 1992. *Nonmanual grammatical markers: an analysis of interrogatives, negations and topicalisations in Sign Language of the Netherlands*. Ph.D. thesis, University of Amsterdam.

Nicolas Courty and Sylvie Gibet. 2010. Why is the creation of a virtual signer challenging computer animation? In *International Conference on Motion in Games*, pages 290–300. Springer.

Bastien Valentin C. David and Pierrette Bouillon. 2018. Prototype of Automatic Translation to the Sign Language of French-speaking Belgium. Evaluation by the Deaf Community. *Modelling, Measurement and Control C*, 79(4):162–167.

Maxence Delorme, Michael Filhol, and Annelies Braffort. 2009. Animation Generation Process for Sign Language Synthesis. *Advances in Computer-Human Interactions, ACHI*, 9:386–390.

Sarah Ebling and John Glauert. 2016. Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community. *Universal Access in the Information Society*, 15(4):577–587.

Ralph Elliott, John Glauert, Vince Jennings, and Richard Kennaway. 2004. An overview of the SiGML notation and SiGML signing software system. In *Workshop on the Representation and Processing of Sign Languages at the Fourth International Conference on Language Resources and Evaluation (LREC)*, pages 98–104.

Lyke Esselink. 2020. Lexical resources for sign language synthesis: The translation of Dutch to Sign Language of the Netherlands. Bachelor's thesis. University of Amsterdam, https://scripties.uba.uva.nl/search?id=715792.

Johannes Fellinger, Daniel Holzinger, and Robert Pollard. 2012. Mental health of deaf people. *The Lancet*, 379(9820):1037–1044.

Sylvie Gibet, Nicolas Courty, Kyle Duarte, and Thibaut Le Naour. 2011. The SignCom system for data-driven animation of interactive virtual signers: Methodology and evaluation. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 1(1):1–23.

John Glauert and Ralph Elliott. 2011. Extending the SiGML notation–a progress report. In *Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*.

Helen Grote and Fizz Izagaren. 2020. COVID-19: The communication needs of D/deaf healthcare workers and patients are being forgotten. *British Medical Journal*, 369.

Wyatte C. Hall. 2017. What you don't know can hurt you: The risk of language deprivation by impairing sign language development in deaf children. *Maternal and Child Health Journal*, 21(5):961–965.

Thomas Hanke. 2004. Hamnosys-representing sign language data in language resources and language processing contexts. In *LREC*, volume 4, pages 1–6.

Tom Humphries, Poorna Kushalnagar, Gaurav Mathur, Donna Jo Napoli, Carol Padden, Christian Rathmann, and Scott Smith. 2016. Avoiding linguistic neglect of deaf children. *Social Service Review*, 90(4):589–619.

Vince Jennings, Ralph Elliott, Richard Kennaway, and John Glauert. 2010. Requirements for a signing avatar. In *Workshop on Corpora and Sign Language Technologies at the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, pages 33–136.

Leonard P Kelly and Dragana Barac-Cikoja. 2007. The comprehension of skilled deaf readers. In Kate Cain and Jane Oakhill, editors, *Children's comprehension problems in oral and written language: A cognitive perspective*, pages 244–280. The Guilford Press.

Richard Kennaway, John Glauert, and Inge Zwitserlood. 2007. Providing signed content on the internet by synthesized animation. *ACM Transactions on Computer-Human Interaction*, 14(3):1–29.

Ulrika Klomp. 2021. *A descriptive grammar of Sign Language of the Netherlands*. Ph.D. thesis, University of Amsterdam.

Harry Knoors and Marc Marschark. 2012. Language planning for the 21st century: Revisiting bilingual language policy for deaf children. *The Journal of Deaf Studies and Deaf Education*, 17(3):291–305.

Rachel I. Mayberry, Elizabeth Lock, and Hena Kazmi. 2002. Linguistic ability and early language exposure. *Nature*, 417(6884):38–38.

Michael McKee, Christa Moran, and Philip Zazove. 2020. Overcoming additional barriers to care for deaf and hard of hearing patients during covid-19. *JAMA Otolaryngology–Head & Neck Surgery*, 146(9):781–782.

Anna Middleton, Alagaratnam Niruban, Gill Girling, and Phyo Kyaw Myint. 2010. Communicating in a healthcare setting with people who have hearing loss. *Bmj*, 341.

Ross E Mitchell and Michaela Karchmer. 2004. Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign language studies*, 4(2):138–163.

Joseph J Murray, Wyatte C Hall, and Kristin Snoddon. 2019. Education and health of children with hearing loss: the necessity of signed languages. *Bulletin of the World Health Organization*, 97(10):711–716.

Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hanke, and Jan Henning. 1989. HamNoSys version 2.0: an introductory guide.

Martin Prins and Joris B. Janssen. 2014. Automated sign language. TNO technical report.

Manny Rayner, Pierrette Bouillon, Sarah Ebling, Johanna Gerlach, Irene Strasly, and Nikos Tsourakis. 2016. An open web platform for rule-based speech-to-sign translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 162–168, Berlin, Germany. Association for Computational Linguistics.

Anika S. Smeijers, Beppie van den Bogaerde, M. Ens-Dokkum, and A.M. Oudesluys-Murphy. 2014. Scientific-Based Translation of Standardized Questionnaires into Sign Language of the Netherlands. *Investigations in healthcare interpreting*, pages 277–301.

Robert G. Smith and Brian Nolan. 2016. Emotional facial expressions in synthesised sign language avatars: a manual evaluation. *Universal Access in the Information Society*, 15(4):567–576.

Connie de Vos, Els van der Kooij, and Onno Crasborn. 2009. Mixed signals: Combining linguistic and affective functions of eyebrows in questions in Sign Language of the Netherlands. *Language and Speech*, 52(2-3):315–339.

Loes N Wauters, Wim HJ Van Bon, and Agnes EJM Tellings. 2006. Reading comprehension of Dutch deaf children. *Reading and writing*, 19(1):49–76.

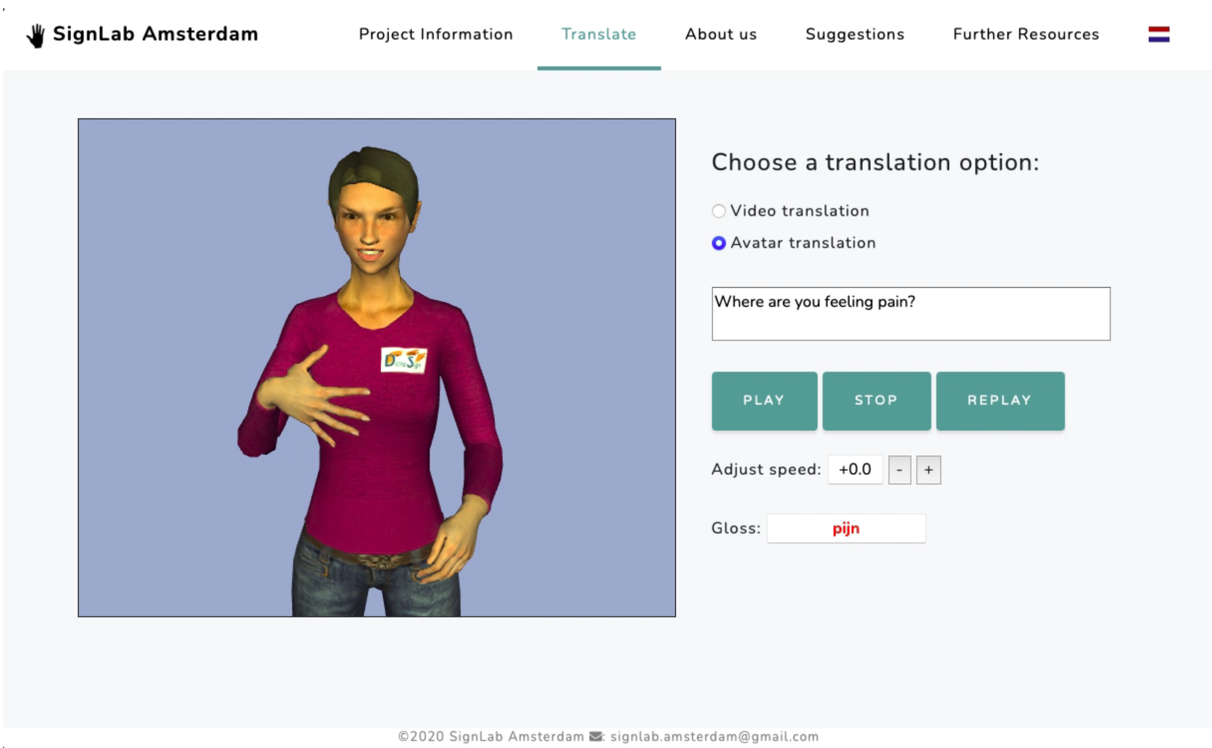
Rosalee Wolfe, Peter Cook, John C McDonald, and Jerry Schnepp. 2011. Linguistics as structure in computer animation: Toward a more effective synthesis of brow motion in American Sign Language. *Sign Language & Linguistics*, 14(1):179–199.

Zhihao Zhou, Kyle Chen, Xiaoshi Li, Songlin Zhang, Yufen Wu, Yihao Zhou, Keyu Meng, Chenchen Sun, Qiang He, and Wenjing Fan. 2020. Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nature Electronics*, 3(9):571–578.

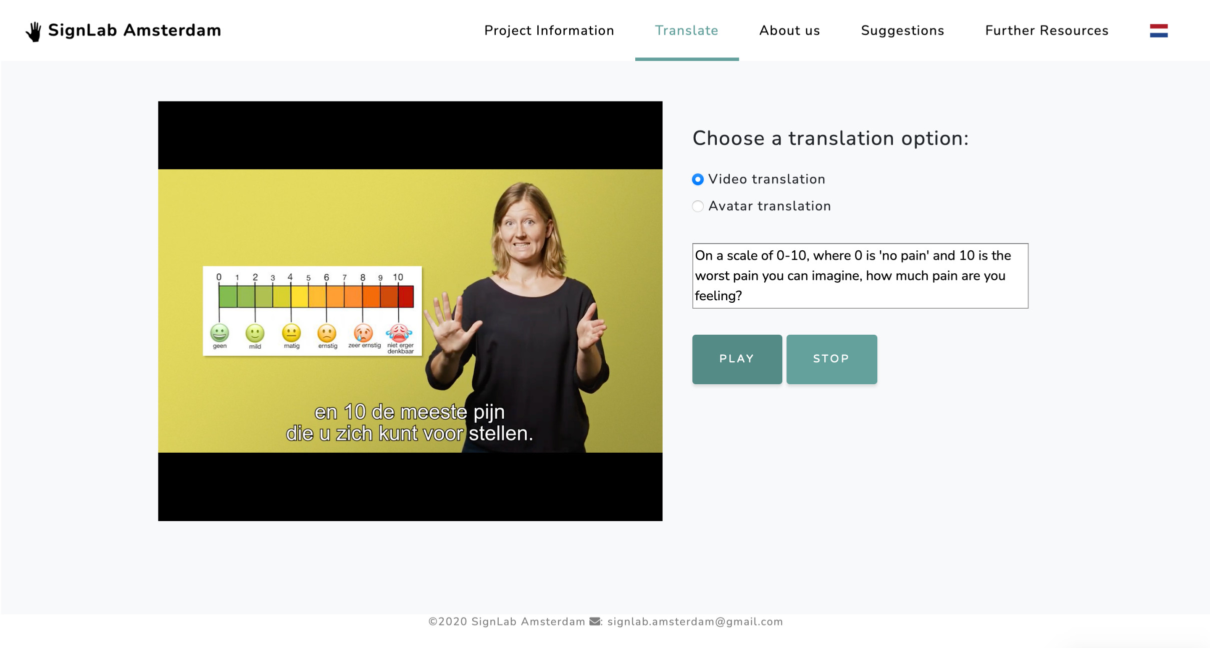
Inge Zwitserlood. 2005. Synthetic signing. In *The World of Content Creation, Management, and Delivery (IBC 2005)*, page 352–357.

Inge Zwitserlood, Margriet Verlinden, Johan Ros, and Sanny Van Der Schoot. 2004. Synthetic signing for the deaf: eSIGN. In *Workshop on Assistive Technologies for Vision and Hearing Impairment*.

A User interface example of avatar signing ’PAIN WHERE?’

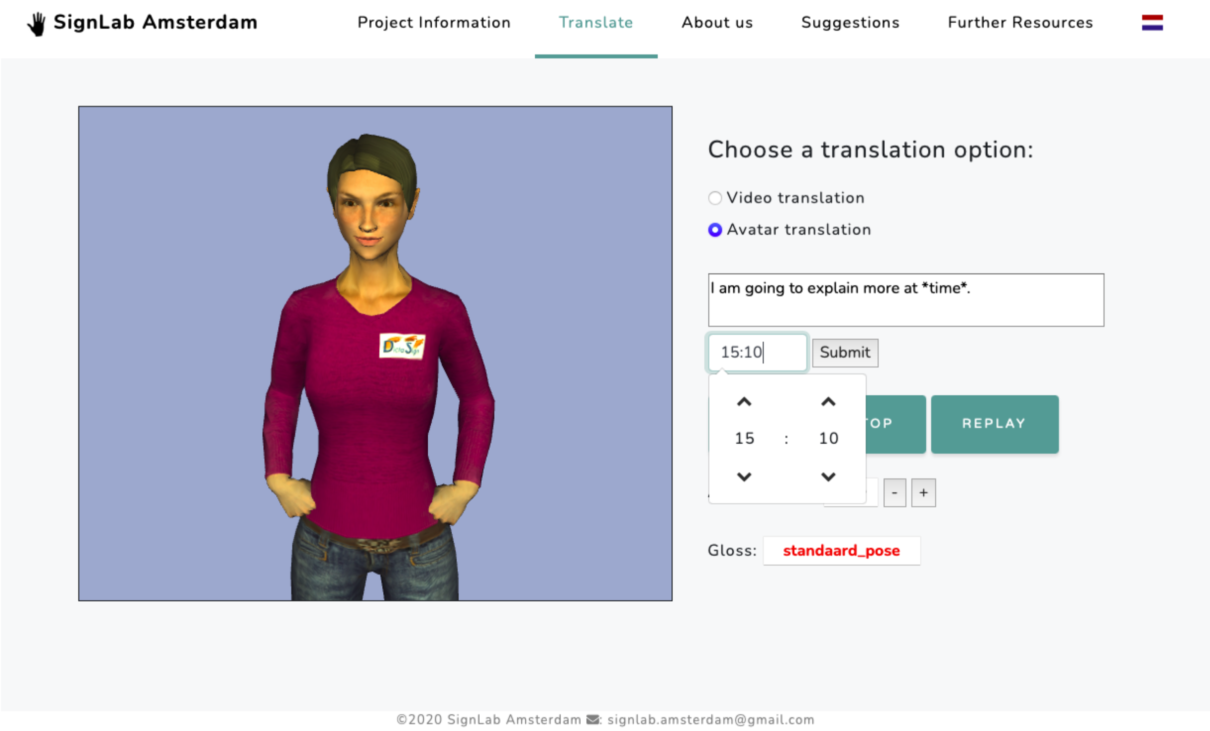


B User interface example of a video translation of a complex/emotional sentence



C User interface example of a template sentence

C.1 Variable input



C.2 Avatar signing template sentence