UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

First Steps Towards a Signing Avatar for Railway Travel Announcements in the Netherlands

van Gemert, B.; Cokart, R.; Esselink, L.; de Meulder, M.; Sijm, N.; Roelofsen, F.

Publication date
2022

Document Version
Final published version

Published in
Proceedings : 7th Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives (SLTAT 7)

License
CC BY-NC

Link to publication

Citation for published version (APA):
van Gemert, B., Cokart, R., Esselink, L., de Meulder, M., Sijm, N., & Roelofsen, F. (2022). First Steps Towards a Signing Avatar for Railway Travel Announcements in the Netherlands. In E. Efthimiou, S.-E. Fotinea, T. Hanke, J. C. McDonald, D. Shterionov, & R. Wolfe (Eds.), *Proceedings : 7th Workshop on Sign Language Translation and Avatar Technology: The Junction of the Visual and the Textual: Challenges and Perspectives (SLTAT 7): LREC 2022 Workshop : Language Resources and Evaluation Conference, 24 June 2022* (pp. 109-116). European Language Resources Association. https://aclanthology.org/2022.sltat-1.17

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:12 Nov 2026

Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology (SLTAT 7), pages 109–116
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022
© European Language Resources Association (ELRA), licensed under CC-BY-NC 4.0

First Steps Towards a Signing Avatar for Railway Travel Announcements in the Netherlands

Britt van Gemert, Richard Cokart, Lyke Esselink, Maartje de Meulder, Nienke Sijm, Floris Roelofsen
Universiteit van Amsterdam, Radboud Universiteit, Hogeschool Utrecht,
Nederlands Gebarencentrum, Nederlandse Spoorwegen
britt-van-gemert@live.nl, r.cokart@gebarencentrum.nl, l.d.esselink@uva.nl,
maartje.demeulder@hu.nl, nienke.sijm@hu.nl, f.roelofsen@uva.nl

Abstract

This paper presents first steps towards a sign language avatar for communicating railway travel announcements in Dutch Sign Language. Taking an interdisciplinary approach, it demonstrates effective ways to employ co-design and focus group methods in the context of developing sign language technology, and presents several concrete findings and results obtained through co-design and focus group sessions which have not only led to improvements of our own prototype but may also inform the development of signing avatars for other languages and in other application domains.

Keywords: text-to-sign translation, signing avatars, co-design, Dutch Sign Language, railway travel information

1. Introduction

This paper presents initial results of a project which aims to develop a sign language avatar for communicating railway travel announcements in Dutch Sign Language (Nederlandse Gebarentaal, NGT), in collaboration with the Dutch national railway company (NS).

For developing responsible and ethical signed language technologies that are adopted by deaf end users, interdisciplinary collaboration between specialists in Deaf Studies, Sign Linguistics, Computer Science, Artificial Intelligence, Human Computer Interaction, Language Policy, and Sign Language Interpreting Studies is essential (Bragg and others, 2019; Bragg and others, 2021; Yin et al., 2021). Such collaboration increases the quality of the developed technologies, ensures that they incorporate deaf communities' demands and values, and guarantees that there is consideration for design and appropriate user interfaces (De Meulder, 2021).

The present project is an example of such interdisciplinary collaboration. Our team consists of three deaf researchers with a background in Applied Sign Linguistics (Cokart, de Meulder) and Deaf Studies (de Meulder, Sijm) and three hearing researchers with a background in AI and Linguistics (Esselink, van Gemert, Roelofsen). Esselink and van Gemert have elementary proficiency in NGT, Roelofsen intermediate. Cokart and Sijm use NGT as their primary sign language and use it in different domains, De Meulder uses NGT primarily in professional contexts. Cokart, De Meulder and Sijm all have knowledge of various other sign languages and are involved in various deaf networks and communities.

The paper makes two contributions. The first is methodological: it exemplifies how co-design and focus group methods can be used effectively in the context of developing sign language technology, and offers some recommendations as to how these methods may be adapted to this specific purpose. The second is technological: it discusses several concrete findings and results obtained through co-design and focus group sessions which have not only led to improvements of our own prototype but may also inform the development of signing avatars for other languages and in other application domains.

2. Brief Background on Sign Languages

Evidently, we cannot provide a comprehensive overview here of the (socio)linguistic properties of sign languages in general (Baker et al., 2016), nor of NGT in particular (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 (Kusters and Lucas, 2022). 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 used on different scales by different deaf and hearing signers (Hou and de Vos, 2022).

Second, although sign languages exist in language ecologies in close contact with spoken languages, 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.

Third, making travel information available in the form of written text does not necessarily make it equally comprehensible for all deaf passengers. Depending on the complexity and time-sensitiveness of the message, textual information may be difficult to process, which may lead to misinterpretation. At the same time, the time-sensitive character of travel information entails specific demands concerning the comprehensibility of avatars.

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. These movement qualities (fluid movement) seem to be a crucial aspect for the rating of avatars by deaf end users (e.g. Quandt et al. (2021)).

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 often 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 *Have you already eaten?*.

(1) $\overline{\text{YOU EAT ALREADY}}^{\text{brow raise}}$

Lexical signs are written in small-caps. They always involve a manual component and often 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 thus 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, will not 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 carrying grammatical information.

Sixth, in the context of machine learning, just like some smaller spoken languages, (most) sign languages belong to the category of 'low-resourced languages', which refers to a lack of available training data and the fragmentation of efforts in resource development (Sayers et al., 2021). For sign languages there is the additional issue of a different language modality, which makes data collection and machine training much more challenging than for most spoken languages.

3. Related Work

We cannot provide a comprehensive overview of all work related to the present project. We restrict ourselves to highlighting some relevant work on (i) signing avatars for NGT, (ii) signing avatars in the railway domain, (iii) co-design and focus group methodologies, and (iv) user feedback on existing avatars.

Signing avatars for NGT Previous research on sign language technology for NGT is rather limited. Prins and Janssen (2014) developed a first prototype signing avatar for NGT to translate an episode of a Dutch TV program for children. Roelofsen et al. (2021) developed an avatar to address concerns in the Dutch deaf community during the COVID pandemic about the difficulty of communicating with healthcare professionals in case sign language interpreters would not be permitted into the hospital (Smeijers and Roelofsen, 2021). This avatar supports basic one-way communication from healthcare professionals to patients, e.g. to inform a patient about the results of their COVID test.

Signing avatars in the railway domain There has been discussion in the literature and the user communities about possible application domains of signing avatars. In general, announcements in public transportation are seen as a 'safe' application domain (Krausneker and Schügerl, 2021; WFD and WASLI, 2018) because their grammar is highly constrained and predictable, and the information that is shared is impersonal. This is different for application domains where the stakes are higher and miscommunication can potentially lead to life-threatening situations.

Prototype avatars for railway travel announcements have been developed for several sign languages, including Italian Sign Language (Battaglino et al., 2015), Swiss German Sign Language (Ebling and Glauert, 2016), and Sign Language of French-speaking Belgium David and Bouillon (2018).

The basic aim of Battaglino et al. (2015) was similar to ours, but the approach quite different. Their project involved a technical development phase and a quantitative assessment of the translation accuracy of the system. Our approach instead involves co-design and focus group methods so as to improve the system through various iterations. The findings we report are qualitative in nature rather than quantitative.

Closer to our project is the work of Ebling and Glauert (2016) and David and Bouillon (2018). These projects involved an initial development phase, a focus group session to collect suggestions for improvements, and a second development phase to implement these suggestions. These projects were similar to ours in that they used a qualitative method for evaluation, and involved multiple (in their case two) development iterations. One difference is that our project did not only include a focus group session, but also several co-design sessions, interleaved with multiple development iterations. Moreover, there is a difference in focus group methodology. These previous projects presented focus group members with a number of sentences signed by an avatar and elicited general feedback on the basis of these sentences. We instead discussed eight specific topics with our focus group members which had arisen during the co-design sessions. In each case we presented three different avatar animations for comparison, in order to make the discussion more targeted and to elicit more specific recommendations.

Co-design and focus group methodologies Co-design of sign language technologies with deaf end users improves the quality of the developed technologies, ensures appropriateness for the intended purpose, and stimulates acceptance. Conversely, lack of co-design may not only lead to sub-optimal technologies but also ones that could negatively impact deaf communities (Bragg and others, 2021).

Community-based co-design has been performed for several sign languages, including South African sign language (Blake et al., 2014). For example, for the design of a Deaf culture website, combining iterative co-design and focus group methods yielded insights in the native point of view and actionable insights on culturally rooted conventions for user experience (Pylvänen et al., 2013). Moreover, it can uncover hidden cultural norms, values, beliefs and attitudes (Chininthorn, 2021). Focus groups are used to elicit perceptions and opinions in early development stages. Young and Hunt (2011) emphasise the importance of avoiding visual distractions in focus group sessions: no busy walls or clothing, and a setting that ensures a good view for all participants.

User feedback User feedback has been collected for several existing signing avatars. We will highlight only some of the most recent studies, which involves state-of-the-art avatars. Krausneker and Schügerl (2021) compared perceptions of avatars vs. human interpreters through focus groups. Deaf participants criticized avatars for "lacking facial expressions, imprecise coordination of manual and non-manual components of a sign, missing phrase melody, jerky, hard, mechanical, wooden, robotic, somnolent, unnatural, incomplete signs and missing transitions between signs." They also mentioned "lack of mobility of upper body, shoulders, cheeks, unclear mouthing, comic face, artificial figure." Younger participants were often more familiar with the uncanny valley effect. Older participants sometimes felt that it was inappropriate that they were informed by a playful cartoon character, because they felt it exacerbated the infantilisation of sign languages — and by extension, deaf people (see also Wolfe et al. (2021)). Participants further reported that "maximal cognitive attention" was needed to understand the avatar.

Quandt et al. (2021) found that deaf respondents rated an avatar based on motion capture significantly more positively than an avatar based on (scripted) keyframe animation, but still not as positively as a human signer. Participants who had learned ASL later in life were more open to signing avatars in general, but also gave more negative ratings to the avatar based on keyframe animation. Participants who learned ASL earlier in life were more sensitive to movement quality issues in the keyframe animation avatar.

4. Design Process

4.1. Phase 1: Initial Design

First, we obtained a list of railway announcement templates from NS (e.g. 'The intercity train to destination X departs at time Y from platform Z'). The Dutch Sign Language Centre (Nederlands Gebarencentrum, NGc) provided NGT translations of these templates (for randomly picked X's, Y's, and Z's).

We created an initial basic system based on these translations. The signing avatar mimicked the video translations as closely as possible, and had the ability to sign several variations of the announcement templates (with different X's, Y's, and Z's). We made use of the JASigning avatar engine for implementation of the avatar (e.g. Ebling and Glauert (2016)). This engine takes phonetic representations of signs as input (specified in the Sign Gesture Markup Language, SiGML for short) and yields an avatar animation as output. This approach allowed us to efficiently create a large number of variations of the given set of templates, without expensive equipment (e.g. motion capture).

In addition, we developed an online user interface to facilitate further development of the basic initial system in subsequent phases of the project.

A general strategy for inclusive collaboration methods was defined through a brainstorm session involving two deaf and two hearing researchers. We opted for a *combination* of multiple co-design sessions and a focus group. The former allows for iterative development with a relatively small, highly engaged and fully dedicated team. The latter ensures input from a larger and more diverse group of potential end users. We envisioned that a combination of the two methods would work particularly well because the co-design sessions could result in specific topics to be discussed in the focus group. Indeed, we feel that this has made the focus group particularly fruitful (see below).

4.2. Phase 2: Iterative Co-design

4.2.1. Method

We held three co-design sessions (2x2 hours on campus with two deaf researchers, two hearing, and two interpreters; 1x1 hour online with one deaf researcher, two hearing, and one interpreter). These sessions focused on improving various aspects of the avatar's signing (e.g., manual movements, facial expressions, mouthing, grammar, transitions between signs) as well as non-linguistic aspects of the animations (e.g., camera angle, speed). Following each session, all suggestions made by the deaf researchers were implemented by the developers, often in several variations, so that they could be reconsidered and possibly further refined during the next session. In some cases, suggested improvements were implemented on the fly and evaluated immediately.

4.2.2. Results

The co-design sessions led to major adjustments of the avatar, pertaining to both linguistic and non-linguistic aspects. Below we discuss a selection of these adjustments.

Greeting All NS announcements start with a greeting, BESTE REIZIGERS *(DEAR PASSENGERS)*. This is not a natural greeting in NGT. At first, we removed the greeting entirely. However, it also functions as a way of getting people's attention and provides a 'time buffer', so that passengers don't miss the first part of the actual announcement. As an alternative, we opted for the greeting HALLO (HELLO), which is commonly used in NGT, both in formal and in informal settings. The avatar initially signed HALLO with a 5-handshape, with all selected fingers stretched. This looked unnatural. We adapted the sign, opting for a handshape that lies between the 5-handshape and the B1-handshape. A subtle difference, but the resulting sign looks substantially more natural.

(a) HALLO - Version 1 (b) HALLO - Version 2

Fig. 1: HALLO - Multiple versions

Mouthings The JAsigning avatar engine offers limited possibilities to produce natural-looking mouthings. The engine requires a specification of the mouthing in SAMPA notation. But SAMPA is a notation system for *phonemes*, and there is no one-to-one mapping between phonemes and mouth movements. For instance, the 's' in 'sun' and the 's' in 'silver' involve different mouth shapes. This makes it difficult to generate correct mouthings for NGT in JAsigning, and in several cases we did not succeed in doing so. For instance, the mouthing for VIJFENVIJFTIG *(fiftyfive)* was initially coded in SAMPA as 'vE_ifv@nvE_iftIx'. After multiple adjustments we ended up with 'vE_ifE_iftI'. While this improved the animation, the last part 'tig' is still unsatisfactory.

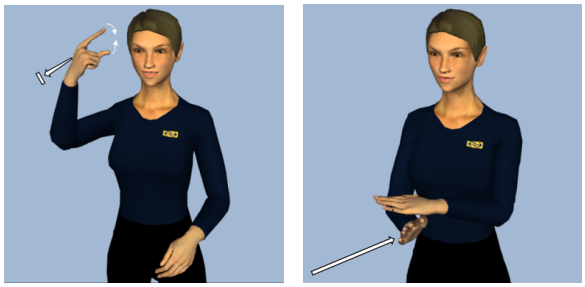
Formal vs informal registers The distinction between formal and informal registers proved to be highly relevant for the perception of the avatar. Clear signing is not sufficient; the avatar's signing style and choice of vocabulary also need to fit the particular context of use. For instance, the sign WEGGAAN (*LEAVE*) which is frequently used in casual interactions (e.g., in 'the train already left') was deemed too informal for official announcements and was replaced by the more formal sign VERTREKKEN (*DEPART*).

(a) WEGGAAN (b) VERTREKKEN

Fig. 2: WEGGAAN and VERTREKKEN

Intensity Preferences regarding the intensity of body movements and facial expressions varied. On the one hand, for station names such as ENSCHEDE and UTRECHT CENTRAAL, the manual movements and facial expressions of the avatar were considered too exaggerated, even aggressive, and had to be 'toned down'. On the other hand, for certain other signs (e.g. BIJNA *(ALMOST)*) they were considered too subtle and had to be intensified.

Transitions In phrases of the form NAAR X *(TO X)*, where X is the name of some destination, the transition between the two signs was sometimes unnatural. For instance, as can be seen in Figure 3(a), the path movement of the sign NAAR ends by default in the upper right corner of the signing space (from the perspective of the signer), but if the initial position of the subsequent sign, e.g. ALMELO in Figure 3(c), is in the upper left corner of the signing space, there is an unnatural prolonged transition between the two signs. This issue was also observed for other destinations, such as MAASTRICHT and AMSTERDAM. This was resolved by manually adapting the sign NAAR whenever needed. Ideally, however, future iterations of the system would be able to automatically adjust the direction

of signs like NAAR, depending on the next sign.

Eye gaze In several phrases, the avatar's eye gaze was too static. For instance, when a destination is signed (e.g., Amsterdam), it is natural for the eyes to be directed at the location of the sign in the signing space and to follow the path movement of the sign.

(a) NAAR (b) Transition (c) First pose of ALMELO

Fig. 3: Transition between NAAR and ALMELO

Camera angle Initially, the camera angle was front-view, the default in JAsigning. This, however, resulted in poor visibility for some signs, e.g., VIJFTIEN *(FIFTEEN)*. By changing the camera angle 13° to the left, and adjusting the head position and eye gaze of the avatar in such a way that she still faced the addressee by default, visibility was significantly improved.

Sentence structure Due to grammatical differences between NGT and Dutch, the signed sentence structure sometimes had to be adjusted. For example, in NGT, certainties should be positioned at the start of the sentence and uncertainties at the end (e.g. 'over een nog onbekende tijd' *(in a yet unknown time)* should be positioned at the end). The structure of sentences containing a final destination and several intermediate destinations was adapted as well: the preference was to mention the final destination before the intermediate destinations. In both cases, the preferred structure in NGT differs from the structure of the original NS announcements in Dutch.

Lists In phrases like *the train to Almelo, Hengelo and Enschede*, the avatar initially used a 'count hand' to list the three destinations, a grammatical construction that is commonly used in NGT for conjunctions and lists. In the present context, however, this gave the wrong impression that the announcement concerned multiple trains. Therefore, all count hand signs were removed.

Indexing There was much discussion about the appropriate use of INDEX signs. For instance, the video translations that served as our point of reference rendered the phrase *The intercity to...* as INDEX INTERCITY NAAR, where the function of the INDEX sign was to place the intercity in the signing space for future anaphoric reference. While grammatically correct, this usage of the INDEX sign seemed superfluous if the announcement did not involve any anaphoric reference to the intercity (which was the case in most announcements). A similar issue arose for phrases like *from platform two*, which were translated as VAN INDEX SPOOR TWEE. No consensus on this issue was established during the co-design process. Moreover, it was suggested that some INDEX signs, if present at all, should be shorter (less prominent) than others. We decided to create several variants for a number of sentences, with index signs present or absent in several positions, and with shorter or longer movements, to be further discussed in the focus group session.

Visual elements If an announcement is a repetition of a previously made announcements, its spoken version always starts with *Herhaling* (*Repetition*). In the signed version, however, starting with the sign HERHALING would not be effective for passengers who miss the beginning of the announcement. We therefore removed the sign and instead added a red bar under the animation displaying the text *Herhaling* to indicate the repetition.

Appearance The avatar's fingers were perceived as being unnaturally long. This affected the appearance of some hand shapes, e.g. in SPOOR *(PLATFORM)*. When properly signed, SPOOR involves a baby-C handshape with extended fingers. Due to the the long fingers of the avatar, this baby-C handshape had an unnatural curved shape.

The deaf researchers in the team also commented on the general appearance of the avatar. It was perceived as somewhat grumpy, not friendly. We added a smile right at the beginning of each announcement, before and during the greeting HALLO. This was an improvement, but a more friendly-looking avatar should be developed/adopted in future work (the JAsigning engine is limited in this regard – it includes some avatars other than the one we used, but not ones that are more suitable for our present purposes). Users should preferably also be able to adapt the clothing of the avatar, and to choose a male, female or androgyn-looking avatar.

Semantic refinement In some cases, a Dutch phrase cannot be univocally translated to NGT without making more specific what its intended semantic interpretation is. For instance, the proper translation of *De trein naar ... rijdt niet* (*The train to ... is cancelled*) depends on whether it's just a single train that is cancelled or the problem is structural. In the first case, the phrase RIJDEN NIET *(DEPART NOT)* is used, where NIET is signed with a 1-handshape moving from a central position in the signing space towards the upper right corner accompanied by a headshake, while in the second case the sign ANNULEREN *(CANCEL)* is more appropriate (drawing a cross in the signing space).

Times and numbers In phrases like INTERCITY NAAR AMSTERDAM TIJD TIEN TWINTIG (*INTERCITY TO AMSTERDAM TIME TEN TWENTY*), it is clear that the numeral phrase TIEN TWINTIG refers to the departure time. The sign TIJD was felt to be redundant and was therefore removed. Instead the preposition VAN *(OF)* was inserted, corresponding to the preposition that is used in Dutch, e.g., *De trein*

van 10:20 (‘*The train of 10:20*’). No consensus was reached on whether to include a sign for the ‘:’ symbol in times like 10:20, and if so, which sign. We created several variations to be discussed further during the focus group session.

Topics for focus group As already alluded to in several places above, we identified a number of specific topics during the co-design sessions that required further discussion in a larger and more diverse group of deaf people (e.g. indexing, time punctuation). For this purpose, multiple variations were created to facilitate comparison and stimulate targeted discussion.

4.3. Phase 3: Focus Group

4.3.1. Method

A 3-hour focus group session with six participants was held. Participants were selected by the deaf team members from their personal and professional network. They represented different regions, age groups and school backgrounds (see Table 1). In advance, participants received a link to an online demo of the avatar. Specific topics for discussion were not sent in advance.

	Age	Home, work and school region
D1	31-40	Noord-Holland, Utrecht, Groningen
D2	41-50	Noord-Holland, Utrecht, Noord-Brabant
D3	18-30	Utrecht, Gelderland, Groningen
D4	51-60	Utrecht, Groningen
D5	18-30	Noord-Holland, Groningen
D6	41-50	Flevoland, Amsterdam

Table 1: Focus Group Participants

The session was held at the University of Amsterdam in a room with a big screen. One team member acted as host and moderator (hearing, intermediate signer), one team member controlled the screen and took detailed minutes (hearing, minimal knowledge of NGT, developer), and two team members took part in the discussion (both deaf). One NGT-English interpreter was present. Having a signing moderator who is familiar with the research terms and project itself is an advantage over solely communicating through the interpreter (less engaging, time lag between communication types) (Orfanidou et al., 2014; Harris et al., 2009).

The discussion concentrated on eight topics determined by the team in advance, including time punctuation, subtitles, animation speed, mouthing, indexing, pauses and choice of vocabulary. In each case, three variants of a sign or a phrase were presented for comparison and participants discussed their perspectives and opinions. At the end of the focus group, participants were asked how and where they would like to see the avatar put to use.

4.3.2. Results

We provide an overview of our main findings.

Subtitles We asked participants whether subtitles might be helpful, and if so, which format would be preferred (per sign vs per sentence). Participants indicated that subtitles could indeed be useful, and had a clear preference for subtitles for entire sentences rather than for individual signs. If the text in the subtitles is used as an information source, displaying the entire sentence at once makes it easier to obtain complete information at once.

Animation speed The JASigning avatar engine offers speed adjustments ranging from 0.00 to +3.00. Participants of the focus group considered an animation speed of +0.40 optimal for comprehension. Lower speed was perceived as too slow. For most participants, a higher speed (e.g. +0.60) was comprehensible as well, but required more cognitive effort.

Indexing We asked participants for their preference concerning the use of INDEX signs (see Section 4.2). They indicated a preference for the use of INDEX sign even if they were strictly speaking redundant, but also indicated that INDEX signs should by default be subtle and not too prominent, often involving just a change of handshape and/or a subtle movement of the wrist, otherwise keeping the body and arms roughly in the same position as where the previous sign ended. For instance, the preferred translation of *The intercity to Almelo departs from platform five* was: INDEX1 (subtle) INTERCITY NAAR ALMELO INDEX1 (subtle) VERTREKKEN SPOOR 5 INDEX2 (subtle).

Time punctuation No consensus was reached for time punctuation, i.e., the sign for the ‘:’ in times like ‘15:31’. In fact, among our six participants, three different signs were used, and preferences seemed to depend on age group.

Personal pronouns In some sentences the avatar used a first person pronoun IK/WIJ *(I/WE)*. The original Dutch announcements of NS involve *impersonal* pronouns instead, but these do not have a direct translation in NGT. However, our focus group participants indicated that the use of first person pronouns was not suitable, as this suggested that the avatar herself was the source of the information, rather than NS.

Pauses In general, animations without any pauses between signs were preferred, or with very short pauses based on the syntactic structure of the sentence (i.e., somewhat longer pauses between conjoined sentences and shorter ones between noun phrases).

User interface Participants made some specific user interface suggestions. They indicated that it would be useful for the avatar to be displayed on screens at train stations and in trains, as well as in the mobile NS app. Drawing passengers’ attention before an announcement starts is essential – otherwise, passengers might miss part of the announcement. At train stations and in trains, flickering lights on the ground could serve this purpose. In the mobile app, a vibrate alert would be a natural choice, and passengers should be enabled

to replay the announcement if they want to. It would be good if deaf users of the mobile app could choose to receive automatic alerts for announcements related to their personal itinerary.

5. Discussion and Conclusion

The combined expertise from various disciplines in the co-design process and the input from a diverse focus group led to significant improvements of our prototype (manual movements, facial expressions, mouthing, grammar, transitions between signs, camera angle, speed), many of which may well be transferable to other languages and application domains.

It is evident, however, that the development of a fully satisfactory signing avatar for railway announcements in NGT requires much further work. Below we highlight some specific limitations of the methodological choices we made and the results obtained so far, as well as some natural avenues for further research.

5.1. Initial design

Reference material As an initial point of reference for our avatar translations, we obtained video recordings of the Dutch Sign Language Centre. However, sign languages are 3D, and videos 2D. In several cases, the video translations, filmed with a front-view camera angle, were not quite sufficient as reference material, since signs could not be viewed from multiple directions. Additional video's with different camera angles could possibly resolve this issue.

Avatar engine The JAsigning avatar engine which we used to generate avatar animations currently has a number of limitations. In particular, important features of the overall appearance of the avatar cannot be adjusted (e.g. excessively long fingers, somewhat unfriendly look) and control over mouth movements and facial expressions is too restricted. Such limitations form a real bottleneck for the development of a truly satisfactory signing avatar. Overcoming them would require substantial further development of the engine, or alternatively, adopting an altogether different approach to generating animations based on motion capture. We aim to explore both routes in future work.

5.2. Co-design

Iterative nature The iterative nature of our co-design process resulted in a thorough analysis of several aspects and considerable changes in the design of the avatar. Nevertheless, many other aspects had to be left for future iterations (e.g. eye gaze direction, facial expressions, topic-marking).

Live vs online sessions Three co-design sessions took place, of which two live and one online. We feel that the live sessions were much more effective, because they ensured a good view of the participants' signing, including their facial expressions and body language. Moreover, they provided the possibility to point out some details on a big screen, write on a chalkboard and point at the screen or at the board while signing the same time. The duration of the live sessions (2 hours) was quite demanding. In the future, more and shorter sessions would be preferable.

Interpreters During the sessions, interpreter(s) were present and some of the communication happened indirectly. Signers sometimes had to wait for interpreters to catch up, become familiar with research terms, or repeat signs so that the person taking notes could see the intended movements. Working with the same interpreters during all sessions is beneficial for familiarity with the relevant terms. However, it should always be kept in mind that if there is no shared language among all researchers and therefore some of the communication has to be mediated by an interpreter, there is always a higher chance of miscommunication. *Iterative* co-designs overcomes this issue to some extent: possible misunderstandings are often identified when suggestions are implemented and re-evaluated.

5.3. Focus Group

Recording vs minutes Detailed minutes were taken during the focus group. However, these minutes only provide a textual transcription, mediated by an interpreter, of what was actually signed during the session. This loss of information could be overcome by capturing the discussion on video, with multiple cameras to ensure a good view of all participants. This would also prevent overlooking information when multiple participants are signing at the same time – in many such cases, interpreter-mediated transcriptions will only capture what one of the participants signed. We should note that in order to make such video data searchable and usable for analysis it would have to be annotated in quite some detail, which would be a labor intensive process. But the information retained in this way could be very beneficial.

Developer presence The presence of the developers during our focus group may well have affected the discussion, as participants may have felt less comfortable criticizing the system. On the other hand, not having developers present during a focus group sessions would result in less direct input and would take away the possibility of directly implementing and evaluating certain suggestions.

Generalising results Our focus group was quite diverse in terms of age group and region. However, for further development it is necessary to organise more focus groups with more diversity in terms of age group, region, educational level, and reading level, among other things. For example, seniors and people from the southern part of the Netherlands were not represented in our focus group. Moreover, use of the avatar in a real-life setting is under-researched. This may affect our current results, especially given the time-sensitiveness of the context in which the avatar needs to provide information.

6. Acknowledgements

We are grateful to the participants of our focus group, and to Martijn van Beek, Babette van Blijenburgh, Stijn van den Brand, Bastien David, Marjon Fonville, John Glauert, Richard Kennaway, and Martha Larson for their help with various aspects of the project. We gratefully acknowledge financial support from NS and the Netherlands Organization for Scientific Research (NWO).

7. Bibliographical References

Baker, A., van den Bogaerde, B., Pfau, R., and Schermer, T. (2016). *The linguistics of sign languages: An introduction*. John Benjamins.

Battaglino, C., Geraci, C., Lombardo, V., and Mazzei, A. (2015). Prototyping and preliminary evaluation of a sign language translation system in the railway domain. In Margherita Antona et al., editors, *Universal Access in Human-Computer Interaction*, pages 339–350.

Blake, E., Tucker, W., and Glaser, M. (2014). Towards communication and information access for deaf people. *South African Computer Journal*, 54:10–19.

Bragg, D. et al. (2019). Sign language recognition, generation, and translation: An interdisciplinary perspective. In *The 21st international ACM SIGACCESS conference on computers and accessibility*, pages 16–31.

Bragg, D. et al. (2021). The fate landscape of sign language ai datasets: An interdisciplinary perspective. *ACM Transactions on Accessible Computing (TACCESS)*, 14(2):1–45.

Chininthorn, P. (2021). Community-based co-design for accessible health information for deaf people in a context with societal complexity.

David, B. V. C. and Bouillon, P. (2018). Prototype of Automatic Translation to the Sign Language of French-speaking Belgium. *Modelling, Measurement and Control C*, 79(4):162–167.

De Meulder, M. (2021). Is "good enough" good enough? Ethical and responsible development of sign language technologies. In *Proc. of the 1st Int. Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 12–22.

Ebling, S. and Glauert, J. (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.

Harris, R., Holmes, H. M., and Mertens, D. M. (2009). Research ethics in sign language communities. *Sign Language Studies*, 9(2):104–131.

Hou, L. and de Vos, C. (2022). Classifications and typologies: Labeling sign languages and signing communities. *Journal of Sociolinguistics*, 26(1):118–125.

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

Krausneker, V. and Schügerl, S. (2021). Best practices protocol on the use of sign language avatars.

Kusters, A. and Lucas, C. (2022). Emergence and evolutions: Introducing sign language sociolinguistics. *Journal of Sociolinguistics*, 26(1):84–98.

Orfanidou, E., Woll, B., and Morgan, G. (2014). *Research methods in sign language studies: A practical guide*. John Wiley & Sons.

Prins, M. and Janssen, J. B. (2014). Automated sign language. TNO technical report.

Pylvänen, S., Raike, A., Rainò, P., et al. (2013). Co-design for accessibility in academia for deaf students. *Co-Create 2013*.

Quandt, L. C., Willis, A., Schwenk, M., Weeks, K., and Ferster, R. (2021). Attitudes toward signing human avatars vary depending on hearing status, age of signed language exposure, and avatar type. Manuscript archived at PsyArXiv, June 25, doi:10.31234/osf.io/g2wuc.

Roelofsen, F., Esselink, L., Mende-Gillings, S., and Smeijers, A. (2021). Sign language translation in a healthcare setting. In *Proceedings of the Translation and Interpreting Technology Online Conference*, pages 110–124.

Sayers, D., Sousa-Silva, R., Höhn, S., et al. (2021). The dawn of the human-machine era: A forecast of new and emerging language technologies. *Report for EU COST Action CA19102 'Language in the Human-Machine Era'*.

Smeijers, A. S. and Roelofsen, F. (2021). Communicatiebehoeften en ervaringen van dove patiënten in Nederland tijdens de COVID-19 pandemie (communication needs and experiences of deaf patients in the netherlands during the covid-19 pandemic). Dutch report available through `https://zorgbeter.info/`, with per-section summaries in NGT.

WFD and WASLI. (2018). World Federation of the Deaf and World Association of Sign Language Interpreters Statement on Use of Signing Avatars. `https://wfdeaf.org/news/wfd-wasli-issue-statement-signing-avatars/`.

Wolfe, R., Cook, P., McDonald, J. C., and Schnepp, J. (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.

Wolfe, R., Mcdonald, J., Efthimiou, E., Fotinea, E., Picron, F., Van Landuyt, D., Sioen, T., Braffort, A., Filhol, M., Ebling, S., et al. (2021). The myth of signing avatars. In *1st International Workshop on Automatic Translation for Signed and Spoken Languages*.

Yin, K., Moryossef, A., Hochgesang, J., Goldberg, Y., and Alikhani, M. (2021). Including signed languages in natural language processing. *arXiv preprint arXiv:2105.05222*.

Young, A. and Hunt, R. (2011). Research with d/deaf people. *Methods Reviews*, 9.