Master's Thesis in Artificial Intelligence

Radboud University Nijmegen

Co-designing a Sign Language Avatar for Railway Announcements

---

Author:
Britt van Gemert
s4555740
B.vanGemert@student.ru.nl

Internal Supervisor:
Prof. M.A. Larson
m.larson@cs.ru.nl
Radboud University

External Supervisors:
Dr. F. Roelofsen
f.roelofsen@uva.nl
University of Amsterdam

Second Reader
Prof. dr. W.F.G. Haselager
pim.haselager@donders.ru.nl
Radboud University

S. van den Brand
stijn.vandenbrand@ns.nl
NS Groep N.V.

April 21, 2022

Abstract

Public transport organisations, such as Dutch railway operator Nederlandse Spoorwegen (NS), generally communicate important messages through spoken announcements in the train or station. As a result, Deaf people miss crucial information. Because of varying levels of reading proficiency, textual information is not always a solution. Moreover, video footage is not scalable and will not yield satisfactory results when individual video parts need an update. Virtual sign language avatars can provide consistent, anonymous and scalable translations. For the development of a comprehensible and fluent sign language avatar, collaboration between experts in language technology, avatar technology and sign language is fundamental. Unfortunately, inclusive approaches are not evident in existing technologies.
To address these problems, we present a case study on co-designing a sign language avatar for the automatic translation of railway announcements into Sign Language of the Netherlands (NGT). For the initial design, video sign language translations of NS announcements created by the Dutch Sign Center were annotated and used for the translation basis. A scripted keyframe animation technique for the JASigning avatar engine makes it possible to efficiently create many variants of a given template, without expensive equipment. Three iterative co-design sessions took place with an interdisciplinary research team including deaf sign language experts. Simultaneously, hearing passengers from NS were consulted for additional remarks on the sign language avatar. Subsequently, multiple phrase variants of the system were evaluated within a diverse focus group audience to account for demographic differences. Combining the various disciplines led to mayor adjustments of the avatar (manual movements, facial expressions, mouthing, grammar, transitions between signs, camera angle and speed). With this case study, we aimed to effectively and inclusively develop sign language avatar technology. More research and focus groups are necessary to ensure high quality translations (improved mouthings, facial expressions and prosody), resulting in enhanced comprehensibility and natural appearance of the avatar.

Keywords: machine translation, signing avatar, co-design, focus group, railway announcements

Acknowledgements

I would like to express my appreciation to my supervisor Floris Roelofsen for all his effort and time that went into this project. With your valuable insights and ideas, you really steered this research in the right direction. I enjoyed working with you and appreciate that you were always there to answer my (lists of) questions.
I also want to thank Martha Larson for her valuable input, advice and feedback throughout this research, and Pim Haselager for his help. Also thanks to Stijn van den Brand and Leo van der Meulen, for their input and support during the project.
I want to thank Lyke Esselink and Shani Mende-Gillings for sharing their knowledge and work on avatar translations.
A special thanks to the team: Maartje de Meulder, Nienke Sijm and Richard Cokart for collaborating and brainstorming with us, I really enjoyed working with you.
Marjon Fonville, Babette van Blijenburgh and Martijn van Beek: thank you for your help and insights in accessibility and railway announcements. Also thanks to the Dutch Sign Centre for their valuable video material that really helped us with the fundamentals of this research, and to Bastien David, who provided us with a basis of SiGML code for the railway domain. I also want to thank Onno Crasborn for his help, and Marijke Scheffener for teaching me NGT: I really enjoyed your classes.
I want to thank Lea for providing me with feedback, and Koen for his feedback and unconditional support throughout the entire project.
Last, but definitely not least, I want to thank all participants and interpreters of this project. I will not state you by name for privacy reasons, but your participation and input was crucial for bringing this research to a success.

Contents

List of Figures

List of Tables

Abbreviations and Notation Conventions

Abbreviations

NGT	Nederlandse Gebarentaal *(Dutch Sign Language)*
NGC	Nederlands Gebarencentrum *(Dutch Sign Centre)*
ASL	American Sign Language
BSL	British Sign Language
HOH	Hard of hearing
HNS	HamNoSys (Hamburg Notation System)
SiGML	Signing Gesture Markup Language
V	Version
NS	Nederlandse Spoorwegen *(Dutch Railways)*

Notation Conventions

- #BLAAK → Name signed by finger spelling
- #D-idam → Name signed by only finger spelling the letter D
- VRIEND *(FRIEND)* → Gloss notation: written description of the manual form of a sign, including its English translation in brackets brackets (if necessary)

Symbols Used in Avatar Figures

Arrows and symbols as used in the descriptive grammar of Sign Language of the Netherlands [51]. These symbols might differ from the symbols in the Van Dale/Dutch Sign Centre dictionary [80].

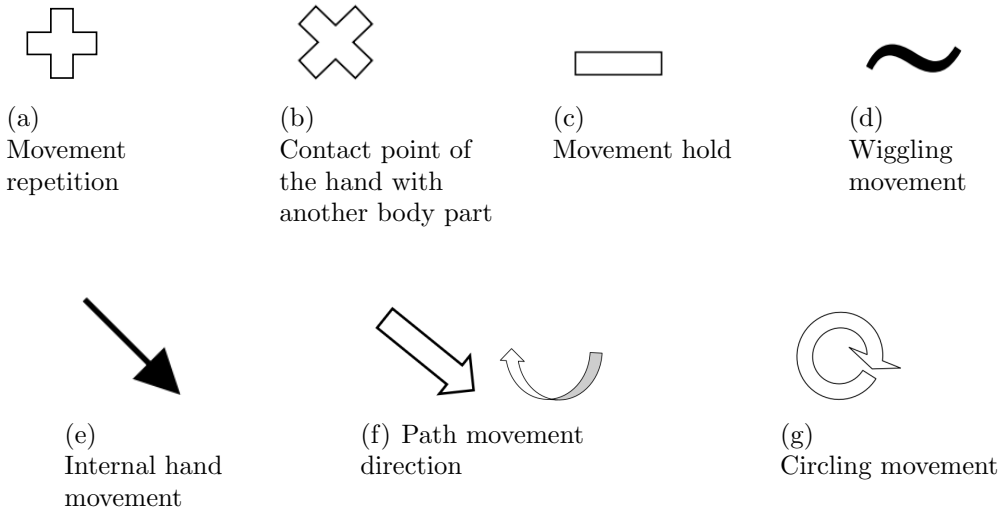
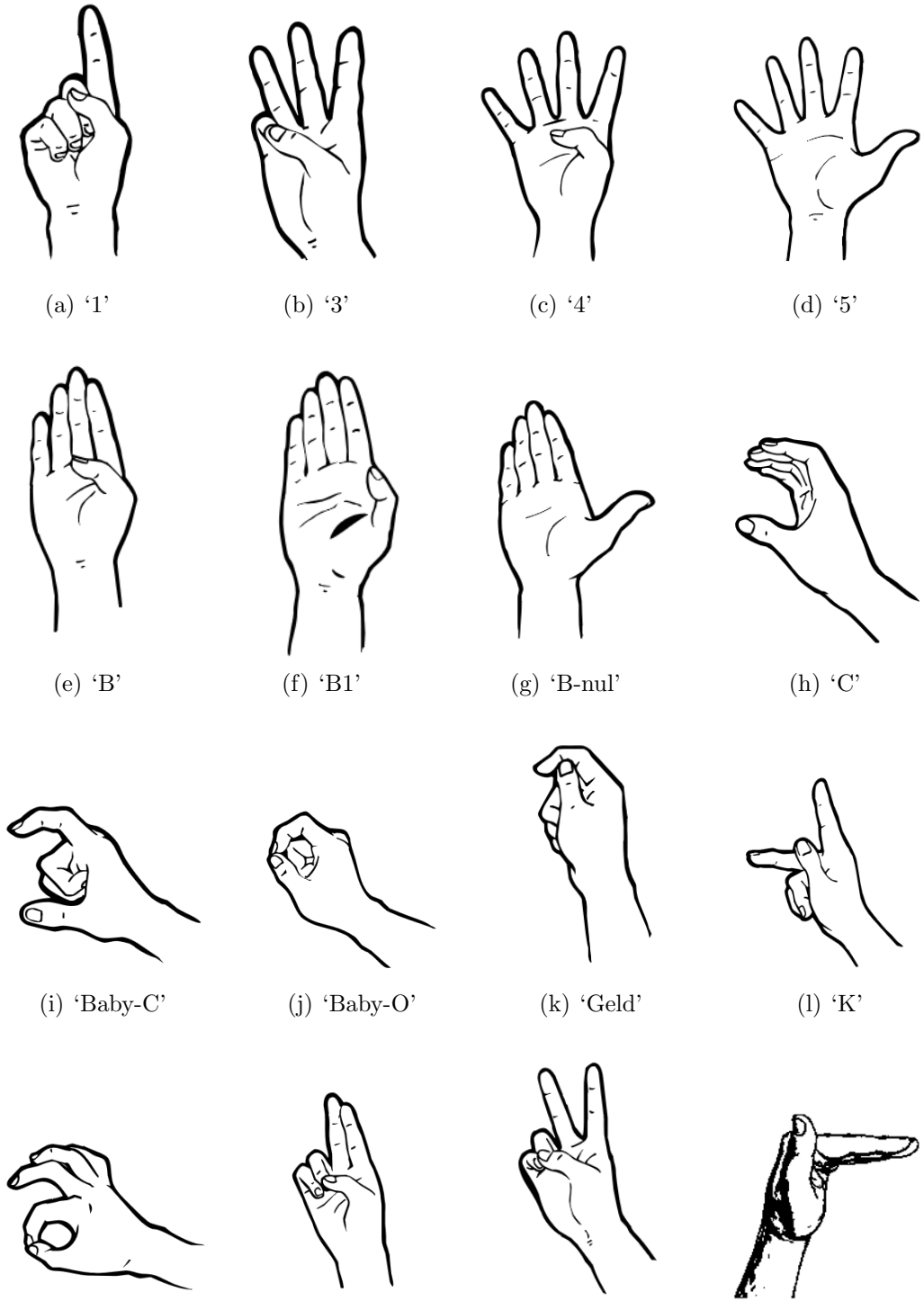


Fig. 1: Symbols used in avatar pictures [51]

Handshapes



(a) ‘1’ (b) ‘3’ (c) ‘4’ (d) ‘5’

(e) ‘B’ (f) ‘B1’ (g) ‘B-nul’ (h) ‘C’

(i) ‘Baby-C’ (j) ‘Baby-O’ (k) ‘Geld’ (l) ‘K’

(m) ‘T-nul’ (n) ‘U’ (o) ‘V’ (p) ‘Hoek’[38]

Fig. 2: Handshapes mentioned in this thesis [87]

Chapter 1 Introduction

Over 5% of the world's population has disabling hearing loss: a loss of more than 35 decibels in their best hearing ear[1]. Sign Language of the Netherlands - in Dutch referred to as 'Nederlandse Gebarentaal' (NGT) - is mainly used by people who are prelingually deaf: an estimated 7 in 10.000 [27]. The official recognition of NGT as a language only took place one year ago [28]. This is an immense achievement compared to the period between 1915 and 1980. Within this period, hearing educators decided that spoken language should be used in education and deaf students were trained to lip-read and speak [81].

More than 95% of Deaf[2] children have hearing parents [61]. As these parents are often not (yet) familiar with sign language and learning resources are scarce, only 10% of hearing parents communicate in sign language with their children [52]. This delays the acquisition of the child's first language (sign language) and affects language proficiency later in life [35, 58]. Because of this, Deaf people may have difficulty understanding the deeper meaning of words and texts: 68% of Dutch Deaf citizens has low literacy, versus 18% of hearing citizens [88, 55]. The fact that Deaf people cannot recognize written languages on its audible phonemes like hearing people do, makes the learning process even more complicated. Moreover, structural differences between spoken and signed languages complicate direct translation. As an example: British and American sign language (BSL and ASL) vary substantially, while their spoken language is (almost) identical. Besides this, the sign for 'VRIEND' *(FRIEND)* differs considerably in Groningen and Amsterdam, because of regional variances. Grammatical distinctions are an important complexity as well. For sign language, non-manual components (facial expressions, head or body movements, etc.) convey grammatical information whereas for spoken languages this is achieved by intonation [79].

As a result of reading difficulties, benefitting from captioning is not evident for Deaf people. Other aids such as hearing aids or cochlear implants (CI) could be viable solutions to some. However, technology may fail and hearing aids can get lost or damaged. Moreover, CI surgery is risky and does not restore or create 'normal' hearing. Background noise or multiple people talking simultaneously decrease the efficacy of these devices as well. Meanwhile, the introduction of such hearing devices affects attitudes of parents, family and teachers: spoken-language approaches will be favored over sign language whenever possible. This may lead to social exclusion and negative health consequences [64, 43, 34].

[1] 432 million adults and 34 million children, World health organisation, 2021

[2] 'Deaf' with a capital D is used to refer to people who are part of the linguistic community of sign language users. 'deaf' without a capital letter is used to refer to the pathological condition: hearing loss

As long as the hearing society is not proficient in sign language, Deaf individuals rely on sign language interpreters and other translation services to participate in society. In research from Roelofsen et al.[3], 17 out of 22 Deaf participants mentioned trains, stations and airports as locations where they were missing crucial information. Although video footage of physical sign language interpreters may be a possible solution to this information deficit, the approach is not scalable nor flexible: combining footage will not yield satisfactory results when individual video parts need an update (e.g. changing a station name within a phrase).

State-of-the-art software introduces virtual sign language avatars for addressing these issues [36, 54]. Using avatars can provide consistent, anonymous and scalable translations within the public transport domain [79, 50, 45, 53]. The one-way conversation structure of railway announcements creates a relatively suitable domain for these techniques. Public transport organizations, such as Dutch railway operator Nederlandse Spoorwegen (NS), can contribute to these recent accessibility advancements by implementing sign language avatars to provide travel information in NGT.

In general, avatar technologies still have a lot of catching up to do when compared to current speech technology, which is usually driven by advanced machine learning approaches. A lack of suitable data sets and the fundamental intricacies of sign language complicate this development. In the history of speech technology, lesser quality was accepted as a starting point and gradual improvements were made after release. However, for sign language technology, prematurely released software is considered unacceptable. Low quality software could influence sign language as a minority language negatively and premature beneficial claims could be perceived as insulting and threatening by deaf signers [53, 10]. Within the railway domain, even little errors may cause problematic situations when passengers rely on this information (e.g. communicating the wrong platform). The creation of more annotated data sets within the railway domain could contribute to error reductions within machine learning approaches.

In order to develop comprehensible and fluent sign language avatars, combining expertise in language technology, avatar technology and sign language is fundamental. Very few individuals possess this highly specific combination of skills. Therefore, collaboration between experts from these various fields is crucial to combine resources and experience, and foster more ideas [10]. Co-design is a collaborative designing method that ensures inclusion of users and the people who deliver or engage with a product by focusing on mutual understanding of the system across all stakeholders [11, 92].
Unfortunately, there has been a lack of involvement of the target group in existing sign language technologies. Teams of hearing people do not have the lived experiences and knowledge of the linguistic complexities of sign language. Evaluation with Deaf members often happens at the very end of the development process, if at all. Moreover, hearing people or sign language interpreters are often the benchmark of translations, even though they hearing people have no competence to assess the actual signing quality and interpreters are not the end users [20]. Co-designing a virtual sign language avatar with the target group ensures an inclusive approach

[3]personal communication, September 14, 2021

and produces additional knowledge, resulting in higher quality technology.

By presenting a case study that explores and brings together several interdisciplinary approaches and expertises, we aim to contribute to effective and inclusive development of sign language avatar technology. The case study regards co-design of a virtual sign language avatar for the railway domain and evaluation across a diverse focus group audience. Hereby, we aim to contribute to enhanced avatar quality and increased travel information accessibility. This thesis describes the phases, challenges and results from the case study and collaboration methods.

1.1 Research Questions

The aforementioned issues and goals lead to the following research question and sub questions:

How can we automatically translate railway travel information from Dutch to animated Dutch Sign Language (NGT)?

1. How and with whom can we efficiently and effectively map Dutch railway announcement phrases to NGT?
2. How can we enhance the naturalness and comprehensibility of a sign language avatar in public transport announcements?

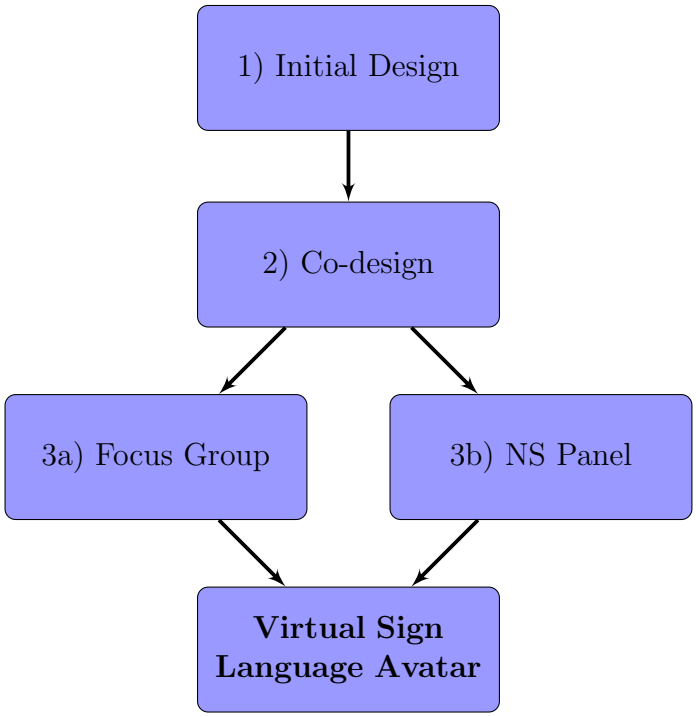


Fig. 1.1: Structure of the research phases

Figure 1.1 shows the order of the various research phases. Phase 1 (RQ 1) explores the best practices for the development of a sign language avatar: interdisciplinary collaboration is explored, as well as the current railway announcement structures and potential techniques for translation and sign synthesis. In phase 2, 3a and 3b (RQ 2), the interdisciplinary consortium will be consulted for co-designing the first prototype. This leads to a preliminary answer about the preferred phrase structures and features of the avatar from a Deaf perspective. To address the research question and co-design results more in-depth, multiple variants of signed phrases will be evaluated within a diverse focus group. Simultaneously, members of the hearing NS panel will be consulted in order to create an overall picture of how the avatar will be received by them. Both RQ 1 and RQ 2 contribute to the overarching goal: the development of a sign language avatar that automatically translates public transport announcements.

1.2 Organization

The research project is performed in collaboration with NS and SignLab Amsterdam. NS is the principal railway operator in the Netherlands. The company operates internationally as well, with its subsidiary company Abellio, mainly in the UK and Germany [66]. SignLab Amsterdam is a cross-faculty sign language research lab located at the University of Amsterdam. They are involved in the development of text-to-sign translation through animated avatars.

The thesis has the following structure: Chapter 2 contains the necessary concepts for the remainder of the thesis and Chapter 3 describes the theoretical foundation for the research. Each phase in the research has its own Chapter, consisting of the methods, results and a small discussion, as previous results influence the consecutive approaches. For phase 1, this includes Chapter 4 and Sections 4.1, 4.2 and 4.3. Phase 2 is described in Chapter 5 and Sections 5.1, 5.2 and 5.3. Chapter 6 and Sections 6.1, 6.2 and 6.3 describe phase 3a. Chapter 7 and Sections 7.1, 7.2 and 7.3 describe phase 3b. To conclude, a general discussion and conclusion for the research questions is presented in Chapter 8.

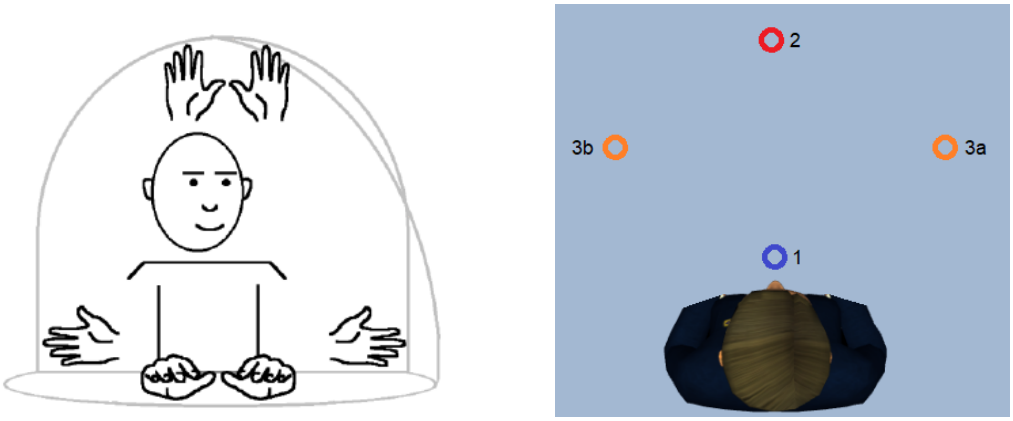
Chapter 2 Conceptual Framework

This chapter explains the necessary concepts to support this research. Section 2.1 discusses Dutch Sign Language (NGT) and its fundamental intricacies, which complicate direct translation from spoken Dutch. Section 2.2 includes approaches for sign synthesis, and their corresponding (dis)advantages.

2.1 Dutch Sign Language (NGT)

2.1.1 The Signing Space

The signing space includes the entire area surrounding the signer (see Figure 2.1(a)). The production of signs is limited to this space.



(a) The Signing Space [5] (b) Indices

Fig. 2.1: The signing space and its indices

Indexing is a pointing movement for referring (back) to entities in a phrase. For this, the location within the signing space is essential, as the pointing movement (INDEX) is performed towards the entity in this space. As can be seen in Figure 2.1(b), INDEX_1 refers to the signer and INDEX_2 to the interlocutor. INDEX_{3a} and INDEX_{3b} do not have to be present by definition and can refer to various entities [26]. By combining multiple INDEX numbers, multiple people can be indicated (e.g. $\text{INDEX}_{1-3a-2-3b}$ would denote *we*). The phrase 'VROUW INDEX_{3a} WINKEL INDEX_{3b} LOPEN_{3a3b}' is an example of how indices can be used in a sentence, and can be translated by 'De vrouw loopt naar de winkel' *(The woman walks to the shop')*.

Indices are also used to denote (argument) sentence topics. For example, the phrase 'BOEK INDEX_{3a}, INDEX_{3b} WEGGOOIEN INDEX_{3a}' *(He threw away the book.)* indicates that the book is the object topic, whereas in the phrase 'MEISJE INDEX_{3b}, INDEX_{3b} BOEK WEGGOOIEN INDEX_{3b}' *(That girl, she threw away the book.)* indicates the girl as the subject topic [18].

2.1.2 Inherent Complexities

Nederlandse Gebarentaal (NGT) has various fundamental intricacies that complicate direct translation from spoken Dutch to NGT. For example, a universal sign language or phonetic alphabet does not exist (such as IPA for spoken languages [3]) and there is no widely adopted writing system. Some systems such as HamNoSys (Hamburg Notation System), Stokoe notation system and Signwriting have been proposed, but these are considered complex and users need training to understand them [72, 86, 59, 57]. As a result, they are mostly used for research purposes. The lack of a writing system is not surprising, as sign languages use a 3D space to convey multiple phoneme directions: signs are composed of *manuals* (hand movements) and *non-manuals* (facial expressions, movements of the mouth, head or upper body) [7]. Spoken languages only have one phoneme direction: a mapping from characters to sounds [93].

The non-manual components span across various signs in a supra-segmental way and convey important grammatical information [79, 60]. For example, a negation or affirmation is denoted by shaking or nodding the head, difficult or big matter is indicated by puffing the cheeks and a brow raise expresses a question. Often, non-manual mouthings resemble the spoken Dutch equivalent of the signed word. In some cases, only some letters are articulated (e.g. 'pm' for 'VERTREKKEN' *(TO DEPART)*). Mouthings can be crucial for minimal pairs: signs that only differ in one of their main characteristic such as handshape, location, orientation, movement or non-manual marker. For example, the signs for 'SITUATIE' *(situation)* en 'OMGEVING' *(environment)* involve the same manual movement, but differ in their mouth movement (both have mouthings equivalent to its Dutch word).

Classifiers represent a group of referents. Conversations are simplified by incorporating classifiers into a movement, palm orientation or location. The falling of an object exemplifies this, as the handshape of this movement varies depending on the object it resembles: a falling *book* is indicated by a 'B1' handshape, while a falling *person* is indicated by the 'V' handshape (see Figure 2 for an illustration of all handshapes) [60, 26].

The word order and conjugation of verbs in NGT also differs from spoken languages. Spoken Dutch is based on the SVO order: subject, verb, object, while NGT is based on SOV: subject, object, verb [51]. For example, the sentence 'Ik loop naar huis' *(I am walking home)* is translated as: 'IK HUIS LOPEN' *(I HOME WALK)*. Word order and topic doubling (X Y X pattern) can denote emphasis: 'WAAROM PANIEK WAAROM' *(WHY PANIC WHY)* [47, 48]. Besides this, instead of using conjugations to denote the time frame, signs such as 'GISTEREN' *(YESTERDAY)* are added to various positions in the phrase.
Directional verbs do not have a fixed position in the signing space (see Figure 2.1(a)), but acquire their start and end positions from the objects they are referring to. In the aforementioned phrase, 'LOPEN' is the last sign of the phrase, but is signed *inbetween* the signs 'IK' *(I)* and 'HUIS' *(HOUSE)* within the signing space. This indicates the spatial relationship between the objects.

For VRAGEN/GEVEN *(TO ASK/TO GIVE)*, the 'B1-handshape' moves between the people involved in the phrase (e.g. from me to you, from you to her, vice versa, etc.).
Several forms of cliticization may occur as well. For example: two signs melting together by merging their movements into one continuous movement, or assimilating the handshapes of adjacent signs.

Other factors that complicate a direct mapping are: plural forms (indicated by repeating the sign or by combining the sign with numerals), absence of participles and function words, multiple signs for the same Dutch word, or combined signs [7]. For example, the sign for 'OUDERS' *(PARENTS)*: is obtained by combining the signs for 'VADER' *(FATHER)* and 'MOEDER' *(MOTHER)*

2.1.3 Prosody: Rhythm

Intonation, stress and rhythm are part of the NGT prosody and can be expressed by both manual and non-manual features. They are of importance for comprehensibility and naturalness. Research by Ormel and Crasborn [70] illustrates this, as they could detect non-native signers by their lack of sign rhythm, and found out that non-natives have more difficulties detecting this by others too. Besides this, signing speed and timing may also be affected by demographic factors or experience [70]. Nonetheless, not all current applications take these prosodic factors into account [56], even though it is seen as an essential factor for reaching comprehensibility scores close to human signers [49].
Eye blinks can be linked to syntactic structure: periodic short blinks can mark higher prosodic phrase endings [89]. However, the distribution over the sentence is the mark, and not the occurrence itself [18].

Pauses in a signed phrase can mark topics [48]. For example, a head nod or short pause often occurs after a time specification (spatio-temporal topic) [18]. According to Al-khazraji et al. [46], the five key concepts of speed and timing in sign language are the words per second, sign duration, signing speed per word (e.g. slower at the end of the sentence), pause insertion and pause duration. In their research, pauses, individual word speed and phrase duration were automatically calculated with machine learning models from ASL motion-capture data. As a baseline, a uniform duration and speed was used, and pauses were placed at the end of each sentence. As a comparison, a 2008 rule-based model of Huenerfauth was used [39, 40]. After providing full syntactic parse trees as input, the new model out-performed the baselines, and participants preferred animations based on this model [46]. In research from Ebling et al., an avatar for Swiss German Sign Language was investigated for the railway domain. They found that some transitions were too abrupt, but were perceived as too long for combined signs (e.g. 'FAMILY WAGON'). Moreover, for distinguishing elements within lists of station names, participants suggested a pause followed by the sign for 'place' and then returning the hands to a neutral position [21]. Adding manual counts between signs is another way to indicate a list structure [31].

Grosjean and Lane found that pause length in sign language is related to the syntactic structure of the sentence (similar to spoken languages). Nonetheless, breathing pauses are not included in this structure, as is true for spoken languages [32]. They also compared the pauses in spoken English with pauses in ASL. The longest holds in ASL occurred at important structural boundaries, such as the end of sentences (+- 229 ms), shorter holds occurred within constituents: +- 134 ms between conjoined sentences, 106 ms between noun phrases (NP) and verb phrases (VP), 6 ms within NPs and 11 ms within VPs. For spoken English, the holds at the end of the sentence were longer (>= 445 ms), as well as all the holds between constituents (245 ms => x < 445 ms) [33].

2.2 Sign Synthesis and Translation

2.2.1 Motion Capture and Keyframe Animation

For sign synthesis, several techniques can be applied. Popular techniques are motion capture and keyframe animation. The best technology depends on the use case requirements. For the highest animation quality, motion capture is the most suitable approach. Motion capture uses sensors (e.g. gloves and suits) to capture data from human poses (e.g. a native signer) and blends them with a virtual avatar to create high quality signs [30]. The downsides of this approach are that it requires expensive equipment, it is a lot of manual work to obtain and clean the data, and the manipulation of existing individual signs can be complex (e.g. when incorporating non-manual markers for question phrases) [14].
Another frequently use technique for creating animations is keyframe animation [4]. Here, individual signs are represented by manually animating each frame. This manual approach requires a lot of work, transitions between signs can be complex, and incorporating non-manual markers can be difficult for individual signs [62]. When using *scripted* keyframe animation generated from phonetic properties of signs, the manual labor can be reduced [79]. These properties include, for example, hand location and shape, body movements, and facial expressions. Nonetheless, the animation quality of (scripted) keyframe animation is lower compared to motion capture.

2.2.2 Intermediate Representations

As mentioned in 2.1, there is no one-to-one mapping between spoken and signed languages. Even when word order is changed according to rules of the target language, crucial grammatical information is not included.

For sign synthesis, either direct or indirect translation can be applied (see Figure 2.2). Direct translation would receive speech or text as input and output avatar coordinates. This has not been attempted yet and requires a lot of motion capture data [77]. For indirect translation, various intermediate representations can be considered. The main approaches are:

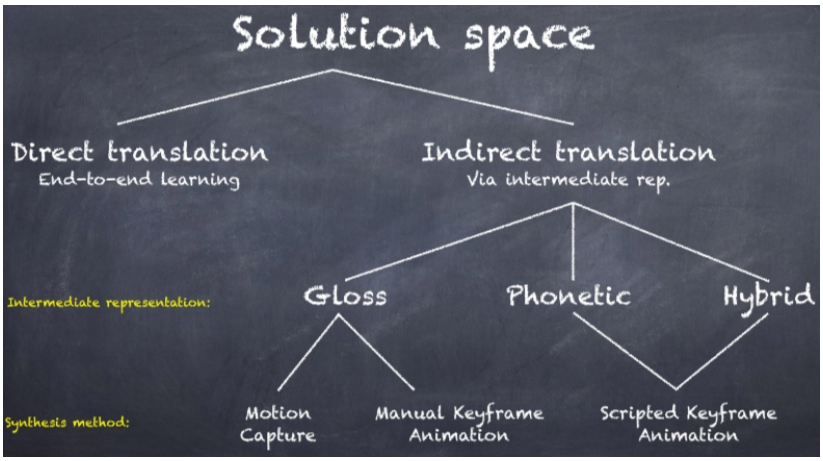


Fig. 2.2: Solution space [77]

- **Gloss approach** Text → gloss → animation
  A gloss is a textual representation of NGT, containing both manual and non-manual components.

$$\frac{\text{whq}}{\text{YOU HOLIDAY GO WHERE}} \tag{2.1}$$

  For example, notation 2.1 shows the gloss representation of the sentence 'Waar ga je op vakantie?' *(Where do you go on holiday?)*. The gloss text represents the corresponding manual signs, and *whq* indicates the non-manual component for questions. The horizontal line shows the duration of this non-manual marker (here: the entire sentence). Depending on domain constraints, the input sentences can be mapped to their corresponding gloss translations in various ways. A rule-base approach is the most suitable for high precision messages (e.g. conveying a doctor's messages), whereas a machine learning (ML) approach might be better suited for broad coverage (e.g. social media messages). An example of a ML approach is the encoder-decoder structure from Amin et al. that automatically translates English text to ASL [2].
  After applying rule-based or ML techniques, the gloss is translated to an animation by using keyframe or motion capture techniques, or a combination of both (see Section 2.2.1). Disadvantages of the gloss approach are that animating the glosses is a lot of manual work, non-manual grammatical elements are not easily integrated, and the translations cannot be transferred to other target languages. Moreover, a gloss-based system currently only exists for other languages than NGT (e.g. HandTalk [36]).

- **Phonetic approach** Text → phonetic repr. → animation
  In sign language, combinations of manual and non-manual features are represented in its phonemes. When applying the phonetic approach, sign synthesis is achieved by translating text into phonetic representations readable for the computer. A common approach for this is the SiGML notation (Signing Gesture Markup Language) in combination with the JASigning software (Java Avatar Signing System). SiGML is a computer-readable XML-based representation, based on HamNoSys (Hamburg Notation System), an annotation method for sign language corpora that describes a sign with a sequence of tokens. SiGML is avatar-independent code that includes both manual and non-manual animation components. It is extensively developed and widely used [44], [22].
  JASigning interpretes SiGML code and translates it into animations. The software has promising results in other studies [24, 25, 85, 45, 23]. For this approach it is not (yet) possible to apply machine learning, because of the lack of parallel corpora (text → phonetic). Manually creating the phonetic representations is a time-consuming process, and expert knowledge of SiGML is needed. Advantages of the phonetic approach are that grammatical non-manual components can be integrated, and the JASigning software can be applied to multiple sign languages.

- **Modular (hybrid) approach**
  Text → gloss → phonetic repr. (SiGML) → animation (JASigning)
  Roelofsen et al. combined the gloss and phonetic into a modular approach (see Figure 2.3). By translating the input sentence into its corresponding gloss translation, the gloss can be transformed into its SiGML representation (containing both manual and non-manual components). Afterwards, the SiGML representation is processed by the JASigning software to generate the avatar animation. With this approach, it is possible to re-use existing SiGML code for distinct phrases.

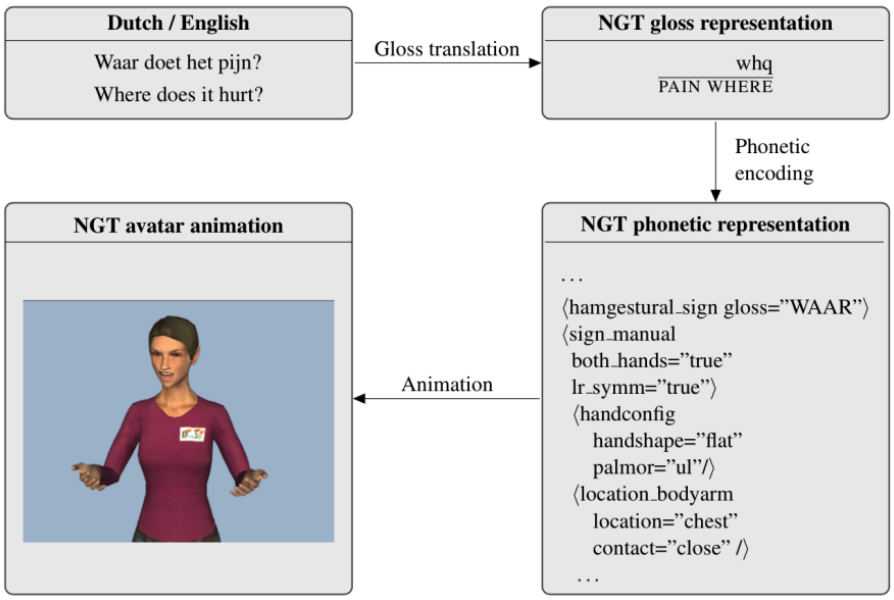


Fig. 2.3: Translation pipeline [79]

Chapter 3 Literature Review

This chapter provides the scientific embedding of the research by discussing existing literature related to the research questions. The first section describes literature on implementation features such as animation length, subtitles and possible applications, which supports the creation of an initial design (RQ 1) while simultaneously improving the signing quality (RQ 2). The second and third section describe co-design and focus group methods from previous studies, which endorses the importance of including these methods in research approaches.

3.1 Avatar Features

In research from Quandt et al., Deaf people perspectives towards various ASL signing videos were compared (keyframe animation avatar, motion capture avatar, and a human signer). According to the resulting correlations, the earlier Deaf people acquired ASL, the more critical they were towards quality issues related to computer movements [75].
Research on a protocol for sign language avatars describes implementation and usage considerations based on 4 focus group discussions with hearing people, and 6 with Deaf people. These focus groups were accompanied by several short videos signed by a motion capture (SIMAX) avatar and Deaf interpreters. The suggested maximum animation length in this study varies from one minute to five sentences, as concentrating on avatars can be quite cognitively intense [53]. Other research suggests a restriction of approximately 7 signs, as this is the maximum item capacity of the short-term memory [78]. Moreover, it was suggested that although future quality will improve, the avatar content is only fully comprehensible when supported with subtitles [53]. Participants in research from Kennaway et al. confirm this by indicating they prefer written text alongside the signing [45]. Nonetheless, subtitles cannot replace the avatar entirely, as they are necessary for people who are HOH but not sign language competent. Additionally, subtitles can provide information to hearing people from foreign countries as well, especially during chaotic situations where train passengers and personnel start talking (in Dutch) very fast.

Other suggestions included the appropriate applications of avatars: whenever real humans or voices are used (e.g. dialogues or emotional contexts), a human interpreter should be present. For one-way communication where computer voices are heard (e.g. on train stations or museums), avatars can be deployed [50]. For videos and films, animated fantasy characters created by motion capture techniques are considered a suitable option. It should be noted, however, that inappropriate use of childish avatars can damage sign language perception in the long run, so adapting the avatar to the situation is very important [53]. For the animations, variations in facial expressions are important (eyebrows, eye lids/gaze and mouthings). However, participants in research from Kipp et al. noted that the duration of signs and mouthings is of importance too: it should be aligned well. Additionally, permanent

avatar eye contact felt uncomfortable to them. Moreover, the unnatural exaggeration of movements was discouraged and a high contrast between skin, clothes and background colour was preferred for clarity. Lastly, the signing space should be optimally used (so not only pointing forward or sideways, but also to other areas) [50].
Another essential factor regarding avatar appearance is the uncanny valley: making avatars so human-like that they become eerie. Realistic eyes and facial expressions are one way to overcome an this scary appearance [29]. Having a moderate human-like appearance is considered a safe option [63]. For enhancing a natural appearance in sign language avatars specifically, a neutral end position can be added [19]. Overall, grammatical structures and correctness are important components that should receive input from signers [41].

3.1.1 Existing Translation Systems

Various systems exist for the translation of spoken languages to signed ones. For example, TESSA uses speech recognition to translate spoken English into British Sign Language (BSL) in a constrained post office setting. The animations were based on captures from a human's hands, body and face. Transactions with TESSA took nearly twice as long to complete and communication acceptability was rated low [15]. SIMAX avatars are 3D animations based on written texts. A prototype study where 188 participants answered questions from a comprehensibility quiz resulted in an average of 52% correct answers [71]. In research from Ebling et al., various types of avatar were investigated in a focus group amongst learners of Swiss German Sign Language (DSGS). These avatars were based on different technologies (XML, motion capture). In general, participants considered facial expressions to be very important. Still, the XML based avatar was considered the most appropriate for public information settings, as it was not too expressive.
Handtalk, a text-to-sign application, translates English or Brazilian-Portuguese text into sign language (American or Brazilian) by using machine learning techniques in combination with manual keyframe animation [36].

3.2 Co-design

Co-design is a stakeholder centered designing method that ensures inclusion of users and the people who deliver or engage with a product, by focusing on mutual understanding of the system and collaborating on an equal footing [11, 92]. For the process of sign language avatar creation, thorough quality control and a strong (long-term) scientific involvement of specialists, sign language staff and the Deaf community is essential [53, 49, 9, 90]. However, that goal is often not met in practice [83]. Hearing non-signers usually have no competence to assess the quality and intelligibility of avatar translations and lack familiarity with deaf everyday life [53]. This may cause misconceptions or inaccuracies about the communities' needs or language [37].
Co-designing with Deaf communities prevents the development of detrimental systems that negatively impact communities [10]. Inclusion of the Deaf community will stimulate acceptance, while providing assurance that the system is suitable for its intended purposes [1]. To involve Deaf individuals in an ethical way, the findings of the study should be communicated back in their languages via e.g. presentations or news letters [84].

Community-based co-design has been performed for several sign languages, including South African sign language [8, 13]. 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 [74].

3.3 Focus group

In a focus group, the goal is to elicit perceptions and opinions about a concept or product early in the development and design process. An agenda sets the topics for discussion and a free communication setting must be ensured (sign language) [50]. Issues that are not familiar for the researchers can be specified and discussed in depth, and participants express their expertise on the topic. The moderator can ask questions for clarification and guides the discussion [6].
For evaluating a sign language system, it is beneficial to include a diverse group of deaf people to appropriately represent the signing population by including various perspectives. This minimizes the risk of harmful technology [10]. Participant selection can be based on hearing status, education level, sign language(s) proficiency, regional differences, parental hearing status, access and ability from auditory-enhancing technologies, type of education and age of sign language acquisition [10, 37]. For screening purposes, previous studies found that asking questions such as 'Are you a native signer?', or 'Is ASL your first language?' may be misinterpreted for questioning the participants' ASL skills, or asking whether the person feels connected with the Deaf culture [41, 42].

Balch et al., researched focus group design and dynamics for court experiences with a variety of participant compositions (deaf, HOH, blind, etc.). They found that in general, focus groups with deaf people took at least double the time compared to the 'normal' focus groups they had in the past. This extra time was necessary to make sure everybody understood each other. Hence, repetitions could be necessary [6].
Related work emphasizes the importance of avoiding visual distractions in focus groups: no busy walls or clothing, and a setting that enables a good view of all participants. [91].

Within evaluative settings, all collaborators/participants would ideally be native signers. During these sessions, only exposing them to fluent sign language (e.g. by using native interpreters) prevents them from code-switching to spoken-language-forms of signing [41]. Nonetheless, it is plausible that the interpreter is less accustomed to the research terms and methodology. Moreover, this type of communication creates an inevitable time lag: the translated message will always lag behind facial expressions and body language [91]. Having a hearing lead researcher that is fluent in sign language will stimulate the engagement of Deaf collaborators more (compared to communicating via an interpreter) [69].

Chapter 4 Initial Design

This chapter describes the methods and findings of phase 1: creating the initial design of the sign language avatar. In the first section, the collected materials and implemented software will be discussed. The second section describes some observational results and the third section addresses the limitations of these results. In a later phase, the initial design will be enhanced by co-design sessions and a focus group. Eventually, the final prototype will serve as a proof of concept (PoC) for the resulting automatic translations. The initial design and later versions of the prototype were developed by the author.

4.1 Methods

The initial design will be presented in the form of a website (see Figure 4.7). By choosing a sentence and entering variables, the animation can be played. The phases for developing the initial design are as follows:

1. **Define current NS announcement phrases** (Section 4.1.1)
   The travel announcement (*'omroep'*) department of NS specified a set of 57 railway announcement templates.

2. **Obtain translations from 'Nederlands Gebarencentrum' (NGC)** (Section 4.1.1)
   The NGC was consulted for the creation of signed video translations of the NS sentences to NGT. These translations are used as a basis for the avatar translations.

3. **Annotate translation videos** (Section 4.1.1)
   To create the translation basis, the NGC videos were annotated with corresponding glosses from the Global Signbank [82].

4. **Create avatar translations** (Section 4.1.2)
   For sign synthesis, SiGML code was implemented into the JASigning avatar engine. The translations could be played and edited via the user interface (see Figure 4.7). This first version mimicked the NGC video translations as closely as possible.

5. **Optimize avatar translations** (Section 4.1.2)
   Some optimizations were added to the initial translations by changing the speed of some phrases and adding pauses in between. Moreover, subtitles were added to the animations.

4.1.1 Materials

A lexicon of railway-related SiGML scripts for NGT was created in advance by Bastien David (University of Geneva). This lexicon includes station names and NS-related words such as *'intercity'* and *'bagagerek'* (luggage rack), etc.
Besides this, an online dictionary from the NGC consisting of more than 16.000 videos of individual signs[1] could be consulted. As sign language evolves over time (similar to spoken languages) new signs and dialectic variations are constantly added to the online dictionary. When searching for signs, a Dutch word can be entered and the corresponding NGT sign is shown in the form of a video. Additionally, a Corpus NGT is available and contains video material of 100 signers of different ages and regions in the Netherlands. A small subset of these videos is annotated with Signbank glosses [17], [16].

NS Station Announcements

As this thesis is performed in collaboration with NS, communication with the travel announcement and accessibility departments was initiated at the start of the project and continued during all research phases. For creating the prototype, the travel announcement department of NS provided a list of phrases that are communicated at stations. The phrases are based on input from their current text-to-speech system for 'Karin': the current voice for announcements at the station [67]. All phrases had a template-like structure and included various variables. By combining phrases and filling in variables, multiple new expressions can be created. An example of a phrase structure is the following:

```
The <trainType> to <intermediateStation(s)><endStation> departure
time <departTime> departs in <waitTime> from platform <platformNr>
```

NGC Translation Videos

The NGC created video translations for the railway announcement sentences, signed by a Deaf person. The template sentences were completed by choosing variables that represented a diverse collection of train types, numbers and station names. This way, we obtained an accurate impression of a variety of sentence-variable combinations. Out of the 57 templates, 38 sentences in total were send to the NGC for translation (almost identical sentences were omitted).

For annotating the translation videos, ELAN software [73] was used. Because there is no universal written notation of sign language, ID-glosses from the Global Signbank [82] were used as a lexical database. These glosses are not translations, but pointers to the manual form of the sign. Examples of ID-glosses are PT-1hand for indexing (see Section 2.1.1) and BESTE-A for BESTE *(DEAR)*. Multiple gloss annotations were created according to the conventions for the corpus NGT to ensure consistency. These conventions, for example, describe how to define the start and end boundaries of a sign and how to indicate negations. Moreover, it describes the use of alphabetical markers to denote sign variants (e.g. the letter A in BESTE-A) [16].

[1] www.gebarencentrum.nl

The Signbank contains approximately 4000 signs at the moment of writing. If no appropriate ID-gloss could be found in the Global Signbank, new ID-glosses were created according to the basic annotation conventions[2]. These ID-glosses were:

- SLASH is signed by performing a slice movement with the B-nul handshape.

Fig. 4.1: SLASH [65]

- STOPPEN-TOT *(STOPPING-UNTIL)* is signed with the B-nul handshape for both hands. Three short upward movements are made towards the non-dominant hand.

Fig. 4.2: STOPPEN-TOT *(STOPPING-UNTIL)* [65]

[2]One-word glosses are preferred (if not possible use '-' between words), neutral word choice (related to meaning), written in capitals, nouns: written in singular form, verbs: written in infinitive form [16]

- RIJDEN-TOT *(DRIVING-UNTIL)* is signed by bumping the tip of the B-shaped dominant hand into the inner palm of the B-nul shaped hand (see Figure 4.3).

Fig. 4.3: RIJDEN-TOT *(DRIVING-UNTIL)* [65]

- STOPPEN-VANAF *(STOPPING-FROM)* is signed with the B1 hand. The 1-shaped hand is a reference to an aforementioned station. Three short upward movements are made starting at the non-dominant hand.

Fig. 4.4: STOPPEN-VANAF *(STOPPING-FROM)* [65]

Besides the gloss tier that described the manual forms of the signs, extra tiers were added to include non-manual information, as these are very important for our avatar translations.
The resulting annotation tiers were the following:

- **Gloss:** Manual description of the sign
- **Head movement:** Headshakes, head sways, tilted movements, etc.
- **Body movement:** Leans, rotations, etc.
- **Facial expression:** Eye gazes, eye lid movements, eye brow movements, etc.
- **Mouth:** If the mouth movement belonging to the sign can be expressed by (part of) a Dutch word, this word is written in lowercase. Other mouth movements that can be articulated by solely using letters are written in lowercase as well (e.g. the mouthing *'pm'* for VERTREKKEN *(DEPARTING)*).
- **Referent:** Specifies the referents of pointing signs (ID-gloss: PT), this pointing behavior is called 'indexing' in this thesis.
- **TranslationNarrow** and **TranslationFree:** TranslationNarrow is the translation that stays close to the source language, whereas TranslationFree is the translation close to the spoken language. An example gloss phrase illustrating the differences is: 'INTERCITY GAAN-NAAR-A ALMERE TIJD-A 12:05' *(INTERCITY GOING-TO ALMERE TIME 12:05)*. Its narrow translation would be: 'intercity naar Almere tijd 12:05' *(intercity to Almere time 12:05)* whereas the free translation would be: 'intercity naar Almere van 12:05' *(intercity to Almere departing at 12:05)*.

For some mouthings, facial expressions and body or head movements, official conventions do not exist yet. For these, their corresponding SiGML element is used as ID-gloss for simplicity (e.g. C04: one cheek puffed, FU: furrowed eyebrows, DN: eye gaze down, TR: head tilted right). Appendix B provides an overview of all possible SiGML elements.
Because creating the annotations was very time consuming (especially the non-manual elements) and the focus of the study was on enhancing the animations, only the demo sentences include all annotation tiers described above. The rest of the sentences are annotated with tiers describing only their manual and mouth activity: Gloss, Mouth, TranslationNarrow and TranslationFree.

4.1.2 Machine Translation

Fig. 4.5: Avatar Francoise in neutral end position - V1

For the sign synthesis, we opted for the scripted keyframe animation approach mentioned in Section 2.2.1, because it has the advantages of both the gloss and phonetic approaches (less time consuming, incorporating non-manuals etc.). Besides this, it allows us to build on the already available resources (see Section 4.1.1) and expertise (e.g. Roelofsen et al. [79]). This makes it realistic to expect that we can develop a first prototype system within the duration of the project.

Although the existing NS lexicon in SiGML included many individual signs relevant for the railway domain, additional signs had to be created by the author. For this, again, the NGC video translations and online dictionary were consulted. For the SiGML code, two implementation variants are possible: gestural SiGML (G-SiGML) and HamNoSys SiGML (H-SiGML). Although both originate from HamNoSys (HNS), G-SiGML extends H-SiGML with more precise gestures [26]. As our existing NS lexicon already included G-SiGML code (instead of H-SiGML), we continued to apply this type of SiGML for the creation of additional signs. The JASigning avatar that was chosen for the animations is 'Francoise' (see Figure 4.5). Compared to other available JASigning avatars, Francoise performed the signs slightly better and had the most human-like appearance. For clarity, the contrast between the avatar's clothing and hands was increased.

```
<hamgestural_sign gloss="EEN PAAR">
    <sign_manual>
        <handconfig handshape="finger2345" thumbpos="out"/>
        <handconfig extfidir="u" palmor="u"/>
            <location_bodyarm contact="close" location="shoulders"
            side="right_at"/>
        <par_motion>
            <directedmotion curve="u" curve_size="small" direction="ur"/>
            <fingerplay/>
        </par_motion>
    </sign_manual>
    <sign_nonmanual>
        <mouthing_tier>
        <mouth_gesture movement="C01" amount="0.8" speed="1.7"/>
            <mouth_picture picture="pa:" speed="1.8"/>
        </mouthing_tier>
        <facialexpr_tier>
            <eye_lids movement="SB" speed="0.8"/>
            <eye_gaze direction="AD" speed="0.8"/>
            <eye_brows movement="RB" amount="0.2" speed="0.8"/>
        </facialexpr_tier>
    </sign_nonmanual>
</hamgestural_sign>
```

Fig. 4.6: EEN PAAR *(A FEW)* - V1[3]

[3]This is not the final version of the sign EEN PAAR and only serves as an illustration for the corresponding SiGML code

The SiGML scripts include code for both the manual (<sign_manual>) and the non-manual (<sign_nonmanual>) components of an individual sign. The manual elements describe the movements of the hands and arms. The script illustrated above is a SiGML script for the sign EEN PAAR *(A FEW)* (see Figure 4.6 for an illustration of the resulting animation). This sign has a 5-handshape (handshape="finger2345") with the thumb pointing outwards. The index finger and palm are pointing upwards (<handconfig extfidir="u" palmor="u"/>). At the start of the sign, the arm is positioned close to the right side of the shoulders (<location_bodyarm contact=" close" location="shoulders" side="right_at"/>). The movement of the arm is an upward curved movement with a direction to the upright corner of the signing space (<directedmotion curve="u" curve_size="small" direction="ur"/>). Simultaneously (<par_motion>), the fingers are wiggling (<fingerplay/>).
The non-manual elements define the facial expressions, body and head movements. The mouth movements are based on the codes from Appendix B and SAMPA (Speech Assessment Methods Phonetic Alphabet), a phonetic alphabet readable by the computer [12]. Multiple combinations can be made from the codes and SAMPA notations to create the final mouthing. For example, in the script above, mouthing code C01 (cheek puff) is applied with an amount of 0.8 and speed of 1.7. Subsequently, the avatar pronounces the SAMPA symbols pa: with a speed of 1.8. Simultaneously, the eye lids are narrowed (SB) and the gaze direction is towards the addressee (AD). The brows are slightly raised (RB). Despite all these options, the signing is not perfect yet. Especially for the mouthings, the articulation of letters is not clear enough, as SAMPA code is based on spoken phonetics instead of visual articulation. Moreover, the limited level of details in the avatar engine is an important factor as well.

The core (grammar) of the machine translation system will mostly be rule-based, as this is a promising approach within restricted domains [76]. Moreover, within the railway domain, high accuracy is considered more important than broad coverage (receiving a wrong travel advice has direct consequences for the user). Considering the lack of parallel corpora for this domain, machine learning approaches would not yield satisfactory results.

User Interface (UI)

A subset of the aforementioned railway announcements will be displayed on a website (see Figure 4.7 for an illustration). The subset was chosen such that the variety of structures and content was represented as much as possible. To work with the UI, the user has to first choose an input language (English or Dutch). Second, an announcement sentence should be chosen from the drop-down menu. Third, the default variables can be optionally changed to the user's preference. This way, many sentence variations can be created. Fourth, by clicking *play*, the animation that corresponds to the current choices is played. The user can *stop*, *pause* and *resume* an animation. Furthermore, when clicking *pause*, the user can navigate to previous and subsequent frames by clicking the *frame -1* and *frame +1* buttons. Besides this, the signing speed of the avatar can be changed by clicking the + and - buttons next to *'Snelheid +0.00'*.

Fig. 4.7: User Interface design - V1

The UI was created with HTML, CSS and JavaScript code and functioned as follows: After clicking *play*, the chosen sentence is detected and its corresponding variables are automatically identified. By extracting the variable values, the entire sentence is split into separate phrase parts that used to surround that variable. To translate these entire phrase parts, individual SiGML scripts are automatically combined. Simultaneously, SiGML scripts of the chosen variable values are extracted from the JSON dictionaries and entered into their corresponding placeholders in between the phrase parts. The result is a long string of combined SiGML scripts, which is send to the avatar engine to play the animation. The sentences and corresponding variable options are listed below. Illustrations of the avatar signing the UI sentences and variables are illustrated in Appendix A.

Fig. 4.8:
Changing the UI departTime variable
Clicking the left arrows will add or subtract 1 hour, clicking the right arrows will add or subtract 5 minutes

UI Sentences

1. Beste reizigers, de treinType naar eindStation van vertrekTijd rijdt niet. *(Dear passengers, the trainType to endStation of departTime is not departing.)*

2. De treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation van vertrekTijd vertrekt over wachtTijd van spoor spoorNr. *(The trainType to interStation1, interStation2, interStation3, interStation4 and endStation of departTime departs in waitTime from platform platformNr.)*

3. Herhaling. De treinType naar eindStation van vertrekTijd rijdt via een andere route. *(Repetition. The trainType to endStation from departTime runs through another route.)*

4. Beste reizigers, de treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation stopt niet op tussengelegen stations. *(Dear passengers, the trainType to endStation does not stop at intermediate stations interStation1, interStation2, interStation3, interStation4.)*

5. Beste reizigers, de treinType naar eindStation vertrekt van spoor spoorNr. *(Dear passengers, the trainType to endStation departs from platform platformNr.)*

UI Variable Options

• **treinType** *(trainType)*	intercity, sprinter
• **tussenStation1-4** *(interStation1-4)*	Zwolle, Arnhem, Deventer, Breda
• **eindStation** *(endStation)*	Almelo, Nijmegen, Enschede, Maastricht, Schiedam, Utrecht Centraal
• **vertrekTijd** *(departTime)*	0:00 - 23:55 (see Figure 4.8)
• **wachtTijd** *(waitTime)*	enkele minuten, ongeveer 5 minuten, ongeveer 35 minuten, ongeveer anderhalf uur, een nog onbekende tijd
• **spoorNr** *(platformNr)*	5, 1a, 1b, 11, 15

Optimization

After finalizing the basic functionality of the UI, some initial optimizations were added. This included topic-marking: adding a small head tilt to the signs for SPOOR *(PLATFORM)* and TIJD *(TIME)*.
Moreover, pauses were added according to the punctuation of the Dutch sentences (commas), after time specifications (spatio-temporal topics), and at the end of a sentence. A pause included an eye blink, brow raise and holding the hand in its current position. The signing speed of some coherent sub phrases was increased (e.g. NAAR-tussen/eindStation *(TO-inter/endStation)*) to make their transitions more natural.
For the list of station names, a 'count hand' (preceded by a pause) was added before each name to denote the fact that it is a list of names (see Figure 4.9). This is a grammatical construction that is commonly used in NGT for conjunctions and lists. To give the avatar a more natural appearance, a neutral end pose was added after each sentence (see Figure 4.5).

Subtitles in Dutch will be shown below the avatar word-for-word. This way, Deaf people who are not familiar with a certain sign (e.g. specific station names such as ALMELO) can still understand this crucial information. Moreover, having subtitles will have additional value for HOH/Deaf people who have no reading difficulties and/or have no competence in Dutch sign language.

4.2 Results

The initial design serves as input for the next stages: enhancing the current prototype via co-design sessions, a focus group and the NS panel. The creation of the initial design yielded insights in the sentence structure that NS uses for its railway announcements, how various software and code can be combined for creating the resulting animations (SiGML, JASigning, UI) and possible prosodic approaches for optimizing the current animations. Moreover, the annotation process of the NGC video translation data set provided us with useful information about NGT and its varieties regarding translations.

4.2.1 Annotated NGT video's

The annotated data set can be found on:

`https://github.com/froelofs/signlab/tree/master/ns-demo/dataset_NGC`

NGT Translation Varieties

- Some translations cannot be separated into individual glosses. For example, TUSSENGELEGEN in the context of 'stoppen op tussengelegen stations' *(stopping at intermediate stations)* did not have an individual translation but was outlined by combining features of the signs STOPPEN *(STOPPING)* and STATIONS. As a result, this phrase was annotated with a combined ID-gloss: STOPPEN-TUSSENGELEGEN-STATIONS). This also applies to RIJDEN-TOT, RIJDEN-VANAF, STOPPEN-TOT and STOPPEN-VANAF, for which new ID-glosses were created (see Section 4.1.1).
- Some signs had distinct translations within various sentences. For example, NAAR was signed one-handed with the 1-handshape or B1-handshape, or two-handed with a combination of both.
- Although in the optimization step of the development some topic markings were added (head tilts for SPOOR *(PLATFORM)* and TIJD *(TIME)*, the signed video translations did not include these markings.
- Station name translations depended on (as explained by the NGC):
  - Existing standard sign (e.g. Amsterdam Centraal)
  - No standard sign exists, but relatively known station name → sign only first letter with mouth movement (e.g. Didam)
  - Otherwise → finger spell whole name (e.g. Olst)

4.3 Discussion

The publicly available annotated NGC video translation data set spurs and contributes to future developments within sign language technology. Nonetheless, these video utterances do not include any spontaneity, which results in a translation structure that is very similar to the source language (Dutch). This could affect the data set's versatility for adoption in future studies. Moreover, although the Signbank conventions helped with consistency, the data set annotations were performed by just one labeler (the author). This means that no inter-annotator agreement could be formed and biases could be embedded in the data. Future research could address this by deploying multiple annotators.
Signers all have their own signing style and dialect. Therefore, to assess the absence of topic markers in the sign translation videos, future research should compare multiple translation video's performed by multiple people.

(a) First

(b) Second

(c) Third

(d) Fourth

(e) Fifth

Fig. 4.9: 'Count Hands' notation within lists of station names

Chapter 5 Co-design

This chapter describes the methods and findings of phase 2: co-designing a sign language avatar. This phase brings together various expertises and disciplines (technical, research and sign linguistics/Deaf studies) for NGT. The goal is to maximize the naturalness and comprehensibility of the sign language avatar. In the first section, the initial steps for building the multi-faceted consortium are described. The second section describes the results of the three subsequent iterative sessions. The third section addresses the limitations of these results.

5.1 Methods

Initial Brainstorm

As mentioned in Chapter 1, inclusion of the Deaf community is essential for the development of sign language avatars. Therefore, multiple Deaf researchers and experts were approached for collaborating in this study. In order to build the consortium and enhance the initial design, a general strategy had to be defined. For this, an initial brainstorm was organised with three hearing researchers (Roelofsen, van Gemert and Esselink) and two Deaf researchers (Sijm and de Meulder). Before the meeting, the initial design and an agenda was sent to Sijm and de Meulder as a guideline. During the meeting, van Gemert wrote minutes. Communication happened via interpreters. Sometimes, this resulted in a communication lag, especially when they were not familiar with the research terms (e.g. 'SiGML').
The goal of the meeting was to discuss the avatar translation progress so far and brainstorm about the interpretation of several collaboration methods. The team agreed on organising multiple co-design sessions with one or two collaborators, focus group(s) with a more diverse group of participants, and an online questionnaire that would serve as a quantitative evaluation of the comprehensibility.

By showing the initial design to the Deaf researchers, some suggestions for improvement naturally arose from the conversation. These suggestions were:

- Attract attention to the start of the phrase by tapping or waving the hands (a strategy that Deaf people apply in interactions), or just signing HALLO *(HELLO)* instead of BESTE REIZIGERS *(DEAR PASSENGERS)*. This is because BESTE REIZIGERS *(DEAR PASSENGERS)* is not a natural greeting in NGT.

- Regarding the avatar appearance: the trousers should be higher and the avatar should look more friendly (i.e. adding some smiles).

- Obtain the current NGC videos not only from a forward-view, but also from a different camera angle to make sure the translation basis for animation is correct.

- Regarding the co-design: two people should have a lot of knowledge about linguistics.
- Regarding the focus group: invite participants that make use of trains. Maybe one session is not enough. For the animations, comparing various signing styles could be considered.
- Evaluating the avatar should happen in real life, not just online. Moreover, the appropriate location for displaying the information is important: Where is the video when you are in a hurry? The animation speed is a factor here as well: it should not be too slow, but should not influence comprehensibility either (e.g. fast animations can complicate the comprehension of mouth movements). Leiden central station was suggested as an appropriate location for real life evaluation.

Co-design After the meeting, hearing researchers created a first shared draft for organising the co-design session. This draft included who will lead and participate in the sessions, which communication will happen via interpreters, which subjects should be addressed, etc. The goal was to enhance the animation quality with experts and show them the technical (im)possibilities. For this, phrases from the initial design would be discussed and combined with several variables. Whenever possible, improvements would be directly implemented in the system by van Gemert. Otherwise, they would be implemented for discussions in subsequent sessions. Deaf researchers evaluated the draft and added some details. For the upcoming collaboration sessions, Sijm wanted to participate herself. de Meulder did not, as her native sign language is Flemish instead of NGT. Some members of the NGC were approached to participate, of which Cokart agreed to join. Thus, the co-design team would consist of Sijm, Cokart, van Gemert and Roelofsen.

Focus group Based on the goals and ideas for the co-design session, Deaf researchers added suggestions for organising the focus group(s). Approximately five participants was considered optimal. Researchers agreed that a specific age or living region division should be considered. The goal of the focus group(s) would be to enhance the signing quality and obtain insights in the most appropriate way for implementing the avatar: via which channel (screen, train, phone, website), the location of the avatar within this channel, the avatar speed versus comprehensibility, animation combinations with text, etc.

If any questions were left unanswered, a quantitative (online or offline) study could serve as an additional source of information. The background of team members from both the initial brainstorm and co-design sessions can be found in Table 5.1.

Team member	D/H**	Affiliation
Richard Cokart	D	Nederlands Gebarencentrum
Lyke Esselink	H	Radboud University Nijmegen / University of Amsterdam / SignLab Amsterdam
Britt van Gemert	H	Radboud University Nijmegen / SignLab Amsterdam / Nederlandse Spoorwegen
Maartje de Meulder	D	Utrecht University of Applied Sciences / Institute of Sign, Language and Deaf Studies
Floris Roelofsen	H	University of Amsterdam / SignLab Amsterdam
Nienke Sijm	D	Utrecht University of Applied Sciences / Institute of Sign, Language and Deaf Studies

Table 5.1:
Team ** Deaf or hearing

Interface

Some suggestions of the initial brainstorm session were not implemented yet for the upcoming co-design session, because this was not feasible (e.g. obtaining additional videos from the NGC with a different camera angle), or not applicable to this phase (e.g. suggestions about the final (online) evaluation). The suggestions that were implemented included: changing BESTE REIZGERS *(DEAR PASSENGERS)* to HALLO *(HELLO)*, and providing Francoise with some new clothing (see Figure 5.1). The darker shades of clothing increased the contrast between hands and body even more, which resulted in clear detection of hand movements. For friendliness, a smile was added before the sign for HALLO. However, more smiles were not added, as this was disturbing and created a contradictory effect for sentences that included negations (e.g. smiling when the train is not departing).

Fig. 5.1: Avatar Francoise - V2

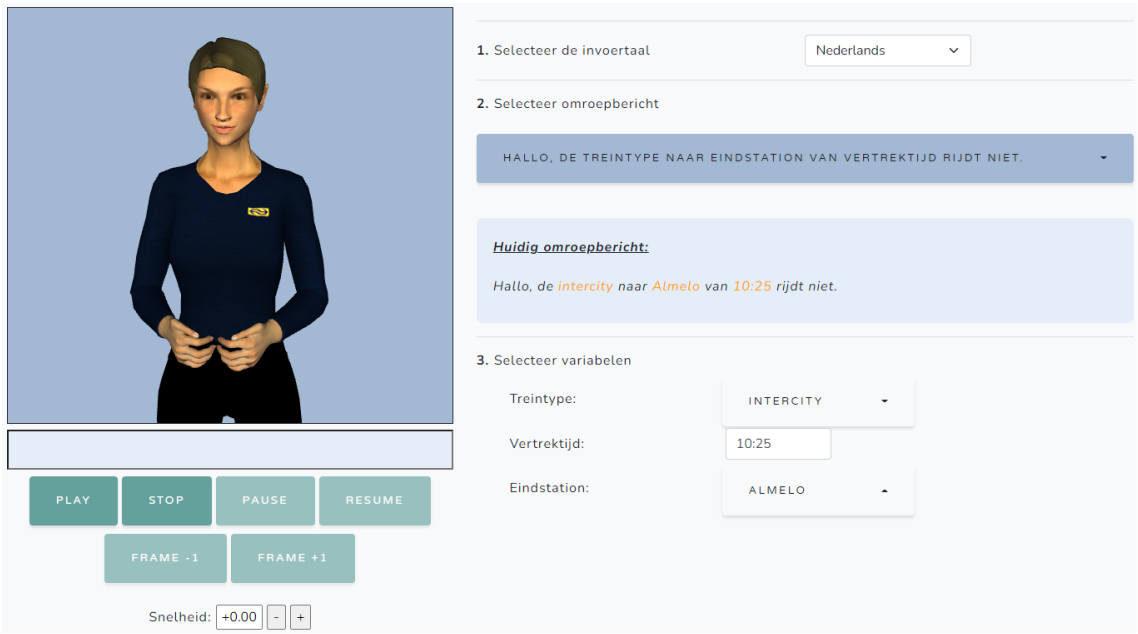


Fig. 5.2: Co-design UI - V2

The UI that was used during the co-design sessions (see Figure 5.2) can be found via the following URL:

`https://www.signlab-amsterdam.nl/ns-demo/index.html`

The password will be made available upon request.

Co-design Setting

Before the co-design sessions, collaborators received an e-mail with a description of the project and the initial brainstorm session. The goal of the co-design was communicated: collaborating to enhance the current prototype (and *not* for evaluation) and dates were proposed for the sessions.

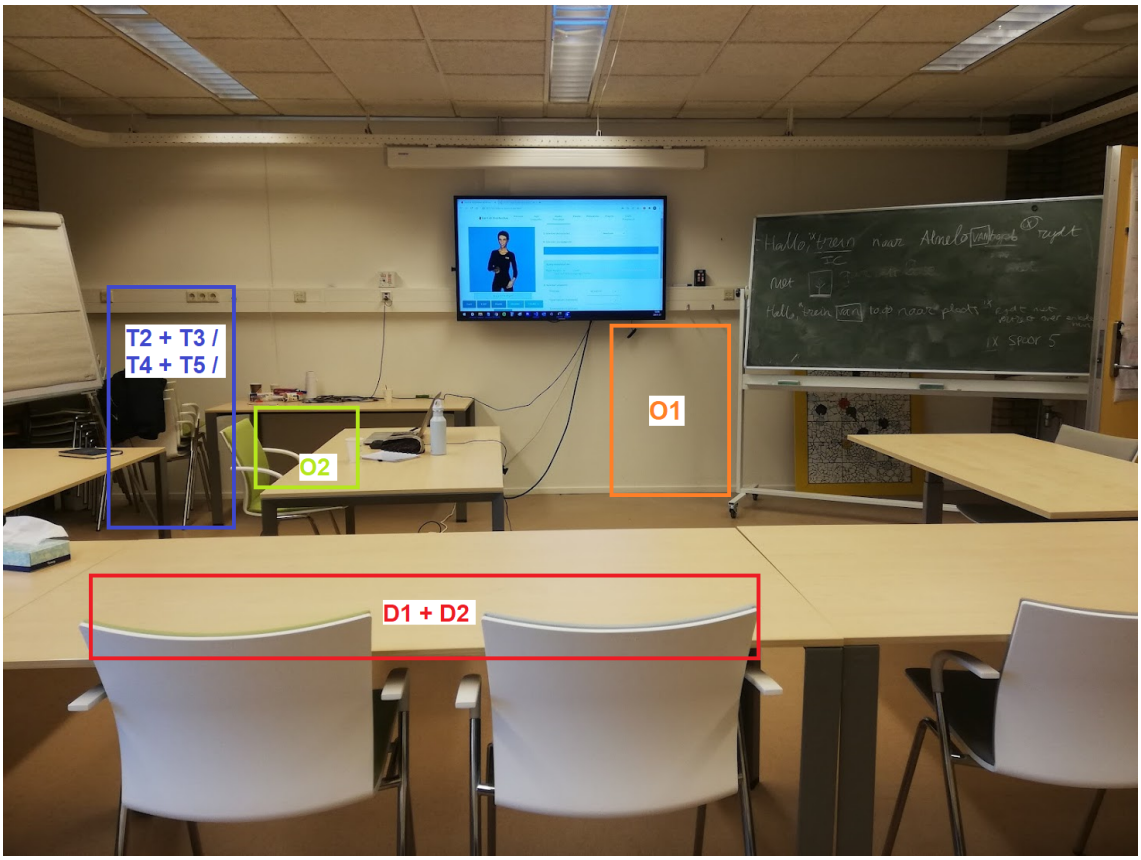


Fig. 5.3: Co-design setting

During the sessions, Roelofsen guided the conversation in sign language, while van Gemert presented the initial design and wrote the minutes. These minutes included the main subjects, discussion points, recommended changes and action points for the next session(s). The entire document was checked by Roelofsen and all team members gained access to it.
Interpreters were present to assist the communication between Deaf and hearing researchers. Nonetheless, Roelofsen communicated via sign language as much as possible. This way, for Deaf researchers, having to 'code-switch' to a more spoken Dutch form of signing was prevented as much as possible.
Interpreters were preferably approached via connections of the Deaf researchers, as they are already familiar with each other's signing styles. However, due to busy schedules and COVID-19, this was not always an option. In these cases, 'Stichting RTGS' was consulted for other available interpreters[1].

Two of the co-design sessions (2 hours per session) took place at Science Park Amsterdam. Two interpreters (T2 + T3 or T4 + T5) were present. In the physical setting (see Figure 5.3), everybody had a good view of the interpreters and each other, which is essential for understanding the facial expressions and body language. On location, animations could be watched on a big screen and whiteboards and chalkboards were present to elaborate on comments further. Sijm and Cokart (D1

[1]https://www.stichtingrtgs.nl/

+ D2) were positioned at the table, which could be used for making notes. Roelofsen (O1) was positioned in front of the screen, having the option of giving additional explanations and pointing to things on the screen. Van Gemert (O2) was positioned in the middle, in order to connect the laptop to the screen and have a good view of the animations and of Sijm and Cokart. The interpreters were positioned behind van Gemert, such that Sijm and Cokart had a good view of the interpreter, and van Gemert could hear the translations well. During the planned breaks, van Gemert implemented small improvements to the avatar such that these could be to directly discussed in the sessions. For example, trying various SAMPA mouthings for long numbers such as '55'. More complicated improvements were written down for the next session.
The last co-design session (1 hour) was held online and served for discussing some minor last adjustments. Interpreter T6 and Sijm were present.

Co-design Sentences and Variables

A general agenda with discussion points (see below) was created as a guideline for the first session. The agenda contained pre-selected sentences-variable combinations to ensure that all variable values would be discussed. The expectation was that other important improvements would arise automatically from the discussions and examples. The demo was shared again with the Deaf collaborators after the first co-design session. This way, they could experiment with it themselves.

Agenda

1. Structurally analyse animations (pre-selected combinations as a guideline) and write down comments
2. Depending on the time: have a break + implement small adjustments
3. Extra discussion points:
    - Subtitles
    - RIJDEN(-NIET) *(NOT DEPARTING)* or VERTREKKEN *(DEPARTING)*
    - Prosody: pauses, rhythm, etc.
4. Depending on the time left: implement small changes and discuss them again

For thorough observation of the animations during the co-design sessions, each sentence was played in a loop after clicking *play* and could be paused or stopped at any time. The neutral end pose and a pause marked the end of the sentence.

Pre-selected Sentence Combinations

1. 'Hallo, de intercity naar Almelo van 5:15 rijdt niet.' *(Dear passengers, the intercity to Almelo departure time 5:15 is not running.)*

2. 'De sprinter naar Zwolle, Maastricht, Deventer en Utrecht Centraal van 12:00 vertrekt over enkele minuten van spoor 1a.' *(The sprinter to Zwolle, Maastricht, Deventer and Utrecht Centraal departure time 12:00 departs in a few minutes from platform 1a.)*

3. 'Herhaling. De intercity naar Enschede van 13:30 rijdt via een andere route.' *(Repetition. The intercity to Enschede departure time 13:30 runs via another route.)*

4. 'Hallo, de sprinter naar Zwolle en Arnhem stopt niet op tussengelegen stations.' *(Dear passengers, the sprinter to Zwolle and Arnhem does not stop at intermediate stations.)*

5. 'Hallo, de intercity naar Schiedam vertrekt van spoor 11.' *(Dear passengers, the intercity to Schiedam departs from platform 11.)*

6. 'De intercity naar Utrecht Centraal van 9:05 vertrekt over ongeveer 1.5 uur van spoor 5.' *(The intercity to Utrecht Centraal of 9:05 departs in approximately 1.5 hours from platform 5.)*

7. 'De intercity naar Breda, Deventer en Schiedam van 0:45 vertrekt over een nog onbekende tijd van spoor 1b.' *(The intercity to Breda, Deventer and Schiedam departure time 0:45 departs in a yet unknown time from platform 1b.)*

8. 'De sprinter naar Maastricht, Deventer, Zwolle, Breda en Arnhem van 23:55 vertrekt over ongeveer 35 minuten van spoor 15.' *The sprinter to Maastricht, Deventer, Zwolle, Breda and Arnhem departure time 23:55 departs in approximately 35 minutes from platform 15.*

9. 'De sprinter naar Almelo van 10:10 vertrekt over ongeveer 5 minuten van spoor 1a' *(The sprinter to Almelo departure time 10:10 departs in approximately 5 minutes from platform 1a.)*

5.2 Results

The initial brainstorm session was very useful for bringing together various expertises and disciplines, and using these for creating a first outline of the collaboration strategies. Besides building the team, it also resulted in first improvements for the initial design of the avatar (although not all of these were feasible to implement within the limited time and budget).

5.2.1 Co-design Sessions

A report (in Dutch) of all three co-design sessions can be found on:

`https://github.com/froelofs/signlab/tree/master/ns-demo/notulen`

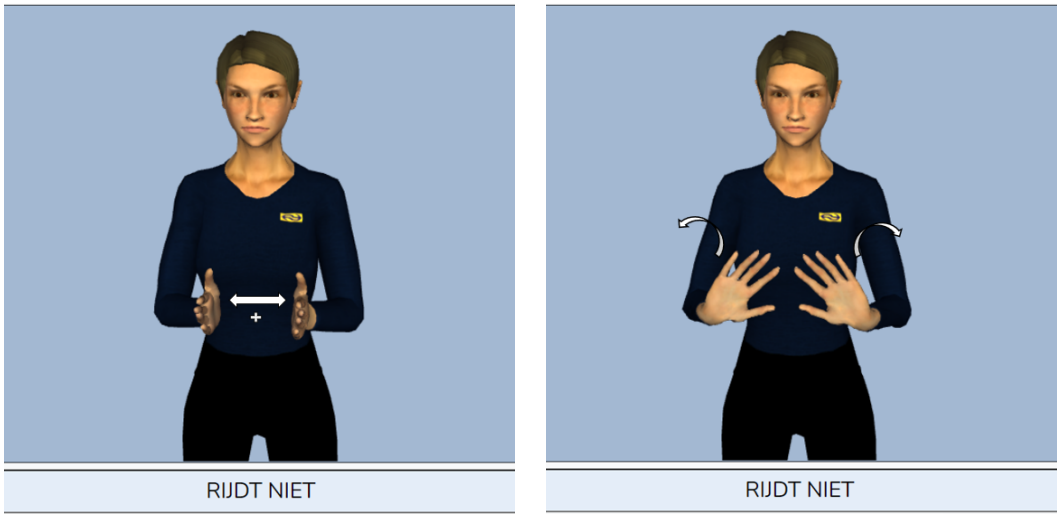
The most important findings will be discussed in this section. If 'left' or 'right' are mentioned in combination with the avatar, this always implies a view from the avatar's perspective (see Figure 2.1(b)).

Overall, the Deaf researchers were positive about the idea of closely collaborating through co-design sessions and focus groups. Constituting a small group for the co-design process was seen as an advantage: it will clearly show the process and creates room for discussing technical (im)possibilities as well. Deaf researchers advised to document the entire process well for future research.

Due to time constraints, not all pre-selected sentences were discussed during the first co-design session. Session 2 discussed the omitted sentences from the first session and elaborated on the phrases and suggestions from session 1. For session 3 (online), only one of the two collaborators was available. Some small avatar adjustments and details for the focus group were discussed here.

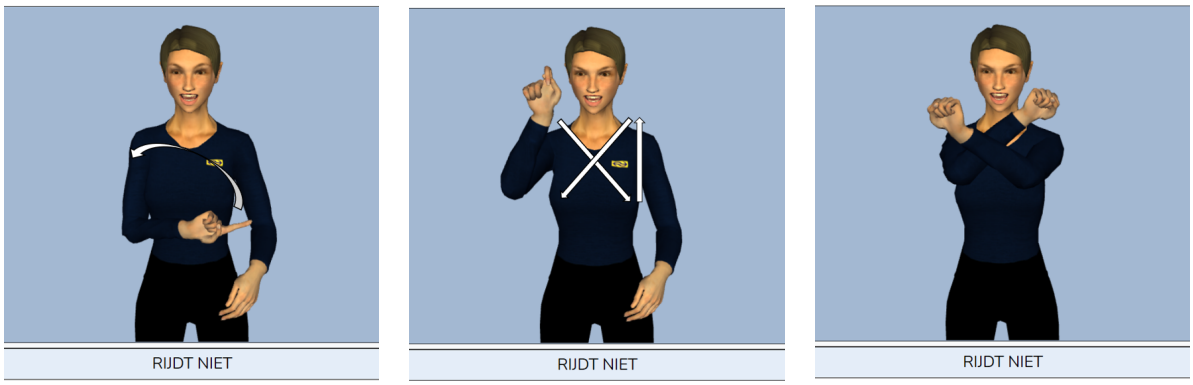
RIJDEN-NIET *(NOT DEPARTING)*

Session 1
Collaborators thought the sign for RIJDEN-NIET was a bit peculiar (see Figure 5.4). The sign for NIET *(NOT)* with two hands was considered too mitigating for the context (see Figure A.12(k)). Suggestions were: a more strict movement with the 1-handshape (see Figure 5.5(a)), a crossed arm posture (see Figure 5.5(c)), a sign similar to ANNULEREN *(TO CANCEL)* with a geld-handshape [65] (see Figure 5.5(b)).



(a) RIJDEN (b) NIET

Fig. 5.4: RIJDEN-NIET *(NOT DEPARTING)* - V1



(a) 1-handshape (b) Geld-handshape (c) Crossed Arms

Fig. 5.5: NIET *(NOT)* - Suggested variants

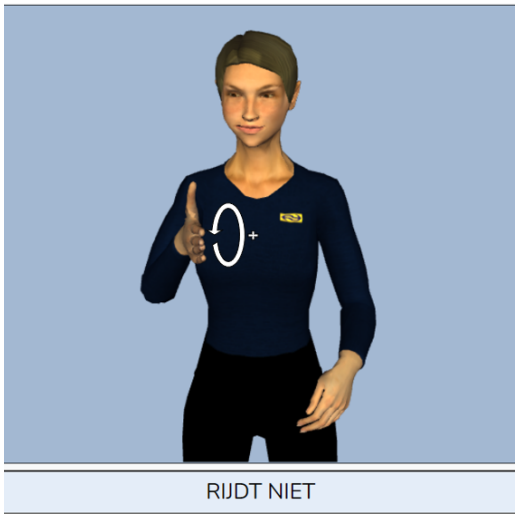


Fig. 5.6: RIJDEN-NIET *(NOT DEPARTING)* - One-handed - V1

Session 2
The sign for RIJDEN(-NIET) *((NOT) DEPARTING)* was changed to a version more similar to the video translations from the NGC (see Figure 5.6). This newer version was received more positively.
NIET with the 1 handshape was an improvement, but not yet optimal: it should contain a bit more wrist movement. The crossed arms movement from Figure 5.5(c) was rejected. The sign similar to ANNULEREN *(TO CANCEL)* (see Figure 5.5(b)) should be used when there is a structural problem and trains do not depart because of this. Otherwise, the 1 handshape movement from Figure 5.5(a) should be used (e.g. for incidental failures).

Session 3
The sign for RIJDEN(-NIET) *((NOT) DEPARTING)* (see Figure 5.6) was considered redundant in combination with ANNULEREN *(TO CANCEL)*. Thus, RIJDEN NIET should be removed here. For ANNULEREN the collaborator was indecisive whether the head shake should stay, and which handshape was correct (i.e. geld or U). The sad mouth picture was fine.

INTERCITY/SPRINTER or TREIN *(TRAIN)*

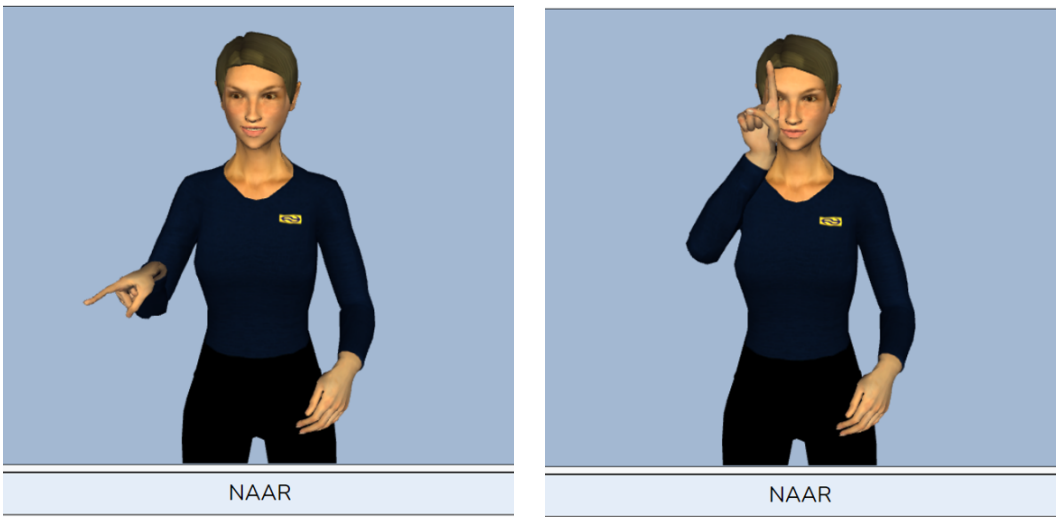
Session 1
There was some speculation about the use of the terms INTERCITY/SPRINTER or TREIN *(TRAIN)*. Consensus was that using the same terms as for hearing announcements was fine (intercity/sprinter).

(GAAN-)NAAR tussen/eindStation *(GOING-)TO inter/endStation)*

Session 1
In general, the NAAR sign should be signed in a more forward direction (instead of rightward, see Figure 5.7(a)). Moreover, it was considered necessary to improve transitions between the NAAR *(TO)* sign and station names. Especially for ALMELO (see Appendix A for the full animation), the prolonged assembly of the hands before the onset of the sign gave the phrase an unnaturally lengthy appearance: 'naaaar Almelo' *(toooo Almelo)* (see Figure 5.7(b)). Ideally, signs adjust their handshape and position to previous and upcoming signs. For NAAR in combination with names of stations, this means the direction of the NAAR path movement should adjust itself to the onset location of the following sign. For NAAR MAASTRICHT (see Figure 5.8(a)), this means a lower endpoint compared to NAAR ALMELO (see Figure 5.8(b)). This is because MAASTRICHT itself (see Figure 5.9(a)) is signed at a lower point within the signing space, corresponding to its topographic location on an imaginary country map. Eye gaze directions for NAAR should correspond to the direction of the path movement.

(a) NAAR (b) ALMELO - Hand assembly before sign onset

Fig. 5.7: NAAR *(TO)* and ALMELO - V1

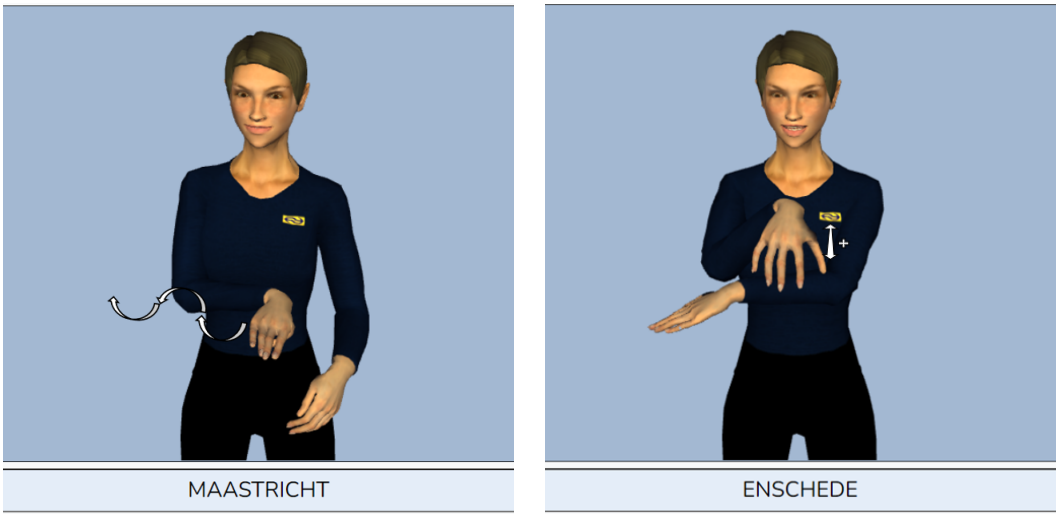


(a) NAAR - Maastricht version (b) NAAR - Almelo version

Fig. 5.8: NAAR *(TO)* - Endframe - V1 - Multiple variants

Session 2

The NAAR movement before the onset of ALMELO is slightly too brief. The assembly of both hands for ALMELO is still too lengthy. Changing the handshape earlier within the transition may improve this.



(a) MAASTRICHT (b) ENSCHEDE

Fig. 5.9: MAASTRICHT and ENSCHEDE - V1

Station Names

Session 1

The horizontal movement for ENSCHEDE should be less extreme (see Figure 5.9(b)). It was compared to the sign for HAMSTEREN *HOARDING*.

Session 2

The current sign for ENSCHEDE has 3 repetitive movements, collaborators noted that 2 repetitions are sufficient. Moreover, the head movement should be more subtle. Instead, the eye gaze should follow the hand movement.

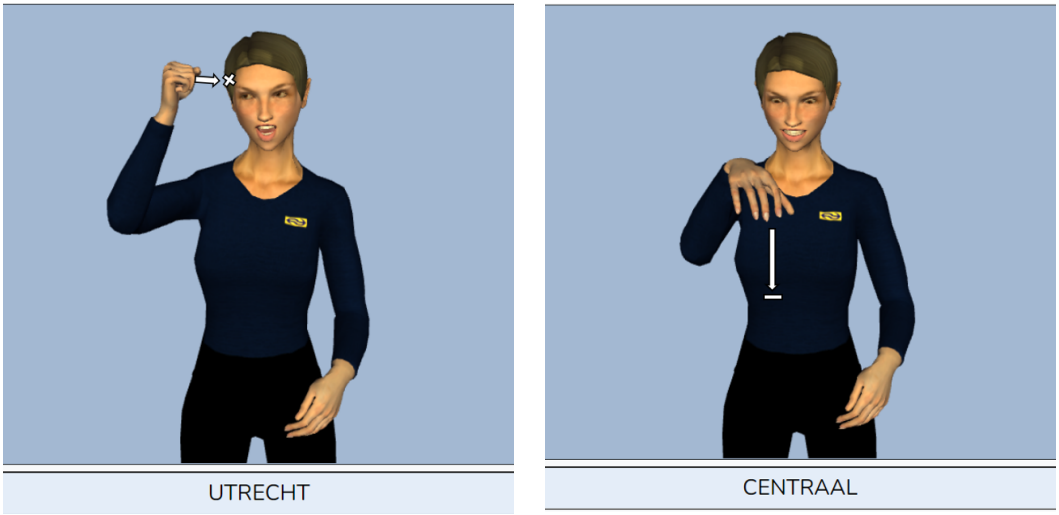
Session 3

The improvement of ENSCHEDE was accepted.
The movements of UTRECHT CENTRAAL were considered too extreme (see Figure 5.10). The arm should stay closer to the body and the head should have a more subtle side-wards tilt.

'Vertrekken over enkele minuten' *(Departing in a few minutes)*

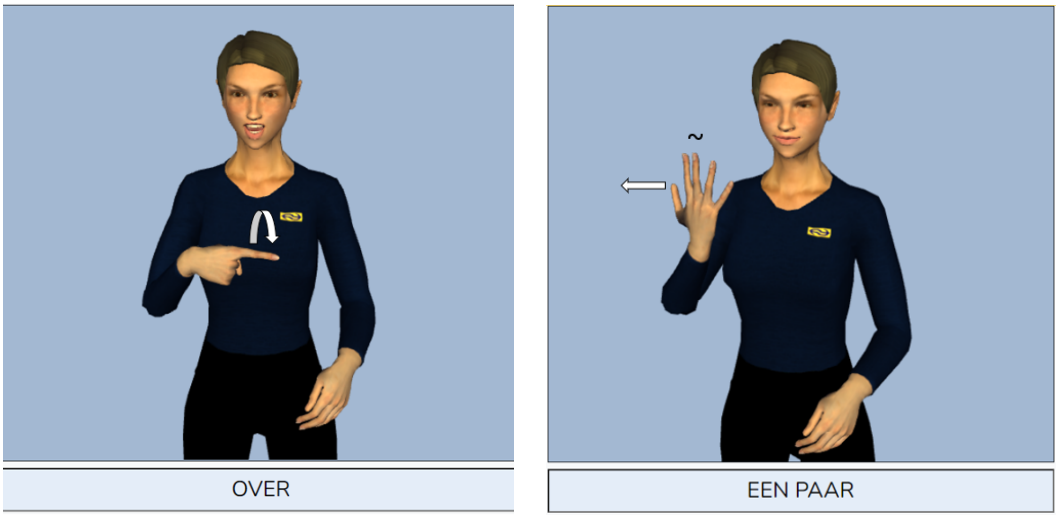
Session 1

OVER (see Figure 5.11(a)) was considered too literal and not used by the collaborators. The path movement of PAAR should be horizontal instead of diagonal/upwards. Moreover, the amount of cheek puff should be reduced (see Figure 5.11(b)).
There was no consensus about the correct version for 'Vertrekken over enkele minuten'. The suggestion was to create the following variants: VERTREKKEN BIJNA *(DEPARTING ALMOST)*, 'VERTREKKEN BINNEN EEN PAAR MINUTEN' *(DEPARTING WITHIN A FEW MINUTES)* and WEGGAAN BIJNA *(LEAVING ALMOST)*.



(a) UTRECHT (b) CENTRAAL

Fig. 5.10: UTRECHT CENTRAAL - V1



(a) OVER - V1 (b) EEN PAAR - V2

Fig. 5.11: OVER *(IN)* and EEN PAAR *(A FEW)* - Muliple versions

Session 2

The hand position for BIJNA (see Figure 5.12(b)) was considered too flat and the movement should be larger. The sign itself may need more emphasis, as there is an action required from the observers (catching the train). To achieve this, a double handed version was experimented with, but eventually rejected by collaborators. BINNEN (see Figure 5.12(a)) was considered a better option compared to OVER (see Figure 5.11(a)). Nonetheless, the BINNEN movement should be improved by a fast repetition (2x) and shifting its position to the right. Overall, the current camera angle prevents a clear overview of the avatar's movements: a more side ward view from the left would improve this. This requires changing all head and eye gaze movements as well. BINNENKORT (see Figure 5.13) was viewed as a good sign, but not appropriate for this context.

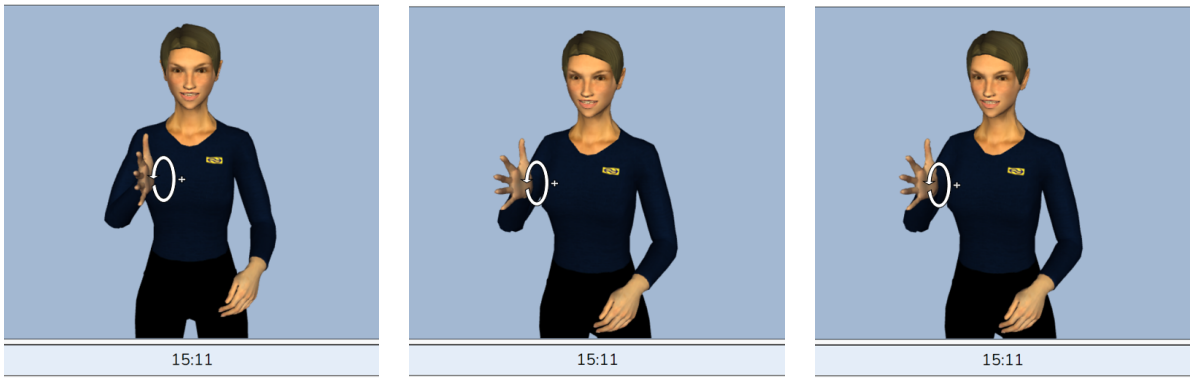
(a) BINNEN (b) BIJNA

Fig. 5.12: BINNEN *(WITHIN)* and BIJNA *(ALMOST)* - V1

Fig. 5.13: BINNENKORT *(SOON)* - V1

Session 3

The camera angle was changed 13° to the left (see Figure 5.14(b)). This new angle was an improvement as opposed to the previous angle (see Figure 5.14(a)). The collaborator preferred a slight increase in head movement, such that the avatar is facing the viewer even more. See Figure 5.14(c) for an illustration of this small adjustment. The improvements of BINNEN *(WITHIN)* and BIJNA *(ALMOST)* were accepted.



(a) V1 (b) V2 (c) V3

Fig. 5.14: 15

'Vertrekken over nog onbekende tijd' *(Departing in an unknown time)*

Session 2
Suggestions for the phrase NOG ONBEKENDE TIJD *(YET UNKNOWN TIME)* were: 'Wanneer? Geen idee' *(When? No idea)* with a shoulder raise, 'Wij/ik weet het niet' *(We/I don't know)*, and 'Tijd is nog niet duidelijk' *(Time is yet unclear)*. The sign for NOG-NIET *(NOT-YET)* was important here, because it implies that clarity will occur soon. Collaborators also noted that the signs for ONBEKEND *(UNKNOWN)* and WETEN NIET *(KNOWING NOT)* are not the same, but the current subtitles imply this similarity. The preferred sentence order would be to first announce the certainties (e.g.: the train departs from platform 5), and afterwards potential uncertainties (e.g.: the time is yet unknown).

'Rijden via een andere route' *(Running through another route)*

Session 1
The avatar's eyes should follow the path movement of the sign ROUTE *(ROUTE)*. The currently used sign for ROUTE gave the impression that the direction of the route was to the north (e.g. to Groningen) (see Figure 5.15). One suggestion was to change this to a horizontal movement instead of a vertical one. For example, the sign for OMWEG. However, this sign was considered more informal.

Session 2
According to the dictionary of the NGC [65], the official sign for OMWEG is also signed in a horizontal path movement. Therefore, we could not create a distinct variant here and skipped this one for the focus group.

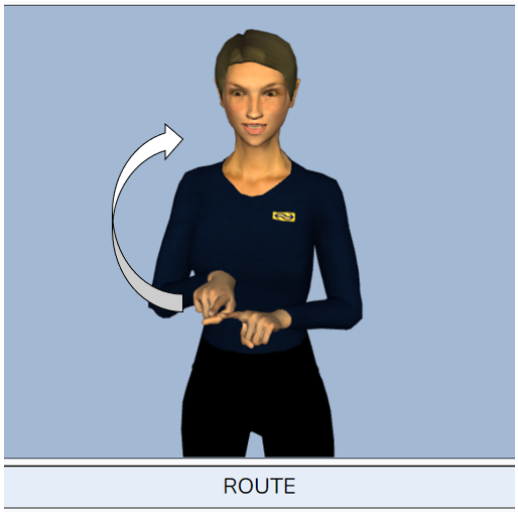


Fig. 5.15: ROUTE *(ROUTE)*

(a) STOPPEN

(b) NIET

Fig. 5.16: STOPPEN-NIET *(NOT STOPPING)* - V1

'Niet stoppen op tussengelegen stations' *(Not stopping at intermediate stations)*

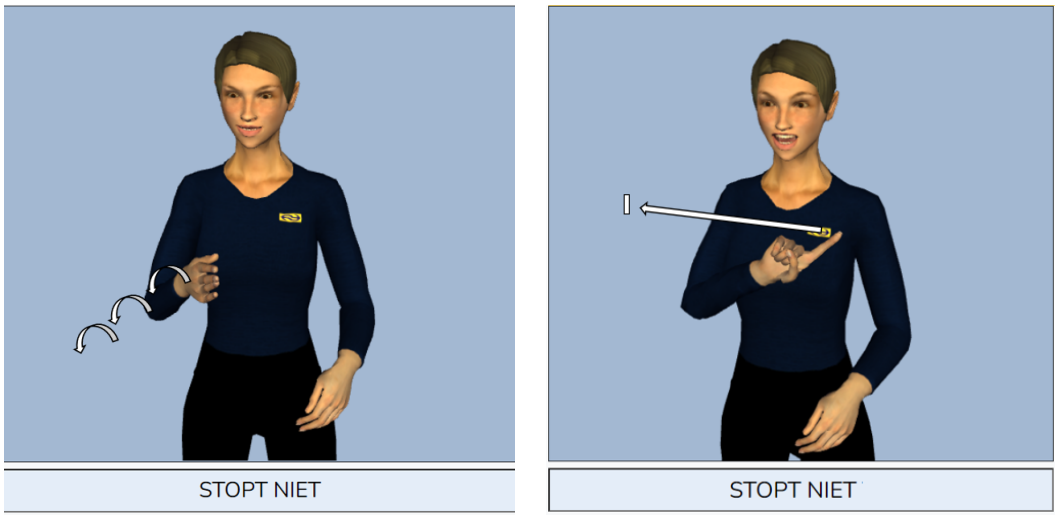
Session 2
Collaborators did not understand this sentence at all. The 'count hands' between station names (see Figure 4.9) gave the impression of 3 individual trains, and the finger tapping movement was distracting. Thus, this counting should be removed. Collaborators considered it more sensible when the end station was mentioned first, and then the intermediate stations where it does not stop. Suggestions were:

- 'De trein naar X, die stopt niet daar *(tussenStation1)*, daar *(tussenStation2)* en daar *(tussenStation3)*' *(The train to X, does not stop there (interStation1), there (interStation2) and there (interStation3))*)

- 'De intercity naar Enkhuizen:'
  - 'Tot Hoorn stopt deze niet bij tussenStation1, tussenStation2 en tussenStation3' *(Until Hoorn it does not stop at interStation1, interStation2 and interStation3)*
  - 'Vanaf Hoorn stopt deze niet bij tussenStation4, tussenStation5 en tussenStation6' *(From Hoorn onwards it does not stop at interStation4, interStation5, interStation6)*

It was also noted that, if you look away for a second, you fail to notice where the train is not stopping. The sign for NIET *(NOT)* was preferred with the 1-handshape here (see Figure 5.16(b)). For the handshape of STOPPEN *(STOPPING)*, several options were suggested: a B1-handshape pointing forward, a B1-handshape pointing diagonally, or the hoek-handshape (similar to BINNEN *(WITHIN)*). The preferred version also depends on the new camera angle for the avatar.

Session 3
The collaborator argued that the phrase TUSSENGELEGEN STATIONS *(INTERMEDIATE STATIONS)* could potentially be removed here, as it mostly induces confusion. Moreover, it is important to indicate the negation within these phrases (e.g. '<negation> NOT STOPPING interStation1 <negation>'). The diagonal movement of NIET *(NOT)* should be changed to a horizontal movement (see Figure 5.17(b)) and should not be too similar to the sign for NAAR *(TO)*. This can be achieve by starting the sign from the elbow instead of the wrist.



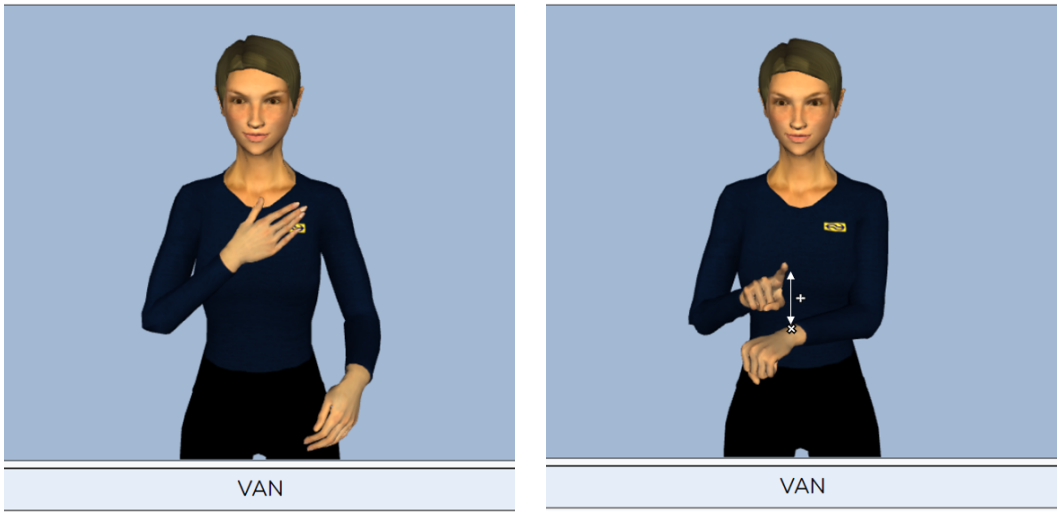
(a) STOPPEN (b) NIET

Fig. 5.17: STOPPEN-NIET *(NOT STOPPING)* - V2

Time Notations and Numbers

Session 1

The sign for TIJD *(TIME)* (see Figure 5.18(b)) was considered as redundant, as it is obvious that a railway announcement is related to time. The preference was to change this sign to VAN *(FROM)* (see Figure 5.18(a)). As VAN and EINDELIJK *(FINALLY)* are a minimal pair, it is important that the mouth picture is distinct.



(a) VAN (b) TIJD

Fig. 5.18: TIJD *(TIME)* and VAN *(FROM)* - V1

For double numbers such as 33 and 55 the motion should origin from the underarm (elbow) instead of the wrist. The non-manual movements were often lengthier compared to the manual movements. Shortening the SAMPA notation for 55 from 'vE_ifv@nvE_iftIx' to 'vE_ifE_iftI' improved this slightly. While this improved the animation, the last part 'tig' is still unsatisfactory. For hours ending on 00 only signing the first number is sufficient. So 11 PUNT 00 *(11 DOT 00)* becomes 11.

(a) Nederlands Gebarencentrum [65] (b) V-handshape

Fig. 5.19: ':' - Multiple variants

Session 2

Opinions about the exact position of the hand for the VAN *(FROM)* varied. For women, this position appears to be slightly higher compared to men. A lower height was thus not necessary for the feminine avatar.

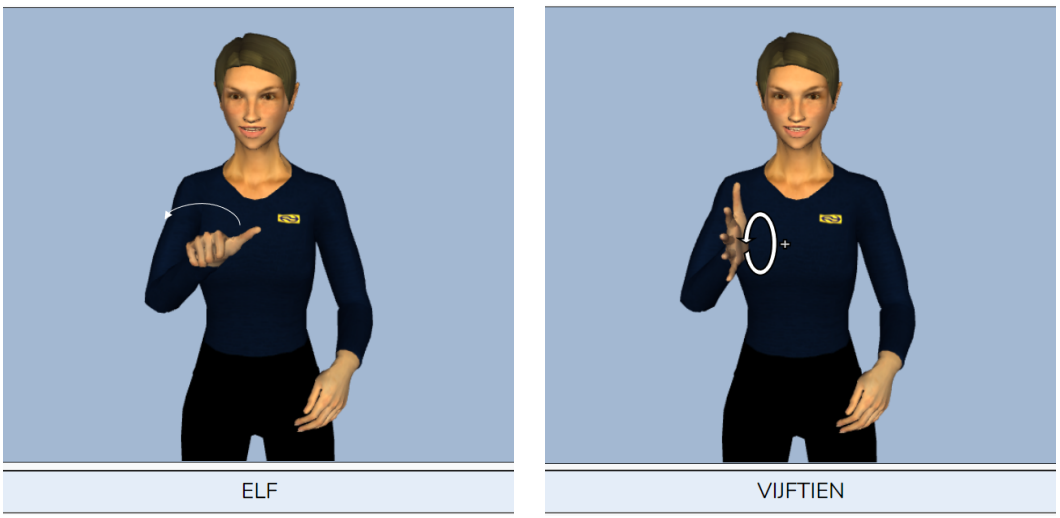
For hours ending on 00, an additional small wiggling movement added to the first number was preferred. This should be a slightly smaller movement compared to double numbers such as 55. 11:00 and 12:00 do not need any additional wiggling movements, because these numbers already have inherent movements.

In general, the time notation with the ':' was fine as it suits the train domain. However, sign preferences for ':' varied. The reference videos from the NGC applied the 1-handshape (see Figure 5.19(a)), whereas collaborators also use the V-handshape [82] (see Figure 5.19(b)). Multiple variants will be created for the focus group: the 1-handshape, V-handshape and baby-O handshape.

The numbers 1A, 1B and 11 were evaluated as sufficient (see Figure 5.20(a), 5.20(b), 5.20(c) and 5.21(a)). For 15, the angle was a problem (see Figure 5.21(b)), but this could be solved by changing the camera view, as mentioned in Section 5.2.1. The wrist motion for 20 (see Figure 5.22(a)) should stay upwards instead of horizontal (see Figure 5.22(b)).

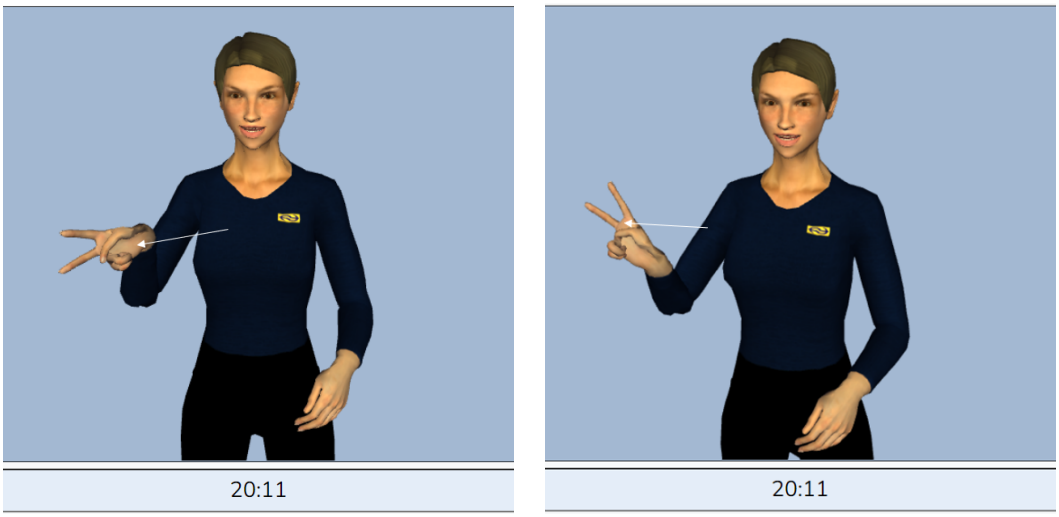
(a) 1 (b) A (c) B

Fig. 5.20: 1, A and B - V1



(a) 11 (b) 15

Fig. 5.21: 11 and 15 - V1

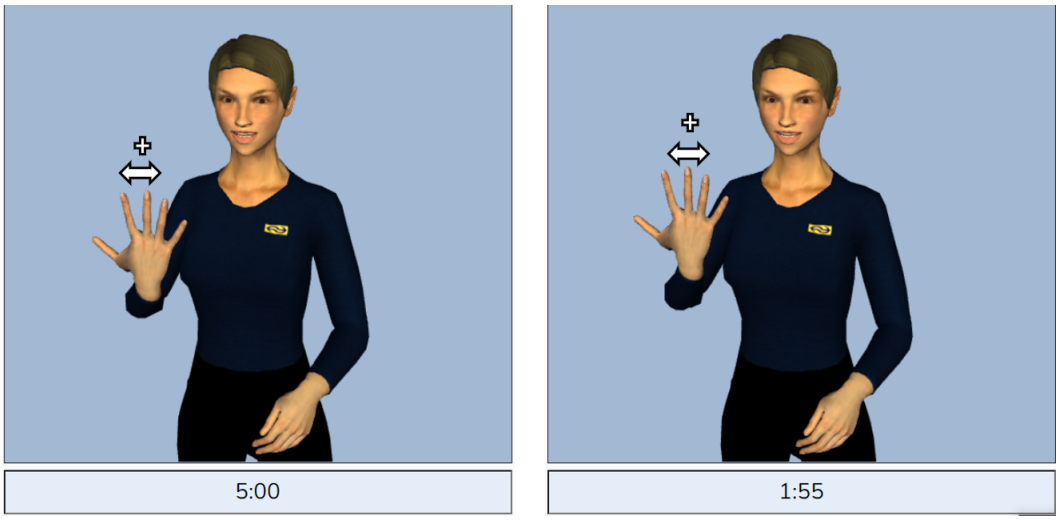


(a) V1 (b) V2

Fig. 5.22: 20 - Multiple versions

Session 3

For hours ending on 00, the movement duration can be slightly extended. This means that '5:00' and 55 will only differ in their hand position (see Figure 5.23). According to the collaborator, this is not a problem: a clear movement within the context prevents misconceptions.



(a) 5:00 (b) 55

Fig. 5.23: 5:00 versus 55 - V3

Optionally, the palm angle for 15 could be left or right oriented, depending on the signers preference. The fluency of the transition from 0 to 5 in notations such as '13:05' should be increased. It should look like one sign instead of two separate signs. The avatar's fingers are frequently inexact: the fingers for 0 do not touch each other (see Figure 5.24(a)), even though they should. Moreover, the thumb and pinky for the 30 have a peculiar ratio (see Figure 5.25(a). Some improvements were made in their second versions (see Figures 5.24(b) and Figure 5.27(b)), but no perfection could be reached.

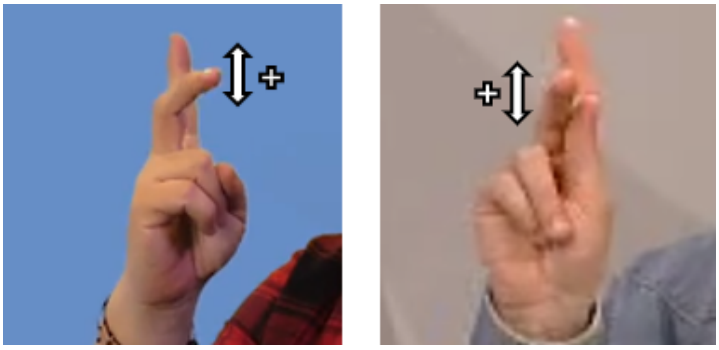
(a) V1 (b) V2

Fig. 5.24: 0 - Finger position - Multiple versions

(a) V1 (b) V2

Fig. 5.25: 30 - Finger position - Multiple versions

There are multiple ways to sign ANDERHALF UUR *(ONE AND A HALF HOUR)* (see Figure 5.27 for an illustration). During the focus group, other variants to sign ANDERHALF *(ONE AND A HALF)* were demonstrated as well (see Figure 5.26). It depends on the exact region which of these two variants is the most commonly used. In the collaborator's experience, the finger sign from Figure 5.26 is not always understood correctly. Therefore, the current way of signing ANDERHALF UUR is sufficient. Some small remarks, however, are that the movement should be clockwise instead of counter-clockwise and the transition between 1 UUR and 30 should be more brief.



(a)
Standard - V-handshape [65]

(b)
Groningen - K-handshape [65]

Fig. 5.26: Regional variances for ANDERHALF

(a) 1 UUR *(ONE HOUR)* - V1 (b) 30 - V2 (c) MINUTEN *(MINUTES)* - V1

Fig. 5.27: ANDERHALF UUR *(ONE AND A HALF HOUR)* - V2

SPOOR *(PLATFORM)*

Session 1
The handshape of SPOOR should be changed to the baby-C [82] and the path movement should turn slightly upwards at the end. The eye gaze should follow the movement of the dominant hand (see Figure 5.28).

Fig. 5.28: SPOOR *(PLATFORM)* - V2

Session 2
The fingers (and especially thumb) of Francoise were considered awkwardly long. Especially for the SPOOR sign this was noticeable. We tried to circumvent this, but did not succeed unfortunately.
Turning the head towards the SPOOR movement was not necessary.

Session 3
The thumb position and appearance was still an issue.

(a) INDEX3a (b) INDEX3b

Fig. 5.29: INDEX3a INDEX3b - V1

Session 1
One suggestion was to add an extra INDEX sign that refers back to the train:

INDEX3a INTERCITY VAN 10:00 INDEX3a GAAN-NAAR-A ...

As SPOOR concerns a new location, this should be indexed as well. The following structure was suggested:

INDEX3b SPOOR spoorNr *(platformNr)*

Session 2
Indexing SPOOR as well was considered correct (see Figure 5.29(b)). However, there were some doubts about the necessity of the double indexing. In general, this is performed more subtly in natural signing. Collaborators suggested to create three variants of this indexing (shorter movement, longer movement, different INDEX position). The head may turn towards the INDEX movement.

Session 3
As an INDEX sign normally points to the location of the object it refers to, the sign for INDEX3b (referring to SPOOR *(PLATFORM)* was relocated to the right side of the avatar. Additionally, two variants per INDEX sign were created: a longer and shorter movement (see Figure 6.4) for illustrations of the longer variants.
According to the collaborator, the position of the INDEX was good. For the focus group, various extra variants with regard to sentence structure will be created. It was noted that capturing the sentence rhythm is important here.

Greeting

Session 1
The handshape of HALLO *(HELLO)* should be formed with a combination of the 5-handshape and B1-handshape to make it more natural (see Figure 5.30).

Fig. 5.30: HALLO *(HELLO)* - V2

Session 2
Adding an extra sign for drawing attention to the avatar at the onset of a new sentence was not necessary (e.g. tapping or waving hands), just HALLO itself 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. Ideally, some flickering lights would be added to the avatar screen.

Announcement Repetitions

Session 1
For the deaf collaborators, the specific way of announcing train messages was yet unknown. This meant they were not aware of the fact that announcement repetitions are always preceded by 'herhaling' *(repetition)*. From a deaf perspective, signing this word did not have any additional value. If you miss the start of the animation (e.g. on a screen when you walk by), you are not aware of this repetition. A better, suggested approach would be to show a red bar below the animation that indicates this repetition for the duration of the entire animation (see Figure 5.31).

Session 2
The current version was accepted and the position below the animation was preferred.

Subtitles

Session 1
A suggestion for the subtitles was to include the entire sentence, and indicate in this sentence what the current signed word is (similar to karaoke).

Fig. 5.31: HERHALING *(REPETITION)* of a sentence - V2

Other

Session 1
Some general comments were made about the avatar's appearance: the translation should be taken seriously so a fictional character is not suitable, but the avatar shouldn't become creepy either (uncanny valley). Other comments encompassed certain grammatical aspects: one collaborator mentioned that sometimes theoretical and common sign language differ. For example, in theoretical ASL, auxiliary verbs (like in spoken English) are generally not used. However, in common ASL people do use them sometimes for clarification. It might be useful to analyse the additional value of these words in NGT for the avatar's natural appearance as well.

It was noted that the mouthings all looked quite similar (i.e. 'bababa'), especially for double (longer) numbers such as 55 (SAMPA: vE_ifv@nvE_iftIx). The origin of the mouthings was explained to the collaborators (SAMPA, see Section 4.1.2) and the mouthings was shortened to: vE_ifE_iftI. Similar changes were applied to other double numbers.

The limitations and possibilities of the current software were discussed as well. Motion capture is preferred amongst the collaborators. However, this approach is currently not scalable enough for this specific domain (containing many variables), and it is costly. If the details get better (e.g. realistic face, enhanced non-manuals and more realistic fingers), collaborators considered the current scripted keyframe animation approach sufficient in the future. To achieve this, continuing the collaboration with Deaf individuals was emphasized once again, as hearing people sometimes have a blind spot for sign language details.

Session 2
The thumbs of Francoise were considered awkward and unnaturally long. This was especially apparent in the sign for SPOOR *(PLATFORM)* (see Figure 5.28).

Session 3

There are still many improvements possible with regards to prosody. However, at this point the priority is that the signs itself should be correct.
The current end pose was considered awkward (see Figure 5.32(a)). The collaborator noted that normally, hands would be folded over each other and positioned on the belly during rest position (see Figure 5.32(b)).

(a) V1 (b) V2

Fig. 5.32: Neutral end poses - Multiple versions

Collaborators indicated that having multiple variants of the animations was desired, as they sometimes had distinct preferences. Evaluation with multiple participants would yield useful insights in these preferences. This conclusion corresponded to our plan from the initial brainstorm session to organise a focus group with a varied group of participants. For this upcoming focus group, the collaborators' preference would be to show three avatars side-by-side, each signing a distinct version of the same phrase. This way, variants can be evaluated in an efficient and clear way. It was also noted that there should be some focus on prosody and the transitions between signs here.

5.3 Discussion

Previous research mentions the fact that animations can be cognitively intense [53]. This may have led to a decrease in concentration for signed sentences at the end of the 2-hour sessions. Future research could pre-select smaller phrases that require less concentration. At a later stage, when these smaller sub phrases are improved, longer sentences containing these sub phrases can be evaluated to address the larger context as well.

Although detailed minutes were created, these are not a literal report of the signed utterances. Moreover, the textual nature cannot capture personal variances of signs. Besides this, interpreters were not always familiar with the research terms and the indirect communication creates an inevitable time lag. This might influence the perception of body language and facial expressions, which prevents an entirely equal participation of Deaf team members. In future research, having only hearing researchers who are fluent in sign language may circumvent this problem. Moreover, approaching interpreters that have already attended previous sessions may also result in more fluent communication and familiarity with the terms, if possible.

Although some topic-marking was added to SPOOR *(PLATFORM)* and TIJD *(TIME)*, other factors (e.g. the sign for TIJD not being informative enough) caused these signs to change drastically before the markings were evaluated. Future research could add multiple types of topic-marking (e.g. backward tilt, head nod, side tilt) for thorough comparison.

Chapter 6 Focus Group

This chapter describes the methods and findings for phase 3a: a focus group to asses the signing quality of the avatar and compare various phrase variants within a diverse group of Deaf participants. This will provide insights in various preferences and perspectives. In the first section, the session outline, participants and subject are discussed. The second section describes the results of each subject. The third section addresses the limitations of these results.

6.1 Methods

6.1.1 Interface

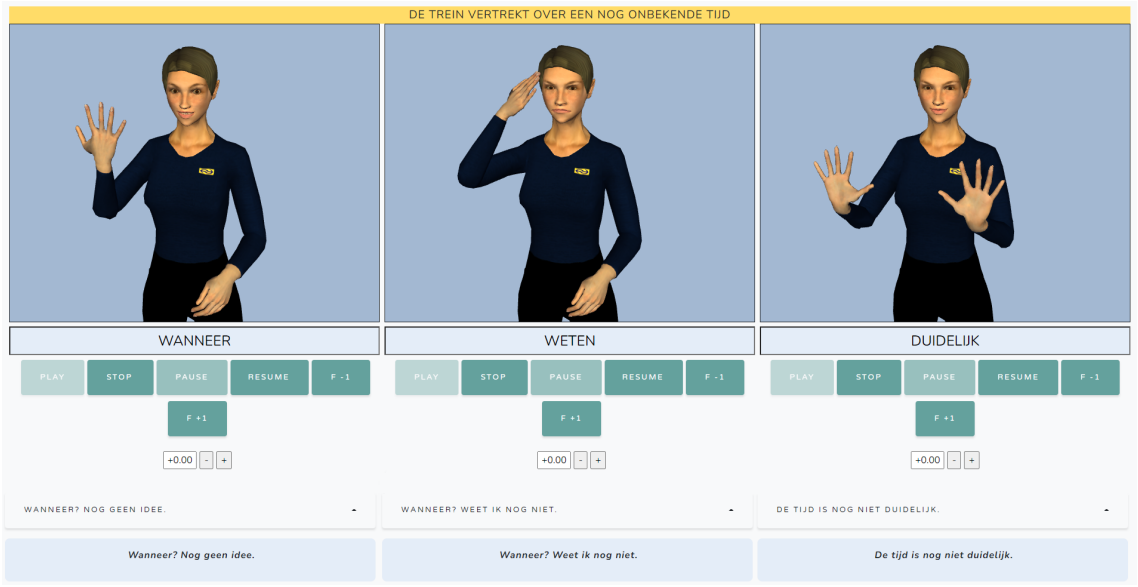


Fig. 6.1: Focus group interface

The interface used for the focus group (see Figure 6.1) can be found via the following URL:

`https://www.signlab-amsterdam.nl/ns-demo/index-focus-groep.html`

The password will be made available upon request.

During the co-design sessions, it was suggested to have 3 variants of a phrase displayed side-by-side. This created a more efficient and clear comparison of phrase variants. Moreover, collaborators noted that the current end pose was considered awkward (see Figure 5.32(a)). Therefore, for the focus group, the current end pose was changed to the new pose illustrated in Figure 5.32(b).

6.1.2 Participants

From the co-design sessions it became clear that age, and personal or regional variances could influence the preferred avatar signing. Especially when evaluating technical applications, age can be an important factor. Therefore, our aim was to create a diverse group of participants for the focus group, with a focus on diverse range of ages (see Table 6.1). The co-design team helped with recruiting possible participants that possessed these features, as they are most familiar with members of the community.

	Age	**Gender**	**Hws* regions**	**Ads****	**Use of NGT**	**Use of written Dutch**
D1	31-40	woman	Noord-Holland Utrecht Groningen	yes	daily, hws	daily, hws
D2	41-50	man	Noord-Holland Utrecht Noord-Brabant	yes	daily, hws	daily, hws
D3	18-30	woman	Utrecht Gelderland Groningen	yes	daily, hws	daily social media
D4	51-60	woman	Utrecht Groningen	yes	daily, hw	daily, hw
D5	18-30	man	Noord-Holland Groningen	yes	daily, h social media	daily, hws
D6	41-50	woman	Flevoland Amsterdam	no (HOH)	daily, hw	daily, hws

Table 6.1:
Focus Group Participants
* = Home, work and school. ** = Attended deaf school (yes/no)

All participants were born Deaf. The level of education varied from 'MBO' *(intermediate vocational education)* to 'WO' *(university)* degrees.

Two collaborators of the co-design session (Sijm and Cokart, referred to as D1 and D2) participated in the focus group as well. This way, they could inform other participants about the enhancement process thus far, share their personal experiences and elaborate on previous choices during group discussions.
Other participants did not have any previous knowledge about the research.

Participants were compensated with two NS daytrip tickets (first class, with a value of 55,80).

6.1.3 Information Beforehand

Information about the session was communicated to participants beforehand via e-mail. The mail included information about the three previous co-design sessions, the reason for an additional focus group, a location and date, and a short description of the focus group agenda. Moreover, it was communicated which interpreters would be present and that Sijm and Cokart would participate (again) as well. If participants accepted the invitation, they received an additional questionnaire about their demographic information. These questions were the following:

- Wat is uw leeftijd? *(What is your age?)*
- Wat is uw geslacht? *(What is your gender?)*
- In welke regio woont u? *(In which region do you live?)*
- In welke regio bevindt uw werk/opleiding zich? *(In which region is your work/education situated?)*
- Wat is uw hoogst genoten opleiding? *(What is your highest level of education?)*
- Op welke leeftijd werd u doof? *(At what age did you become deaf?)*
- Zat u op een dovenschool, en zo ja, welke (naam/regio)? *Did you attend a deaf school? And if so, which name/region?*
- Hoe vaak gebruikt u NGT? (dagelijks/een paar keer per week/een paar dagen per maand/een paar dagen per jaar) *(How often do you use NGT? (daily/a few times a week/a few days a month/a few days a year)*
- Waar gebruikt u NGT? (thuis/werk/opleiding) *(Where do you use NGT? (at home/work/education))*
- Hoe vaak gebruikt u Nederlands? (dagelijks/een paar keer per week/een paar dagen per maand/een paar dagen per jaar) *(How often do you use Dutch? (daily/a few times a week/a few times a month/a few days a year))*
- Waar gebruikt u Nederlands? (thuis/werk/opleiding)? *(Where do you use Dutch? (at home/work/education))*

It was communicated to participants that this demographic information would be anonymized, if published in the thesis or ancillary work. Moreover, it was communicated that no video recordings would be made. The same demo that was used in the co-design sessions was sent to participants beforehand, so they could experiment and get used to the avatar. This demo did not contain the subjects that would be discussed during the focus group, in order to guarantee spontaneity during the actual session.
After the session, it was communicated to them that they would receive the thesis when finished, and would receive an invitation to the final presentation.

6.1.4 Setting

Fig. 6.2: Focus group setting

The focus group (3 hour session) took place at the Roeterseiland complex at the University of Amsterdam (UvA). Unfortunately, one interpreter could not attend due to COVID-19. The short notice, focus group location and 3 hour duration made it impossible to find a new interpreter (despite many attempts). Therefore, the session was held with one interpreter and extra breaks were added to compensate for this. This did not cause any changes to the agenda of the session. Nonetheless, sometimes it was hard for the interpreter to keep up with the discussions when multiple people were signing at the same time.

Roelofsen guided the conversation while van Gemert presented the initial design and wrote the minutes. These minutes included the main subjects, discussion points, recommended changes and action points for the next session. The entire document was checked by Roelofsen, and the entire team had access to it.
As can be seen in Figure 6.2, hearing researcher Roelofsen was positioned next to the screen, which created the opportunity of giving additional explanations and pointing to things on the screen. Hearing researcher van Gemert was positioned in the back of the room, to have a good view of the screen and the participants. Interpreter T1 was positioned next to van Gemert, in order to clearly see all participants and translate loud enough. All participants were positioned at the tables, which were positioned in an L shape such that they could see each other's body language and expressions well during discussions. The table position in the picture differs slightly: the red line indicates the correct position. The screen, Roelofsen, van Gemert and T1 were also visible from the participants' position.

6.1.5 Agenda

The focus group started by introducing the goal of the research and giving a description of what has already been done (initial design and iterative co-design sessions). Subsequently, the agenda was presented on screen and discussed by van Gemert. This way, the order of discussions would be logically structured. During the initial brainstorm session, the idea was to evaluate the avatar in a on-site setting (e.g. Leiden central station) to get a good idea of the most appropriate implementation. However, within the co-design process, it became clear (and was mentioned by collaborators) that animations should be improved with regards to prosody and transitions before any on-site testing on stations could take place (e.g. testing various locations for the avatar in the station). Therefore, prosody was addressed during the focus group, as well as the signing quality for various phrases. At the end of the session, participants could still express their view on the preferred implementation for the avatar. After each subject, the main conclusion was briefly repeated by Roelofsen for confirmation.
In general, variables within phrases were kept the same (e.g. ALMELO and INTERCITY) to ensure consistency and keep the focus on the current subject. If necessary for clarification, there was a possibility to play other variations as well.

1. **Subtitles**
   In the co-design sessions, researchers preferred to see the entire sentence, with an indication of the currently translated word. Therefore, two versions that met these criteria were added. Moreover, the default version (just showing the currently translated word) was added as well for comparison.

2. **Avatar animation speed**
   The JASigning avatar engine offers speed adjustments ranging from 0.00 to +3.00. The default speed is 0.00. The default setting for the three avatars was: -0.20, 0.00 and +0.20. These are just a first setting and can be altered to values < -0.20 or > +0.20 (in steps of -/+ 0.10).

3. **INDEX**
   Two variants for the INDEX sign were created: a longer and shorter movement. Moreover, several positions within the sentence were tested. The fluency of the sentence part 'INDEX3b SPOOR 5 INDEX3b' was enhanced by shortening the duration of the individual signs.

4. **'Vertrekken over nog onbekende tijd'** ***(Departing at an unknown time)***
   Three versions were created, according to the various preferences from the co-design sessions: WANNEER? (NOG)GEEN IDEE *(WHEN? NO IDEA YET)*, WANNEER? NOG-NIET WETEN IK *(WHEN? NOT-YET KNOW I)* and TIJD NOG-NIET DUIDELIJK *(TIME NOT-YET CLEAR)*.

5. **'Vertrekken over enkele minuten'** ***(Departing in a few minutes)***
   Three versions were created, according to the various preferences from the co-design sessions: VERTREKKEN BIJNA *(DEPARTING ALMOST)*, VERTREKKEN BINNEN EEN PAAR MINUTEN *(DEPARTING WITHIN A FEW MINUTES)* and WEGGAAN BIJNA *(LEAVING ALMOST)*.

6. **'Niet stoppen op tussengelegen stations'** ***(Not stopping at intermediate stations)***
   As mentioned during the co-design sessions, it was important to indicate the end station early in the sentence, and potential intermediate stations afterwards. Various handshapes were considered an option for the phrase TUSSENGELEGEN STATIONS *(INTERMEDIATE STATIONS)*. However, removing this entire phrase was a possibility as well. This resulted in the following 3 phrase options:

   Diagonal B1-handshape 'Hallo, de intercity naar Almelo stopt niet op tussengelegen stations Arnhem (hoofdschudden), Breda (hoofdschudden), Zwolle (hoofdschudden).' *(Hello, the intercity to Almelo does not stop at intermediate stations Arnhem (head shake), Breda (head shake), Zwolle (head shake))*,

   Hoek-handshape 'Hallo, de intercity naar Almelo stopt niet op tussengelegen stations Arnhem (hoofdschudden), Breda (hoofdschudden), Zwolle (hoofdschudden).' *(Hello, the intercity to Almelo does not stop at intermediate stations Arnhem (head shake), Breda (head shake), Zwolle (head shake))* and

   Removed phrase 'Hallo, de intercity naar Almelo stopt niet op Arnhem (hoofdschudden), Breda (hoofdschudden), Zwolle (hoofdschudden).' *(Hello, the intercity to Almelo does not stop in Arnhem (head shake), Breda (head shake), Zwolle (head shake))*

7. **Time punctuation**
   Three handshape versions were created voor PUNT *(POINT)*, according to the various preferences from the co-design sessions: V-handshape, baby-O handshape and 1-handshape. The phrase '15:31' was used to demonstrate the versions.

8. **Pauses**
   Previous research found that for ASL, pauses and syntactic structure are related (similar to spoken English) [33]. Therefore, in this study, the rhythm of the current sentences was analysed based on the X-bar syntax theory. The goal was to optimize the rhythm of the sentences by adding pauses based on their POS-tags (Part Of Speech). This resulted in 134 ms pauses between conjoined sentences, 106 ms between Noun Phrases (NP) and Verb Phrases (VP), 6 ms within NP and 11 ms within VP. For comparison, the first avatar contained these aforementioned pause duration (in ms), the second avatar did not contain any pauses, and the third avatar contained half of the pause duration from the first avatar.

9. **Other**
   At the end of the session, some general questions were asked about the preferred application of a potential sign language avatar:

   (a) What would be your preferred location for the avatar, if you are positive about the idea at all?

   (b) For which type of announcements would you prefer to see the avatar?

6.2 Results

A detailed report (in Dutch) can be found on:

`https://github.com/froelofs/signlab/tree/master/ns-demo/notulen`

The most important findings will be discussed in this section.

Most animations were played one at a time and were replayed if considered necessary or desired by participants. After playing, the animations were evaluated and participants expressed their preferred choice. In some cases, all three animation were played at the same time to thoroughly compare small details (e.g. time notation 15:31, where the ':' notation was evaluated). An overview of the analysed phrases and variants for all sessions can be found in Section 6.1.5.

6.2.1 Subtitles

1. Hallo, de intercity naar Almelo vertrekt van spoor 5. (see Figure 6.3(a))
2. Hallo, de intercity naar Almelo vertrekt van spoor 5. (see Figure 6.3(b))
3. Hallo, de intercity naar Almelo vertrekt van spoor 5. (see Figure 6.3(c))

In general, participants considered subtitles as an addition to the animation under the condition that the entire Dutch sentence was presented (similar to subtitles on TV). A gloss notation of the signs, as opposed to spoken Dutch translations, was not necessary. A prerequisite was that the avatar should follow its own (signed) translation and not the Dutch translation, because of differences in word order. This difference was also the reason that participants rejected the idea of highlighting the currently signed word within the entire Dutch sentence. They were convinced this would not function as desired. Displaying solely the Dutch words corresponding to the current sign (as opposed to the full sentence) was considered too fast and did not convey the full information demand (e.g. knowing which train to take at a quick glance).

6.2.2 Animation Speed

The default speed (JASigning 0.00) was considered very slow by all participants. Opinions varied between a preferred speed of +0.20, +0.40 or +0.60. Participants assumed that for some Deaf people, +0.40 and higher would probably be too fast. One participant noted that increasing the speed levels to +0.40 and +0.60 resulted in more 'staccato' animations. Others discussed that for fluent animations, transitions between signs were important (as also discussed during the co-creation session for NAAR-tussen/eindStation *(TO-inter/endStation)* (see Section 5.2.1). This triggers others to start a small conversation about the location of certain signs (e.g. GRONINGEN is signed higher in the signing space compared to MAASTRICHT, according to an imaginary map of the Netherlands). It became clear that the size of this imaginary map varied between participants and influenced the endposition of the previous sign NAAR (*TO*). Handedness also played a role in these positions. The optimal speed was also perceived to be influenced by the presence of subtitles.

Faster rates would be harder to follow without textual support and clarifying signs could be necessary here. After consensus, the default speed was set to +0.40 for all future animations during the focus group.

(a) Entire phrase with current Dutch word marked

(b) Current Dutch word

(c) Entire phrase with current Dutch word

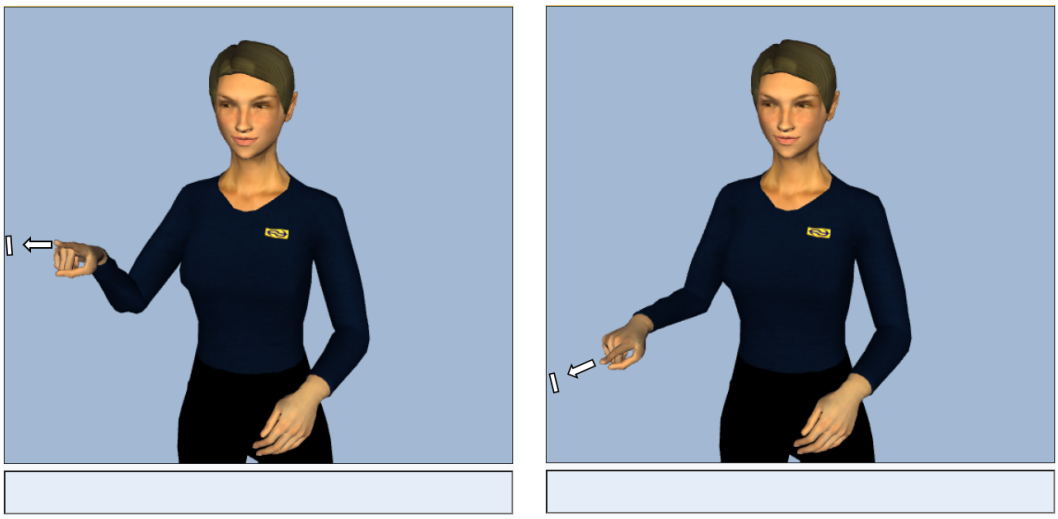
Fig. 6.3: Subtitle options for avatar 1, 2 and 3

6.2.3 INDEX

1. Hallo, **INDEX3a_lang** intercity naar Almelo **INDEX3a_lang** vertrekt van **INDEX3b_lang** spoor 5.

2. Hallo, **INDEX3a_kort** intercity naar Almelo **INDEX3a_kort** vertrekt van **INDEX3b_kort** spoor 5.

3. Hallo, **INDEX3a_kort** intercity **INDEX3a_kort** naar Almelo **INDEX3a_kort** vertrekt van **INDEX3b_kort** spoor 5 **INDEX3b_kort**.



(a) INDEX3a_lang *(INDEX3a_long)* (b) INDEX3b_lang *(INDEX3b_long)*

Fig. 6.4: INDEX3a_lang and INDEX3b_lang - V2

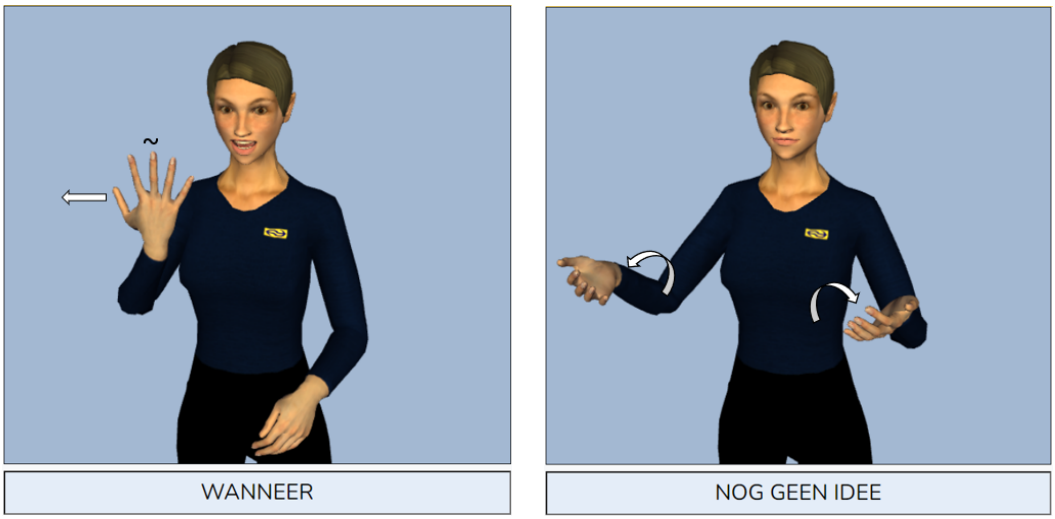
For INDEX3a_short and INDEX3b_short, the endpoint of the movement occurs earlier: around the center of the arrow.

All three animations were played twice. In general, the long INDEX movement was dismissed and participants decided that avatar 3 included too many INDEX movements. The more subtle INDEX motions of avatar 2 and 3 were received positively. However, the preferred suggestion was not to perform the whole movement, but only alter the handshape (in the case of SPOOR 5 *(PLATFORM 5)*: from the 5-handshape to the 1-handshape). A small wrist rotation could be added as well. Using INDEX3b_kort *(INDEX3b_short)* twice in the phrase SPOOR 5 from avatar 3 elicited a discussion about the necessity of this backreference. Some participants argued that this was grammatically correct, while others thought it was redundant. Eventually, they all agreed that using INDEX once, before SPOOR 5 or afterwards, would be sufficient.
The extra fluency for individual signs in the phrase 'INDEX3b_kort SPOOR 5 INDEX3b_kort' for avatar 3 was received positively.
A combination of avatar 2 and 3, combined with only performing a subtle hand/wrist movement was considered optimal. The resulting preferred sentence (in gloss notation) would be as follows:

> INDEX3a_kort (subtle) INTERCITY GAAN-NAAR-A ALMELO INDEX3a_kort (subtle) VERTREKKEN SPOOR 5 INDEX3b_kort (subtle)

6.2.4 'Vertrekken over nog onbekende tijd' *(Departing in an unknown timeframe)*

Avatar 1: WANNEER? (NOG)GEEN IDEE *(WHEN? (YET) NO IDEA)*



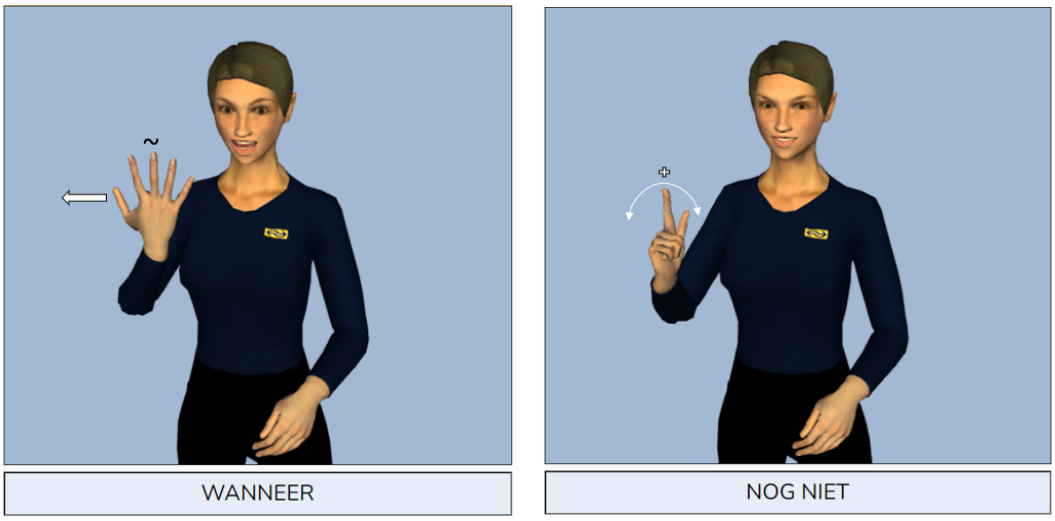
(a) WANNEER (b) (NOG) GEEN IDEE

Fig. 6.5: WANNEER (NOG) GEEN IDEE *(WHEN (YET) NO IDEA)* - V1

Participants suggested changing the sign for WANNEER *(WHEN, see Figure 6.5(a))* to a two-handed version with a more intense wiggling movement to clarify the meaning. Furthermore, they agreed that the non-manual features of NOG GEEN IDEE *(NO IDEA YET)* needed improvements. Briefly, the SiGML structure that composes the signs was explained and, as suggested by the participants, a shoulder and eye brow raise were implemented directly. The new sign for NOG GEEN IDEE is illustrated in Figure 6.5(b). This version was received as an improvement.

Fig. 6.6: NOG GEEN IDEE *(YET NO IDEA)* - V2

Avatar 2: WANNEER? NOG-NIET WETEN IK *(WHEN? NOT-YET KNOW I)*



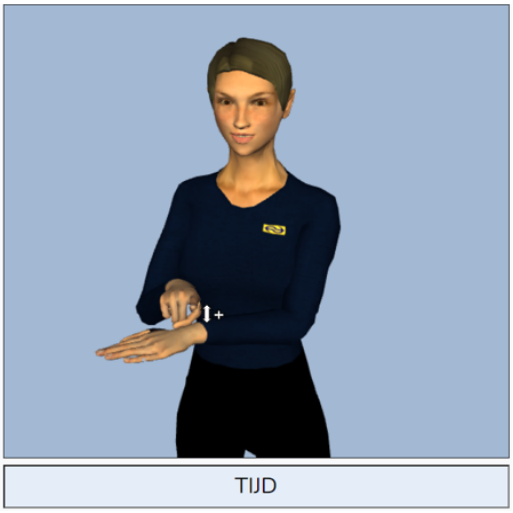
(a) WANNEER (b) NOG-NIET

(c) WETEN (d) IK

Fig. 6.7: WANNEER? NOG-NIET WETEN IK *(WHEN? NOT-YET KNOW I)* - V1

The order of the individual signs was considered incorrect. NOG-NIET WETEN *(NOT-YET KNOW)* was corrected to WETEN NOG-NIET *(KNOW NOT-YET)*. In general, this version did not trigger enthusiasm.

Avatar 3: TIJD NOG-NIET DUIDELIJK *(TIME NOT-YET CLEAR)*



(a) TIJD

(b) NOG-NIET

(c) DUIDELIJK

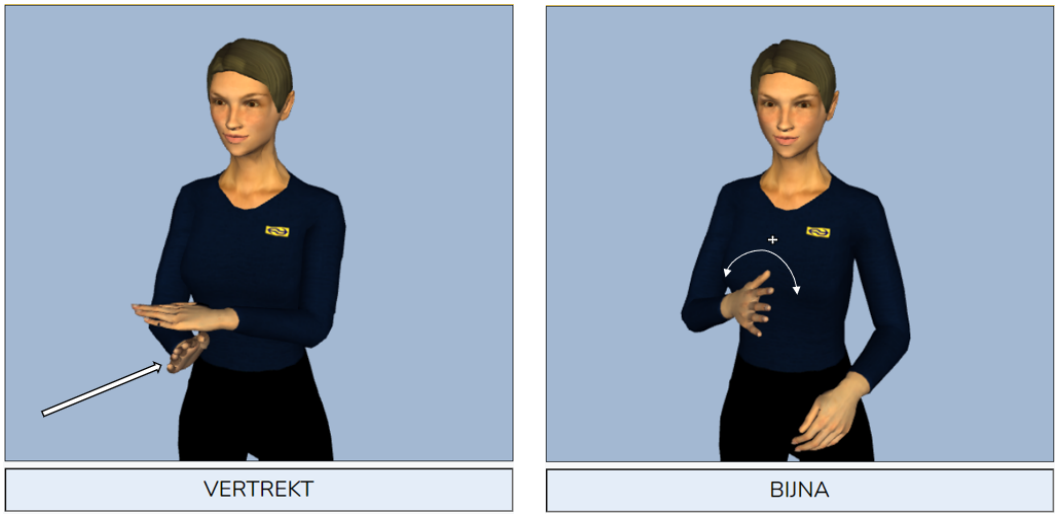
Fig. 6.8: TIJD NOG-NIET DUIDELIJK *(TIME NOT-YET CLEAR)* - V1

The sign for DUIDELIJK *(CLEAR)* (see Figure 6.8(c)) should be signed higher in the signing space and a shoulder raise should be added. Participants agreed that it was a good thing that this phrase did not contain any pronouns, as this might not be suitable for this domain and adds extra signs to the phrase.

Other suggestions for this phrase were: INFORMATIE NOG-NIET BESCHIKBAAR *(INFORMATION NOT-YET AVAILABLE)* and SPOOR NOG-NIET DUIDELIJK *(PLATFORM NOT-YET CLEAR)*. However, the first version was considered a very 'Dutch' translation. Participants preferred the version of avatar 3 over the other variants. Especially because it has a focus on the subject of time itself.
They agreed that the specific order of the phrase (DUIDELIJK NOG-NIET *(CLEAR NOT-YET)* or NOG-NIET DUIDELIJK *(NOT-YET CLEAR)*) depends on what you want to emphasize. This emphasizing approach is common in NGT.

6.2.5 'Vertrekken over enkele minuten' *(Departing in a few minutes)*

Avatar 1: VERTREKKEN BIJNA *(DEPARTING ALMOST)*

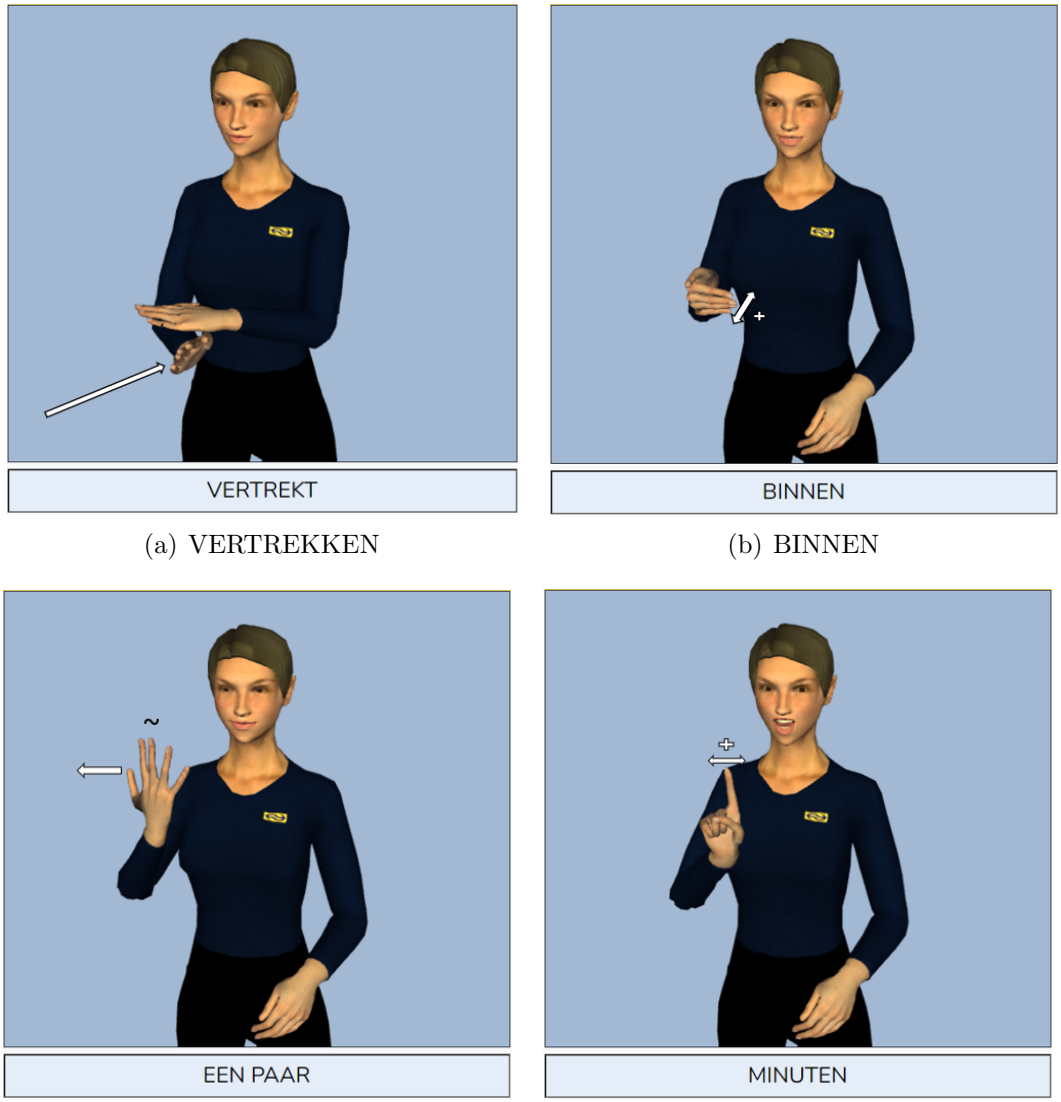


(a) VERTREKKEN (b) BIJNA

Fig. 6.9: VERTREKKEN BIJNA *(DEPARTING ALMOST)* - V1

Due to the adjusted speed from Section 6.2.2, BIJNA (see Figure 6.11(b)) was perceived as extremely fast. Manually resetting this sign to +0.00 resolved this. Additionally, non-manuals were insufficient for this sign. Participants indicated that a brow raise, shoulder raise and (slight) forward head movement were preferred here.

Avatar 2: VERTREKKEN BINNEN EEN PAAR MINUTEN *(DEPARTING WITHIN A FEW MINUTES)*



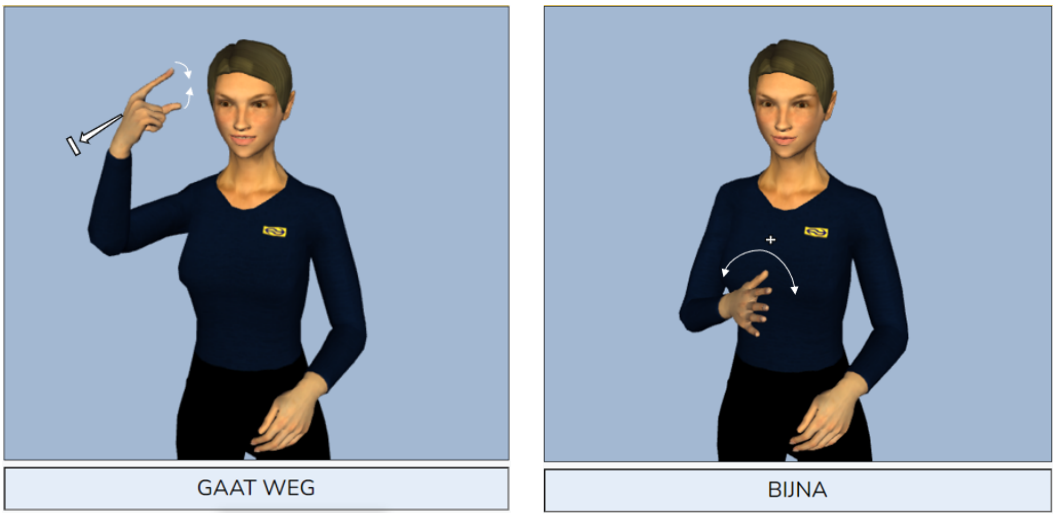
(a) VERTREKKEN (b) BINNEN

(c) EEN PAAR (d) MINUTEN

Fig. 6.10: VERTREKKEN BINNEN EEN PAAR MINUTEN *(DEPARTING WITHIN A FEW MINUTES)* - V1

For MINUTEN *(MINUTES)* the mouth picture 'min' was suggested as well, other participants noted that the picture 'minuten' was grammatically necessary. However, the quality (SAMPA) was not sufficient. Additionally, the order of the phrase was reconsidered: One participant noted that, grammatically speaking, the time notation should be placed at the start of the phrase. However, the structure 'De treinType vertrekt over wachtTijd van spoor spoorNr' *(De trainType departs in waitTime from platform platformNr)* was preferred over 'De treinType van spoor spoorNr vertrekt over wachtTijd' *(The trainType from platform platformNr departs in waitTime)*. In their opinion, this presents the most urgent information first (whether the train is departing or not).

Avatar 3: WEGGAAN BIJNA *(LEAVING ALMOST)*



(a) WEGGAAN (b) BIJNA

Fig. 6.11: WEGGAAN BIJNA *(LEAVING ALMOST)* - V1

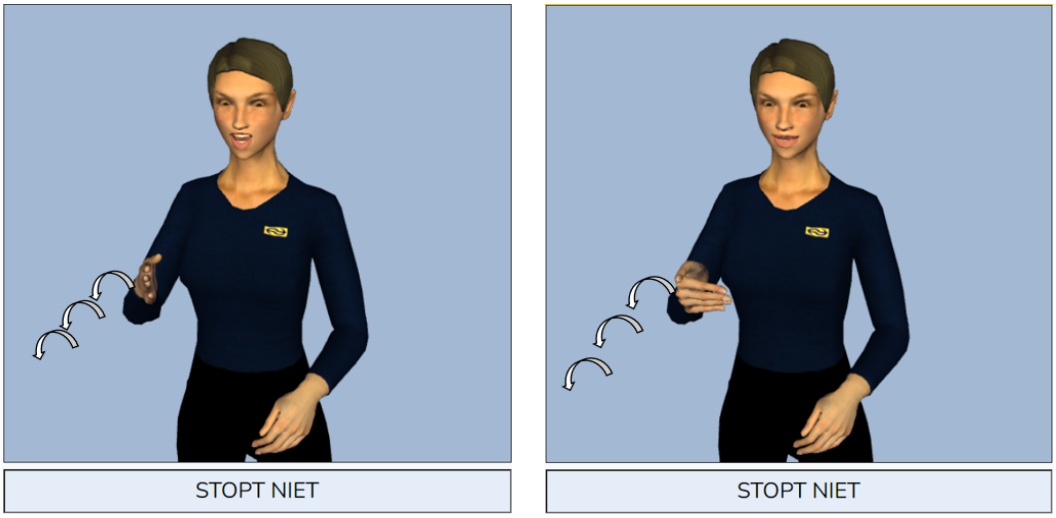
Avatar 3 elicited a discussion about the use of WEGGAAN *(LEAVING)* versus VERTREKKEN *(DEPARTING)*. WEGGAAN was perceived as an easier option when signing with one hand. One participant made the comparison that for flying it is common to sign WANNEER VLIEGEN VLIEGTUIG *(WHEN FLYING AIRPLANE)* instead of WANNEER VERTREKKEN VLIEGTUIG *(WHEN DEPARTING AIRPLANE)*. Participants questioned the formality of WEGGAAN and concluded that the comparison was a matter of preference.

6.2.6 'Niet stoppen op tussengelegen stations' *(Not stopping at intermediate stations)*

1. **Diagonal B1-handshape** Hallo, de intercity naar Almelo stopt niet op tussengelegen stations Arnhem (hoofdschudden), Breda (hoofdschudden), Zwolle (hoofdschudden). *(Hello, the intercity to Almelo does not stop at intermediate stations Arnhem (head shake), Breda (head shake), Zwolle (head shake))* (see Figure 6.12(a))

2. **Hoek-handshape** Hallo, de intercity naar Almelo stopt niet op tussengelegen stations Arnhem (hoofdschudden), Breda (hoofdschudden), Zwolle (hoofdschudden). *(Hello, the intercity to Almelo does not stop at intermediate stations Arnhem (head shake), Breda (head shake), Zwolle (head shake))* (see Figure 6.12(b))

3. Hallo, de intercity naar Almelo stopt niet op Arnhem (hoofdschudden), Breda (hoofdschudden), Zwolle (hoofdschudden). *(Hello, the intercity to Almelo does not stop in Arnhem (head shake), Breda (head shake), Zwolle (head shake))*



(a) B1-handshape (b) Hoek-handshape

Fig. 6.12: STOPPEN NIET *NOT STOPPING*- Multiple handshapes

Fig. 6.13: STOP *(STOP)* - V1

In general, the animation where the phrase TUSSENGELEGEN STATIONS was removed was preferred. Participants mentioned that for NGT the handshape often depends on the signed object. For example, a B1-handshape with the palm facing down and fingers pointing forward represents a car.

One participant also noted that the forward motion of the sign for STOPPEN *(TO STOP)* gave the impression that the train is still moving forward. The hoek-handshape did not give this impression. Another participants felt that the sign for STOP with a B1-handshape moving forward should be present (see Figure 6.13). The rest agreed. A headshake should be included as well.
It was mentioned that manually counting within a list of station names is common (grammatically speaking), although it was not perceived appropriate for the avatar. A new version was suggested where station names were made visible in a horizontal line (see Figure 6.16 for an illustration). Here, rhythm is very important, as was

illustrated by participants on the whiteboard (see Figure 6.14 for this illustration). The suggested structure was as follows:

> <hoofdschudden> STOPT (Figure 6.13) NIET (Figure 5.16(b)) BREDA, ARNHEM, ZWOLLE </hoofdschudden>

<hoofdschudden> and </hoofdschudden> indicate that the headshake spans across the entire phrase.

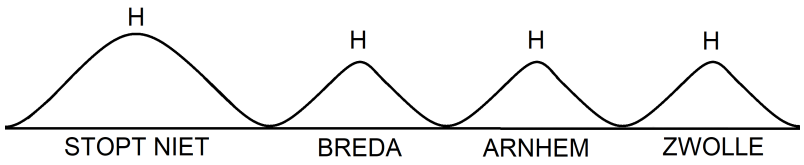


Fig. 6.14:
Suggested rhythm
H: The line indicates the intensity of the head shake: the minimum denotes a subtle movement and the maximum an intense movement.

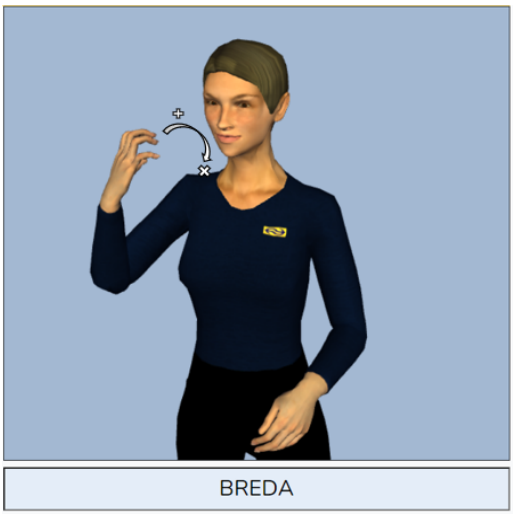
6.2.7 Time punctuation

(a) V-handshape (b) baby-O handshape (c) 1-handshape

Fig. 6.15: PUNT - Multiple handshapes

Multiple improvements were suggested for the V-handshape: a more upward onset of the sign by turning the wrist, placing the sign a bit higher in the signing space, and viewing the sign even more sideways for clarity. Overall, participants rejected the baby-O handshape. Other preferences varied between the V-handshape and 1-handshape [82]. In general, younger participants ($\leq$ 44 years) preferred the V-handshape, whereas older participants ($>$ 45 years) preferred the 1-handshape.

(a) INDEX3a



(b) BREDA

(c) INDEX3b

(d) ARNHEM

(e) INDEX3c

(f) ZWOLLE

Fig. 6.16: Suggested station list notation

6.2.8 Pauses

1. Hallo **134**, de intercity **6** naar Almelo **6** van 13:05 **106** rijdt niet.
2. Hallo **0**, de intercity **0** naar Almelo **0** van 13:05 **0** rijdt niet.
3. Hallo **67**, de intercity **3** naar Almelo **3** van 13:05 **53** rijdt niet.

During the co-design, it was noted that transitions between numbers within time notations should be more fluent. Due to the adjusted speed (+0.40), these signs were not identifiable anymore. Overall, participants preferred the avatar without breaks between signs. However, eventually, short breaks (syntactically-based breaks / 2) were considered a safer option for people who are less fluent in sign language. In the case that breaks are omitted, the time notation numbers should be slowed down. For the short breaks, this was not considered necessary.

6.2.9 Other

The mouthing quality and awkwardly long thumbs of the avatar were mentioned here again.
A few suggestions for the greeting at the start of a phrase were given. As mentioned in Section 4.2, NS' BESTE REIZIGERS *(DEAR PASSENGERS)* phrase was changed to HALLO *(hello)*. Focus group participants suggested GOEDENDAG/ GOEDEMORGEN (*GOOD DAY/GOOD MORNING*) or just waving without articulating anything.
Participants agreed that the sign ANNULEREN *(TO CANCEL)* in the sentence 'Hallo, de intercity naar Almelo van 13:05 rijdt niet.' *(Hello, the intercity to Almelo from 13:05 is not departing.)* needed a stronger forward wrist motion. Moreover, the headshake was considered ambiguous and should be removed, because it was not clear whether the train was canceled or not canceled (double negation).
Sometimes, participants would engage in discussions that deviated from the current subject. Discussions still related to the avatar (e.g. comments about the greeting) were left to run their course. Unrelated discussions were interrupted by the researchers.
Participants suggested that in the future, having the possibility to choose between various avatar genders within the NS application would be preferred. For example choosing between a woman or man, or making the character androgynous (e.g. by removing the breasts for the current character, as suggested by a participant).

Setup

D1 indicated that it would be better to record future focus groups on video, to fully understand participants' signing variations and dialects. For example, the same Dutch word communicated by the interpreter could be signed differently by distinct participants. Moreover, some recommendations were difficult to explain by solely using words (e.g. the discussion about performing the same sign at various location in the signing space).
Eventually, the optional quantitative (online/offline) evaluation was omitted, because of the restricted time frame that was left for the study.

Participants recommended to recruit people from the Deaf community from certain communes as well, as this would provide additional insights into other reading capabilities.

Application Preferences

Participants indicated that they would like to see the avatar on both their phones as well as screens in and around the trains. Especially at the entrance of stations, so announcements are immediately visible. They prefer to see the avatar within the regular information frames and not separately in a corner of the screen. Having this information accessible for everyone within one frame/screen creates an inclusive atmosphere.

Preferably, the type of avatar translations should be based on the same information that is communicated through the speakers (incidental announcements). Animations on screens should activate simultaneously with these announcements. Participants currently use the behavior of others within trains (e.g. people suddenly looking up) as a visual cue for important information, but prefer flickering lights or visual cues on the ground to catch their attention. One participant mentioned an example from the Washington metro, where lights at the platform edges are activated once a metro enters the station.

Regarding the mobile application, push notifications for important announcements/animations are desired. This extra information source is considered especially useful when announcements on screen are overlooked. Participants also suggested a possibility to personalize these notifications to your personal trajectories within the app.
Some ideas that were mentioned but require more complex adaptations were: buttons on chairs to indicate that you are deaf, resulting in specialized chair vibrations that indicate the onset of holographic announcements.

6.3 Discussion

The focus group emphasized the importance of context and how it could influence the comprehensibility. Nonetheless, due to the many necessary enhancements to the signing quality itself, the specific application of the avatar on-site is still underresearched. As a result, analyzing the avatar in an actual on-site setting could affect the current results, especially given the time-sensitivity of information. Future focus groups should evaluate the avatar in on-site settings.

Although detailed minutes were created, the lack of a written form of sign language prevents a literal transcript of the signed utterances. Moreover, a written form cannot capture personal variances of signs. Besides this, communication via an interpreter creates an inevitable communication lag, which might influence the perception of body language and facial expressions. For future research, making video recordings would be preferred, as it participants captures individual signing variances and prevents missing information or body language when multiple people are signing at the same time. Having hearing researchers who are able to fully communicate in sign language circumvents the communication lag as well.

The preferred pauses within a sentence differed from the syntactic basis acquired from the syntax trees and ASL research. However, individual sign duration and speed may have influenced the general perception of speed. Moreover, although most participants preferred animations without pauses, they assumed adding pauses would be more suitable for people who are less fluent in sign language. Future research in a on-site setting with participants that possess this characteristic is necessary to confirm this assumption.

During the discussion of the animation speed, the preferred speed was immediately applied to all future animations, in order to evaluate signs in the new setting. Nonetheless, this caused some signs to appear unexpectedly fast (e.g. BIJNA *(ALMOST)*), and it affected previously cliticized transitions between numbers for time notations (e.g. 05 in 13:05). This shows that even small adjustments can have mayor effect on the comprehensibility. This complicates the development of a scalable translation system, as manual fine-tuning may still be necessary to provide comprehensive and natural animations.

Within focus groups, participants should feel comfortable criticizing the system. On the one hand, the presence of the developers and interpreter during the focus group might affect the discussion, as participants could have felt less comfortable criticizing the system. On the other hand, not having developers present during the focus group would lead to less direct input and would take away the possibility for direct implementation and evaluation of certain suggestions.

Based on their daily use of written Dutch (see Table 6.1), the reading level of participants might not be representative for the entire Deaf community. This might have influenced their preferences (e.g. subtitles).

For sentences such as 'niet stoppen op tussengelegen stations' *(not stopping at intermediate stations)*, the order of the signs, rhythm and optional number of intermediate stations make the sentence complex. For such sentences, more focus groups have to be conducted in order to achieve the maximum comprehensibility and naturalness.

Chapter 7 NS Panel

This chapter describes the methods and findings of phase 3b: consulting the NS Panel to create an overall picture of how the avatar will be received in a public setting with both Deaf and hearing passengers. In the first section, the chosen statements, open questions and corresponding analysis is described for the questionnaire. The second section describes the results for each statement and open question. The third section addresses the limitations of these results. The resulting data set is confidential and cannot be shared.

7.1 Methods

The NS Panel is an online collaboration between the NS and its passengers. Members can voluntarily sign up and give their opinion in various research topics from NS. The studies vary from online questionnaires to polls and discussions. After collecting and analysing responses, the results of these studies are communicated back to the panel via an online news letter [68].

Responses were collected between 28th of January and the 6th of February 2022. The desired number of responses was *n=500*. The criteria for participation were that the respondent should minimally have 3-5 trips per year. This guaranteed that social, recreational and study or work related travel behavior was included, as all of these groups are confronted with the avatar. Participating in the research was on a voluntary basis (self-selection), no payment or compensation was included.

7.1.1 Questionnaire

The study was executed by MWM2, an independent market research company. The invitation mail for the study can be found in Appendix B. The questionnaire can be found in Appendix B. Because of the exploratory nature of the research, we could not present examples of how or where exactly this avatar will be shown (e.g. position on screens and/or in the app), as it is not clear yet whether the avatar will be put into practice. A YouTube URL with a video of the avatar was included in the questionnaire, to give a broad impression of the system. The signed phrase was: 'De intercity naar Almelo van 11:55 rijdt niet' *(The intercity to Almelo of 11:55 is not departing)* After showing the video, participants could indicate their opinion on a 4-point Likert scale. The 'neutral' option was labeled 'I don't know', because having a neutral option may yield confusing responses. The questions for gauging the opinion of participants about the avatar were as follows:

- Het lijkt mij een goed idee om NS berichten ook in Nederlandse Gebarentaal (NGT) aan te leveren. *(I think it would be a good idea to also provide NS announcements in Dutch Sign Language (NGT).)*
- Deze toepassing lijkt mij nuttig voor NS reizigers die Nederlandse Gebarentaal (NGT) beheersen. *(This application seems useful to me for NS passengers that are proficient in Dutch Sign Language (NGT).)*
- De virtuele avatar zou mij inspireren om meer te leren over de Nederlandse Gebarentaal. *(The virtual avatar would inspire me to learn more about Dutch Sign Language.)*
- Ik vind het nuttig dat de video met gebarentaal ondertiteling bevat. *(I think it is useful that the sign language video contains subtitles.)*

Possible advantages and disadvantages of the avatar were asked, and general comments:

- 'Wat ziet u als de grootste voordelen van gebarentaal communicatie d.m.v. een virtuele tolk?' *(What are the biggest advantages of sign language communication by a virtual avatar?)*
- 'Wat ziet u als mogelijke nadelen van gebarentaal communicatie d.m.v. een virtuele tolk?' *(What could be possible disadvantages of sign language communication by a virtual avatar?)*
- 'Heeft u nog andere opmerkingen over deze toepassing?' *(Do you have any other comments about this application?)*

The participants' proficiency in NGT and technology was addressed, as well as their use of the NS application:

- 'In welke mate beheerst u Nederlandse Gebarentaal (NGT)?' *(To what extent are you proficient in Dutch Sign Language?)*
- 'Bent u vaardig met technologische apparaten zoals smartphones en computers?' *(Are you proficient with technical devices such as smartphones and computers?)*

7.1.2 Analysis

For analysis of the multiple-choice statements, graphs were created. For each option, the percentage of participants that chose the option was indicated. The open answers were analysed by colour-coding: green indicated a generally positive opinion about the usefulness of the application, red indicated a general negative opinion. Moreover, the occurrence of some often mentioned topics was counted. These were: 'Deaf people can read', 'the avatar is also useful for hearing people', 'the avatar raises awareness for NGT', ideas about attracting attention for the animation, ideas about the eventual application of the avatar, 'distracting/cluttered information' or 'overload of information' mentioned as a disadvantage, 'confusing' mentioned as a disadvantage, and for whom they envisioned this disadvantage (Deaf, hearing or unknown). The number of unusable open answers (e.g. empty answers) was counted as well. The results will be communicated back to the respondents via the monthly news letter of the NS Panel and the analysed data set was shared with NS.

7.2 Results

In total, the questionnaire provided us with 641 responses. 52% men and 48% women attended the study. Only 30% of participants was aged < 55, 70% had an age >= 55 years.

7.2.1 Multiple Choice Questions

Het lijkt mij een goed idee om NS berichten ook in Nederlandse Gebarentaal (NGT) aan te leveren. *(I think it would be a good idea to also provide NS announcements in Dutch Sign Language (NGT).)*

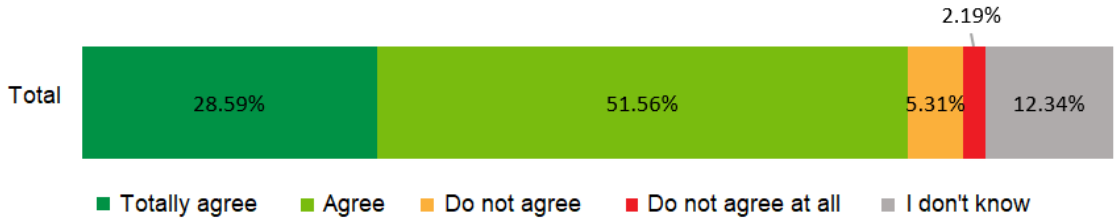


Fig. 7.1: Panel results statement 2.1

The majority of the respondents (80.15%) was positive about the idea of providing NS announcements in Dutch Sign Language (NGT), 7.50% was not positive and 12.34% did not know (see Figure 7.1).

Deze toepassing lijkt mij nuttig voor NS reizigers die Nederlandse Gebarentaal (NGT) beheersen. *(This application seems useful to me for NS passengers that are proficient in Dutch Sign Language (NGT).)*

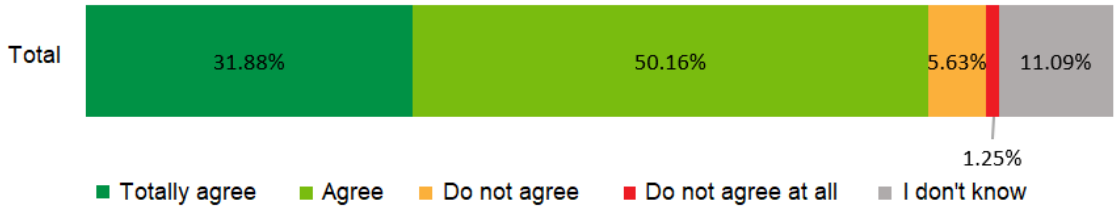


Fig. 7.2: Panel results statement 2.2

The majority of the respondents (82.04%) thought the virtual sign language avatar was useful for passengers that are proficient in Dutch Sign Language (NGT), 6.88% did not agree with this and 11.09% did not know (see Figure 7.2).

A minority of respondents (34.69%) was (totally) convinced that the virtual sign language avatar would inspire them to learn more about Dutch Sign Language (NGT), 46.09% was not convinced (at all) and 19.22% was unsure (see Figure 7.3).

De virtuele avatar zou mij inspireren om meer te leren over de Nederlandse Gebarentaal. *(The virtual avatar would inspire me to learn more about Dutch Sign Language.)*

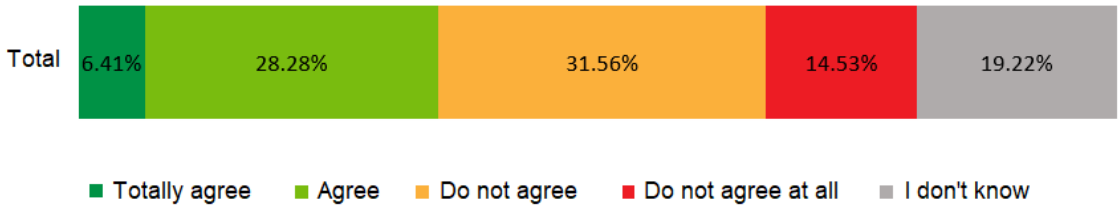


Fig. 7.3: Panel results statement 2.3

Ik vind het nuttig dat de video met gebarentaal ondertiteling bevat. *(I think it is useful that the sign language video contains subtitles.)*

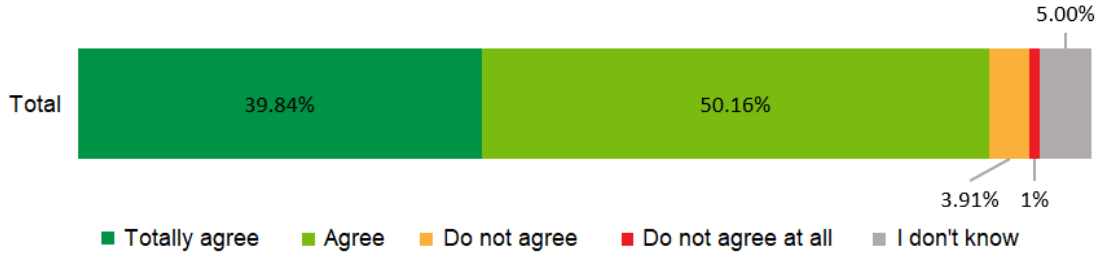


Fig. 7.4: Panel results statement 2.4

The vast majority considered it useful to deploy subtitles for the virtual sign language avatar (90%), 4.91% did not think it was useful and 5.00% was unsure (see Figure 7.4).

Proficiency with sign language was not a prerequisite for the questionnaire. As a result, less than 1% of respondents indicated a high level of competence in sign language, 7.97% was moderately proficient and 91.25% was absolutely not proficient (see Figure 7.5).

The vast majority of respondents was proficient with technical devices such as smartphones and computers (89.06%), 9.84% was moderately proficient and 1.09% was not proficient at all.

In welke mate beheerst u Nederlandse Gebarentaal (NGT)? *(To what extent are you proficient in Dutch Sign Language?)*

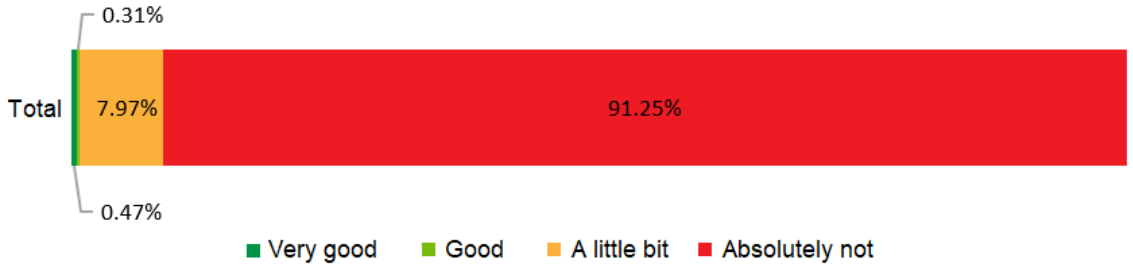


Fig. 7.5: Panel results statement 6

Bent u vaardig met technologische apparaten zoals smartphones en computers? *Are you proficient with technical devices such as smartphones and computers?*

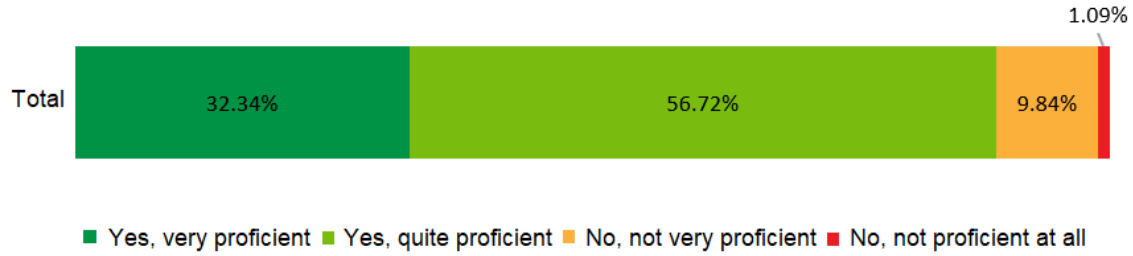


Fig. 7.6: Panel results statement 7

7.2.2 Open Questions

Seven answers were omitted from this analysis, because they did not contain any (topic related) content. In total, 29% of respondents did not observe any disadvantages of the virtual avatar. Empty responses were omitted from this percentage. Out of all disadvantages, the avatar being a possible distraction or overload of information (when displayed on the currently existing screens) was mentioned most frequently (12% of all responses that mentioned disadvantages). However, some respondents noted that voice announcements may be distracting as well for hearing people. General confusion was also seen as a possible disadvantage (4% of all responses that mentioned disadvantages). The main causes for this confusion were: information communicated by the avatar that contradicts with voice announcements or the NS application, and bad animation quality leading to confused deaf people. The vast majority of respondents mentioned accessibility for everyone as the main advantage.

10% of all responses included the assumption that all deaf people are able to read textual information. In total, 3 respondents expressed their presumption that deaf people prefer sign language over spoken Dutch.

> "Voor doven een uitkomst, want Nederlands is voor hen vaak een tweede taal." *(For deaf people a solution, because spoken Dutch is often a second language for them.)*

Application

Overall, the current information screens were not seen as a suitable option for the sign language avatar, as these screens are often small and not visible from everywhere. Especially for urgent information where the passenger has to act immediately, it is important that the deaf passenger gets appropriately notified when a new announcement starts. Flickering lights on stations or vibrating chairs in trains were suggested as attention mechanisms. The duration of the animation was a concern in rapidly changing situations too: respondents often assumed interpreting textual information would be more efficient for Deaf people. Often, voice announcements notifying passengers about unexpected delays or platform changes were assumed to be the most suitable for the avatar translations. Announcements by the conductor within the train were suggested as well.
For displaying the avatar, the NS application was mentioned as well, where (personalized) animations can be shown preceded by a notification. One respondent proposed to create a separate application for the animations.

As mentioned in Section 7.2.1, subtitles were considered useful by the vast majority (90%). Within the open answers, subtitles were often mentioned as an advantage as well. Respondents considered it an addition to the current system for hearing people, as announcements are often inaudible or not loud enough. Passengers that are partially deaf may experience these benefits as well, they noted.
According to some respondents, the subtitles will also increase awareness and understanding about the Deaf while teaching hearing people the most important signs within the train domain.

Quality and Appearance

Some respondents preferred videos of a human interpreter over an avatar. Some were aware that the displayed technology provided more ad-hoc possibilities for communication (as opposed to videos). Having a choice between multiple avatar appearance variants was suggested.

> *About the possible advantages of the virtual avatar:* "Ten eerste natuurlijk handig voor mensen die gebarentaal gebruiken, en slechthorend zijn. En ten tweede: een virtuele tolk kan je - denk ik - zodanig programmeren dat ie alles kan zeggen (dit in tegenstelling tot van te voren opgenomen filmpjes met een echt persoon.)." *(First of all, it is of course useful for people who use sign language, or are hard of hearing. Second: you can program a virtual interpreter to say everything you want - I think - (as opposed to pre-recorded videos of real people).*

> "Met een virtuele tolk komt de boodschap beter over; het uiterlijk van een "echte" persoon leidt mogelijk af." *(A virtual interpreter conveys the message in a better way; the appearance of a real person may distract.)*

In general, the avatar was considered a bit stiff, unrealistic and impersonal. The appearance and movements could be more friendly. Some respondents mentioned that the lack of facial expressions could influence the level comprehensibility. One respondent noticed that the hands of the avatar were awkward. Remarkable was that some respondents - who did not posses a high proficiency level in NGT - described the signing speed as slow.

Other

Some respondents interpreted the avatar incorrectly. For example, they assumed the avatar also includes a speaking function, or will be used to intermediate between the conductor and hard of hearing passenger.

> "Wat is een virtuele tolk? Die spreekt wel? Dat horen doven niet." *(What is a virtual interpreter? Does it speak? Deaf people cannot hear that.)*

> *About the possible disadvantages of the virtual avatar:* "Langer wachten tot de conducteur aanspreekbaar is als hij met een slechthorende via de virtuele tolk communiceert" *(Having to wait for a longer time to talk to the conductor when he is communicating with a hard of hearing person through the virtual interpreter.)*

Others used the occasion for additional suggestions with relation to accessibility. For example, the suggested flickering lights for announcing animations could also be applied to indicate that a train leaves (the final whistle). Teaching personnel on the stations sign language was considered an option as well. Moreover, other languages were discussed as well. Some wondered whether there was a universal sign language or not, or suggested providing the subtitles in other languages as well:

> "Waarom niet ook ondertiteling in Engels toevoegen. Ik zie vaak buitenlanders die niet alle informatie/wijzigingen meekrijgen." *(Why not add subtitles in English too. I often see foreigners not being aware of all information/changes.)*

Remarkable was that multiple participants expressed their concerns about passengers possibly making fun of the avatar. Additionally, some worried that the avatar would be used for PR purposes.

7.3 Discussion

As the NS Panel questionnaire was send to participants before the co-design feedback was completely processed, the demo sentence in the questionnaire did not include this feedback. On the one hand, this may have caused a higher number of critical responses due to the lower signing quality (e.g. about the original camera angle). On the other hand, participants voluntarily choose to fill in the questionnaire, which may have caused a higher number of positive responses due to self-selection bias. Randomization of the participant group could circumvent this bias in future studies. Due to the fact that we could not provide any information about if and where the avatar would be implemented (e.g. on screens or in the app), participants might have formed their own (incorrect) conclusions. For example, many expressed their worries about getting distracted by the avatar, assuming it would be displayed on screens that already encompass many visual or textual elements. Moreover, only one avatar sentence example was included in the questionnaire. This might not have been enough, as some participants assumed that all sentences communicated by the avatar would be similar to this demo sentence. Future research should clarify the implementation purposes before reaching out to the panel and include more examples of these implementations.
Apparently, many misconceptions still exist about Deaf community (e.g. about text being just as accessible for all deaf people as sign language) which might have influenced the participants' answers. For example, many hearing participants thought the animation speed was too slow for urgent situations, based on a comparison between the comprehension rates of a signed versus written announcement. Future research might address this by including a small section in the questionnaire that informs participants about common misconceptions of the Deaf community.

Chapter 8 Discussion and Conclusion

This thesis performed a case study to explore two collaborative methods for enhancing a sign language avatar that automatically translates Dutch to NGT animations in the public transport domain. The aim was to automatically create efficient and effective translations, and enhance the signing comprehensibility and naturalness. From this process, researchers and members of the Deaf community learned more about sign language technology and its potential. The current system is able to translate a subset of railway sentences relatively well, but certainly does not (yet) cover a full-fledged announcement system. Nonetheless, the research is a first step towards more comprehensive communication in the railway domain. The entire team (van Gemert, Roelofsen, Esselink, de Meulder, Sijm and Cokart) is planning to submit a paper of this study to the SLTAT (Sign Language Translation and Avatar Technology) 2022 conference.

8.1 RQ1: Consortium and Efficient and Effective Translations

An interdisciplinary consortium of experts is essential for efficient and effective development of a sign language avatar: Deaf researchers and experts have the linguistic knowledge and lived experiences, which ensures incorporating their demands and values. Nonetheless, Deaf researchers were not included in the software choices for the initial design (e.g. JASigning). This could have influenced their attitude towards the avatar. For example, the appearance of the avatar was considered 'staccato' and unfriendly by both the hearing and deaf community. Adding smiles at several locations in the sentence was not the solution for this and could be perceived as creepy or contradictory to the sentence content. In future work, deaf researchers should be included at the very start of the research.
In addition to the unfriendly look, the JAsigning engine is limited with regard to other aspects of avatar appearance: the excessively long fingers cannot be adjusted and control over mouth movements and facial expressions is restricted. Efficient and effective translations require a substantial further development of the engine to overcome these limitations. Alternatively, future research could explore motion capture techniques (and how to make these scalable), to ensure good quality animations.

A good basis of video translations that fit the particular context is fundamental for choosing appropriate versions of signs for the avatar. The publicly available annotated NGC video translation data set gave insights in possible translations and sign variants, which contributed to the efficiency of future developments within sign language technology. Nonetheless, the lack of spontaneity affects its versatility in future studies and annotator bias could be embedded. Furthermore, as sign language is used in a 3D space, the 2D translations might not be sufficient enough for proper translations. Future research should record translation videos from a different camera angle as well to create a complete foundation for translation. Besides

this, obtaining videos from various signers will give insights in signing styles and dialects, and creates the opportunity to assess these styles for a variety of detailed aspects (e.g. topic-marking, gaze direction, facial expressions).

The JASigning avatar engine makes it possible to efficiently create many variants of a given template, without expensive equipment. Nonetheless, besides the inherent intricacies of sign language, the differences between the railway sentence structure and preferred structure from the collaborative sessions complicate efficient translation and possible connections to existing audio systems (i.e. NS' text-to-speech system).

8.2 RQ2: Naturalness and Comprehensibility

Using collaborative methods resulted in mayor enhancements of the sign language avatar. The co-design sessions and focus group contributed to both linguistic (manual movements, facial expressions, mouthing, grammar, transitions between signs) as well as non-linguistic aspects of the animations (e.g. camera angle, speed). Physical sessions provided a clear overview of body language and facial expressions and gave the opportunity to point out details. Nonetheless, discussing animations in 2- or 3-hour sessions can be cognitively intense and may affect the assessment. Future research should consider more and shorter sessions.
The specific constitution of team members is of high importance for communication during sessions: having interpreter(s) and the developer present in the sessions can be advantageous for communication and direct implementation, but can prevent comfortably criticizing the system. Working with the same set of interpreters across sessions is beneficial for their familiarity with the research terms and prevents miscommunication. Moreover, making video recordings of the sessions (from various angles) will help assessing body language and facial expressions in a way that is synchronized with the utterances. However, manually annotating these recordings in order to use them for analysis would be a labor intensive process.

Natural and comprehensible translations are not only based on the accuracy and clarity of translations, but also on preferences in signing style, choice of vocabulary (BESTE REIZIGERS *(DEAR PASSENGERS)* versus HALLO *(HELLO)*), formality (WEGGAAN *(LEAVING)* versus VERTREKKEN *(DEPARTING)*) and visual aspects (e.g. HERHALING *(REPETITION)*). Moreover, changes in speed may have a mayor effect on comprehensibility (especially for cliticized signs) and may result in having to manually fine-tune animations. This complicates the development of an automatic, scalable and potentially personalized system.
Although speed, duration, amount, SAMPA and mouthing codes was experimented with, the lack of detail for mouthings was an overarching problem that affected the avatar's perceived comprehensibility. Future research should consider the adoption of an improved system based on visual features rather than phonetics. This will increase the level of distinction between mouthings, which results in more comprehensible and natural animations.

For enhancing naturalness and comprehensibility, several aspects were important. First of all, the preferred intensity of body movements and facial expressions varied. For some station names (e.g. ENSCHEDE and UTRECHT CENTRAAL) they had to be 'toned down', whereas for other signs (e.g. BIJNA *(ALMOST)*) they had to be intensified. Second, list notations easily caused confusion about the number of trains involved, affecting comprehensibility. Third, the order and rhythm of signs were important: there was a clear preference to mention certainties at the start of the sentence, and distinct indexing intensities affected the rhythm of the sentence. Fourth, pause duration was addressed within the sentences: long pauses for increasing naturalness were not preferred by participants. Nonetheless, the added brow raises and blinks within these pauses might have influence this preference (e.g. disturbing or distracting). Above all, the effect of prosody on natural and comprehensible signing should be researched more in-depth. For example, by adding multiple types of topic-marking (e.g. backward tilts, head nods, side tilts) and comparing blink or brow raise patterns.

As only 5 out of 57 railway sentences were thoroughly developed and analysed in the co-design sessions and focus group, more sentence types should be discussed in order to asses the naturalness and comprehensibility of the complete system. Moreover, future focus groups should include distinct reading levels, age groups and regions (e.g. seniors and participants from the southern part of the Netherlands). Besides this, the on-site setting of the avatar should be researched in order to adapt the avatar's comprehensibility to the situation appropriately (e.g. time-sensitiveness). Moreover, this will allow for more in-depth expertise about possible application areas and implementation details (e.g. on screens at stations or within the NS application). This knowledge will also help to obtain a complete impression about the opinion of hearing people on a sign language avatar, as these answers were influenced by the lack of detailed information. Moreover, in such evaluations, misconceptions about the Deaf community should be addressed beforehand, to increase the applicability of the answers. Future deployment of sign language avatars in public areas may raise awareness and stimulate inclusion.

In the future, various avatar genders could be considered, giving the user a choice between their preferred appearance. Moreover, various sign languages and dialects could be implemented to make translations comprehensible everywhere in the world. This also benefits hearing passengers when combined with subtitles. To conclude, sign language avatars can be useful for multiple domains: in museums to facilitate routes and to translate airport announcements, or it may help hearing parents of deaf children to learn sign language while creating awareness for the community.

Bibliography

[1] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. "User-centered design". In: *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37.4 (2004), pp. 445–456.

[2] Mohamed Amin, Hesahm Hefny, and Ammar Mohammed. "Sign Language Gloss Translation using Deep Learning Models". In: *Sign* 12.11 (2021).

[3] International Phonetic Association, International Phonetic Association Staff, et al. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, 1999.

[4] Kamel Ayadi, Yahya OM Elhadj, and Ahmed Ferchichi. "Prototype for Learning and Teaching Arabic Sign Language Using 3D Animations". In: *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*. IEEE. 2018, pp. 51–57.

[5] Anne Baker et al. *The linguistics of sign languages: An introduction*. John Benjamins Publishing Company, 2016.

[6] George I Balch and Donna M Mertens. "Focus group design and group dynamics: Lessons from deaf and hard of hearing participants". In: *American Journal of Evaluation* 20.2 (1999), pp. 265–277.

[7] Jetske Beks. "Vertaling van Nederlands naar Nederlandse Gebarentaal". Bachelor's thesis, Universiteit van Amsterdam, 2018.

[8] Edwin Blake, William Tucker, and Meryl Glaser. "Towards communication and information access for Deaf people". In: *South African Computer Journal* 54.si-2 (2014), pp. 10–19.

[9] Danielle Bragg et al. "Sign language recognition, generation, and translation: An interdisciplinary perspective". In: *The 21st international ACM SIGACCESS conference on computers and accessibility*. 2019, pp. 16–31.

[10] Danielle Bragg et al. "The FATE Landscape of Sign Language AI Datasets: An Interdisciplinary Perspective". In: *ACM Transactions on Accessible Computing (TACCESS)* 14.2 (2021), pp. 1–45.

[11] Ingrid Burkett. "An introduction to co-design". In: *Sydney: Knode* (2012).

[12] Dominic Chan et al. "EUROM-A spoken language resource for the EU". In: *Proceedings of the 4th European Conference on Speech Communication and Speech Tecnology, Eurospeech'95*. 1995, pp. 867–880.

[13] P Chininthorn. "Community-based co-design for accessible health information for Deaf people in a context with societal complexity". In: (2021).

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

[15] Stephen Cox et al. "Tessa, a system to aid communication with deaf people". In: *Proceedings of the fifth international ACM conference on Assistive technologies*. 2002, pp. 205–212.

[16] O Crasborn et al. "Annotation Conventions for the Corpus NGT, version 4". In: *Centre for Language Studies & Department of Linguistics, Radboud University Nijmegen* (2020).

[17] O. Crasborn et al. "Corpus NGT". In: (2020, 4th edition).

[18] Onno Crasborn et al. "Topic agreement in NGT (Sign Language of the Netherlands)". In: (2009).

[19] Bastien David and Pierrette Bouillon. "Prototype of Automatic Translation to the Sign Language of French-speaking Belgium. Evaluation by the Deaf Community". In: *Modelling, Measurement and Control C* 79.4 (2018), pp. 162–167.

[20] Maartje De Meulder. "Is "good enough" good enough? Ethical and responsible development of sign language technologies". In: *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*. 2021, pp. 12–22.

[21] Sarah Ebling and John Glauert. "Building a Swiss German Sign Language avatar with JASigning and evaluating it among the Deaf community". In: *Universal Access in the Information Society* 15.4 (2016), pp. 577–587.

[22] Ralph Elliott et al. "An overview of the SiGML notation and SiGMLSigning software system". In: *Fourth International Conference on Language Resources and Evaluation, LREC*. 2004, pp. 98–104.

[23] Ralph Elliott et al. "Linguistic modelling and language-processing technologies for Avatar-based sign language presentation". In: *Universal Access in the Information Society* 6.4 (2008), pp. 375–391.

[24] Ralph Elliott et al. "The development of language processing support for the ViSiCAST project". In: *Proceedings of the fourth international ACM conference on Assistive technologies*. 2000, pp. 101–108.

[25] Ralph Elliott et al. "Towards the integrationn of synthetic sl annimation with avatars into corpus annotation tools". In: *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Valletta, Malta*. Citeseer. 2010, p. 29.

[26] Lyke D Esselink. "Lexical resources for sign language synthesis". Bachelor's thesis, Universiteit van Amsterdam, 2020.

[27] Johannes Fellinger, Daniel Holzinger, and Robert Pollard. "Mental health of deaf people". In: *The Lancet* 379.9820 (2012), pp. 1037–1044.

[28] Nederlands Gebarencentrum. *Erkenning Nederlandse Gebarentaal nu officiëel*. https://www.gebarencentrum.nl/Erkenning%20NGT. 2021.

[29] Tom Geller. "Overcoming the uncanny valley". In: *IEEE computer graphics and applications* 28.4 (2008), pp. 11–17.

[30] Sylvie Gibet et al. "The Signcom System for Data-Driven Animation of Interactive Virtual Signers: Methodology and Evaluation". In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 1.1 (2011), pp. 1–23.

[31] Ann Grafstein. “HandSpeak: A Sign Language Dictionary Online”. In: *Reference Reviews* (2002).

[32] François Grosjean. “A study of timing in a manual and a spoken language: American Sign Language and English”. In: *Journal of Psycholinguistic Research* 8.4 (1979), pp. 379–405.

[33] François Grosjean and Harlan Lane. “Pauses and syntax in American sign language”. In: *Cognition* 5.2 (1977), pp. 101–117.

[34] Wyatte C Hall. “What you don’t know can hurt you: The risk of language deprivation by impairing sign language development in deaf children”. In: *Maternal and child health journal* 21.5 (2017), pp. 961–965.

[35] Wyatte C Hall, Leonard L Levin, and Melissa L Anderson. “Language deprivation syndrome: A possible neurodevelopmental disorder with sociocultural origins”. In: *Social psychiatry and psychiatric epidemiology* 52.6 (2017), pp. 761–776.

[36] *Hand Talk American Sign Language Application.*

[37] Raychelle Harris, Heidi M Holmes, and Donna M Mertens. “Research ethics in sign language communities”. In: *Sign Language Studies* 9.2 (2009), pp. 104–131.

[38] *https://www.lifeprint.com/.*

[39] Matt Huenerfauth. “A linguistically motivated model for speed and pausing in animations of american sign language”. In: *ACM Transactions on Accessible Computing (TACCESS)* 2.2 (2009), pp. 1–31.

[40] Matt Huenerfauth. “Evaluation of a psycholinguistically motivated timing model for animations of American Sign Language”. In: *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility.* 2008, pp. 129–136.

[41] Matt Huenerfauth and Hernisa Kacorri. “Best practices for conducting evaluations of sign language animation”. In: *Journal on Technology and Persons with Disabilities* 3 (2015), pp. 20–32.

[42] Matt Huenerfauth et al. “Evaluating American Sign Language generation through the participation of native ASL signers”. In: *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility.* 2007, pp. 211–218.

[43] Tom Humphries et al. “Avoiding linguistic neglect of deaf children”. In: *Social Service Review* 90.4 (2016), pp. 589–619.

[44] Leonard P Kelly and Dragana Barac-Cikoja. “The comprehension of skilled deaf readers”. In: *Children’s comprehension problems in oral and written language: A cognitive perspective* (2007), pp. 244–280.

[45] JR Kennaway, John RW Glauert, and Inge Zwitserlood. “Providing signed content on the Internet by synthesized animation”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 14.3 (2007), 15–es.

[46] Sedeeq Al-khazraji et al. "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*. 2018, pp. 259–270.

[47] Vadim Kimmelman. "Doubling in RSL and NGT: a Pragmatic Account". In: *Information structure: empirical perspectives on theory* 17 (2013), pp. 99–118.

[48] Vadim Kimmelman. "Information structure in Russian sign language and sign language of the Netherlands:(University of Amsterdam, 2014)". In: *Sign Language & Linguistics* 18.1 (2015), pp. 142–150.

[49] Michael Kipp, Alexis Heloir, and Quan Nguyen. "Sign language avatars: Animation and comprehensibility". In: *International Workshop on Intelligent Virtual Agents*. Springer. 2011, pp. 113–126.

[50] Michael Kipp et al. "Assessing the deaf user perspective on sign language avatars". In: *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. 2011, pp. 107–114.

[51] Ulrika Klomp. *A descriptive grammar of Sign Language of the Netherlands*. LOT, 2021.

[52] Harry Knoors and Marc Marschark. "Language planning for the 21st century: Revisiting bilingual language policy for deaf children". In: *The Journal of Deaf Studies and Deaf Education* 17.3 (2012), pp. 291–305.

[53] Verena Krausneker and Sandra Schügerl. "Best Practices Protocol on the Use of Sign Language Avatars". In: (2012).

[54] D Manoj Kumar et al. "EasyTalk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence". In: *2020 2nd International Conference on Advancements in Computing (ICAC)*. Vol. 1. IEEE. 2020, pp. 506–511.

[55] *Laaggeletterdheid Nederland*. 2018.

[56] Imen Lagha and Achraf Othman. "Understanding Prosodic Pauses in Sign Language from Motion-Capture and Video-data". In: *2019 7th International conference on ICT & Accessibility (ICTA)*. IEEE. 2019, pp. 1–4.

[57] J Martin. *A linguistic comparison. Two notation systems for signed languages. Stokoe Notation and Sutton sign writing. Sign writing web site*. 2000.

[58] Rachel I Mayberry, Elizabeth Lock, and Hena Kazmi. "Linguistic ability and early language exposure". In: *Nature* 417.6884 (2002), pp. 38–38.

[59] Amy L McCarty. "Notation systems for reading and writing sign language". In: *The Analysis of verbal behavior* 20.1 (2004), pp. 129–134.

[60] Shani E Mende-Gillings. "The signing space for the synthesis of directional verbs in NGT". Bachelor's thesis, Universiteit van Amsterdam, 2020.

[61] Ross E Mitchell and Michaela Karchmer. "Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States". In: *Sign language studies* 4.2 (2004), pp. 138–163.

[62] Robyn Moncrief. "Generalizing a model for animating adverbs of manner in American Sign Language". In: *Machine Translation* 35.3 (2021), pp. 345–362.

[63] Masahiro Mori, Karl F MacDorman, and Norri Kageki. "The uncanny valley [from the field]". In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100.

[64] Joseph J Murray, Wyatte C Hall, and Kristin Snoddon. "Education and health of children with hearing loss: the necessity of signed languages". In: *Bulletin of the World Health Organization* 97.10 (2019), p. 711.

[65] *Nederlands Gebarencentrum*.

[66] *Nederlandse Spoorwegen*.

[67] *NS Karin*.

[68] *NS Panel*.

[69] Eleni Orfanidou, Bencie Woll, and Gary Morgan. *Research methods in sign language studies: A practical guide*. John Wiley & Sons, 2014.

[70] Ellen Ormel and Onno Crasborn. "Prosodic correlates of sentences in signed languages: A literature review and suggestions for new types of studies". In: *Sign Language Studies* 12.2 (2012), pp. 279–315.

[71] Sandra Pauser. "Prototypentest SiMAX im Rahmen des Innovationsschecks". In: *Retrieved from Equalizent: https://www. equalizent. com/fileadmin/ user_upload/News/2019_0 4_Avatar_Projektbericht. pdf* (2019).

[72] S Prillwitz et al. "et-al. 1989."Hamburg Notation System for Sign Languages–An introductory Guide"". In: *International Studies on Sign Language and the Communication of the Deaf* 5 ().

[73] Nijmegen: Max Planck Institute for Psycholinguistics The Language Archive. *ELAN (Version 6.2)*. 2021. URL: `https://archive.mpi.nl/tla/elan`.

[74] Suvi Pylvänen, Antti Raike, Päivi Rainò, et al. "Co-design for accessibility in academia for Deaf students". In: *Co-Create 2013* (2013).

[75] Lorna C Quandt et al. "Attitudes toward signing human avatars vary depending on hearing status, age of signed language exposure, and avatar type". In: (2022).

[76] Emmanuel Rayner et al. "An OpenWeb platform for rule-based speech-to-sign translation". In: *54th annual meeting of the Association for Computational Linguistics (ACL)*. 2016.

[77] Floris Roelofsen. *Solution Space*. https://osf.io/maqz8/. 2021. (Visited on 09/09/2021).

[78] Floris Roelofsen et al. "Online Evaluation of Text-to-sign Translation by Deaf End Users: Some Methodological Recommendations (short paper)". In: *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*. 2021, pp. 82–87.

[79] Floris Roelofsen et al. "Sign language translation in a healthcare setting". In: *Translation and Interpreting Technology* (2021).

[80] Gertrude Maria Schermer and Corline Koolhof. *Van Dale Basiswoordenboek Nederlandse Gebarentaal*. Van Dale, 2009.

[81] Trude Schermer. "Sign Language Planning in the Netherlands between 1980 and 2010". In: *Sign Language Studies* 12.4 (2012), pp. 467–493. ISSN: 03021475, 15336263. URL: `http://www.jstor.org/stable/26190876`.

[82] *Signbank*.

[83] Soraia Silva Prietch et al. "The Human and the Context Components in the Design of Automatic Sign Language Recognition Systems". In: *Iberoamerican Workshop on Human-Computer Interaction*. Springer. 2019, pp. 369–380.

[84] Jenny L Singleton, Gabrielle Jones, and Shilpa Hanumantha. "Toward ethical research practice with deaf participants". In: *Journal of Empirical Research on Human Research Ethics* 9.3 (2014), pp. 59–66.

[85] Robert G Smith and Brian Nolan. "Emotional facial expressions in synthesised sign language avatars: a manual evaluation". In: *Universal Access in the Information Society* 15.4 (2016), pp. 567–576.

[86] William C Stokoe, Dorothy C Casterline, and Carl G Croneberg. "A dictionary of American Sign Language on linguistic principles". In: *Washington, DC: Gallaudet College* (1965).

[87] *The handshape fonts are created by CSLDS, CUHK*.

[88] LN Wauters, C Tijsseling, and H van Gelder. "Leesprofielen voor dove en slechthorende volwassenen". In: *Van Horen Zeggen* 58 (2017), pp. 10–19.

[89] Ronnie Wilbur. "Eyeblinks & ASL phrase structure". In: *Sign Language Studies* 84.1 (1994), pp. 221–240.

[90] Kayo Yin et al. "Including signed languages in natural language processing". In: *arXiv preprint arXiv:2105.05222* (2021).

[91] Alys Young and Ros Hunt. "Research with d/Deaf people". In: (2011).

[92] Theodore Zamenopoulos and Katerina Alexiou. *Co-design as collaborative research*. Bristol University/AHRC Connected Communities Programme, 2018.

[93] Inge Zwitserlood. "Sign language lexicography in the early 21st century and a recently published dictionary of Sign Language of the Netherlands". In: *International Journal of Lexicography* 23.4 (2010), pp. 443–476.

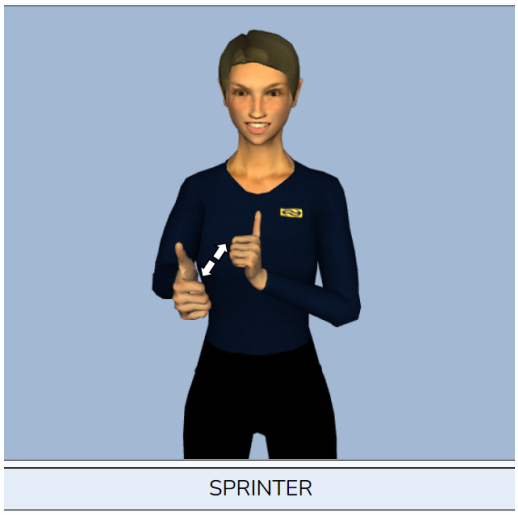
Chapter A Initial Design Phrases

A.1 Variables

A.1.1 treinType *(trainType)*

(a) INTERCITY *(INTERCITY)* - Part 1

(b) INTERCITY *(INTERCITY)* - Part 2



(c) SPRINTER *(SPRINTER)*

Fig. A.1: treinType *(trainType)* - V1

A.1.2 tussenStation1, ..., tussenStation4 *(interStation1, ..., interStation4)*

(a) ZWOLLE

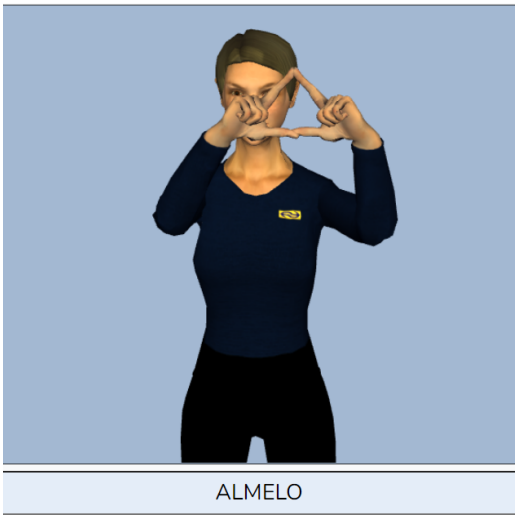
(b) ARNHEM

(c) DEVENTER

(d) BREDA

Fig. A.2: tussenStation1, ..., tussenStation4 *(interStation1, ..., interStation4)* - V1

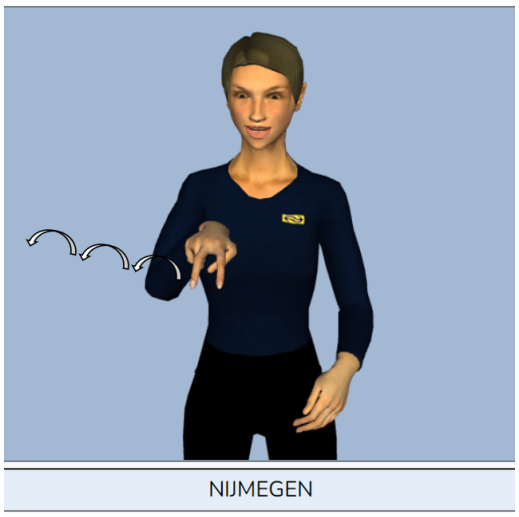
A.1.3 eindStation *(endStation)*



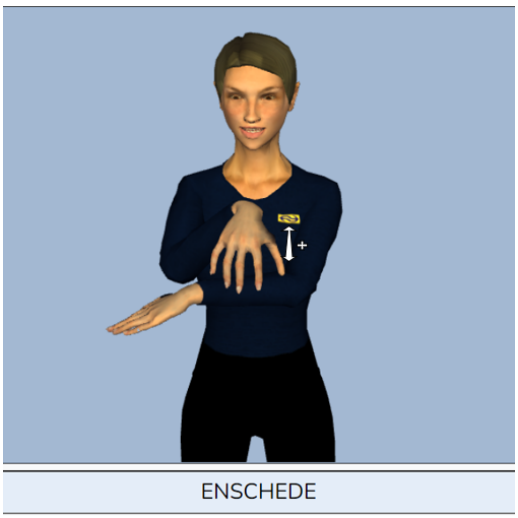
(a) ALMELO - Part 1

(b) ALMELO - Part 2

(c) ALMELO - Part 3



(d) NIJMEGEN

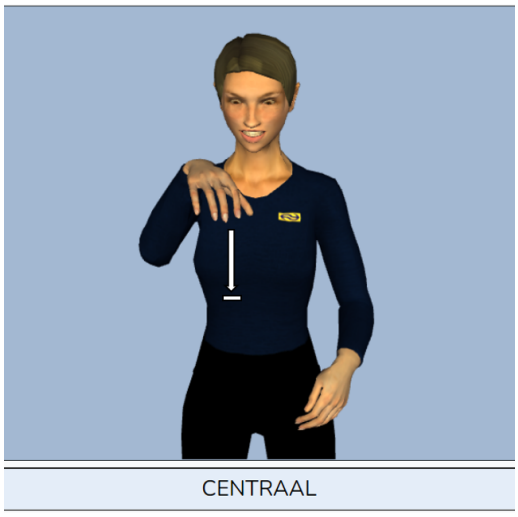


(e) ENSCHEDE

(f) MAASTRICHT

(g) SCHIEDAM

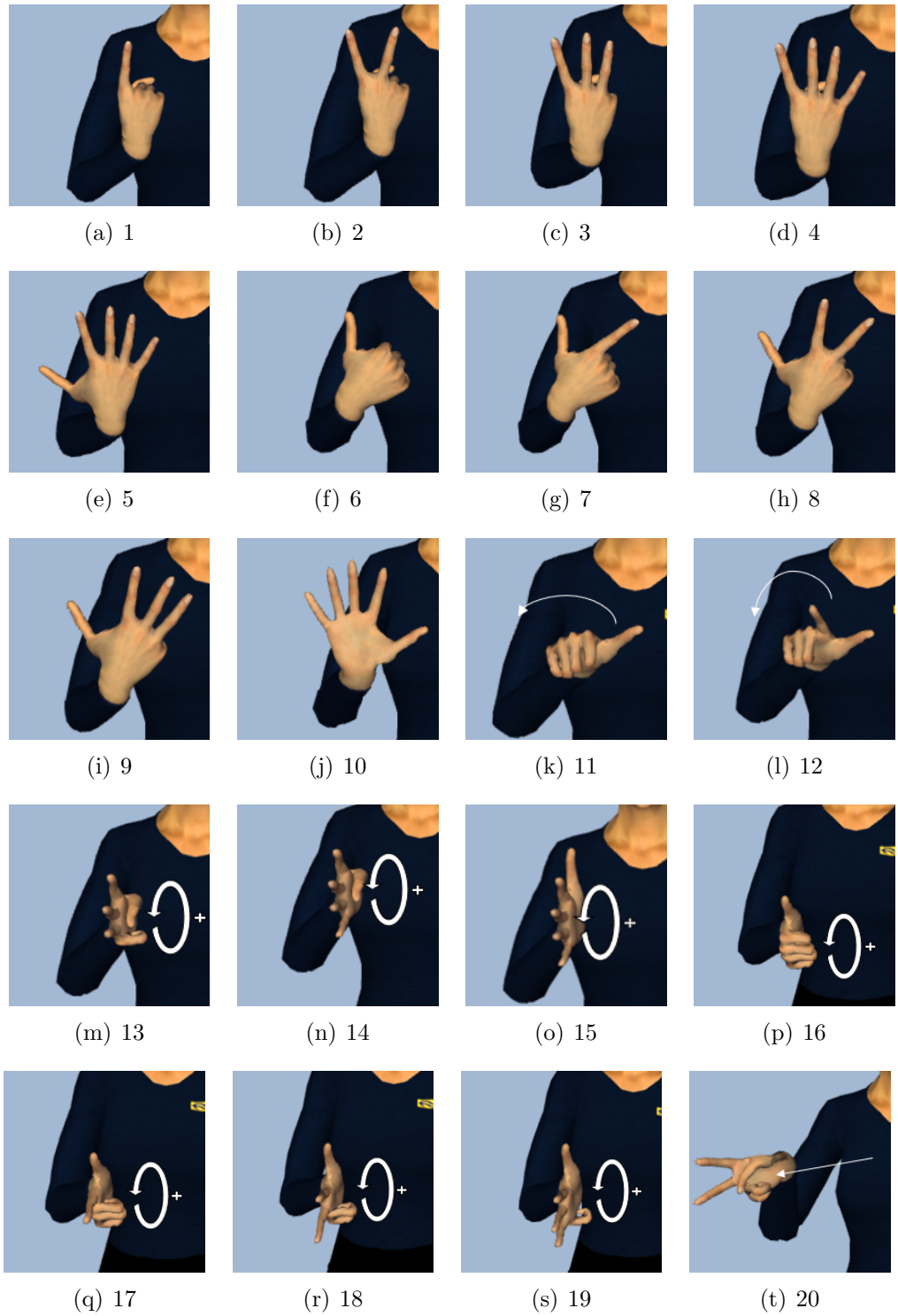
(h) UTRECHT



(i) CENTRAAL

Fig. A.3: eindStation *(endStation)* - V1

A.1.4 vertrekTijd *(departTime)*



(a) 1 (b) 2 (c) 3 (d) 4

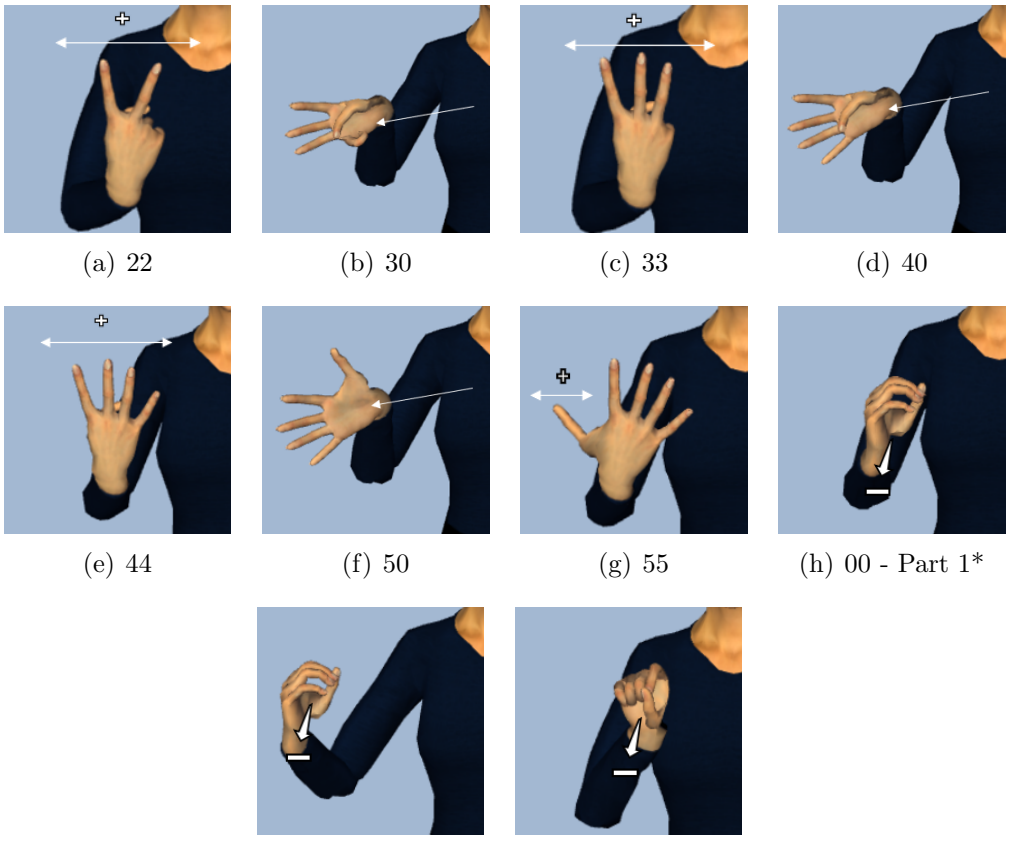
(e) 5 (f) 6 (g) 7 (h) 8

(i) 9 (j) 10 (k) 11 (l) 12

(m) 13 (n) 14 (o) 15 (p) 16

(q) 17 (r) 18 (s) 19 (t) 20

Fig. A.4: Elements used in time notations - Part 1 - V1



(a) 22 (b) 30 (c) 33 (d) 40

(e) 44 (f) 50 (g) 55 (h) 00 - Part 1*

(i) 00 - Part 2 (j) PUNT *(POINT)*

Fig. A.5:
Elements used in time notations - Part 2 - V1

* Also used individually for '0'

A.1.5 wachtTijd *(waitTime)*

(a) OVER *(OVER)*

(b) EEN PAAR *(A FEW)*

(c) MINUTEN *(MINUTES)*

Fig. A.6: wachtTijd: Enkele minuten *(A few minutes)* - V1

(a) OVER *(OVER)*

(b) ONGEVEER *(APPROXIMATELY)*

(c) 5

(d) MINUTEN *(MINUTES)*

Fig. A.7: wachtTijd: Ongeveer 5 minuten *(Approximately 5 minutes)* - V1

(a) OVER *(OVER)*

(b) ONGEVEER *(APPROXIMATELY)*

(c) 5

(d) 30

(e) MINUTEN *(MINUTES)*

Fig. A.8: wachtTijd: Ongeveer 35 minuten *(Approximately 35 minutes)* - V1

(a) OVER *(OVER)*

(b) ONGEVEER *(APPROXIMATELY)*

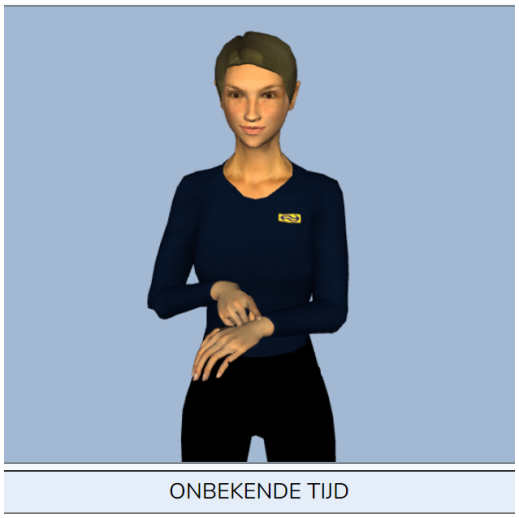
(c) 1 UUR *(ONE HOUR)*

(d) 30

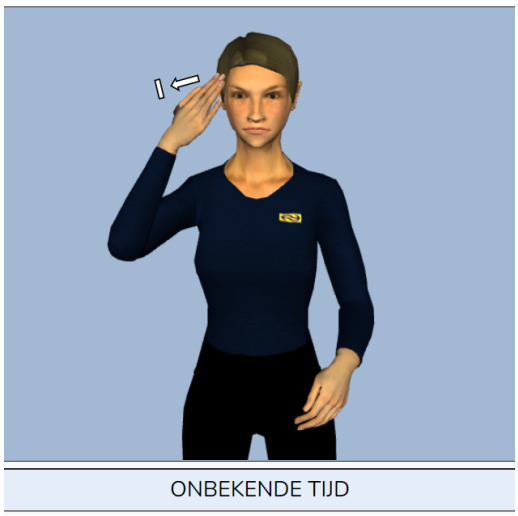
(e) MINUTEN *(MINUTES)*

Fig. A.9: wachtTijd: Ongeveer anderhalf uur *(Approximately one and a half hour)* - V1

(a) HOE LAAT *(WHAT TIME)* - Part 1



(b) HOE LAAT *(WHAT TIME)* - Part 2



(c) WETEN NIET *(DO NOT KNOW)*

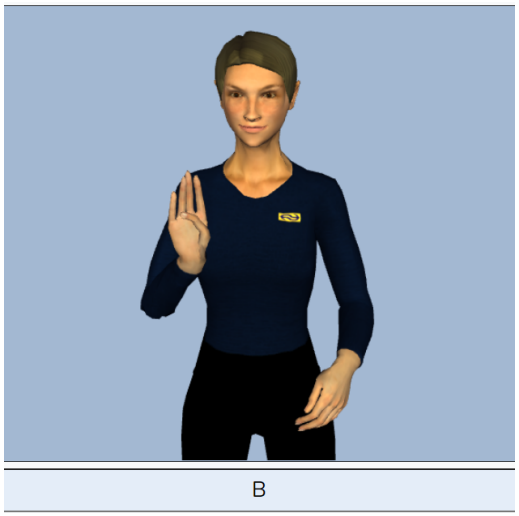
Fig. A.10: wachtTijd: Een nog onbekende tijd *(A yet unknown time frame)* - V1

A.1.6 spoorNr *(platformNr)*

(a) 5

(b) 1

(c) A



(d) B

(e) 11

(f) 15

Fig. A.11: spoorNr *(platformNr)* - V1

A.2 Sentences

A.2.1 Beste reizigers, de treinType naar eindStation van vertrekTijd rijdt niet. *(Dear passengers, the trainType to endStation of departTime is not departing.)*

(a) BESTE *(DEAR)*

(b) REIZIGERS *(PASSENGERS)* - Part 1

(c) REIZIGERS *(PASSENGERS)* - Part 2

(d) INDEX3a

trainType

(e) treinType

(f) GAAN-NAAR *(GOING-TO)*

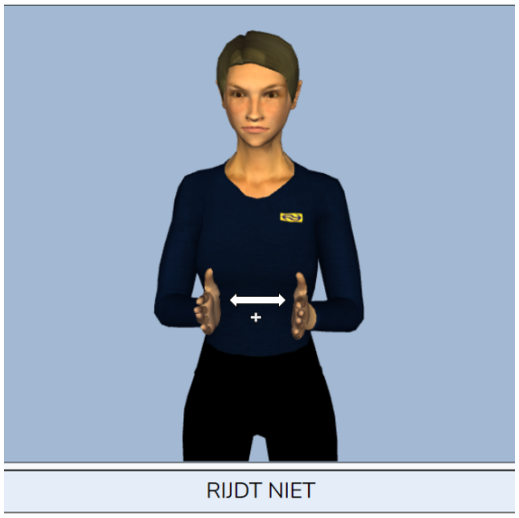
endStation

(g) eindStation

(h) TIJD *(TIME)*

departTime

(i) vertrekTijd



(j) RIJDEN-NIET *(NOT DEPARTING)*

(k) RIJDEN-NIET *(NOT DEPARTING)* (l) Neutral End Pose

Fig. A.12: Sentence 1: Beste reizigers, de treinType naar eindStation van vertrekTijd rijdt niet. *(Dear passengers, the trainType to endStation of departTime is not departing.)* - V1

A.2.2 De treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation van vertrekTijd vertrekt over wachtTijd van spoor spoorNr. *(The trainType to interStation1, interStation2, interStation3, interStation4 and endStation of departTime departs in waitTime from platform platformNr.)*

(a) BESTE *(DEAR)*

(b) REIZIGERS *(PASSENGERS)* - Part 1

(c) REIZIGERS *(PASSENGERS)* - Part 2

(d) INDEX3a

trainType

(e) treinType

(f) GAAN-NAAR *(GOING-TO)*

interStation1-4

(g) tussenStation1, ..., tussenStation4

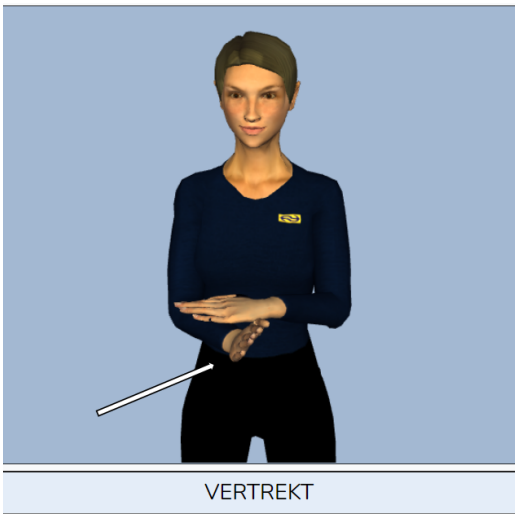
endStation

(h) eindStation

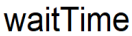
(i) TIJD *(TIME)*

departTime

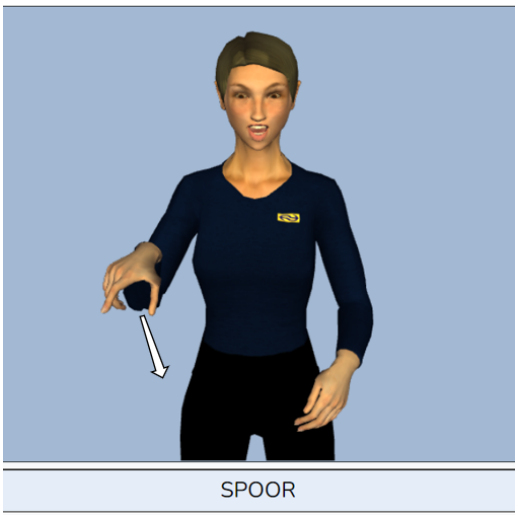
(j) vertrekTijd



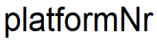
(k) VERTREKKEN *(LEAVING)*



(l) wachtTijd



(m) SPOOR *(PLATFORM)*



(n) spoorNr

(o) Neutral End Pose

Fig. A.13: Sentence 2: De treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation van vertrekTijd vertrekt over wachtTijd van spoor spoorNr. *(The trainType to interStation1, interStation2, interStation3, interStation4 and endStation of departTime departs in waitTime from platform platformNr.)* - V1

A.2.3 Herhaling. De treinType naar eindStation van vertrekTijd rijdt via een andere route. *(Repetition. The trainType to endStation from departTime runs through another route.)*

(a) HERHALING *(REPETITION)* - V1

(b) INDEX3a

trainType

(c) treinType

(d) GAAN-NAAR *(GOING-TO)*

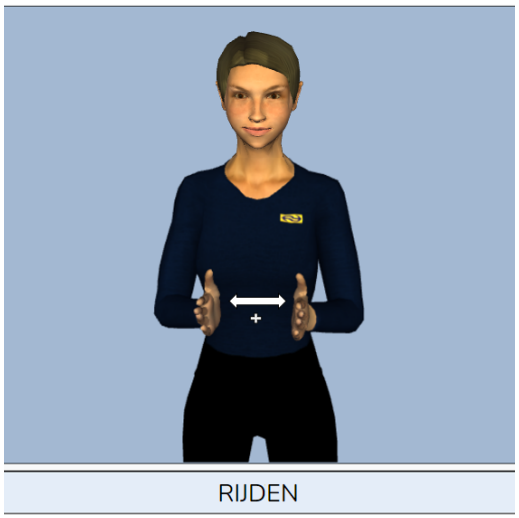
endStation

(e) eindStation

(f) TIJD *(TIME)*

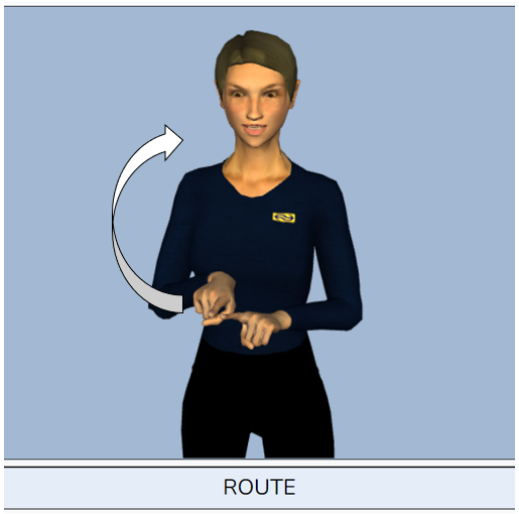
departTime

(g) vertrekTijd



(h) RIJDEN *(DEPARTING)*

(i) ANDERE *(OTHER)*

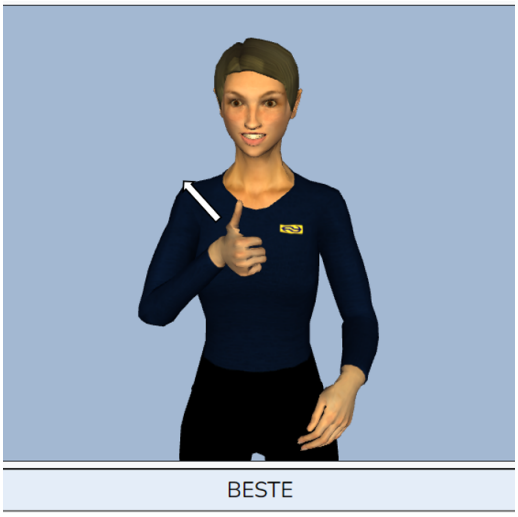


(j) ROUTE *(ROUTE)*

(k) Neutral End Pose

Fig. A.14: Sentence 3: Herhaling. De treinType naar eindStation van vertrekTijd rijdt via een andere route. *(Repetition. The trainType to endStation from departTime runs through another route.)* - V1

A.2.4 Beste reizigers, de treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation stopt niet op tussengelegen stations. *(Dear passengers, the trainType to interStation1, interStation2, interStation3, interStation4 and endStation does not stop at intermediate stations.)*



(a) BESTE *(DEAR)*

(b) REIZIGERS *(PASSENGERS)* - Part 1

(c) REIZIGERS *(PASSENGERS)* - Part 2

(d) INDEX3a

trainType

(e) treinType

(f) GAAN-NAAR *(GOING-TO)*

(g) TELHAND-1

interStation1

(h) tussenStation1

(i) TELHAND-2

interStation2

(j) tussenStation2

(k) TELHAND-3

interStation3

(l) tussenStation3

(m) TELHAND-4

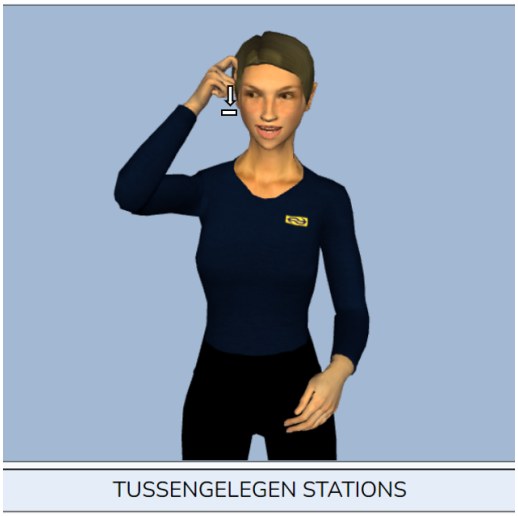
interStation4

(n) tussenStation4

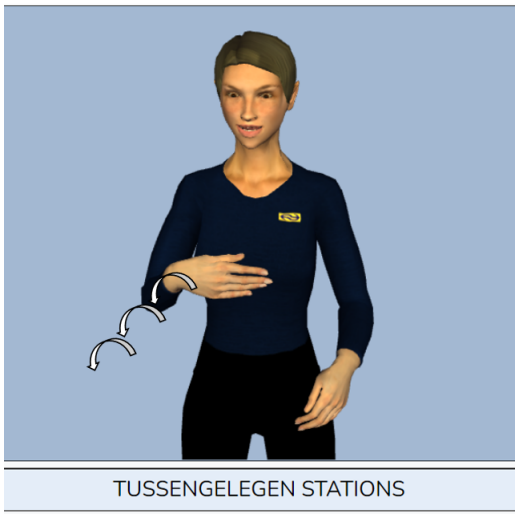
(o) TELHAND-5

endStation

(p) eindStation



(q) STATIONS *(STATIONS)*



(r) TUSSENGELEGEN *(INTERMEDIATE)*

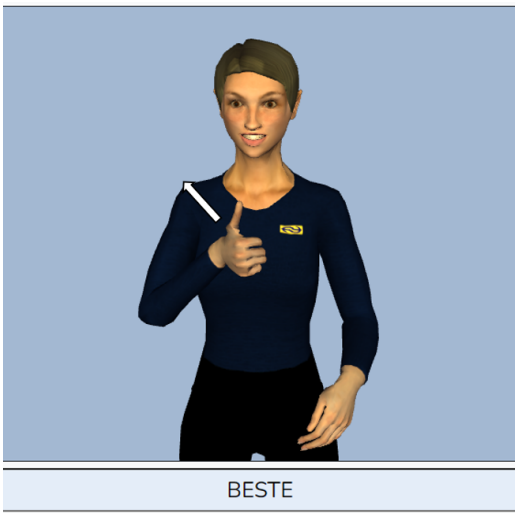
(s) STOPPEN-NIET *(STOPPING-NOT)*

(t) NIET *(NOT)*

(u) Neutral End Pose

Fig. A.15: Sentence 4: Beste reizigers, de treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation stopt niet op tussengelegen stations. *(Dear passengers, the trainType to interStation1, interStation2, interStation3, interStation4 and endStation does not stop at intermediate stations.)* - V1

A.2.5 Beste reizigers, de treinType naar eindStation vertrekt van spoor spoorNr. *(Dear passengers, the trainType to endStation departs from platform platformNr.)*



(a) BESTE *(DEAR)*

(b) REIZIGERS *(PASSENGERS)* - Part 1

(c) REIZIGERS *(PASSENGERS)* - Part 2

(d) INDEX3a

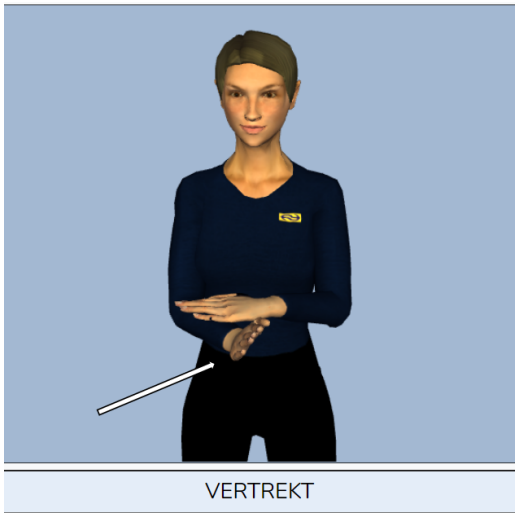
trainType

(e) treinType

(f) GAAN-NAAR *(GOING-TO)*

endStation

(g) eindStation



(h) VERTREKKEN *(LEAVING)*

(i) SPOOR *(PLATFORM)*

platformNr

(j) spoorNr

(k) Neutral End Pose

Fig. A.16: Sentence 5: Beste reizigers, de treinType naar eindStation vertrekt van spoor spoorNr. *(Dear passengers, the trainType to endStation departs from platform platformNr.)*

Chapter B Additional Annotation Codes

Body movement

RL	rotated left
RR	rotated right
TL	tilted left
TR	tilted right
TF	tilted forwards
TB	tilted backwards
SI	sigh
HE	heave
ST	straight
RD	round

Eye brows

RB	both eyebrows raised
RR	right eyebrow raised
RL	left eyebrow raised
FU	eye brows furrowed

Head movement

NO	nodding up and down
SH	shaking left and right
SR	turned right
SL	turned left
TR	tilted right
TL	tilted left
NF	tilted forward
NB	tilted back
PF	pushed forward
PB	pushed backward
LI	head movement linked to eye gaze

Eye gaze

AD	towards addressee
FR	far
HD	towards the signers own hands
HI	towards the signers own dominant hand
HC	towards the signers own non-dominant hand
UP	up
DN	down
LE	left
RI	right
NO	no target unfocussed
RO	rolling eyes
LU	left up
LD	left down
RU	right up
RD	right down

Eye lids

WB	wide open eyelids
WR	wide open right eyelid
WL	wide open left eyelid
SR	narrowed almost closed eyelids slits
SL	narrowed almost closed right eyelid
CB	narrowed almost closed left eyelid
CR	closed eyelids
CL	closed right eyelid
TB	closed left eyelid
TR	tightly shut eyelids
TL	tightly shut left eyelid
BB	eye blink at the very end of a sign

Mouth gesture

D05	clattering teeth
D06	clattering teeth with raised upper lip
D07	one bite resulting in closed teeth
D08	one bite lips stretched teeth visible
D09	teeth on lower lip open almost close tongue behind upper teeth
J01	lower jaw moves sideways left and right
J02	lower jaw chews mouth remains closed
J03	mouth open jaw forward teeth visible
J04	mouth open jaw 'gagaga' at pharynx
L04	pursed lips
L05	'o oa' open o
L06	'ooo' closed o
L24	'mmm' while holding breath
L26	one side of upper lip raised
L27	mouth slightly open tongue to upper close lips hidden
L28	tongue on upper lip close mouth lips hidden
L29	lips closed hidden mouth corners curved down
L30	lips pursed curved down
L31	lips closed corners of mouth curved down
L32	mouth slightly open blow lips vibrate initially
L33	mouth open close sh with teeth showing
L34	lips closed stretched strongly
L35	blow out air through slightly open lips
C01	puffed cheeks
C02	cheeks and lip area puffed
C03	gradually puffing cheeks
C04	one cheek puffed
C05	one cheek puffed while briefly blowing out air
C06	one cheek puffed briefly blowing air cheek pushed
C07	cheeks sucked in
C08	cheeks sucked in sucking in air
C09	tongue pushed visibly into cheek
C10	tongue repeatedly pushes into cheek
C11	one cheek puffed blow out briefly at corner several times
C12	lips closed tongue pushed behind lower lip
C13	cheeks slightly in jaw down blow closed lips several times
T02	tip of tongue slightly protruding
T04	tongue sticks out briefly
T06	tongue sticking out repeatedly
T12	tip of tongue touches one corner of the mouth
T13	tongue tip between lower lip lower teeth middle tongue showing
T14	tip of tongue is protruded and moving side-wards
T15	oval circling movement of tongue in open mouth
T16	lips pursed with tip of tongue protruding
T17	mouth open tongue protrudes briefly

Chapter C NS Panel Invitation Mail

Afzender: NS Panel
Onderwerp: NS is benieuwd naar uw mening over gebarentaal en taaltechnologie!

NS Panel

[{panel_aanspreekvorm}] [{panel_aanhef}] [{panel_tussenvoegsel}] [{panel_achternaam}],

Graag nodigen wij u - als lid van het NS Panel - uit voor een onderzoek over gebarentaal en taaltechnologie.

Nederlandse Gebarentaal (NGT) is sinds een jaar erkend als officiële taal. Bij NS zijn we aan het kijken hoe we hier invulling aan kunnen geven. We willen u daarom graag een paar vragen stellen, ook als u niet bekend met gebarentaal.

Dit onderzoek wordt in opdracht van NS uitgevoerd door onafhankelijk marktonderzoeksbureau MWM2. We werken hiervoor samen met SignLab Amsterdam. SignLab Amsterdam zal ook een deel van de resultaten analyseren.

Vanzelfsprekend is uw deelname volstrekt anoniem. Het invullen van de vragenlijst duurt ongeveer 4 minuten. U kunt tot en met 7 februari a.s. uw mening geven.

Klik op de onderstaande knop om het onderzoek te starten:

Start onderzoek >

Heeft u technische vragen over het onderzoek? Dan kunt u contact opnemen met Britt van Gemert via britt.vangemert@ns.nl.
Bij andere vragen kunt u via ons contactformulier (link) contact met ons opnemen.

Alvast hartelijk dank voor uw deelname!

Britt van Gemert
Data, Innovatie & Analyse
Nederlandse Spoorwegen

Valerie Severens
Commercie & Ontwikkeling
Nederlandse Spoorwegen

Deze e-mail is verzonden door marktonderzoeksbureau MWM2. MWM2 beheert in opdracht van NS het NS Panel.
MWM2 is aangesloten bij de MOA. Zij houden zich aan de internationale ESOMAR/ICC code voor marktonderzoek en de gedragscode inzake de bescherming van privacy.

Wanneer u problemen ondervindt met het openen van de vragenlijst, dan heeft u wellicht baat bij één of meerdere van onderstaande suggesties:
1. Het kopiëren van deze link uit uw mail naar de adresbalk van uw internet browser
https://net09.mwm2.nl/go.aspx?uu=##user_uniek##&vp=##ond_uniek##&alin=ja
2. Eventuele pop-up killers uitzetten
3. Het uitvoeren van Javascript in uw browser toestaan
4. Wanneer bovenstaande tips niet helpen kunt u altijd mailen met info@nspanel.nl

Kent u anderen die ook lid willen worden van het NS Panel? Wijs hen dan op: http://www.nspanel.nl/

Chapter D NS Panel Questionnaire

Klantonderzoek NS

Start of Block: Voorbeeld

Welkom! NS doet in samenwerking met SignLab Amsterdam onderzoek naar gebarentaal en taaltechnologie. Omdat Nederlandse Gebarentaal (NGT) pas sinds een jaar erkend is als officiële taal, is onderzoek hiernaar erg belangrijk. We willen u graag een paar vragen stellen over dit onderwerp.

In de onderstaande video is een virtuele gebarentolk te zien die reisinformatie communiceert. U kunt de video zo vaak bekijken als u wilt, daarna volgen een paar vragen hierover. U kunt de video ook op YouTube bekijken door op de titel of YouTube tekst te klikken.

https://www.youtube.com/watch?v=yIQBWShISbg

In hoeverre bent u het eens met de volgende stellingen?

	Helemaal niet mee eens (1)	Niet mee eens (2)	Mee eens (3)	Helemaal mee eens (4)	Weet ik niet (5)
Het lijkt mij een goed idee om NS berichten ook in Nederlandse Gebarentaal (NGT) aan te leveren. (1)	○	○	○	○	○
Deze toepassing lijkt mij nuttig voor NS reizigers die Nederlandse Gebarentaal (NGT) beheersen. (2)	○	○	○	○	○
De virtuele avatar zou mij inspireren om meer te leren over de Nederlandse Gebarentaal. (3)	○	○	○	○	○
Ik vind het nuttig dat de video met gebarentaal ondertiteling bevat. (4)	○	○	○	○	○

Wat ziet u als de grootste voordelen van gebarentaal communicatie d.m.v. een virtuele tolk?

_____

Wat ziet u als mogelijke nadelen van gebarentaal communicatie d.m.v. een virtuele tolk?

_____

Heeft u nog andere opmerkingen over deze toepassing?

_____

End of Block: Voorbeeld

Start of Block: NGT/technologie

	Helemaal niet goed (1)	Niet zo goed (2)	Goed (3)	Zeer goed (4)	Weet ik niet (5)
In welke mate beheerst u Nederlandse Gebarentaal (NGT)? (1)	○	○	○	○	○

	Ja, zeer vaardig (1)	Ja, redelijk vaardig (2)	Nee, niet zo vaardig (3)	Nee, helemaal niet vaardig (4)	Weet ik niet (5)
Bent u vaardig met technologische apparaten zoals smartphones en computers? (2)	○	○	○	○	○

End of Block: NGT/technologie

Start of Block: NS

Hoe vaak maakte u vóór de COVID-19 pandemie gebruik van de NS reisplanner app?

- ○ Dagelijks (1)
- ○ Regelmatig: een paar keer per week (2)
- ○ Af en toe: een paar keer per maand (3)
- ○ Zelden: een paar keer per jaar (4)
- ○ Weet ik niet

End of Block: NS

Chapter E Syntactic Trees (X-bar Theory)

General

- Labels correspond to the parts of speech (POS) tags in traditional grammar.
- Tree structures based on X-bar syntax theory.
- Tags based on Signbank glosses [2] [1].
- Numbers correspond to milliseconds.

Variable Abbreviations

- **tt** **traintype**
- **wt** **waittime**
- **dt** **departtime**
- **pn** **platformnumber**
- **is*** **interstation1, ..., interstation4** (* → 0 or more)
- **es** **endstation**

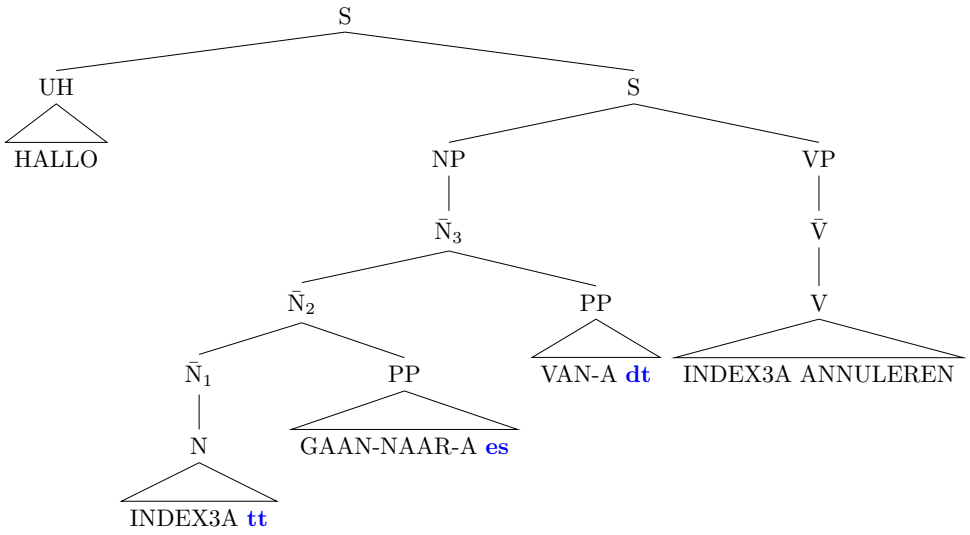
POS-tags

- UH Interjection
- S Sentence
- NP Noun Phrase
- VP Verb Phrase
- PP Prepositional Phrase

Original: Beste reizigers, de treinType naar eindStation van vertrekTijd rijdt niet.

English: Dear passengers, the traintype to endstation from departtime is not departing.

Gloss after co-creation: **HALLO, INDEX3A, traintype, GAAN-NAAR-A, endstation, VAN-A, departtime, INDEX3A, ANNULEREN**



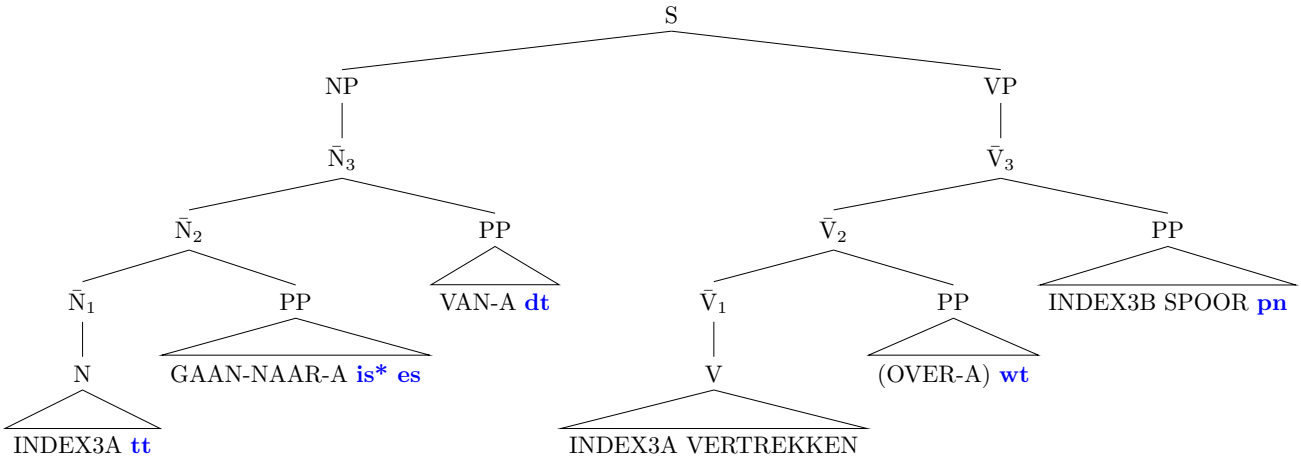
Original: De treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation van vertrekTijd vertrekt over wachtTijd van spoor spoorNr.

English: The traintype to interstation1, interstation2, interstation3, interstation4 and endstation from departtime departs in waittime from platform platformnumber.

Gloss after co-creation: **INDEX3A, traintype, GAAN-NAAR-A, interstation1, interstation2, interstation3, interstation4, endstation, VAN-A, departtime, INDEX3A, VERTREKKEN, (OVER-A,)2**[0] **waittime, INDEX3B, SPOOR, platformnumber**

OR[1]

INDEX3A, traintype, GAAN-NAAR-A, interstation1, interstation2, interstation3, interstation4, endstation, VAN-A, departtime, INDEX3A, VERTREKKEN, INDEX3B, SPOOR, platformnumber, (OVER-A,)[2] **waittime.**

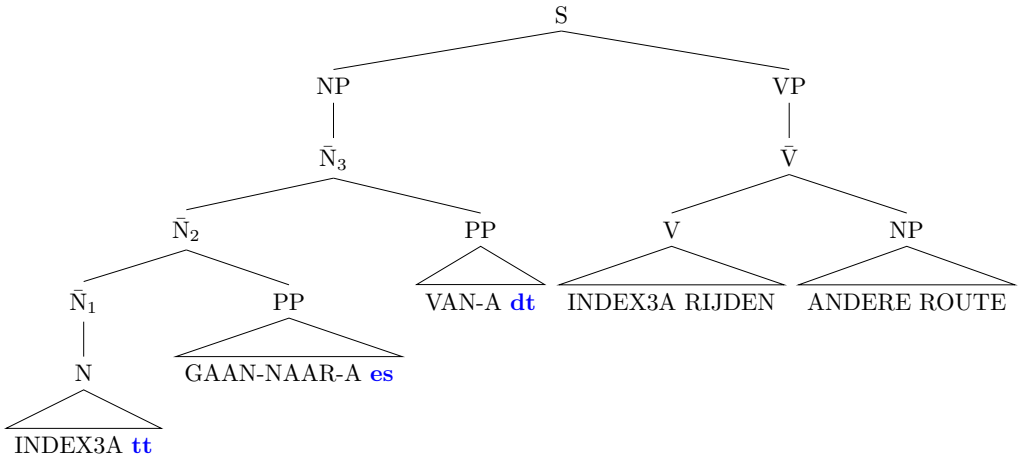


[1]From co-design results: The latter structure is preferred if the waittime is yet unknown
[2]May be skipped depending on the definition of waittime

Original: Herhaling. De treinType naar eindStation van vertrekTijd rijdt via een andere route.

English: Repetition. The traintype to endstation from departtime runs through another route.

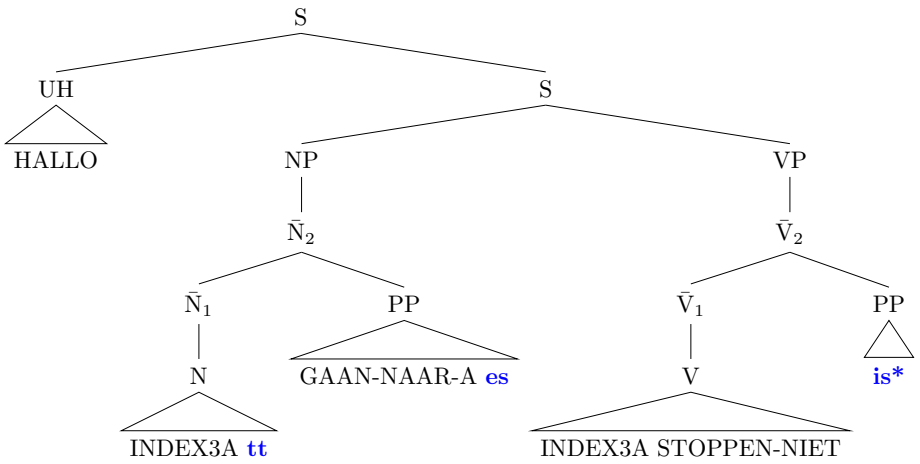
Gloss after co-creation: **INDEX3A, traintype, GAAN-NAAR-A, endstation, VAN-A, departtime, INDEX3A, RIJDEN, ANDERE, ROUTE**



Original: Beste reizigers, de treinType naar tussenStation1, tussenStation2, tussenStation3, tussenStation4 en eindStation stopt niet op tussengelegen stations.

English: Dear passengers, the traintype to interstation1, interstation2, interstation3, interstation4 and endstation does not stop at intermediate stations.

Gloss after co-creation: **HALLO, INDEX3A, traintype, GAAN-NAAR-A, endstation, INDEX3A, STOPPEN-NIET, interstation1, interstation2, interstation3, interstation4**



Original: Beste reizigers, de treinType naar eindStation vertrekt van spoor spoorNr.

English: Dear passengers, the traintype to endstation departs from platform platformnumber.

Gloss after co-creation: **HALLO, INDEX3A, traintype, GAAN-NAAR-A, endstation, INDEX3A, VERTREKKEN, INDEX3B, SPOOR, platformnumber**