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

Britt van Gemert, Richard Cokart, Lyke Esselink, Maartje de Meulder, Nienke Sijm, Floris Roelofsen

Universiteit van Amsterdam, Radboud Universiteit, Hogeschool Utrecht, Nederlands Gebarencentrum, Nederlandse Spoorwegen

britt-van-gemert@live.nl, r.cokart@gebarencentrum.nl, l.d.esselink@uva.nl, maartje.demeulder@hu.nl, nienke.sijm@hu.nl, f.roelofsen@uva.nl

Background

Problem statement Making railway travel announcements available in sign language would make railway services more equally accessible for deaf and hearing passengers. Recording videos of a human signer for every possible announcement is not a scalable solution. Signing avatars are in an early stage of development, but the synthesis of railway travel announcements is expected to be relatively tractable: communication in this domain is only in one direction, and the announcements have a highly predictable structure and content (see the full paper for discussion of earlier work in this area).

Approach Strong involvement of deaf researchers and an interdisciplinary team is essential to ensure that deaf communities' demands and values are taken into account. We report results from several co-design and focus group sessions towards a signing avatar for railway travel announcements in NL.

Phase 1: Initial Design
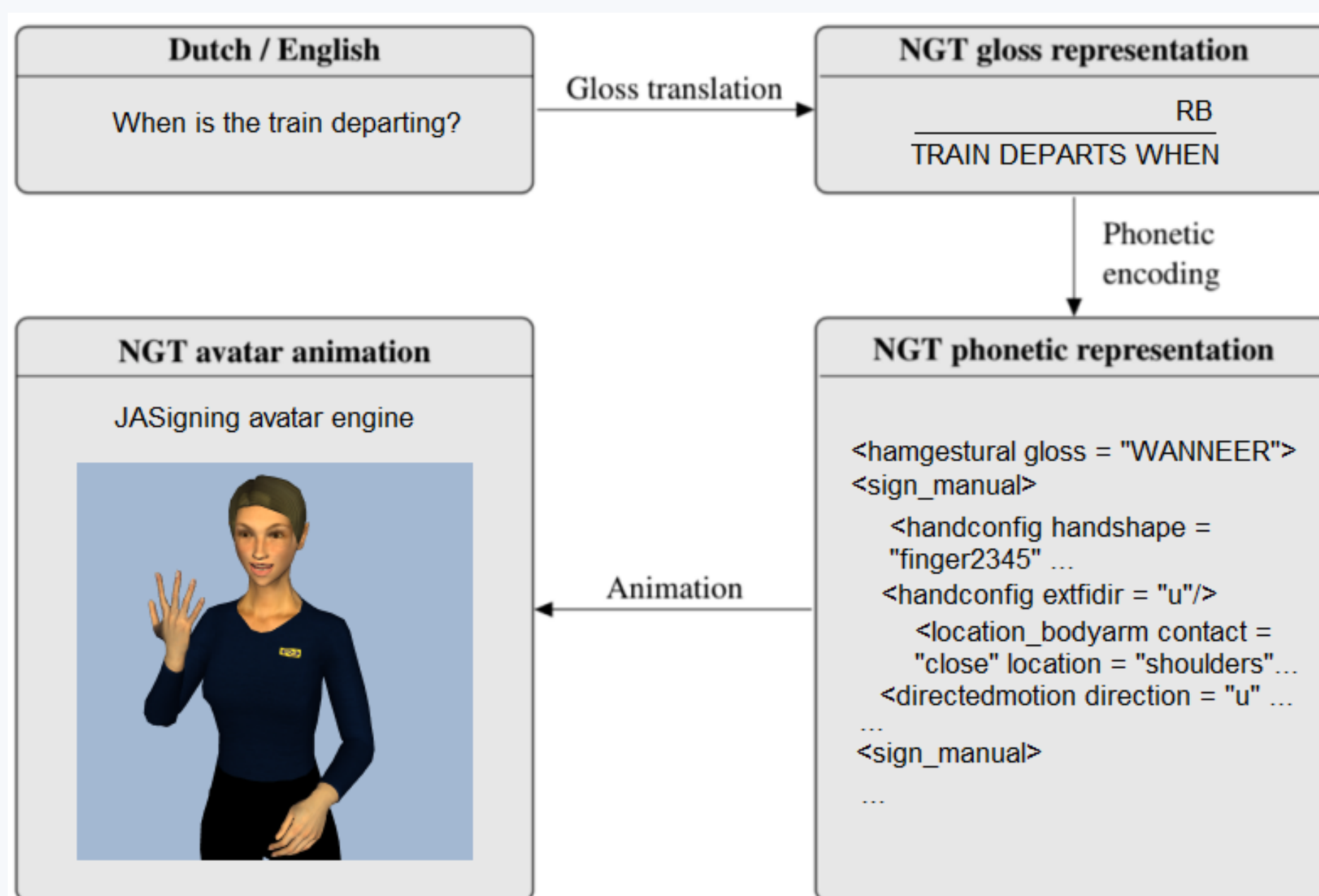
Steps

1. Textual dataset: all possible announcements (from NS)
2. Video dataset: NGT translations of representative sample of textual announcements (by Nederlands Gebarencentrum)
3. Annotation of video translations
4. Avatar animations (scripted keyframe animations, using the JASigning engine)

Phase 2: Co-design

The initial design was further developed through three **iterative** co-design sessions with two deaf and two hearing researchers, and interpreters. Many aspects of the avatar translations were refined.

- Facial expressions, mouthing, eye gaze
- Sentence structure, lists, indexing
- Greeting, visual elements
- Formal vs informal registers
- Sign intensity
- Times and numbers
- Sign transitions (Fig. 1)
- Camera angle (Fig. 2)
- Avatar appearance (Fig. 3)
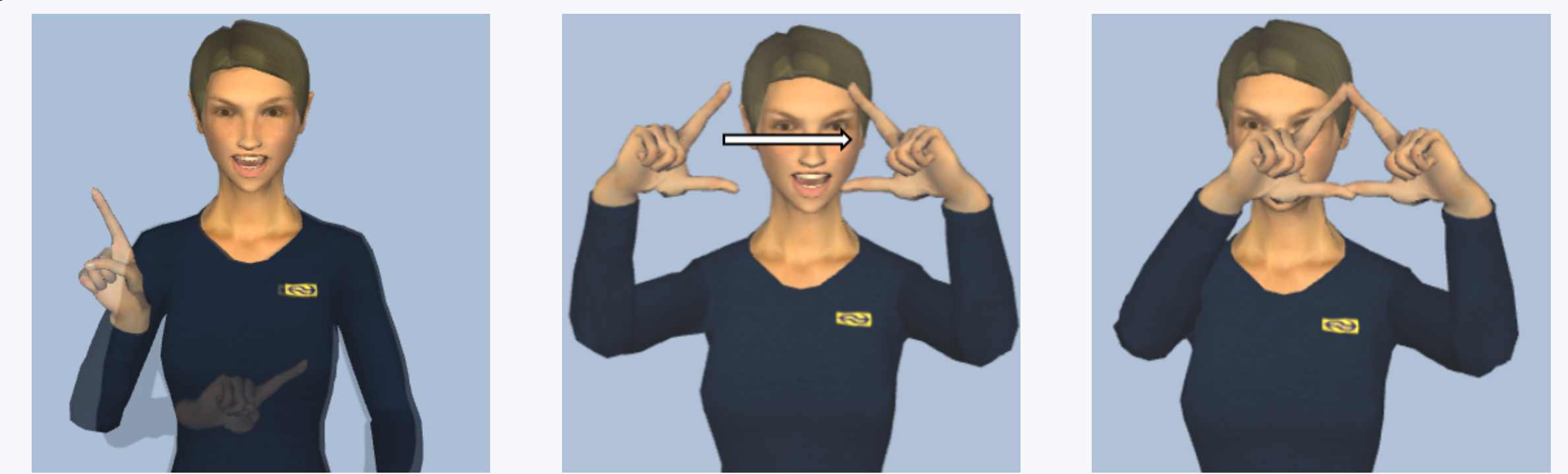
(a) NAAR (b) Transition (c) ALMELO, onset

Fig. 1: Sign transitions

(a) Before (b) After

Fig. 2: Camera angle

(a) Friendly smile (b) Handshapes

Fig. 3: Avatar appearance

Discussion

Combined expertise from various disciplines led to significant improvements of our prototype. Some of these may be transferable to other languages and application domains. That said, the present study merely constitutes a first step and much further work is needed.

Initial Design In future work, reference videos should preferably involve multiple camera angles The JAsigning avatar engine we used is good for fast prototyping, but the resulting animations are not quite natural, something that will be difficult if not impossible to ever achieve with scripted keyframe animation. Facial expressions, mouthing, prosody, and transitions between signs are particularly challenging. A combination with motion capture techniques seems to be a promising route to explore.

Co-design and focus group The co-design sessions identified specific issues for the focus group, and specific alternatives to present for comparison, to stimulate targeted discussion. So we feel that the *combination* of co-design and a focus group was particularly beneficial.

In future work, all sessions should ideally be recorded on video from various angles to avoid the information loss that occurred in our textual transcription of these sessions, mediated by an interpreter. Ideally, the original signed discussion should remain available for analysis.

Phase 3: Focus Group

A 3-hour focus group was organized with six deaf participants, a hearing signing moderator, hearing developer and interpreter. Participants represented different regions, age groups, genders, and school backgrounds. The discussion concentrated on eight topics determined based on the co-design sessions, including animation speed, mouthing, indexing, pauses, choice of vocabulary, time punctuation, and subtitles. In each case, three variants of a sign or a phrase were presented for comparison.

Selection of results

- Indexing (Fig. 4)
- Animation speed +0.40 (JASigning)
- Time punctuation (Fig. 5)
- No first person pronouns
- No pauses

(a) INDEX1 (b) INDEX2

Fig. 4: Indexing

INDEX1 INTERCITY NAAR ALMELO
INDEX1 VERTREKKEN SPOOR 5 INDEX2.

(a) Type V (b) Type 1

Fig. 5: Time punctuation

Acknowledgments