

Online Evaluation of Text-to-sign Translation by Deaf End Users: Some Methodological Recommendations



Floris Roelofsen¹, Lyke Esselink¹, Shani Mende-Gillings¹,
Maartje de Meulder², Nienke Sijm², Anika Smeijers³

1. University of Amsterdam, 2. Univ. of Applied Sciences Utrecht, 3. Amsterdam Medical Centre

Automatic Translation for Signed and Spoken Languages (AT4SSL)
MT Summit 2021

Motivation

- No generally accepted methodology for evaluating comprehensibility of avatars for text-to-sign translation
- Previous methods usually involve on-site experiments
(Gibet et al. 2011; Smith and Nolan 2016; Ebling and Glauert 2016; David and Bouillon 2018; Huenerfauth 2006; Kacorri et al. 2015)
- COVID-19 pandemic calls for online evaluation
(Quandt et al. 2021; Schnepf et al. 2011)

Aim for today

- Discussing some of the **methodological issues** we ran into in designing an online evaluation methodology
- Sharing some of the lessons we learned
- Specific use case: translating frequently occurring phrases in diagnosis and treatment of COVID-19 from Dutch to Dutch Sign Language (NGT)
- But aim here is not to discuss the system itself or the evaluation results --- we focus on methodological issues related to online evaluation

Evaluation goals

Main goals

1. **Individual sign recognition:** To what extent do deaf NGT users recognise the individual signs that the avatar translations consist of?
2. **Sentence comprehension:** To what extent do deaf NGT users understand the avatar translations as intended at sentence level?
3. **Clarity:** How clear are the avatar translations that the system produces?

Secondary goals

1. **Attitude:** How do members of the deaf community in the Netherlands view avatar technology for text-to-sign language translation
2. **Use cases:** What do they see as potentially beneficial use cases for such technology?

Main lessons learned

Three lessons concerning the **design** of the questionnaire

1. **Structure**: Crucial to include video items to obtain baseline
2. **Format**: Ideally all questions and instructions are presented both in NGT and in text format, so that participants can choose their preferred format
3. **Length**: Both the number of items and the length of each item have to be restricted, see below for more specifics

Two lessons concerning the **online execution** of the questionnaire

4. **Individual sign recognition**: Task is not straightforwardly understood --- needs to be clarified with examples and structured response form
5. **Transcription of responses**: Important to include feedback loop --- participants check textual transcription of their signed responses

Structure

- Introduction, informed consent
- Background questions (mother tongue, demographic data,...)
- Comprehension of avatar translations
- Comprehension of video translations by deaf signer
- General perception of avatar technology and potential use cases

Important:

- Including video translations is crucial to obtain **baseline**
- Video items should follow avatar items to **avoid learning effect**
- Learning effect in the other direction (avatar => video) is not forestalled


Format

- All questions and instructions were formulated both in **NGT** (by means of pre-recorded videos) and **Dutch** text
- Participants chose preferred format
 - Most preferred videos, some text
 - Choice of format was greatly appreciated
 - Signer in videos was deaf, this was also appreciated
- Participants reported that questions and instructions were very clear

Length

- Aim was to keep the overall duration of sessions under 45 minutes
 - 10 minutes for introduction, consent, background questions, and example items illustrating the task
 - 10 minutes for questions about perception of technology and potential use cases at the end
 - So: **25 minutes** for actual test items, both avatar and video
- As a consequence, the **number of test sentences** had to be limited:
 - **12** avatar translations
 - **12** corresponding video translations
- The **length of test sentences** also had to be restricted to avoid short term memory overload (esp. in the individual sign recognition task)
 - Around **7** signs per sentence

Recognition of individual signs



What are the individual signs in this sentence?

What is the meaning of this sentence?

How clearly was this sentence signed?

Not clear Very clear

0	1	2	3	4	5	6	7	8	9	10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

←→

- In pilot experiment, participants did not understand the first task

Recognition of individual signs

Two adjustments

- More structured response form
- Example videos

What are the individual signs in this sentence?

What is the meaning of this sentence?



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

What is the meaning of this sentence?

Recognition of individual signs

Two adjustments

- More structured response form
- Example videos



Recognition of individual signs

Two adjustments

- More structured response form
- Example videos

What are the individual signs in this sentence?


Sign 1	Sign 2	Sign 3	Sign 4
YOU	EAT	WHAT	QUESTION

What is the meaning of this sentence?

What would you like to eat?



Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

What is the meaning of this sentence?

How clearly was this sentence signed?

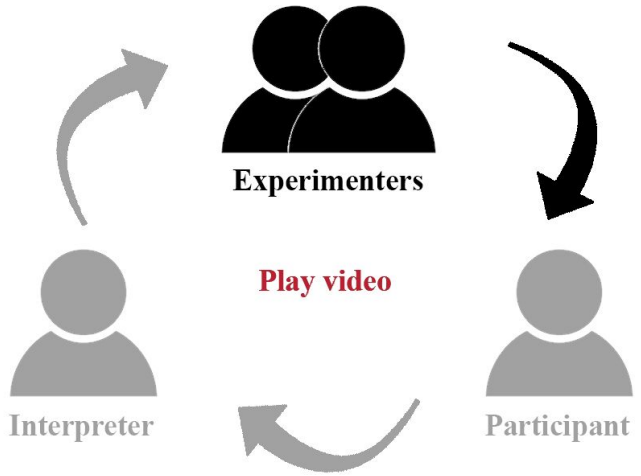
Not clear Very clear

0 1 2 3 4 5 6 7 8 9 10

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

←

→




Experimenters

Interpreter

Participant

Play video

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

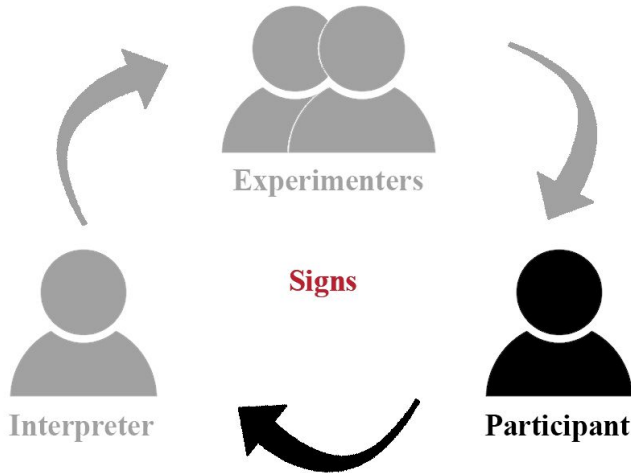
What is the meaning of this sentence?

How clearly was this sentence signed?

Not clear


012345678910

Very clear



```
graph TD; Exp[Experimenters] -- "Signs" --> Int[Interpreter]; Exp -- "Signs" --> Part[Participant]; Part -- "Responses" --> Exp
```

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

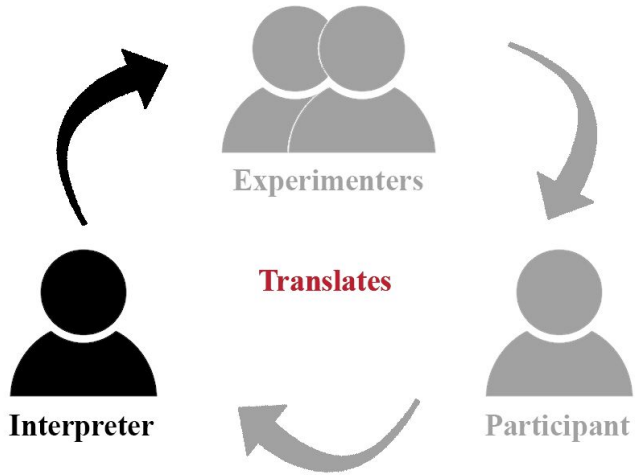
What is the meaning of this sentence?

How clearly was this sentence signed?

Not clear


0 1 2 3 4 5 6 7 8 9 10

Very clear



```
graph TD; Interpreter[Interpreter] --> Experimenters[Experimenters]; Experimenters --> Participant[Participant]; Participant --> Interpreter; Experimenters --- Translates[Translates];
```

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
YOU	PILLS	USE	QUESTION

What is the meaning of this sentence?

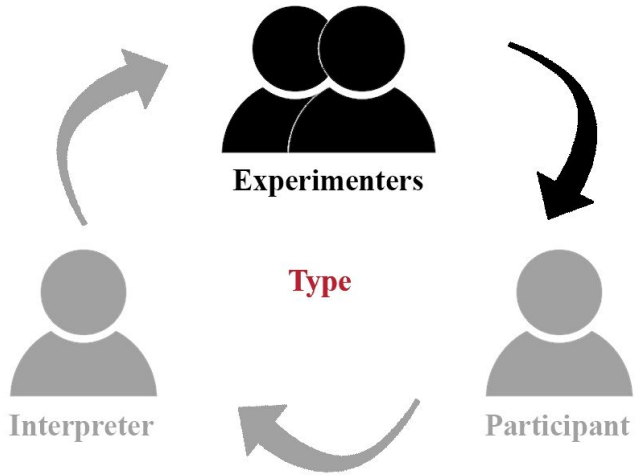
Do you take any pills?

How clearly was this sentence signed?

Not clear


012345678910

Very clear



```
graph TD; Experimenters --> Participant; Participant --> Type; Type --> Interpreter; Interpreter --> Experimenters;
```


Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
YOU	PILLS	USE	QUESTION

What is the meaning of this sentence?

Do you take any pills?

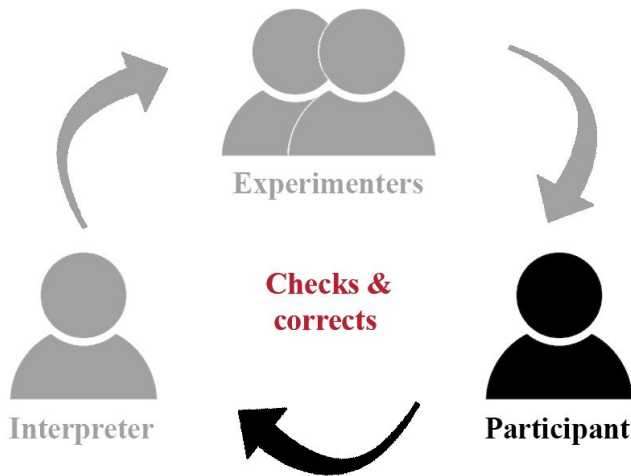
How clearly was this sentence signed?

Not clear

0 1 2 3 4 5 6 7 8 9 10

Very clear

8




Experimenters

Interpreter

Participant

Checks & corrects

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
YOU	PILLS	USE	QUESTION

What is the meaning of this sentence?

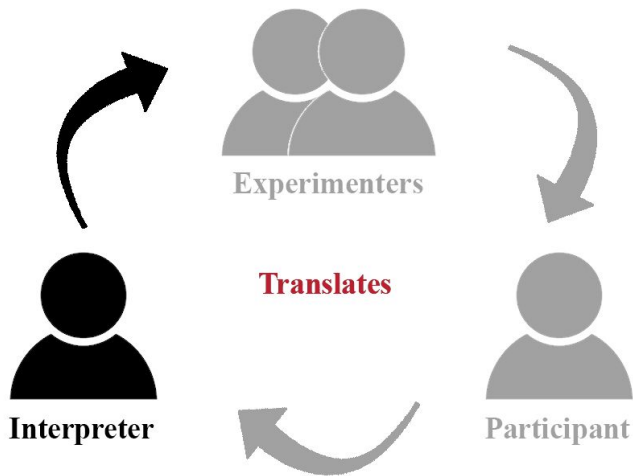
Do you take any pills?

How clearly was this sentence signed?

Not clear


0 1 2 3 4 5 6 7 8 9 10

Very clear



```
graph TD; Interpreter --> Experimenters; Experimenters --> Participant; Participant --> Interpreter; Experimenters --- Translates[Translates];
```

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
YOU	MEDICINE	USE	QUESTION

What is the meaning of this sentence?

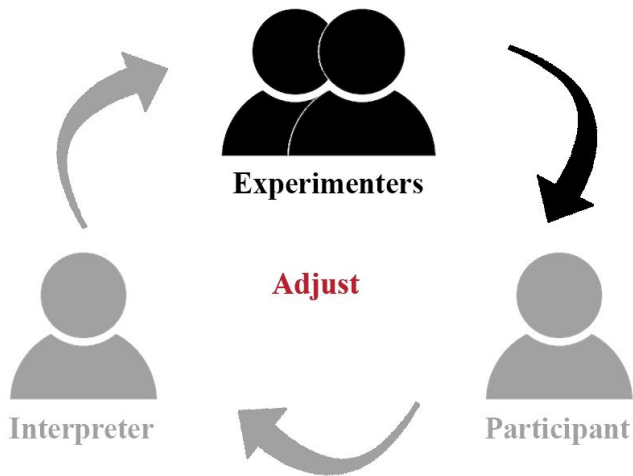
Do you take any **medications**?

How clearly was this sentence signed?

Not clear Very clear

0 1 2 3 4 5 6 7 8 9 10


☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☒ ☐ ☐



```
graph TD; Experimenters[Experimenters] --> Interpreter[Interpreter]; Interpreter --> Participant[Participant]; Participant --> Experimenters; Adjust[Adjust] --- Experimenters;
```

The diagram illustrates a feedback loop involving three roles: Experimenters, Interpreter, and Participant. Experimenters are at the top, connected by a curved arrow to the Interpreter on the bottom left. The Interpreter is connected by a curved arrow to the Participant on the bottom right. The Participant is connected by a curved arrow back to the Experimenters. In the center of the loop, the word "Adjust" is written in red, indicating a feedback mechanism.

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
YOU	MEDICINE	USE	QUESTION

What is the meaning of this sentence?

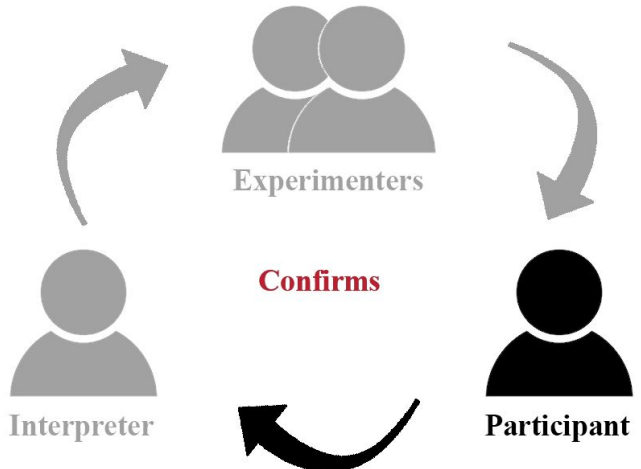
Do you take any medications?

How clearly was this sentence signed?

Not clear


0 1 2 3 4 5 6 7 8 9 10

Very clear



```
graph TD; Experimenters --> Interpreter; Interpreter --> Participant; Participant --> Experimenters; Confirms
```

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
YOU	MEDICINE	USE	QUESTION

What is the meaning of this sentence?

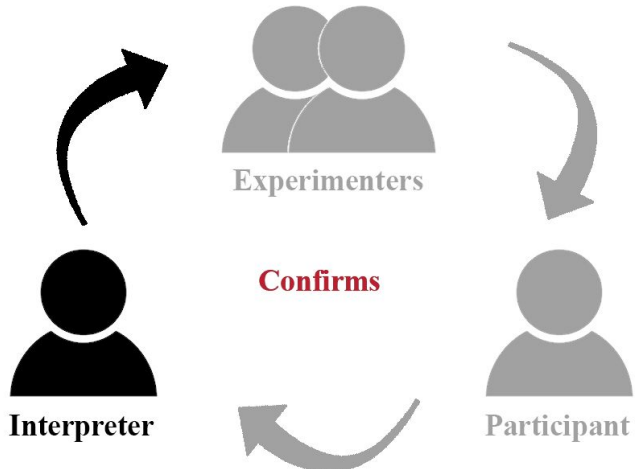
Do you take any medications?

How clearly was this sentence signed?

Not clear Very clear


0 1 2 3 4 5 6 7 8 9 10

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☒ ☐ ☐



```
graph TD; Interpreter --> Experimenters; Experimenters --> Participant; Participant --> Interpreter; Confirms((Confirms))
```

Transcribing responses: feedback loop is crucial



What are the individual signs in this sentence?

Sign 1	Sign 2	Sign 3	Sign 4
YOU	MEDICINE	USE	QUESTION

What is the meaning of this sentence?

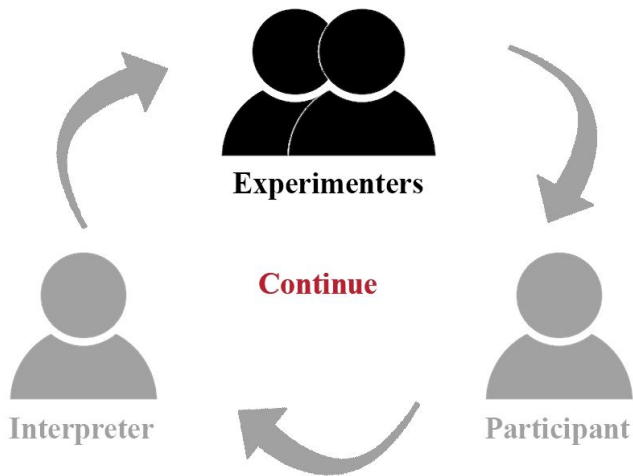
Do you take any medications?

How clearly was this sentence signed?

Not clear

0 1 2 3 4 5 6 7 8 9 10

Very clear



Experimenters

Continue

Interpreter

Participant

Conclusion

Three lessons concerning the **design** of the questionnaire

1. **Structure**: Crucial to include video items to obtain baseline
2. **Format**: Ideally all questions and instructions are presented both in NGT and in text format, so that participants can choose their preferred format
3. **Length**: Both the number of items and the length of each item have to be restricted

Two lessons concerning the **online execution** of the questionnaire

4. **Individual sign recognition**: Task is not straightforwardly understood --- needs to be clarified with examples and structured response form
5. **Transcription of responses**: Important to include **feedback loop** --- participants check textual transcription of their signed responses