Sign Language Translation Mobile Application and Open Communications Framework

Deliverable 1.6: Quality Assessment Report

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 101017255

Project Information
Project Number: 101017255
Project Title: SignON: Sign Language Translation Mobile Application and Open Communications Framework
Funding Scheme: H2020 ICT-57-2020
Project Start Date: January 1st 2021

Deliverable Information
Title: Quality Assessment Report
Work Package: WP1- Co-creation and user response
Lead beneficiary: UPF
Due Date: 30/11/2023 (Revised date)
Revision Number: V1.0
Authors: Josep Blat, John O'Flaherty, Eva Valls, Victor Ubieto, Jaume Pozo
Dissemination Level: Public
Deliverable Type: Report

Overview: As foreseen in the Description of Work of the project, this deliverable on *Quality Assessment* focuses on two aspects: on one hand, the summative evaluation by stakeholders of the performance of the SignON service and the SignON app, and on the other hand on the sign language production by means of an avatar. Other deliverables report on the quality of partial aspects, such as Sign Language Recognition, or translation.

Revision History

Version #	Implemented by	Revision Date	Description of changes
v1.0	Josep Blat	28/11/2023	Integration of the comments from partners
V0.2	Josep Blat	15/11/2023	Focusing the deliverable on aspects complementary to other deliverables, and finalising it. Version for Review
V0.1	Josep Blat	Working version during first half of November	Refining and focusing the different contributions, some of them mature
shell	Josep Blat	Working version from June till end October	Initial outline for contributions from partners

The SignON project has received funding from the European Union's Horizon 2020 Programme under Grant Agreement No. 101017255. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the SignON project or the European Commission. The European Commission is not liable for any use that may be made of the information contained therein.

Approval Procedure

Version #	Deliverable Name	Approved by	Institution	Approval Date
V0.2	D1.6	Shaun O'Boyle	DCU	17/11/2023
Vx.x	D1.6	Vincent Vandeghinste	INT	20/11/2023
V0.2	D1.6	Adrián Núñez-Marcos	UPV/EHU	21/11/2023
V1.0	D1.6	John O'Flaherty	MAC	28/11/2023
V0.2	D1.6	Euan McGill, Santiago Egea Gómez	UPF	21/11/2023
V0.2	D1.6	Lorraine Leeson	TCD	16x/11/2023
Vx.x	D1.6	Mathieu De Coster	UGent	20/11/2023
Vx.x	D1.6	Caro Brosens	VGTC	21/11/2023
Vx.x	D1.6	Henk van den Heuvel	RU	16/11/2023
V0.2	D1.6	Lisa Rombouts Myriam Vermeerbergen	KU Leuven	17/11/2023 23/11/2023
V0.2	D1.6	Davy Van Landuyt	EUD	23/11/2023
V0.2	D1.6	Dimitar Shterionov	TiU	20/11/2023

Table of Contents

1. Introduction

This Deliverable D1.6, *Quality Assessment Report*, continues to present the results from the testing and validation by stakeholders of the performance of the SignON service and the SignON app with specific attention to the quality of the virtual signer, i.e. the 3D avatar. As the project nears conclusion, the results presented are based on the most advanced / final version of the app. Thus, the evaluation is summative, although much work remains to be done. In Section 2 we present two different aspects: how the current version of the app meets the KPIs of the project; users found that the app has positive aspects related to ease of use, while the accuracy and effectiveness of the sign language translation and the avatar's signing accuracy was criticised by several users. In the conclusion, we provide specific recommendations to support improvements of the app with regard to SL Machine Translation.

Section 3 assesses the quality of the SL production via a 3D virtual character. Here, we discuss the overall strategy of signing avatar synthesis, within the goals of real-time translation of the project, comparing different alternatives. Then, we present qualitative comparison with systems oriented to SL production in the framework of translation and research support, showing that the SignON system has a coverage at least as wide as existing systems, while providing improvements in some aspects of Manual and Non-Manual Features (MFs and NMFs, respectively). This Section concludes by discussing ongoing improvements to the sign synthesis system and future work to make it more semantically powerful as well as flexible.

The project set itself very ambitious targets, and the research has met more challenges than foreseen, resulting in an app that still does not meet the translation needs of users. While in this deliverable we summarise our evaluation with respect to the application and the avatar, other deliverables discuss the SignON advances with respect to the different components and pipelines to meet those users' needs. For example, D3.2 presents the research related to SL recognition along with evaluation of the individual component; D5.2 presents the final version of the virtual signer; while D4.5 covers the advances along the translation pipeline from recognition (of either audio or sign language) to synthesis. Other relationships of this deliverable to other ones are discussed within the different Sections.

2. SignON Summative Evaluation addressing the KPIs given in D1.13

2.1 Summative Evaluation Survey

The SignON Summative Evaluation Survey addressing the SignON KPIs given in D1.13, as discussed in Annex B, was based on the following analysis:

KPI	***Ways to measure***	**SE Survey Questions**
1. At least 75% average user satisfaction rating with the overall operation of the SignON service.	*Communication with the stakeholders via co-creation events (e.g. interviews, surveys, round tables, workshops, etc.).*	How satisfied are you with the overall operation of the SignON service ? 1. Very dissatisfied 2. Not satisfied 3. Satisfied 4. Very Satisfied
2. Respond with user-acceptable accuracy for 75% of users.	*Anonymous surveys using structured evaluation tasks, as in D1.4.*	Is the SignON translation accuracy acceptable & useful? 1. No 2. Mildly No 3. Mildly Yes 4. Yes
3. Capability to represent full meaning in meaning transfer between languages, according to human evaluation.	*Perform human evaluation such as direct assessment, ranking and/or A/B Testing.*	Does the SignON translation get the meaning of the message across ? 1. No 2. Mildly No 3. Mildly Yes 4. Yes
4. Language richness, expressivity and intelligibility judged by human assessor.	*Construct a survey and involve human linguists and/or professional translators to indicate the lexical richness, expressivity & intelligibility.*	Are the SignON translations generally understandable ? 1. Mostly No 2. Sometimes No 3. Sometimes Yes 4. Mostly Yes
5. Analysis and linguistic description of sign languages (SLs) relating to the project. Completion of a broad linguistic analysis on a phased basis working through the specified SLs, Irish Sign Language (ISL), Sign Language of the Netherlands (NGT), Flemish Sign Language (VGT), Spanish Sign Language (LSE) and British Sign Language (BSL)	*Feedback on linguistic quality of avatar communication when synthesising a specific SL.*	How would you rate the linguistic quality of the SignON Avatar's communications? 1. Very Bad 2. Bad 3. Good 4. Very Good

To minimise "research fatigue[1]" a structured Cognitive Walkthrough[2] UX evaluation script of using the latest SignON Mobile App V3.0 needed to be complete, but be as short, quick and straightforward as possible. So, the following survey was developed and iteratively simplified:

At SignON Mobile App V3.0 Evaluation - Google Forms each Approved SignON User[3] was asked for their feedback in 3 quick steps:

1. Basic **anonymous information** about the user.
2. **Use of the SignON SLMT App**'s 4 main functions by imagining themselves in a real situation - such as – "*You are In the Netherlands on a long train journey in conversation with a local person, and one of you is using sign language." – and* record how they got on as follows:

Use the SignON Translation App's main functions	To what extent did you find it difficult to complete the task, on a scale from 1 to 4?	Please explain
1. Run the SignON App & check its Information & Preferences screens	□ 1 □ 2 □ 3 □ 4	
2. Text or Audio record & translate a Message	□ 1 □ 2 □ 3 □ 4	
3. Display SignON's Avatar Signing your Message	□ 1 □ 2 □ 3 □ 4	
4. Record & translate a Sign Language Message	□ 1 □ 2 □ 3 □ 4	

3. Fill in 2 **short forms** on:
   a. Scoring the system Usability[4].
   b. The user's overall feedback[5].

2.2 SignON Summative Evaluation Results

As discussed in Annex B[6], the Summative Evaluation (SE) results based on the D1.3 KPIs are as follows:

[1] `We're Over-Researched Here!': Exploring Accounts of Research Fatigue within Qualitative Research Engagements - Tom Clark, 2008 (sagepub.com) and Accounting for research fatigue in research ethics (florenceashley.com)
[2] As described in the SIgnON "Technical requirements and user research (UX design)" reports , D1.4 (June 2021), D1.9 (June 2022) and D1.10 (June 2023)
[3] Explained and discussed in D1.4, D1.9 & D1.10
[4] As described in D1.4, D1.9 and D1.10
[5] As per the above analysis
[6] Annex B explains that the SE was addressed to all 68 SignON Approved Users, of whom 33% (22) responded. Of these 27% were non-partner external users, 27% SL users, and 73% used an Android phone.

D1.3 KPI	Overall	SE Results
*1. At least **75%** average **user satisfaction** rating with the overall operation of the SignON service.*	**60%**	Very Good, but short of target
*2. Respond with **user-acceptable accuracy** for **75%** of users.*	**55%**	Good to Very Good, but short of target
*3. Capability to represent full meaning in **meaning** transfer between languages, according to human evaluation.*	**53%**	Good
*4. Language **richness, expressivity and intelligibility** judged by human assessor.*	**45%**	Fair to Good, which is reasonable for a TRL6 prototype
*5 Analysis and **linguistic description** of SLs relating to the project. Completion of a broad linguistic analysis on a phased basis working through the specified SLs, Irish Sign Language (ISL), Dutch Sign Language (NGT), Flemish Sign Language (VGT), Spanish Sign Language (LSE) and British Sign Language (BSL)*	**39%**	Fair, which is reasonable for a TRL6 prototype

These are based on the results of the following summary of the users' feedback on the SignON App as discussed in Annex B:

Overall Feedback on the SignON SLMT APP *1 Very Bad, 2 Bad, 3 Poor, 4 Fair, 5 Good, 6 Superior, 7 Excellent, 8 Outstanding, 9 Superb, 10 Very Good*	**SL Users**	**Other Users**	**All Users**	**Comments**
1. How satisfied are you with the overall operation of the SignON service ?	43%	66%	60%	*Overall Very Good, but Fair for SL Users, both short of the 75% target*
2. How useful is the SignON translation accuracy ?	33%	64%	55%	*Good to Very Good Overall, Poor for SL users, both short of the 75% target*
3. How well generally does the SignON translation get the meaning of the message across ?	30%	61%	53%	*Good+ Overall, but Poor for SL users*
4. How would you rate the linguistic quality of the SignON Avatar's communications?	27%	51%	45%	*Poor for SL Users, but Overall Fair to Good - reasonable for a TRL6 prototype*

5. Would this App would be useful for you to better communicate & manage your work?	20%	46%	39%	*Bad for SL Users, but Overall Fair to Good - reasonable for a TRL6 prototype*
6. How likely would you recommend this App to a colleague?	53%	53%	53%	*Fair to Good, which is reasonable for a TRL6 prototype*
AVERAGE	**34%**	**57%**	**51%**	*Overall Good, but Poor for SL Users !*

The Users' rating of using the SignON SLMT App V3.0 was as follows (details in Annex B):

Use of the App / Average Difficulty (1. Irrelevant to 5. Very High)	**SL users**	**Other users**	**Overall Average**
Overall difficulty Rating	2.4	2.2	2.3
System Usability Score (SUS) Rating	49.2	39.7	42.7

Users overall rated the difficulty of using the App as 2.3 (low), with no significant difference between SL users (2.4) and Other users (2.2). This "Low" difficulty score indicates that a usable TRL 6 prototype app has been developed and is a good foundation for the future evolution of the SignON service, for all users.

A SUS Rating of 68.0 is the threshold to indicate acceptable usability, so the overall average of 42.7 (n=22), and especially 49.2 for SL Users (n=6), is good for a TRL6 prototype especially given its limited SLT functionality.

2.3 Summary and Recommendations

Users found that the app has positive aspects related to ease of use, but the accuracy and effectiveness of the sign language translation and the avatar's signing accuracy was criticised by several users. Some users also had issues with the app's instructions and interface. Improvements in the accuracy and reliability of sign language translation are recommended, based on the following main points of the users' feedback:

Positive Aspects of the App:	Negative Aspects of the App:
• Easy to use & navigate. • Clear how to do a translation. • Very useful speech to text & sign language for a choice of languages. • Useful speech-to-text & sign language support for multiple languages. • Quick & clear when working (audio/text/voice only).	• Poor signing accuracy of the avatar. • Inaccurate & incomplete sign language translations. • Limited or ineffective sign language input. • Slow response from the service. • Avatar/Signing function did not work for some users.

Thus the status of the TRL6 prototype SignON SLMT App V3.0 is summarised as follows:

App V3 Features	User Technical Requirements from D1.9	Current status
A. User's Mobile Device	1.Easy & intuitive to use. Simple but powerful. Standard modern phones & tablets.	Yes[7]
	2.Will be free & on Android & Apple phones.	Yes
B. System Performance	1.Translation/conversion will be unidirectional - users take turns to input messages	Yes
	2.Translation/conversions respond – 2 to 5 seconds, user-acceptable accuracy 75%.	Not yet
	3.Users >75% average satisfaction rating with the overall operation of the service.	Not yet
C. User Preferences	1. Video, audio & text User Interface (UI) modalities of communication.	Yes
D. SL Translation	1.User SL Input - SLR lexical accuracy & operation acceptable to 75% of users:	Not yet
	a.Performance better than best available automatic SL translation for all SLs.	Not yet
	2. User SL output: Avatar must adhere to Vienna Best Practice SL Avatars	80% - 29 of 35
	a.Overall performance better than best available avatars.	Not yet
E. Speech & Text Translation	1.Normal & atypical, formal & informal, speech.	Yes, partially
	2.Performance is better than best available..	Not yet

This indicates that the SignON SLT, SLR & SLS requires further R&D beyond the current RAI project's TRL6 prototype. However the SignON open Framework and Apps provide an excellent foundation for that

[7] Colour legend: Green = done, Light Green= partially done, Light Orange=not quite yet done, Orange=not yet done, Red=not done at all.

future R&D, as D2.5 "Final release of the Open SignON Framework'' concludes, a flexible and scalable architecture and production infrastructure (software and hardware) for the SignON services Framework is in place and operational. Having overcome many technical challenges through its user-centric design and validation, the SignON open prototype SLMT apps and framework services are a valuable resource for the DHH community.

In summary: “A lot done but more to do” for SLMT.

3. SignON Synthesis: Quality of the different elements and integrated perspective

This Section aims to provide qualitative evaluations of several synthesis elements that have an impact on the final result of the SignON main application, the signing avatar. We have structured this Section by separating the evaluation along three concepts: evaluation of computer animation techniques; evaluation of our system compared with similar existing ones; and qualitative evaluation of independent behaviours of the resulting animation.

3.1 Evaluation of animation strategies for virtual signers

Esselink et al. (2022) note that the ability to create animations featuring a signing avatar is a crucial prerequisite for a scalable automated text-to-sign translation, a view we share. We use the same structure presented in Esselink et al's paper to classify sign synthesis strategies, as this approach supports the automated translation goal. In order to ground the evaluation, we start by understanding the advantages and limitations of the different synthesis strategies that lead to our choice of the optimal strategy that we have used to generate our animations inside the SignON mobile app. Nevertheless, we also provide support to other animation strategies. By providing support to all these strategies, each of which with their particular benefits and drawbacks, we target generalisability and increase the potential for future adoptions of the SignON approach(es).

Keyframe animation is a classical technique, where an animator creates by hand (on a drawing or a computer) the key positions, while those in between are created interpolating among them, so that the illusion of continuous movement is obtained by playing the sequence; each position is technically a frame, continuous movement requires 25-30 frames per second, the more the better. It does not require expensive equipment but considerable manual and highly skilled labour, leading to very good animations. However, this approach cannot be used for real-time animation generation, and its high cost limits the extent of the corpus that can be generated. Moreover, coming from an animator's hand and created within a specific context, this implies a certain degree of individual style and variety that could potentially lead to diverse contextual interpretations which would limit re-use.

Motion capture (MoCap) is the process or technique of recording patterns of movement digitally, specifically the recording of an actor's movements for the purpose of animating a digital character in a film or video game. It stands out as the method that achieves the highest quality animations, as it collects the data from the movements of a real person. MoCap can be subdivided into different subcategories, which share the same methodology but use different techniques: marker-based, sensor-based, and machine learning (ML) techniques. *Marker-based* methods obtain data as positions of markers placed on the person and are based on computer vision techniques; commercial systems are based on expensive equipment and software and require specific labs; additionally, a big amount of manual post-processing is needed to clean the noisy data obtained. These reasons make a marker-based approach unfeasible in many applications. *Sensor-based* methods capture data using inertial or electromagnetic sensors; the systems are quite intrusive, and increasing precision (resolution) requires significant investments; their intrusiveness is a major barrier for DHH users. More recently, with the success of Deep Learning (DL), using ML techniques combined with computer vision seems a way of obtaining high-quality at a fraction of the previous cost, as webcams can be used as input, with good results and little manual cleaning. ML/DL models are trained with a large amount of real human body movement data. When used to estimate poses and movements, they deal very well with occlusions without the need to be intrusive to users when capturing data. Among the well-known architectures, we performed tests using OpenPose 2D (Cao et al., 2019) and MediaPipe (Lugaresi et al., 2019).

During the early stages of SignON, we tested the different techniques mentioned above, with the outcomes we just indicated. Our system provides support for sign synthesis based on ML/DL, as we presented in D5.8 (some progress is reported in D1.12 due on m12 of the project). It is able to generate animations from different inputs, including video or webcam. To support its use in the context of SL synthesis, our system includes a highly visual manual editor to correct the errors that can come in the estimation of the animation derived from the ML-based model.

In order to understand the errors from ML-based models we need to start from the state-of-the-art approaches of this method, which generally aim at estimating 3D positions of the bones per frame. They keep being improved at the moment of writing this deliverable and these improvements can be re-used in our system. However, these positions results need to be processed in a further stage to turn them into 3D rotations (which is the standard for animating skeletons).

Most current systems achieve this through inverse kinematics (IK), which does not ensure a quick convergence, if any at all, and can be computationally expensive, leading to performance bottlenecks. These issues might be avoided taking into account that a continuous smooth stream of frames has a relatively small variation rate between frames. However, IK-based animations are dependent on the chain of bones defined to move a skeleton, but this chain can be specified in multiple variations, separating the whole chain in order to give more natural results but at the same time requiring more 3D positions in the space. IK might give anatomically impossible solutions. Some approaches to fix this issue are hybrid techniques where IK is used as an indicator of a predefined animation, but making it not usable for other avatars with different skeletons.

During the time of this project, we proposed an alternative to this approach by directly estimating quaternions (4D representations of rotations, which are more compact than the usual 3x3 rotation matrices, and can be more easily interpolated) from single camera images. Through this alternative approach we skip the problems related to the use of IK. However new problems arise, as estimating from 2D images generates 4D representations lacking depth information that makes it difficult to predict for the ML models. In addition, as it happens with keyframe animations, in both approaches semantic information is not considered, and thus the intended message can probably be not fully understandable by signers. This would make it an unfeasible approach for scalable SL production and supporting research. In this context, linguistic understanding of SLs is needed, in the same way as for the synthesis of spoken languages, leading us to the complementary animation strategy that we discuss next.

Scripted animations are procedurally generated based on structured specifications of the phonetic descriptions of signs. Despite results not achieving the same quality as the previous approaches, this strategy does not need expensive equipment; the phonetic descriptions are SL independent and potentially lead to more re-usability and less manual skilled labour. This makes the approach very suitable for scalable sign synthesis systems.

In order to review the state of the art of this strategy we need to firstly present the existing sign representations. Very briefly, current representations are evolutions of the initial SL notations (such as Stokoe's notation and HamNoSys) that proposed / made use of the components representing the SL phonetics. Moreover, as Naert et al. (2020) point out, these notations only deal with the representation of static and independent signs or phonemes, but fail to correctly represent dynamic signs, the concatenation of static signs, and the synchronisation of the different components. Within the rapid evolution of information technologies, SL notations following an XML-like structure format were proposed. These representations can be compiled by computers, are more standard and interchangeable, and were intended to support SL production via signing avatars. Among the different proposals, we consider it especially relevant to mention EMBRScript (Héloir and Kipp, 2010), as it is similar to our approach; it is related to the Behavioural Markup Language (BML), which is used within conversational characters research as an exchange format and which we discuss further later. On the other hand, SiGML (Kennaway, 2004) is arguably more extensively used. It was originally developed within the European projects ViSiCAST (2000-03) and eSIGN (2002-04), as an XML version of HamNoSys, with which it co-evolved. It was expanded in the JASigning project (2010-16) (Glauert and Elliott, 2011), which provided explicit timing control, synchronisation between elementary motions and direction specification in various contexts. It was designed for animating virtual signers, which explains why some timing attributes or very precise orientations can be specified in SiGML and not with HamNoSys. Indeed, JASigning turns SiGML code into character animations. As SiGML is based on HamNoSys, the existing phonetic transcriptions in HamNoSys could be re-used by the JASigning system. An example of a recent tool automatically turning HamNoSys into SiGML is the one by Neves et al. (2020).

Beyond isolated signs and their phonetic representations, more complex representation is needed to express meaningful SL utterances. For example, the annotations for the manual part of a sign language utterance (i.e. the "signs") can be supplemented by a transcription for the simultaneously produced non-manual features, e.g., sentence marking for negation, topicalisation, mouthings, and so on. For this purpose, words from the spoken language or symbols can be used. However, this representation does not provide grammatical context and lacks the 3D information to support SL utterance synthesis. Filhol et al. (2014) proposed AZee as an improvement of Filhol's Zebedee work (Filhol, 2009). AZee specifies linguistic input which includes semantic meaning in the representation by adding some syntactic mechanisms such as the relationship between entities, without fixed lexical signs which would make it very SL dependent. Other more recent work by Murtagh (2019) also follows the idea of including more grammar complexity in the representations, namely proposing a framework made of RRG+Sign_A, which combines Role and Reference Grammar (RRG), a structural functionalist theory of grammar and a functional model of language, with Sign_A, a notation that includes the five phonetic components of SL with the addition of temporal information of each feature.[8]

Our first attempts to base our virtual signers on SiGML descriptions, which focused on NMFs, specifically mouthing, revealed its limitations: co-articulation models were not possible within this representation. Other limitations were in terms of timing information and synchronisation, especially when compared with BML (Behaviour Markup Language), a specification used by researchers in conversational virtual

[8] The work by Murtagh has been employed to generate SL representations and SL repositories within SignON. See deliverable D5.4 First Sign language-specific lexicon and structure and D5.5 Second Sign language-specific lexicon and structure

characters, which we had used in previous projects. We proposed an extension of BML, which added support to new SL approaches that include semantic information such as Sign_A (D5.7 “A planner for translating from Sign_A to BML-based script”).

The following Section discusses in detail existing signing synthesis systems which are symbolically based, relating them to ours, which has been implemented in the second half of the project to support the SignON app as the best scalable approach, and complements the ML/DL strategy, which is also part of our system.

3.2 Comparative Evaluation of our Symbolic Animation System

The AZee system previously mentioned evolved from a design proposal to “i) realise animations that feature interleaving communicative channels, and ii) ... be part of a generator interchanging synthesis with pre-animated sequences” (Nunnari et al., 2018), p. 1 (p. 155)) to a generator system through their initial implementation of “a bottom-up animation solution” based on an IK solver (Nunnari et al., 2018). This allowed for executing SL notations and synthesising signing avatars. The AZee system allows for the combination of low-level features (such as phonetic components) into larger ones (such as signs). The paper uses a classification of synthesis platforms into bottom-up and top-down systems, which we adopt here.

Bottom-up systems take the most elementary features (e.g., phonetic representations) to generate signed animations and build the intended, complex, meaningful output from them. The AZee representation system allows for the resolution of meaningful utterances into parallel elementary features. The drawback of these bottom-up systems is that they render robotic animations. On the contrary, top-down systems start from larger meaningful chunks, representing them through pre-recorded animation sequences, thus offering significantly more understandable and likeable results. However, as the paper mentions, “it is never possible to have everything prerecorded” and it argues that a bottom-up solution is necessary as a fallback for existing top-down systems, so that they can be used to produce signing avatars, despite the risk that the bottom-up solution may appear robotic.

JASigning (Glauert and Elliott, 2011) offers a web platform to execute SiGML orders in real-time, its neat interface allows users to write their own SiGML code in a window and see the resulting animation performed by an avatar in an adjacent one. While it attempts to create a simple and intuitive notation, the SiGML representation has significant complexity derived from its HamNoSys relationship, and it lacks user-friendly documentation. Its JASigning rendering counterpart is not Open Source and also lacks sufficient documentation. Despite this difficulty for third parties to use SiGML/JASigning, the availability of immediate results is a noteworthy contribution to the SL community and research and indeed, SiGML/JASigning has been recently used by several projects as the SL synthesis tool to display procedural avatar animations. Some examples are BabelDr (Strasly et al., 2018) originated in 2016 as a platform to help communication in the medical field, uses a speech based UI layer which provides doctors or nurses with a sentence selector tool from a list of limited possibilities, which have quality-controlled SL translation, through signed videos and avatar synthetic animations. Other examples are (Esselink et al., 2022), applied in healthcare, and (Van Gemert et al., 2022), to provide railway travel announcements. A very interesting part of this work is the effort on relating each gloss phonetic representation with its corresponding rule-based grammar representation, in order to derive more correct utterances. It is also worth mentioning that all the examples presented required a previously crafted repository of all the signs/utterances in SiGML, in the Swiss version of the French Sign Language (LSF-SR) for BabelDr, and in the Sign Language of the Netherlands (NGT) for the other two.

From the previous discussion, we see that bottom-up systems based on a phonetic representation, for instance, are necessary to provide systems that can be used in practice and to support SL research, despite their robotic results, as the scalability is a significant advantage, to which real-time rendering is a potentially additional one. On the other hand, the use-cases require repositories with this phonetic representation, which are scarce as they currently require extensive labour, and whose quality needs to be tested in practice. The quality of the rendering and characters of existing systems and their documentation to facilitate third party use can be improved.

The real-time procedural/symbolic open system we have been recently developing within SignON takes into account these different aspects: while it is based on BML representations to achieve maximal quality, it supports SiGML representations, to facilitate repositories re-use. The representation and rendering fixes some of the identified shortcomings of JASigning (e.g. in mouthing or transitions). Besides being Open Source, extensive documentation is available.

Up to this point, we have reviewed the most significant symbolic systems which go beyond limited research prototypes and evaluated their strengths and weaknesses, and now indicate several aspects where our system shows improvements or alternative strategies, building on their experience.

1. Using an extension of BML instead of SiGML adds support for a wide facial expressive range (which is required by SLs) (we will see this in Subsection 3.3), and in giving more control to the timing of animations. Further, BML has been created within a very active research community for virtual characters that are conversational, thus fitting closely the communication objective of

signing avatars. Indeed, the research field goes far beyond that strict label, as seen in the two recent handbooks reflecting 20 years of research, (Lugrin et al., 2021) and (Lugrin et al., 2022), and adopting this approach makes it possible to potentially integrate its diverse research outcomes into signing avatars. It is also closely related to the SAIBA framework (Kopp et al., 2006), where intention, planning, and realisation of (conversational) behaviours are separated and interrelated, which facilitates abstraction and re-usability. In addition, as we previously discussed, our representation is well documented and can be found online, which will facilitate users to create signs and avoid possible misuse of attributes when building a sign.

2. Regarding the execution of NMFs, our approach follows a one-to-one relation to the Sign_A classification. Our facial animation is based mainly on blendshapes, which is quite standard for animation, and complementary use of Action Units (AUs), which correspond to actual facial muscles, and are widely used in research, especially for facial recognition; see for instance the recent approach to NMFS (Wolfe et al., 2022) and the extensive state-of-the-art discussion of Wolfe and McDonald (2021). A valence-arousal emotional model (see, e.g. Tellamekala et al. (2023)) is supported, which makes it easier to generate a very wide range of facial expressions. When considering NMFs, mouthings, which "are silent articulations of (a part of) a corresponding spoken word of the surrounding language" (Pfau and Quer, 2010) and heavily used by some SLs, require a specific attention, since mouthing follows a speech methodology involving several phonetic aspects. JASigning (SiGML) has limited support for realistic mouthing, with the consequence of very reduced intelligibility in SLs that heavily use them, such as NGT (Esselink et al., 2022). The key reason is the articulation model used to turn phonemes (sounds accompanying letters) into visemes (visual appearance of the phonemes): JASigning maps single phonemes to their corresponding visemes, while our system is based on coarticulation during phoneme transitions, i.e., the visemes of the neighbouring phonemes are used to compute the viseme (its blendshape values) for the current phoneme (Pozo, 2022). Our mouthing system uses a set of six facial blendshapes (one of them driving the tongue) to generate the different visemes, instead of relying on specific blendshapes for each phoneme. This supports coarticulation of specific blendshapes (muscles) when transitioning between visemes, as some require certain muscles not to differ too much from their original values. For example, the phoneme /p/ does not need to specify the state of the lips (stretched or narrowed) as long as the lips are touching each other, which can be problematic to achieve with a blendshape per phoneme strategy. Furthermore, our system uses the ARPABET (Shoup, 1980) phonetic

transcription of words, which is available in multiple languages. While SAMPA transcription is more extensive, it is also more complex to use. ARPABET encoding is already sufficient to cover all the mouthing requirements since many phonemes differ among them in sound rather than in visual appearance. Further explanation about each NMF component and some visual comparison is included in the Subsection 3.3.

3. Our preferred input representation is BML because of its expressiveness and extensibility. But as previously mentioned, corpora/lexicon/dictionaries with phonetic description of signs are scarce, and are mostly based on HamNoSys and/or its computer-readable SiGML counterpart. Thus our system includes a SiGML/BML bridge, so that SiGML representations can be automatically realised (i.e., rendered as avatar animations). As JASigning, our real-time system interface includes a text input widget where SiGML or BML representations can be typed/pasted and the corresponding signing avatar visualised, to check the correctness of representations, or eventually improve them. Furthermore, there are already existing parsers from HamNoSys to SiGML, so at some point, we would also extend our system to support HamNoSys input representations.
4. Lastly, the recent significant advances in the render techniques and the expected visual quality has affected "old" systems. For instance, MetaHumans introduction represents a big leap with respect to previous existing systems and current avatar technology, which we tried to reduce by providing high quality characters (such as EVA) that supports Physically Based Rendering among other rendering techniques. Again, visual quality comparisons can be seen in Subsection 4.3, and a more in depth discussion on the characters' quality, and their customisation will be provided in D5.2 "A Virtual Character".

3.3 Manual Features qualitative evaluation

Manual features (MFs) are generally acknowledged as a core characteristic of sign languages. As stated before, we have followed a phonetic approach, acknowledging its potential limitations with regards to animations looking robotic. Beyond this, besides the internal testing of our IK system generating symbolically sign animations that we have extensively carried out, we think that at the current stage, the priority is to assess its comparative coverage of MFs. Indeed, as Nunnari et al. (2018) mention, improving the MFs in the context of bottom-up systems (and the phonetic approach) implies "using larger dedicated procedures or even full play-back or pre-recorded chunks", which itself "advocates against bottom-up approaches". In terms of comparative evaluation of the coverage, we start by reminding that we proposed a BML extension to cover the full Sign_A description at the early stages of SignON. On the other hand, guaranteeing that the instructions define a complete set that covers all possible signs across any language is not trivial. Thus, at later stages of SignON, our strategy has been to compare (and refine) our system with the existing one that has been more generally used in use cases where rendering is based on phonetics, HamNoSys and its computer readable and avatar playable counterpart SiGML. Indeed, researchers have been steadily working on and improving both to apply them in different use cases, while other projects, such as JASigning, have already exploited and exposed them. This strategy has the additional advantage of not having to introduce for already existing users a new notation system different from the already existing ones, as this might deter potential users from learning as they will stick to what they already know. Thus, our extended BML instructions have been developed to resemble and match the capabilities of HamNoSys and SiGML; so existing repositories and their research can be reused, and different renderings compared. While more extensive testing is needed to confirm it fully, our current understanding is that our coverage of SiGML is complete. On the other hand, the extended BML and our system allows the users more freedom such as full control of the timing of instructions or normalised but continuous possible values to describe attributes instead of a restricted set of tags, which would represent improvements. For instance, setting the distance of the hand from the body while in a location can be set through a value in a range of 0 to 1 being 0 contacting the body and 1 fully extending the arm instead of only having three tags: touch, close and arm extended. While more extensive testing, using the existing repositories repositories, or extending them to new use cases, might uncover limitations of our system, as well as of HamNoSys and SiGML (and of the phonetic approach itself), especially with regards to prosody, the existing ones can be used as a starting point for more extensive evaluation.

We turn next to more detailed comparisons of several aspects of MFs.

3.3.1 Locations and Hand constellations

The set of body *locations* to which a hand can be moved within our extended BML, originates from the set SiGML offers, which, in turn, comes from the HamNoSys specification. Although some refactoring of labels has been carried out, labelling remains essentially the same. On the other hand, our extended BML definition of the point of the hand that should be moved or which should be reached, differs from SiGML, as this specification follows an XML format, while BML uses JSON. This difference is in format, but not in information contained. In BML the whole information is contained in the instruction itself instead

of having children nodes which contain additional information. BML attribute names are more intuitive as the name itself hints at what is being modified (either as origin or as target).
Hand constellations in SiGML define how both hands should be positioned with respect to each other. Our extended BML exposes an attribute to define whether one or both hands are moved when reaching each other is the goal.

3.3.2 Hand configuration and handshapes

A *hand configuration* in SiGML defines the handshape, where it is pointing to (if the index finger were fully extended) and where the palm is facing (hand twisting). In our extended BML approach these are three separate instructions for more clarity (although it is possible to pack them into a single instruction). Hand and palm orientation remain the same as in SiGML and we use the same technique to determine palm orientation given a hand orientation (HamNoSys is quite ambiguous regarding how to determine the result of each palm orientation combined with each possible hand orientation).

Handshapes are more complex instructions containing many attributes that can overlap each other. The basic set of available handshapes is already supported in our extended BML implementation and most of the attributes that modify such base shapes that are described both in HamNoSys and SiGML are present in our extended BML. Some attributes such as finger crossing are not fully supported or not supported at all in the current implementation. Such attributes have currently been left out as they are not used in any of the glosses of the repositories we have been using. Yet, we expect to implement them in the near future. Some attributes such as splay of fingers have been already implemented and indeed set the base for the upcoming support of such properties as finger crossing.

3.3.3 Motions

Motion instructions are mostly used once a base pose is established through other MFs and are divided into several subtypes. Regarding spatial displacement, a directed (linear) and a circular movement exist. Both instructions offer the same capabilities that SiGML provides, while additionally numeric input instead of predefined tags can be used. Moreover, some attributes such as the zig-zag oscillation speed have been included and exposed to the user.
The fingerplay, which involves wiggling the fingers, and the wrist motion, which involves nodding, shaking and/or twisting the wrist, expose not only the intensity of movement but also their oscillation

speed. As usual, some default values are implemented, so that the user only needs to specify intensity and oscillation speed when convenient.

3.4 Non Manual Features qualitative evaluation

In this Subsection, we evaluate qualitatively several aspects of the NMFs. HamNoSys defined NMFS as “multi-tier representations or schemes” (Hanke, 2004). They encompass several body aspects such as the shoulder, body, head, and gaze movement, facial expressions, and mouth. Our implementation follows these categories, and adds an “emotions” one. As previously discussed the JASigning system has been used in several projects as the final render result of the sign animations, and a degree of DHH users’ acceptance shown in them. For this reason we compare our results for each aspect with those of JASigning.

3.4.1 Shoulders

Shoulder raise and hunch can be relevant in situations where a question is posed or to exaggerate some intended meaning. Both JASigning and our system expose both attributes, so that a gloss might include the instruction to move the shoulder. Nonetheless, our approach includes an automatic raise and hunch derived from the simple act of moving the arm; it is added on top of the mentioned attributes when they are present. Although subtle, using multiple bones to perform an action, as in this example, adds realism to the final animation, because real movements are complex (meaning that involves not one, but multiple parts of the body). Indeed, the avatar looks more natural and less robotic as the amount of bones that remain completely static are less.

3.4.2 Body, head and eye movement (gaze)

The SiGML instructions related to these aspects, derived from HamNoSys, seem to be complete enough. Our system can reproduce them, as seen in the mapping we performed of SiGML and our functionalities, included in the documentation of our system[9]. The following image shows a comparison of results of the execution of a similar instruction using JASigning and ours.

[9] https://github.com/upf-gti/SignON-realizer/blob/main/docs/InstructionsBML.md

Figure 1 Avatars tilting their head to the left and looking to the right

Left: JASigning. Right: SignON

3.4.3 Facial expressions and mouthing

Both systems diverge considerably in how they manage facial expressions. JASigning divides them into four categories: eyebrows, eyelids, nose states and mouthings, where mouthings are both simple facial expressions around the mouth and movements derived from actual words in their spoken language. The SignON approach uses the Facial Action Coding System (FACS) (Ekman & Friesen, 1978l and intensities to modulate the entire face and a specific mouthing for the spoken words.

The FACS was meant to discretise the possible face gesticulations based on the actual muscles involved in them. Thus every possible movement can then be generated through a combination of such Action Units (AU). Moreover, other kinds of software such as facial recognition most frequently use FACS as well, adding another argument in favour of an approach based on FACS.

Not all AUs have been included in our system as some might seem to be an over specification of already existing AUs. As an example, both the Lid Droop and Eyes Closed AUs are specified as Relaxation of Levator Palpebrae Superioris. Yet the former is simply the latter in a lower intensity (slightly closed/relaxed eyelids).

Our mouthing approach uses the ARPABET encoding, which is simpler to understand and write than SAMPA (what JASigning uses). Furthermore, a system of coarticulation has been implemented, so transitions between “phonemes” are not only smooth but interfere with their neighbours to produce a more organic result. Some examples of such coarticulation, comparing it to JASigning can be seen at https://www.youtube.com/watch?v=40FkJFUXNEA&list=PLdaCikBoa8nBF97lwGQrpZFzl9uyh1_iT&index=11

https://www.youtube.com/watch?v=2LXTKVU4cUI&list=PLdaCikBoa8nBF97lwGQrpZFzl9uyh1_iT&index=12
https://www.youtube.com/watch?v=2LXTKVU4cUI&list=PLdaCikBoa8nBF97lwGQrpZFzl9uyh1_iT&index=13

Finally, an automatic blink is performed in our system instead of having to specify it as is the case of JASigning. This decouples the signs from natural procedures such as blinking.

3.4.4 Emotions

Emotions are not explicitly covered by JASigning at the time of writing this deliverable. Our system includes the provision of emotions, which internally uses several FACS to generate them. A set of 8 basic emotions are hardcoded and can be accessed directly. Further, a valence-arousal approach is also exposed which allows the users to carefully set intermediate emotions, facilitating the animation task, as emotions usually involve several concurrent AU. Within our system signs and emotions can be decoupled from each other. This means that emotions can start mid sign and span several signs, for instance, changing their meaning without practically changing any code.

Figure 2 Avatar showing several emotions

Left: Neutral. Centre: Happiness. Right: Worry

3.5 Remarks, limitations, and future work

The previous quality assessment of our SignON animation system has focused on comparing it with the existing alternatives which go beyond research prototypes and intend to reach sufficient breadth and applicability. As indicated earlier this seems to be the most appropriate approach in the current stage of the different implementations. Furthermore, all existing comparable systems, including ours, need substantial improvements to reach a level of quality that leads to their widespread use. In this sense, our most recent approach has tried to build on them, ensuring that most research can be reused, and providing some improvements. It is convenient to restate that quality and customisability of the characters is another ingredient of acceptability by users. In a complementary deliverable *5.2 A Virtual Character*, we present a proposal to make skeletons and facial structures more interoperable, the consequent adaptations of our system to become truly multi-character in signing, and a pipeline to easily and reliably generate multiple characters.

According to those researchers who have used existing NGT / SiGML repositories for travel or healthcare, DHH user testing revealed that the correctness of the lexical descriptions contained cannot be taken for granted (Van Gemert et al 2022); a similar experience holds for LSF-SR applied to health (Strasly et al 2018). Extensive user testing in applied situations is thus needed to both improve the repositories, and the systems rendering them. This testing is a next stage of our research, to be able to achieve significant reliable improvements. Some improvements that might be less significant, e.g., making blinking more reactive besides automatic, will be carried out regularly.

There are some performance indicators which are relevant for the acceptability / usability of the application which it already satisfies. The various stages of the synthesis need to be executed with as few waits as possible. Preprocesses need to be relatively few and fast, so that users do not think the application has stalled. While executing the animation, the frame rate needs to be kept at a minimum of 24 frames per second (fps) as it is the rate at which humans perceive a set of changing images as motion (videos). Nonetheless a minimum of 30 is usually required to ensure seamless perception. Keeping such a minimum of frames per second ensures also that the application is responsive to user input. Currently, the sign synthesis is performed in real time, at the frame rates indicated, in standard smartphones, once the character has been loaded. As detailed in *D5.2 A Virtual Character*, adaptations of the higher quality models have been made so that the downloading is fast, and the avatar reactive. This is an example of the tradeoffs that need to be considered when developing applications expected to run on most mobile devices, which are already restrictive hardware per se. Furthermore, the animations are based on XML-like commands, and thus, there is no latency in them even in low-connection situations.

Now we turn to aspects more related to linguistic traits of SLs and their understandability. These aspects are more challenging and we discuss the approaches we intend to carry out in the next future to address them. The naturalness and expressivity of the avatar signs and expressions is essential for its acceptance by the DHH community. Communication needs to feel as human-like as possible. It does not only involve well crafted signs but a smooth transition between them. Moreover, the production of a sign might be influenced (sometimes unconsciously) by the signer's personality and preferences, adding small variations that do not alter its meaning. Such variations are what makes the sign unique and human and one of the reasons why synthetic animations often look robotic. Moreover, novel signs need to be taken into account. Some traits can be decoupled from the signs as such and be assigned to the overall sentence meaning. This can facilitate the task of synthesising the animation phonetically as new signs can be added on top of the raw animation. We are currently working on those different aspects, while managing complexity and integration.

Signed languages, as well as spoken languages, require appropriate grammar use to be understood. However, within a rich communication, subtle changes can modify the meaning of an utterance. This means that supporting predefined sentence templates, one of the lines we are currently working on, to facilitate the synthesis of utterances, can be too limiting; yet trying to cover every possible subtlety related to the signing might prove to be unfeasible. Thus we will continue the line of supporting predefined sentence templates.

The phonetic approach is the most scalable from the current set of available approaches. Creation of new animated signs consists only of selecting high level instructions and positioning them correctly in time, as our system supports. Because of the coverage of current representations which are language independent, the inclusion of novel instructions and their corresponding rendering code and editing interface is very infrequent. The symbolic approach does not involve retraining neural networks, which requires generating new raw data to train them with and the computationally expensive actual training, nor specific animation knowledge to manually move each bone and facial action unit through several frames each instant of the animation.

In terms of linguistic evaluation, the signs synthesised are bound to some specific information and meaning. Slight changes in the animation might completely alter their meaning. Thus, the system understandability and acceptance by the DHH communities need to be carefully assessed, and this belongs to our future work.

Nonetheless, another group of potential users need to be accounted for in this future assessment. The phonetic approach relies heavily on existing sign repositories which need to be populated by experienced professionals. The tools for such a task, which can come from our system, need to be complete and robust yet with a smooth learning curve that encourages the professionals to keep pushing the repository. In this regard, supporting previously developed notation systems such as SiGML or

HamNoSys might prove advantageous as the professionals might already be familiar with them and their intricacies.

4. Conclusions and Perspectives

This deliverable presented the summative evaluation of the final version of the SignON app and the virtual signer. These are the visually, tangible outcomes of the product which are accessible by the end users. The SignON application meets in a high degree the project KPIs, achieves a substantial degree of usability but falls short in terms of the translation needs of the users (more details on the pipelines will be presented in D4.5). The SignON signing avatar synthesis system achieves a coverage comparable to the most performant systems; it further offers some improvements over them in different aspects and it is open source and extensively documented, unlike comparable systems.

As indicated at the end of Section 2, the work conducted within SignON on Sign Language Machine Translation is quite broad and extensive and it addresses all related aspects. However, we must acknowledge that more research and development are needed to meet user requirements.

References

Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y .Sheikh. Openpose: Realtime multi-person pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **43**(01):172–186, jan 2021.

Ekman P, Friesen W. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press. 1978

Lyke Esselink, Floris Roelofsen, Jakub Dotlacˇil, Shani Mende-Gillings, Maartje de Meulder, Nienke Sijm, and Anika Smeijers. Exploring automatic text-to-sign translation in a healthcare setting. Unpublished, 2022.

Michael Filhol. *Zebedee: a lexical description model for sign language synthesis*. Technical report, Computer Science Laboratory for Mechanics and Engineering Sciences. French National Centre for Scientific Research, 2009.

Michael Filhol, Annick Choisier, and Mohamed Hadjadj. Non-manual features: the right to indifference. In *Proceedings of the 6th Workshop on the Representation and Processing of Sign Language (LREC)*, 05 2014.

John Glauert and Ralph Elliott. Extending the sigma notation; a progress report. In *Proceedings of the Second International Workshop on Sign Language Translation and Avatar Technology*, 2011.

Hanke, T.. HamNoSys -- Representing sign language data in language resources and language processing contexts. *Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Representation and Processing of Sign Languages Workshop (pp. 1-6). Paris: European Language Resources Association. 2004

Alexis Héloir and Michael Kipp. Real-time animation of interactive agents: Specification and realization. *Applied Artificial Intelligence*, **24**:510–529, 07 2010.

Richard Kennaway. Experience with and requirements for a gesture description language for synthetic animation. In Antonio Camurri and Gualtiero Volpe, editors, *Gesture-Based Communication in Human-Computer Interaction*, pages 300–311, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thórisson, and Hannes Vilhja ́lmsson. Towards a common frame- work for multimodal generation: The behavior markup language. In Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin, and Patrick Olivier, editors, *Intelligent Virtual Agents*, pages 205–217, Berlin, Heidelberg, 2006.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for building per- ception pipelines. *CoRR*, abs/1906.08172, 2019.

Birgit Lugrin, Catherine Pelachaud, and David Traum, editors. *The Handbook on Socially Interactive Agents*: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition, volume 37. Association for Computing Machinery, New York, NY, USA, 1 edition, 2021.

Birgit Lugrin, Catherine Pelachaud, and David Traum, editors. *The Handbook on Socially Interactive Agents*: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application, volume 48. Association for Computing Machinery, New York, NY, USA, 1 edition, 2022.

Irene Murtagh. *A linguistically motivated computational framework for irish sign language*. PhD thesis, Trinity College Dublin. School of Linguistic Speech & Comm Sci, 2019.

Lucie Naert, Caroline Larboulette, and Sylvie Gibet. A survey on the animation of signing avatars: From sign representation to utterance synthesis. *Computers & Graphics*, **92**:76–98, 2020.

Carolina Neves, Lúısa Coheur, and Hugo Nicolau. HamNoSyS2SiGML: Translating HamNoSys into SiGML. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6035–6039, Marseille, France, May 2020. European Language Resources Association.

Fabrizio Nunnari, Michael Filhol, and Alexis Heloir. Animating azee descriptions using off- the-shelf ik solvers. In Mayumi Bono, Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, Julie Hochgesang, Jette Kristoffersen, Johanna Mesch, and Yutaka Osugi, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), SL Workshop*, pages 155–162, Paris, France, May 2018. European Language Resources Association (ELRA).

Roland Pfau and Josep Quer. Nonmanuals: their grammatical and prosodic roles, pages 381–402. *Cambridge Language Surveys*. Cambridge University Press, 2010.

Jaume Pozo. *Procedural Mouthing of Synthetic Characters*. Bachelor’s thesis, Universitat Pompeu Fabra, 2022.

June E. Shoup. Phonological aspects of speech recognition. In Wayne A. Lea, editor, *Trends in Speech Recognition*, pages 125–138. Prentice-Hall, Englewood Cliffs, NJ, 1980.

Irene Strasly, Tanya Sebaï, Evelyne Rigot, Valentin Marti, Jesus Manuel Gonzalez, Johanna Gerlach, Hervé Spechbach, and Pierre Bouillon. Le projet babeldr : rendre les informations médicales accessibles en langue des signes de suisse romande (lsf-sr). 2018.

Mani Kumar Tellamekala, Ömer Sümer, Björn W. Schuller, Elisabeth André, Timo Giesbrecht, and Michel Valstar. Are 3d face shapes expressive enough for recognising continuous emotions and action unit intensities?, 2023.

Britt Van Gemert, Richard Cokart, Lyke Esselink, Maartje De Meulder, Nienke Sijm, and Floris Roelofsen. First steps towards a signing avatar for railway travel announcements in the Netherlands. In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology*: The Junction of the Visual and the Textual: Challenges and Perspectives, pages 109–116, Marseille, France, jun 2022. European Language Resources Association.

Rosalee Wolfe and John McDonald. A Survey Of Facial Non Manual Signals Portrayed By Avatar. *Grazer Linguistische Studien*, **93**:161–223, Frühjahr 2021.

Rosalee Wolfe, John McDonald, Ronan Johnson, Ben Sturr, Syd Klinghoffer, Anthony Bonzani, Andrew Alexander, and Nicole Barnekow. Supporting mouthing in signed languages: New innovations and a proposal for future corpus building. In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology*: The Junction of the Visual and the Textual: Challenges and Perspectives, pages 125–130, Marseille, France, June 2022. European Language Resources Association.

ANNEX A: SignON Project's KPIs

Analysis of responses to each of the KPIs in D1.13 from June 2021.

Objectives KPI	Ways to measure	Outcome	Who
Objective 1 - Co-creation workflow and community			
1. At least 75% average user satisfaction rating with the overall operation of the SignON service.	Communication with the stakeholders via co-creation events (e.g. interviews, surveys, round tables, workshops, etc.).	Summative Evaluation (SE) found Average user satisfaction = 60%	MAC
2. Maintain and build on quality engagement of end users in co-creation events.	Participation in focus group sessions (estimate 30-40 per year); follow up activities and tests with interested target language cohorts.	Partner report	
Objective 2 - SignOn Framework and Mobile application			
3. Application and Service at TRL7 (System prototype demonstration in operational environment) running on standard modern mobile devices without the need for additional special equipment.	A working prototype for Android and iOS freely available for users to download, install and exploit.	Done	
4. Respond with user-acceptable accuracy for 75% of users.	Anonymous surveys using structured evaluation tasks, as per D1.4.	SE found Average user satisfaction = 55%	MAC

5. The SignON Framework, Application and cloud platform, specification, implementation documentation and source code will be published free and open-source.	A publicly-accessible repository containing the code and documentation for the open-source SignON framework is available. As any free (or libre) software, the publicly available and accessible SignON framework will be distributed with no warranty or quality promises as these would depend on use-cases, data, resources, etc.; quality will be measured according to the other KPIs described in this document and presented via planned dissemination processes.	Done	FINCON
Objective 3 - Automated recognition and understanding of sign and oral language input.			
6. At least 15% relative improvement (of Sign language recognition, SLR) over current[10] SOTA on internationally recognised benchmarks.	At the moment of compiling this project, SOTA in SLR is achieved at a WER of 23%. The development of SLR in SignON will reach better performance on these benchmarks. The performance difference will also be measurable in the downstream translation quality.	Lab test	
7. ASR performance of maximum 7% word error rate (WER) for hearing speakers, and below 30% WER for deaf and hard of hearing speakers.	At the end of the 3rd year SignON will reach below 7% WER for hearing speakers and below 30% WER for DHH speakers calculated on relevant test sets for the SignON applications.	Lab test	
8. Text normalisation leading to 15% improved NLU results over unnormalised text.	Comparing quality metrics computed on the same test sets with or without text normalisation.	Lab test	
9. Improvements on general downstream performance for each added component on gold standard data.	Comparing quality metrics computed on the same test sets for each downstream task.	Lab test	

[10] The proposal, where these KPIs are originally defined, is based on 2018 / 2019 SOTA

Objective 4 - Language Independent Meaning Representation			
10. Capability to represent full meaning in meaning transfer between languages, according to human evaluation.	Perform human evaluation such as direct assessment, ranking and/or A/B Testing.	SE found that overall users considered App to be Good at 53%E	MAC
11. Accuracy of the entire processing pipeline a. Analysis from input to representation b. Synthesis from representation to output	Evaluate recognition and analysis based on commonly accepted test sets. Evaluate decoding of text from the InterL, synthesis of audio and 3D virtual character based on commonly accepted test sets.	Lab tests	
12. Robustness against noisy input due to inaccurate NLU.	Evaluate the quality of decoded text (from the InterL), the audio and the 3D virtual character based on artificially created test sets with missing or erroneous information.	Lab tests	
Objective 5 - Sign, speech and text synthesis			
13. Robustness of the tools in case of wrong (or noisy)input.	Evaluate the quality of decoded text (from the InterL), the audio and the 3D virtual character based on artificially created test sets with missing or erroneous information.	Lab tests	
14. Language richness, expressivity and intelligibility judged by human assessor.	1. Conduct manual (human) evaluation of the decoded text such as A/B Testing or Direct Assessment comparing different types of text. 2. Construct a survey and involve human linguists and/or professional translators to indicate the lexical richness, expressivity and intelligibility.	1.Lab Test 2. In the SE, users rated it Fair to Good, at 45%	MAC
15. Scalability measured as the software and hardware utilisation to synthesis requests.	Invoke the system to respond to parallel requests and measure response time. Conduct multiple iterations of the test with	Lab tests	

	incrementally large sets of parallel requests. Conduct load tests.		
16. Reusability measured as the post-deployment efforts for extending to new vocabulary, languages and use-cases.	Measure the efforts of SignON partners and users in terms of human-hours dedicated to the corresponding tasks.	Partner report	
Objective 6 - Wide-range of supported languages and extensibility of the framework			
17. Translation and conversion between ISL, BSL, VGT, NGT and LSE as well as English, Irish, Dutch and Spanish oral languages has been showcased and accepted by DHH as well as hearing users.	Communication with and evaluation by stakeholders via co-creation events (e.g. interviews, surveys, round tables, workshops, etc.).	Co-creation events report.	
18. The SignON framework is trained automatically to support at least one additional sign language and one additional spoken language.	Communication with the stakeholders via co-creation events (e.g. interviews, surveys, round tables, workshops, etc.).	Partner report	

WPs' based KPI	Ways to measure	Outcome
WP1: Co-creation and User Response		

Established regular communication with end users, keeping them informed on progress while updating user requirements, technical user requirements and working towards a community within/outside the SignON project	1. Engagement on the part of end-users on social media, in focus groups/interviews (approx 30-40 annually, across the SL sessions), and in responding to requests for feedback on iterations of the product. 2. 2 workshops with 20-40 international participants in each (1 after each major release of the SignON service and framework). For example, there are already planned: one workshop, to be organised by DCU and TCD in Ireland; another one organised by EUD in Switzerland; AT4SSL workshop (https://sites.google.com/tilburguniversity.edu/at4svl2021/), organised by TiU and other consortium partners which aims to bring scientific communities together. 3. 4 round table sessions (1 in year 1, 1 in year 2, 2 in year 3). One in Ireland, one in Spain, one in Belgium, one in Switzerland. 4. Social media, live webinars/ in-person presentations, posts and publications (academic or non-scientific) to facilitate an open communication between technical partners and end users groups. Also, please see Table 1 for our dissemination KPIs. 5. Transparently incorporate feedback into product over the life of SignON - report back on the changes that have been made to the communities who have engaged. This builds trust and demonstrates our co-creation approach in a visible way. 6. High user acceptance (minimum75%).	Report on each
WP2: SignON Service and Mobile App		
The SignON Mobile App and Framework	SignON Mobile App and Framework providing translation/conversions to users: 1. Within 2 seconds – with a maximum of 5 seconds for SL-to-SL translations, 2. Between any combination of the following sign, spoken and text languages: ○ Flemish Sign Language (VGT), Sign Language of the Netherlands (NGT), Irish Sign Language (ISL), British Sign Language (BSL) and Spanish Sign Language (LSE). ○ English, Irish, Dutch and Spanish spoken and text languages. 3. On standard modern Android and Apple/IOS mobile phones and tablets.	1. SE. 2. SE. 3. SE.
WP3: Source Message Recognition, Analysis and Understanding		
Accuracy and robustness of Sign Language Recognition (SLR)	Accuracy of the recognition, robustness 1. Accuracy of the recognition: Word Error Rate. 2. SLR approach is robust if it works for different sign languages, and if the performance of translation using	Lab tests

	this SLR method is sufficient (see WP4 for translation metrics).	
High performance of the Sign Language Recognition component	1. In terms of isolated sign recognition: Accuracy. 2. In terms of continuous sign language recognition (from video to an ordered sequence of glosses): Word Error Rate.	Lab tests
Enough data is available for the purposes of the different research activities: ● Making sign language corpora available to the research groups, ● Ensuring that GDPR and IPR obligations are upheld, ● Setting up infrastructure so consortium partners can upload/deposit /download their data sets.	1. Corpora for each SL envisaged in the project are available to the consortium partners. 2. Consortium partners can upload and download the necessary corpora as per terms of GA and additional licensing agreements that are in place for third-party data	Lab tests
Automatic Speech Recognition (speech to text modules)	1. Low Word Error Rates (WER). 2. Improvement of WER. 3. Innovative ASR designs (scientific innovation). 4. Publishable Results.	Lab tests
Effectiveness of Natural Language Understanding	1. F1 scores (intent, or specific NLP tasks like named entity recognition, part-of-speech tagging, tokenisation). 2. Accuracy % based purely on intent matching. 3. UD treebank accuracy %. 4. MT evaluation metrics - BLEU, TER, YiSi, COMET, significance testing based on individual experimental settings and methodologies. 5. Human evaluation: for simplification comparing texts/sentences for simplicity, grammaticality and meaning preservation comparing different SoA systems and SignON. 6. For automatic simplification, current metrics will be used such as SARI and BLEU. Measures related to readability will also be incorporated.	Lab tests
Pipelines for language understanding: POS, Named Entity Recognition, Parsing, Coreference, etc.	1. Compute automatic metrics such as precision, recall, accuracy, etc. on benchmarking dataset 2. Compute metrics specifics to the task (based on benchmarking dataset).	Lab Tests

Analysis and linguistic description of sign languages (SLs) relating to the project. Completion of a broad linguistic analysis on a phased basis working through the specified SLs, Irish Sign Language (ISL), Dutch Sign Language (NGT), Flemish Sign Language (VGT), Spanish Sign Language (LSE) and British Sign Language (BSL)	Feedback on linguistic quality of avatar communication when synthesising a specific SL.	In the SE – users rated it Fair at 39%.
WP4: Transfer and InterLingual Representations		
Tools, components	Since MT will be addressed by other partners I can suggest simplification metrics for the simplification work to be carried out. But human evaluation should also be considered.	Partner Report
Machine translation	Compute BLEU, TER and SARI based on commonly accepted benchmark datasets. 1. Expected output for dialog MT is above 40 BLEU and below 50 TER; we consider acceptable quality if the BLEU is above 35 and TER below 60. 2. For text simplification employ SARI, 3. Generation Time, Training Time (for updating the model), 4. Energy Consumption, Model size (Number of weights).	Lab tests
Successful combination of InterL-E and InterL-S into a hybrid InterL.	Compare the overall performance based on only InterL-E, InterL-S or the hybrid InterL.	Lab tests
WP5: Target Message Synthesis		
The virtual character is of a quality of rigging to support Manual and Non Manual features.	The success of the presented tasks is defined by the approachability and validation of the stakeholders so our system produces a 3D virtual signer whose lexicon is accepted by the deaf community.	Partner report
The virtual character is modifiable by users.		Partner report
An interactive system for learning from user generated signed content is available to users.		Partner report

Develop a planner for the 3D virtual character that supports Sign_A input representations.		Partner report
A high visual and comprehension quality of the rendering.		Partner report
Development and implementation of a Sign_A framework to produce Sign_A + RRG logical structures (to drive SL synthesis): Definition of a lexicon architecture capable of catering for linguistic phenomena associated with SLs and development of Sign_A + RRG logical structures, which will be used to generate any of the five associated SLs.	Ability of the lexicon and logical structures to accommodate associated SLs, which in turn will be used to generate SL as an output.[11]	Partner Report
WP6: Communication, Dissemination and Exploitation		
Communication and dissemination	The communication plan goals, detailed in Table 1 below, are achieved.	Partner report
Exploitation Plan	Documented viable SignON sustainable Exploitation, Innovation and IPR plans, agreed by all Partners.	D6.7
WP7: Coordination and Management		
Completion of Tasks	All project deliverables delivered on time and to an acceptable quality.	Final Report
Milestones Reached	All milestones achieved and reported on time.	Final Report
Reports Submitted	All necessary reporting compiled on time and as expected.	Final Report
WP9: Ethics		

[11] The quality of the output will depend on the data available to populate the SL lexicon, this data will be drawn from within the interlingua.

Ethics overview for SignON project	1. Embed a Deaf studies lens into SignON's ethical processes and practices - outlining our expectations in our ethics and data management guidance as appropriate, and building in key principles (as per D9.1) into our practices. Success is also measured at project level - by the perception that stakeholders in the Deaf, Hard of Hearing and interpreting communities have of SignON practices and outputs as being ethical. 2. Success of applications for ethical approval from our partner institutions.	1. Partner report 2. Done

ANNEX B: Detailed Summative Evaluation Survey Feedback

During October/November 2023, the SignON Partners and their users installed, reviewed and provided structured Cognitive Walkthrough summative evaluation feedback on the SignON Mobile App V3.0 on their own Android and Apple phones as described in Section 2.1 using the script at [SignON Mobile App V3.0 Evaluation (google.com)](#) as described in Section 2.1. The resulting detailed results and very rich feedback on users' technical requirements are presented in this Annex.

The SignON App Approved Users Testing Group of 68 people, completed 22 structured feedback walkthroughs with users distributed as follows:

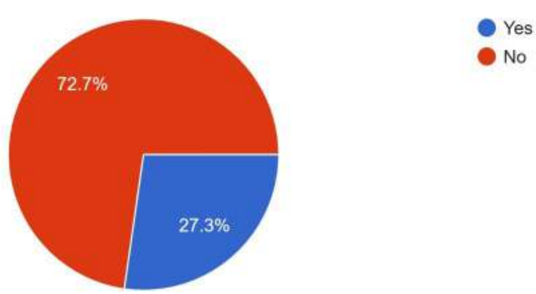
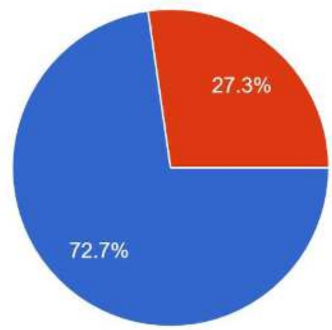
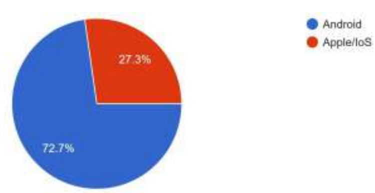
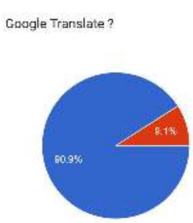
Regular Users of Sign Language	**In a partner of the SignON project**
	

Figure Distribution of Users providing Feedback

So, the received responses represented 33% of the Approved SignON Users. Of those 27% were non-partner external users, giving a reasonable balance between consortium and external users. However, only 14% of the responses came from SL users. While this means that SL Users were represented, a higher percentage would have been preferred.

73% of the respondents use Android Phones, so it was worthwhile providing both Android and Apple/IoS versions of the App, but with the Android version first.





90% of respondents use Google Translate, so they are all familiar with the concept of a MT mobile App, and can thus evaluate and judge the SignON SLMT App.

Users indicated that their main use of the App would be:

- *Communicating with deaf friends, colleagues, and clients*
- *Translating to and from other sign languages*
- *Learning sign language*
- *Translating sign language*
- *Conversing with deaf people who use ISL*
- *Testing speech recognition (ASR) and sign language recognition (SLR) models*

So users' main use of the app would be to communicate with deaf clients, friends, signers, and colleagues in various everyday situations. They also mentioned using the app for personal communication, including speech-to-text interactions, and learning new sign languages. These are in line with the SignON Mobile App's objectives and expectations in the DoA.

B.1 Use of the SignON SLMT App

Using the tasks form in Section 5.1, the users generated the following average difficulty scores and technical requirements feedback for using the SignON Mobile App's main functions:

Use of the SignON SLMT App / Average Difficulty[12]	SL users	Other users	Overall Average
1. Run the SignON App & check its Information & Preferences screens	1.3	1.4	1.36
2. Text or Audio record & translate a Message	2.0	1.63	1.73
3. Display SignON's Avatar Signing your Message	3.3	2.63	2.82
4. Record & translate a Sign Language Message	3.0	3.13	3.09
OVERALL[13]	**2.4**	**2.2**	**2.3**

Table Users' feedback on using the SignON Mobile App V3.0

[12] 1. Irrelevant: the problem does not need to be solved, but it could be improved.
2. Low: the problem causes uncertainties, but it does not prevent the task completion.
3. Medium: the problem can slow down the task execution, but does not prevent task completion.
4. High: the problem causes frustration and prevents the task completion.

[13] This ordinal scale measured the central tendency of users' feedback, similar to the Horizon Europe evaluation scoring, as defined at ef_he-ria-ia_en.pdf (europa.eu). So these overall averages are just descriptive indications that provide a practical summary of this feedback , see https://www.scribbr.com/statistics/ordinal-data/

The overall severity score of 2.3 out of 4.0 (i.e. "Low": *the problem causes uncertainties but it does not prevent the task completion*) and with no statistically significant (P<0.05) differences between SL and Non-SL users on all 4 main functions, indicates that a usable prototype app has been developed and is a good foundation for the future evolution of the SignON service, for all users[14].

Users overall rated the difficulty of using the App as 2.3 (low), with no significant difference between SL users (2.4) and Other users (2.2). Compared to the previous versions of the App: the D1.9 pre-SL App tasks were 10% lower at 1.9 (low), 1.8 and 1.8, respectively, and difficulties with the initial fast prototype App of D1.4 were rated 38% lower at 1.3 (irrelevant), 1.4 and 1.0, respectively. These increasing difficulty levels also probably reflect that the performance of the SL functions need to be improved.

User Feedback and Suggestions:

Did you have problems ? Is the information text clear ? Are your preference options clear ?	
SL Users	1. *I don't have problems* 2. *App is not super intuitive to use*
Other Users	1. *No problems and the text is clear.* 2. *no problems, and the options are clear once I tried the various buttons. Nice and intuitive to use.* 3. *Both information text and preference options are clear* 4. *Could not identify any "Translate windows Only the home, settings and information windows are available.* 5. *I did get a bit confused with the phrase "Go Arrow", as I had already played the voice recording of my text. That may just be me, but perhaps a picture of the Go arrow?* 6. *Infoscreen:* *Dutch interface: link to instruction leads to instruction sheet in English* *There is no information about what happens to your input recordings* 7. *Settings:* *In nightview it is hard to visually distinguish setting items and the options. It iall black white.* *There is no setting for text input language* *Is text not an input mode?* *Select spoken language also determines the language of text output.* *Select spoken language: It is unclear for a user what the difference between standard speech and Natural speech is* 8. *Home:* *Uploading a short mp4 video file takes very long, even with good wifi* *If I use the Google speech recognizer then it recognizes well but after that I cannot*

[14] Note: the SE process assessed how useful the SignON SLMT prototype mobile app was for the user in a typical situation where they might use it. So to be concrete, the communication situation storyline was set in the Netherlands, and most respondents would not be proficient in the local sign and spoken languages. So the results indicate users' subjective perception and utility of the App to them, and not objectively assess the quality of the App's outputs as such. This latter is scientifically addressed in Section 3, and in later deliverables of the project.

	use the SignON recognizer anymore. The microphone icon has been replaced by a dustbin icon

Table Users' Feedback on the SignON Mobile App Tasks

Users provided excellent and rich feedback on various aspects of the app's user interface, functionality, and performance (see later). They mentioned minor difficulty with the UI but found it intuitive overall. Overall, there were mixed opinions on the performance of the app's features.

This feedback mirrors and is consistent with much of the users' evaluation and feedback on the EASIER SL MT app, where "*Participants brought up issues with several aspects of the app, and provided specific feedback on solutions, demonstrating a clear vision of what they wanted the app to look like. Major issues concerned progression through the app, setting up translation parameters, visual choices in icons and images, and general useability. Participants also were able to identify language specific localization issues. Overall, the app feedback provides a clear checklist of issues for developers to work through to increase useability and tailor the app towards user needs"*[15].

The SignON feedback texts identify many issues with the prototype App V3.0 that will now be addressed. They also provide a wealth of detailed input to the future user technical requirements of the SignON SLMT Mobile App, and are reflected in the tabulated user technical requirements of the SignON Mobile App in Section 5.3.

B.2 System Usability Score (SUS) Results

After each trial was completed the form in Section 5.1 the users generated the following SUS results for the SignON Mobile App.

SUS Question & Scores **1 – Strongly Disagree – 5 Strongly Agree**	**SL Users**	**Other Users**	**Overall Average**
1. I think that I would like to use this tool frequently.	2.3	3.3	3.0
2. I found this tool unnecessarily complex.	2.7	3.3	3.0
3. I thought this tool was easy to use.	3.3	2.1	2.3
4. I think that I would need assistance to be able to use this tool.	2.3	3.5	3.5

[15] EASIER D1.3 "Report on Interim Evaluation Study", December 2022, EASIER-D1.3-Report-on-interim-evaluation-study_v1.0_final.pdf (project-easier.eu)

5. I found the various functions in this tool were well integrated.	2.3	2.5	2.5
6. I thought there was too much inconsistency in this tool.	3.0	3.9	3.5
7. I would imagine that most people would learn to use this tool very quickly	3.0	2.4	2.5
8. I found this tool very cumbersome/awkward to use.	3.0	3.5	3.4
9. I felt very confident using this tool.	2.0	2.8	2.8
10. I needed to learn a lot of things before I could get going with this tool.	2.3	3.0	2.7
Overall SUS Rating[16]	**49.2**	**39.7**	**42.7**

Table SignON Mobile App V3.0 SUS results

For this V3.0 App, the users rated the SignON Mobile App at 42.7 overall, which is below the SUS threshold of acceptability (68). While SL users rated it at 49.2, non-signers were more critical of its usability at 39.7. The D1.9 pre-SL App tasks was rated 38% higher at 62.6 (overall), 63.0 (signers) and 62.5 (other), while the initial fast prototype App of D1.4 was rated 77% higher at 80.2 (overall), 77.1 (signers) and 83.8 (other), which were above the acceptability threshold. These diminishing SUS levels probably reflect that while the App format is good, the performance of the SL MT has disappointed.

Comments, Suggestions & Feedback

Users	Any problems? Which combination of speech input & output options are best for you ?	How useful is the Avatar's signing accuracy?	Comment on the accuracy of the Sign Language to text translation ?	Your Comments and Suggestions

[16] The Overall SUS Rating is calculated for positive aspects (odd questions) as score assigned by the user minus 1, and for negative aspects (even questions) as 5 minus score assigned by the user, with summed total multiplied by 2.5 to give a result from 0.0 to 100.0. A Rating of 68.0 is considered the threshold to indicate acceptable usability. See https://webdesignviews.com/measuring-usability-with-system-usability-scale-sus

SL Users	*1. I don't like that the old speech recordings remain there if you don't delete them - it's not super easy to tell which is which without playing them which is cumbersome. However the audio output is not bad - the Spanish audio is good - the English is also not bad. My Dutch is less good but it also sounds okay (a bit more choppy than the other two).* *2. When I look at the translation between written Dutch and output as written Dutch, it is perfectly translated. But, if I look at the avatar, it only signs one part (The train goes to Nijmegen: train is translated), but not correctly.*	*1. Really bad - also there's no non manuals which make it impossible to phrase a question. Sometimes the avatar just mouths a word which is not acceptable. Plus sometimes it only translates one word of the sentence which does not work to convey meaning - and sometimes the one sign that is translated is an inaccurate one that did not appear in the original message.* *2. It is not accuracy enough. It shows only partly the translation, without a correct signs. It is not enough to get the content out of it.* *3. I don't see an avatar. I get the message that 'this function is only available in Dutch Sign Language', but I don't even see an avatar in Dutch Sign Language*[17]*.*	*1. This is really an impossible question for me, because I don't know NGT. I think it's also totally ridiculous to ask someone to copy the avatar's signing and use that as NGT input and then evaluate the output intelligibility -- the avatar output is certainly not good enough and if you don't know NGT you are not qualified to give a rating of translation intelligibility for a translation involving NGT, this just makes no sense* *2. I only got an error as soon as I tried to sign to text.(that the app could not recognize the signs)* *3. I'm not able to select the source our output language to be Dutch Sign Language*	*1. I suggest a status bar for the avatar video so we can see if it's actually playing or not, and easily pause and rewind and repeat.*

[17] Note this is the user feedback text. The correct name is "Sign Language of the Netherlands", not "Dutch Sign Language".

Other Users	1. No, all of the speech I/O options are very similar. I would have a slight preference for the natural speech output, but the defaults are fine. 2. No problems. The default speech recognition was easier to use as it did have the slow upload. For the speech output both options were fine, but the natural option was slightly easier to listen to. 3. Speech to text is best. 4. There is no output. THE "GO" ICON DOES NOT APPEAR 5. Default Recognition with Natural Speech 6. No, I think I like speech input and output best. 7. Translations into the Irish language are not great but I imagine that is due to the lack of Irish language data. 8. See the extensive feedback on the previous page	1. The avatar's movements and quality are great. i cant comment on the accuracy, as I dont know sign language. The translated output text is good to have but it looks odd, with most words capitalised and includes strange words such as "PT-1". Also the replay option is really useful, but the icon was half hidden. 2. Im not sure as I do not understand sign language. However her movements looked clear and easily seen. The white text also helped. I was really impressed and surprised that I could swipe the avatar in 3 dimensions. Perhaps thats useful for reading the signs. I could not see the white replay button, so I pressed the Go arrow, which uploaded my text again. However the avatar seemed to have different signs - so I tried it a third time and she repeated the 2nd signs. Not sure what happened for the first time. Weakens my confidence in her accuracy though 3. Not sure as I am unable to use sign . language. Found that it was too easy to accidentally move the avatar. Would have thought it better just to see the avatar from the front. Dont see why I would want to see the avatars back 4. There is no media output option on the home screen.mThere is no Avatar	1. I dont know sign language, so I tried the hint of copying what the avatar did,, but the translated text was totally different from my very simple signed hallo ! As it could be my signing I scored it 3. 2. I dont know sign language so I used a very simple gestures, however I got back 3 lines of text with a very cryptic message ! It also took about 30 seconds to come back - which is too long for a conversation. I repeated the upload and got the same message so it seems to be consistent. All in all it would probably be easier to converse using short text sentences - hence my score ! 3. Not accurate 4. I can select Sutch sign language. I can record and save the video.There is a "review video option. There is a "text translation" option, which seems to require me to write text. When I write text I can save and upload it. After the upload, the app returns to the initial screen with no other options. There are no output avatars or options, therefore I cannot assess any output. 5. Not Applicable 6. I got an error message so couldn't complete this task. Also, I am confused by what is meant be task 2 & 3? 7. I could not do this test	1. The app was very easy to use, but I am not sure how much it would help me understand what a deaf person would be signing. However it would really help me to with its speech to sign language for the conversation on the train ! But its biggest benefit would probably be its speech to Dutch text, and Dutch text to spoken English to enable our conversation on the train. So its a very useful first step. 2. The text and voice translations, and user interface are all great. The sign language output seems to be quite good, but the sign language input is not useful. 3. It would be good to have a menu of statements or questions that I could pick from to get translated by the app 4. Looking forward to trying out a more completed version. 5. App is way to use but is limited by the sign language outputs 6. I think if I played with this it would be easy but unless I read the instructions incorrectly, I couldn't get the avatar/sign language function? 7. Infoscreen: Dutch interface: link to instruction leads to instruction sheet in English There is no information about what happens to

		option. There is no "Go" arrow icon *5. Not Applicable* *6. It didn't work for me. It said that Dutch is the only language currently in use, so I typed in a Dutch word. However, I couldn't even get the avatar on my screen.it misses large parts of the message*		*your input recordings* *Settings:* *In nightview it is hard to visually distinguish setting items and the options. It iall black white.* *There is no setting for text input language* *Is text not an input mode?* *Select spoken language also determines the language of text output. Select spoken language: It is unclear for a user what the difference between standard speech and Natural speech is* *Home:* *Uploading a short mp4 video file takes very long, even with good wifi* *If I use the Google speech recognizer then it recognizes well but after that I cannot use the SignON recognizer anymore. The microphone icon has been replaced by a dustbin icon*

In summary, users felt that the prototype App shows promise with clear and fluid avatar movements, but SL recognition is slow, and accuracy is uncertain. Improvements are needed in the app's interface, including simplifying media output modes, and enhancing clarity in accessing recorded text or audio. Users encountered difficulties such as slow translation speed, unclear indication of when the avatar finishes signing, limited visibility of the avatar, and potential lack of facial expressions. The video upload times and translation accuracy need to be improved.

B.3 Overall Users' Feedback

Overall Feedback on the SignON SLMT APP *1 Very Bad, 2 Bad, 3 Poor, 4 Fair, 5 Good, 6 Superior, 7 Excellent, 8 Outstanding, 9 Superb, 10 Very Good*	**SL Users**	**Other Users**	**All Users**	**Comments**
1. How satisfied are you with the overall operation of the SignON service ?	43%	66%	60%	*Overall Very Good, but Fair for SL Users, both short of the 75% target*
2. How useful is the SignON translation accuracy ?	33%	64%	55%	*Good to Very Good Overall, Poor for SL users, both short of the 75% target*
3. How well generally does the SignON translation get the meaning of the message across ?	30%	61%	53%	*Good+ Overall, but Poor for SL users*
4. How would you rate the linguistic quality of the SignON Avatar's communications?	27%	51%	45%	*Poor for SL Users, but Overall Fair to Good - reasonable for a TRL6 prototype*
5. Would this App would be useful for you to better communicate & manage your work?	20%	46%	39%	*Bad for SL Users, but Overall Fair to Good - reasonable for a TRL6 prototype*
6. How likely would you recommend this App to a colleague?	53%	53%	53%	*Fair to Good, which is reasonable for a TRL6 prototype*
AVERAGE	**34%**	**57%**	**51%**	*Overall Good, but Poor for SL Users !*

There was a significant ($P<0.05$) difference between average ratings by SL and Non-SL users on questions 1 to 5 above. However, average responses were similar for SL and Non-SL users on whether they might recommend this App to a colleague.

Users	**Main Positive aspects of the App,**	**Main Negative aspects of the App,?**
SL Users	*1. The main positive is that the settings are not difficult to find or navigate* *2. It is clear how to do a translation.* *3. The capability of getting into interaction with both of the societies and different languages.*	*1. Main negative is that the avatar does not sign well, and the translation misses mostly all the information (at best, at worst it misses all information, or adds information that does not belong).* *2. The specifications of the avatar is currently not yet that good for a clean understanding* *3. I can't comment on the avatar our sign language translation, since I haven't seen any.*

Other Users	*1. The big positive is its very useful speech to text and sign language for a choice of languages* *2. The App is very easy and nice to use.* *3. Positives are the text and speech translations and sign language output* *4. Speech to text and text to speech is very good.* *5. App is great and services are very fast 6. It would be excellent to integrate such services into social media* *7. Positive: when working it is quick and clear (audio/text /voice only)*	*1. The big negative is the sign language translations which are slow and highly inaccurate.* *2. The almost useless sign language input,* *3. Slow response of the service* *4. Needs more work on translating to sign language.* *5. Thought the avatar should not be so easy to accidentally move on the screen.* *6. Thought it would be useful to have a drop down menu to choose from for translation* *7. As indicated, I can only input, either speech, video or text. No output is available.* *8. The Avatar/Signing function did not work for me, perhaps the instructions were hard to understand?*

The feedback on the app is mixed. Some users find the app relatively intuitive and see great potential in it, while others find it clunky, slow, and in need of improvement. The interface is generally considered okay, but there are suggestions for optimization and better design. Users express a desire for better SL translation algorithms, SLR, SLS, faster processing, and more accurate translations.

The average SUS rating for the app is lower compared to previous versions, potentially indicating a decline in user satisfaction. However, there is still recognition of the app's potential and appreciation for its current features, such as speech and text translations.

Overall, the SignON app is seen as a useful tool for speech to text and text to speech. However, the SL translations are slow and inaccurate. The app is easy to use, but the avatar's signing is not good. The app is a good first step, but it needs more work on translating to SL. Some other key points from the feedback:

- The main positive is that the settings are not difficult to find or navigate.
- It is clear how to do a translation.
- The text and speech translations and sign language output are good.
- App is great and services are very fast. It would be excellent to integrate such services into social media.
- When working it is quick and clear (audio/text /voice only).
- The big negative is the SL translations which are slow and highly inaccurate.
- The SL input is almost useless.

- The response of the service is slow.
- The avatar should not be so easy to accidentally move on the screen.
- It would be useful to have a drop down menu to choose from for translation.
- The Avatar/Signing function did not work for some users.

Users found that the app has positive aspects related to ease of use, but the accuracy and effectiveness of the SL translation and the avatar's signing accuracy were criticised by several users. Some users also had issues with the app's instructions and interface. Improvements in the accuracy and reliability of SL translation are required.

Based on the user feedback provided, the following are the main points:

Positive Aspects of the App:	Negative Aspects of the App:
• Easy to use & navigate • Clear how to do a translation • Very useful speech to text & sign language for a choice of languages • Useful speech-to-text & sign language support for multiple languages. • Quick & clear when working (audio/text/voice only)	• Poor signing accuracy of the avatar. • Inaccurate & incomplete sign language translations. • Limited or ineffective sign language input. • Slow response from the service. • Avatar/Signing function did not work for some users