Sign Language Avatars: Animation and Comprehensibility

Michael Kipp, Alexis Heloir, and Quan Nguyen

DFKI
Embodied Agents Research Group
Saarbrücken, Germany
firstname.lastname@dfki.de

Abstract. Many deaf people have significant reading problems. Written content, e.g. on internet pages, is therefore not fully accessible for them. Embodied agents have the potential to communicate in the native language of this cultural group: sign language. However, state-of-the-art systems have limited comprehensibility and standard evaluation methods are missing. In this paper, we present methods and discuss challenges for the creation and evaluation of a signing avatar. We extended the existing EMBR character animation system[1] with prerequisite functionality, created a gloss-based animation tool and developed a cyclic content creation workflow with the help of two deaf sign language experts. For evaluation, we introduce *delta testing*, a novel way of assessing comprehensibility by comparing avatars with human signers. While our system reached state-of-the-art comprehensibility in a short development time we argue that future research needs to focus on nonmanual aspects and prosody to reach the comprehensibility levels of human signers.

Keywords: Accessible interfaces, virtual characters, sign language synthesis.

1 Introduction

"Why do deaf people need signing avatars on internet pages? They can *read*, can't they?" To motivate the concept of signing avatars, we have to give a brief introduction on the culture and language of the deaf. Most deaf people communicate in a sign language. Every country has its own specific sign language and each sign language is a proper language in all its complexity [28] and is fundamentally different from a spoken language. Therefore, a German deaf person's native language is German Sign Language (Deutsche Gebärdensprache, DGS) while (spoken) German is only the *second* language. In fact, it is a particularly hard-to-learn second language for deaf individuals: it must be learnt based only on a set of written symbols and based on observations of highly ambiguous mouth patterns, without any auditory cues – an almost impossible task. As a

[1] http://embots.dfki.de/EMBR

H. Högni Vilhjálmsson et al. (Eds.): IVA 2011, LNAI 6895, pp. 113–126, 2011.

consequence, many deaf pupils leave school with significant writing and reading problems[2].

To make written material like internet pages more accessible to deaf users, prerecorded *videos* of human signers are used. However, a video's content cannot be modified after production which makes it impossible to use them in dynamic or interactive scenarios (e.g. train station announcements or question answering). Moreover, production cost is high, appearance parameters cannot be adjusted (clothes, gender, lighting) and videos cannot be anonymized. Therefore, *signing avatars* could complement the current range of human signer videos. With intuitive tools, production cost could be low, post-production adjustments are easily done (even automatically or at runtime), and the identity of the content producer is not disclosed. Sign language avatars could be used for the automatic translation of web pages, interactive e-learning applications, sign language lexicon visualization or simple train/flight announcement services (cf. [1,9]). However, creating signing avatars involves multiple challenges, ranging from content representation, since a universal writing system for sign language does not exist, to realizing a comprehensible animation. Sign language is a highly multichannel/multimodal language where hands/arms, the face and the whole body must be synchronized on various levels. Therefore, state-of-the-art avatars reach rather low comprehensibility levels of 58-62%, with a single study reporting 71% [13]. The diversity of evaluation methods and the variance in test material selection also makes it difficult to conclusively compare results.

To investigate the potentials of signing avatars for the internet, the German Federal Ministry of Labour and Social Affairs (Bundesministerium für Arbeit und Soziales, BMAS) commissioned us to investigate the technical feasibility of signing avatars and the acceptance in the German deaf community [18]. In this paper, we focus on the technical feasibility aspect of this study, consisting of two major parts. First, to explore sign language synthesis we created a signing avatar including necessary tools, an animation workflow and the identification of core challenges. Second, to explore reliable evaluation we developed a novel way to assess comprehensibility. In summary, we consider the following to be our main contributions to the research community:

– Showing how to transform a general-purpose avatar [4] to a sign language avatar, including necessary tools and workflow
– Identifying important challenges and their relative importance
– Introducing *delta testing* as a novel comprehensibility testing method that compares avatars with human signers

In the following, we survey related work (Sec. 2) before presenting our avatar extensions (Sec. 3), our animation technology (Sec. 4) and evaluation method (Sec. 5). We conclude with a summary and future work (Sec. 6).

[2] There have been several studies on deaf pupils' literacy levels. For instance, a US study showed "that deaf students around age 18 have a reading level more typical of 10-year-old hearing students" [7].

2 Related Work

Signing avatars are a relatively young research area with two decades of active research and some significant results. A major prerequisite for sign language synthesis is a representation system or *notation*. In *gloss notation* each sign is denoted with a word (gloss) that most closely corresponds to the sign's meaning [12]. For instance, the sign sequence for "What's your name?" would be YOUR NAME WHAT. Gloss notation, however, does not describe how to execute a sign. The same gloss may be executed in various ways due to grammatical modifications and dialect variations. Historically, a milestone notation for the description of how to execute a sign was *Stokoe notation* [28] which formed the basis of modern notation systems like the widely used *HamNoSys*, the Hamburg Notation System for Sign Language [25].

In the research area of signing avatars, one can distinguish approaches along the articulatory vs. concatenative axis [9]. While concatenative approaches piece together prerecorded chunks of human motion, articulatory approaches compute motion on-the-fly based on a sparse specification. Two influential European projects, ViSiCAST and eSIGN, developed technology for signing avatars based on HamNoSys [1,13], transitioning from a concatenative to an articulatory approach, and advancing HamNoSys to SiGML. The resulting avatar technology is called *Animgen* and was used e.g. for the *Virtual Guido* avatar [1]. Drawbacks of Animgen are that it is not open source but also the reliance on HamNoSys, a relatively high-level language with no transparent way to modify animations. To overcome limitations of HamNoSys the LIMSI institute developed *Zebedee* which is based on geometric constraints and allows parametrizable scripts [2]. However, we find the notation hard to read for humans and potentially hard to realize on the animation side. Another avatar project is called *Paula* with a number of interesting results for the synthesis of fingerspelling, nonverbal components and natural pose computation [31]. Both Paula and Zebedee are not yet mature and integrated enough to be used outside their original labs. On the other extreme, researchers have used *commercial* animation software (e.g. *VCom3D*) for their experiments (e.g. [8]). These packages allow convenient timeline-based editing of animations but are otherwise closed toward further external development. Lastly, there are general-purpose avatars that could be made usable for signing purposes. The *Greta* avatar has been used for sign language research but only at an early prototype stage [23]. The *SmartBody* agent is another well-known avatar technology which focuses on coverbal behaviors [30]. Both general-purpose and signing avatars lack a clean animation interface with the notable exception of EMBR [4] which introduced the animation layer in an effort to decouple behavior specification from low-level animation parameters [15].

In most signing avatar projects, the actual comprehensibility of the produced animation by deaf users has been assessed. This is particularly important because most of the experts working in this field are *not* native speakers of sign language. Most evaluation studies establish a pure sign language environment (instructions and supervision by a native signer) and grant a dedicated warm-up time to get used to the avatar [27,13,11]. For questionnaires, it has been recommended to rely on non-written content like pictograms. In terms of assessment methodology, the subjective rating of understanding by the participant him/herself turns out to be highly unreliable [11]. Instead, outside judgements by experts are taken, based on questions about the content of the communicated [27,13]. Here, mere imitation without understanding may be a problem. Also, asking dedicated questions may give part of the answer away, especially in sign language. [10] made a multiple choice test where similar spatial arrangements were shown. This method may not always be feasible, especially for more complex/abstract utterances, and requires careful decisions on what to ask and how to formulate the options. A more general challenge is to define a control condition, i.e. what is the avatar's signing compared against? [10] suggested *Signed English* (SE) as a control condition. SE is an artificial language that translates the words of spoken English in a one-to-one mapping to signs. Since Signed English and Sign Language are two distinct languages, the former sometimes even harder to understand than the latter, we do not deem this a good option. Instead, we suggest to use the comprehensibility of the human signer as the control condition. Moreover, we suggest to circumvent the theoretical problem of defining optimal understanding by using relative measures (e.g. word/sign counts).

3 Avatar Extensions

In this section we describe what changes to a general-purpose avatar are necessary for sign language synthesis. We decided to use the EMBR [4,15] character animation engine because it offers a high degree of control over the animation and is publicly available[3].

EMBR introduces a declarative layer of abstraction around its animation facilities. Based on the notion of a generalized key pose, arbitrary animation sequences can be specified and edited without resorting to programming. The EMBR animation system has grown out of research on coverbal gesture production [22,16,5] and lacked a number of necessary features for sign language production. These are mainly: range of hand shapes, upper body control, mouth control and gaze control.

In sign language, *hand shape* is a highly meaningful feature, whereas for conversational agents, a very sparse set of 8–10 is sufficient. Hence, we had to implement 50 new hand shapes, including the complete *finger alphabet* (27 hand shapes for the letters A to Z) and the ASL *classifier hand shapes*. Also, upper body control is necessary, like raising the shoulders, and therefore we added IK-based spine controls. Since, during signing, the hands move in a wider space compared to coverbal gesture we relaxed shoulder joint restrictions to make this possible. Also, *facial expression* is more expressive than in verbal communication which made us increase the upper limit of facial expression intensity for our morph targets.

To animate *mouthing*, i.e. the lip movement of words that give a definite cue to the meaning of the manual sign, we used the viseme generation capabilities

[3] EMBR has been released as an open-source software under the *LGPL* license.

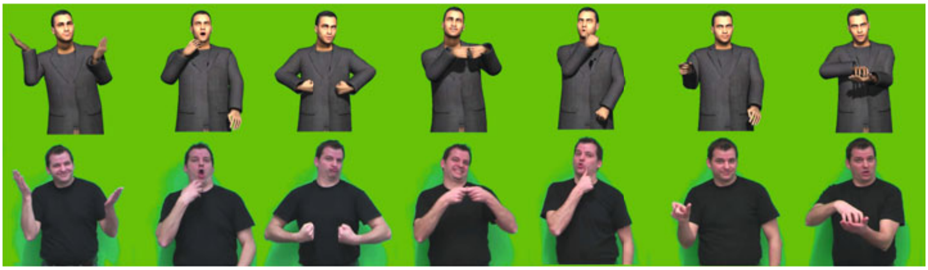

Fig. 1. Based on the EMBR character animation system, we created a signing avatar to explore the technical feasibility and develop evaluation methods

of the OpenMARY[4] speech synthesis system [26]. Note that mouthing implies a number of questions in terms of selection (which word to mouth), timing (when to onset), duration (how much of the word to mouth, often the first part is enough) and how to synchronize mouthing with repeated strokes[5] [24]. We have not yet included all important mouth gestures like puffed cheeks and a flexible tongue.

Another important movement type is *gaze*. We extended EMBR to allow independent control of eye-balls and head because gaze can give important cues to disambiguate two manually equal signs. We stress that our extensions were targeted at German Sign Language (and, to some extent, at ASL) but should also meet most requirements of other sign languages. Fig. 1 shows some of the posing capabilities of the extended EMBR.

4 Animating Sign Language

In this section, we motivate our cyclic, gloss-based animation approach. To test the feasibility of the approach we used videos of human signers. Whereas in prior work very basic utterances were used [27], we selected two videos with more complex content from a German e-learning portal for deaf people called *Vibelle.de*. The videos teach concepts from the hearing world in sign language. We selected two videos, *yellow pages* (37 sec) and *vitamin B* (53 sec), with a total of 11 sign language utterances.

4.1 One Step Forward, Two Steps Back

We found that we had to create overarticulated versions of our original videos in order to compensate for avatar shortcomings. Our first "pilot test" was to imitate a piece of video with a single EMBRScript animation. However, we had to realize that our result was *not comprehensible* by our deaf assistant – not a single sign. The superficial sign language knowledge of our (hearing) animation

[4] `http://mary.dfki.de`

[5] The *stroke* is the most energetic part of a sign and can be repeated [19].

expert was insufficient – we needed a sign language expert before, during and after animation production.

Our initial attempt failed for a number of reasons, some on the level of a single sign, some on the utterance level. For single signs, sign linguistics distinguishes between *manual* and *nonmanual* features [12]. Manual features include hand shape, hand location, hand orientation and movement [20,28]; here, the animator needs to decide which of these is the decisive, so-called *formative*, component that has to be modeled with utmost precision. *Nonmanuals* (facial expression, gaze, torso movements) [24] are even more difficult to capture for various reasons. Nonmanuals may stretch over several signs: for instance, facial expression indicates sentence mode (question vs. assertion) and eyebrows and posture relate to information structure by marking the topic. On the single sign level, nonmanuals are e.g. used for negation/affirmation, adjectival information (face: emotion/attitude) or adverbial modification (face: manner of execution). In German sign language (DGS) the parallel "speaking" of the corresponding word, called *mouthing*, is highly important (less important in e.g. ASL [12]). However, it is unclear in which cases mouthing supports comprehension. In many cases, the lack of mouthing simply introduces irritation. Indeed, sign language listeners need the face as a *focus point for their visual attention* because this allows to see hands, face and body at the same time [29]. The usually static faces of avatars generate so little visual interest that the listener's visual focus jumps from hands to mouth to torso etc., making comprehension harder. Generally, nonmanuals have only recently received more attention from the linguistic side [24] but need more research in the area of sign language synthesis.

Our main conclusion is that current avatars with current animation methods cannot reproduce all the subtleties of a human signer's synchronized body behaviors. Therefore, our *working hypothesis* is that avatars need to start from a different point of departure and suggest to use *overarticulated base material* for guiding our animations.

To create overarticulated video remakes, each video was segmented into utterances and glosses by two DGS experts using the ANVIL annotation tool [14]. This transcription, together with the original video, was the basis for the new video recordings, performed by a deaf DGS native speaker with the following instructions: make single signs as clear as possible, include clear mouthing, separate signs cleanly from each other while maintaining overall fluidity.

4.2 Gloss-Based Animation in EMBRScript

We created a database of single gloss animations based the human signer's videos which were used to assemble utterances. The animation notation EMBRScript was particularly suitable as it allows the specification of so-called *k-pose-sequences* [4], i.e. a collection of generalized poses (including IK constraints, morph targets and predefined skeletal configurations), which elegantly corresponded to single glosses. To add parallel movements that span several glosses, we can use additional, separate k-pose-sequences. We extended the existing *BehaviorBuilder* tool [15] to support the definition of single glosses (i.e. one

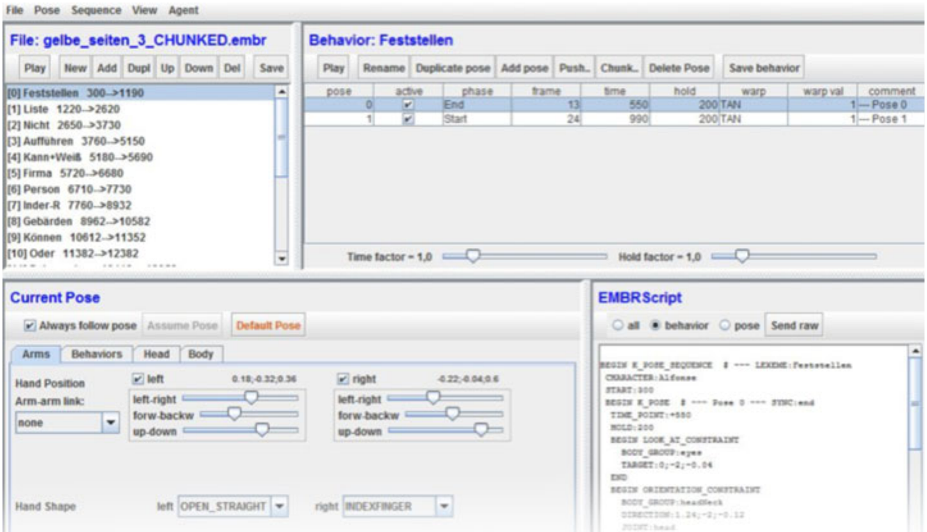

Fig. 2. Screenshot of the BehaviorBuilder tool which allows to create animation on three levels: single pose (bottom left), single gloss (top right) and gloss sequence (top left). The result is a declarative script in the EMBRScript language (bottom right).

k-pose-sequence) and the sequencing of glosses to a complete utterance. Fig. 2 shows the revised tool that allows the interactive creation of single poses, pose sequences (glosses) and gloss sequences. We used the OpenMARY [26] text-to-speech synthesis system to generate viseme animations which were assigned the same start time as the corresponding gloss. Our gloss-based approach is similar to *iLex* [3] which is based on HamNoSys. However, working with HamNoSys requires many tweaks to control the animation so that we suggest to rather use EMBRScript as a separating layer to keep HamNoSys free from low-level animation data.

The animation process followed a tight feedback cycle with a deaf assistant who, in each iteration, provided comments on single signs as well as the overall utterance comprehensibility. This provided the animator with clear priorities whether to optimize hands/arms, facial expression, gaze, upper body or timing. For the future, a formalized feedback with rating scales or fixed feedback categories may be possible. We found that nonmanuals, especially mouthing and gaze, were as important as manual components. Additional feedback concerned general aspects of the 3D scene. Lighting must ensure that face and hands are highly visible, e.g. by using extra light sources. Realistic shadow casting is perceived as a pleasant addition by deaf users as it adds depth and 3D illusion. From interviews with deaf individuals it became clear that the *outer appearance* of the avatar is important. This may be due to the association of avatars as sign language *interpreters* whom deaf people may regard as their representatives.

We created 11 utterances which contained 154 gloss instances, i.e. on average an utterance contained 14 gloss instances. The workload for creating single

Table 1. These are, for each utterance, the number of contained glosses and the duration of the video materials (orginal, remake, avatar)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
contained glosses	17	18	21	10	8	7	12	21	13	14	13
Video (original)	10s	7s	11s	7s	5s	3s	7s	11s	13s	3s	7s
Video (remake)	14s	14s	20s	14s	9s	6s	16s	19s	19s	7s	13s
Avatar	17s	20s	25s	16s	12s	11s	15s	23s	25s	8s	22s

glosses was 5-25 mins (with 1-2 weeks of initial training). The resulting avatar animation sequences were longer compared to the original video. This becomes clear in Table 1 which shows, for each utterance, the number of glosses and the durations of the video(s) and of the avatar version.

4.3 Limitations

The gloss-based approach in its current form is a simplified abstraction where each gloss always looks the same independent of context. This ignores individual or stylistic variations and grammatical flections, e.g. for directed verbs like GIVE or SHOW. Moreover, glosses do not contain information relating to larger units of an utterance or the discourse such as information structure (old/new distinction). However, we consider the gloss-based approach a useful point of departure that must be extended using e.g. parameterized glosses and an added representation layer for utterance-level and discourse-level information.

Regarding scalability, we experienced limited re-use of glosses at this stage of the project. With a database of 95 glosses we created 11 utterances with 154 gloss instances which means that each gloss was used 1.6 times on average. We assume that gloss reuse increases with the larger projects, once a “basic vocabulary” has been established. It is an open question how large such a basic vocabulary has to be such that the number of new glosses per new utterance is minimal in a specific domain.

5 Comprehensibility Testing

As material we used a corpus of 11 utterances from two e-learning videos (Sec. 4). For every utterance, we wanted to compare the avatar animation (A), the original video (V_{org}) and the overarticulated remake (V_{re}). We invited 13 native signers (6m, 7f), of age 33–55, to the experiment which took 1.5 – 2 hours per subject and was supervised by a deaf assistant. Every subject was compensated with 10 Euro plus travel cost. Since all sessions had to be videotaped for later analysis, subjects had to sign an agreement to grant us scientific usage of the material.

5.1 Method

We set up the following frame conditions: we provided a sign-language-only environment and made the users feel comfortable to criticize the system by having

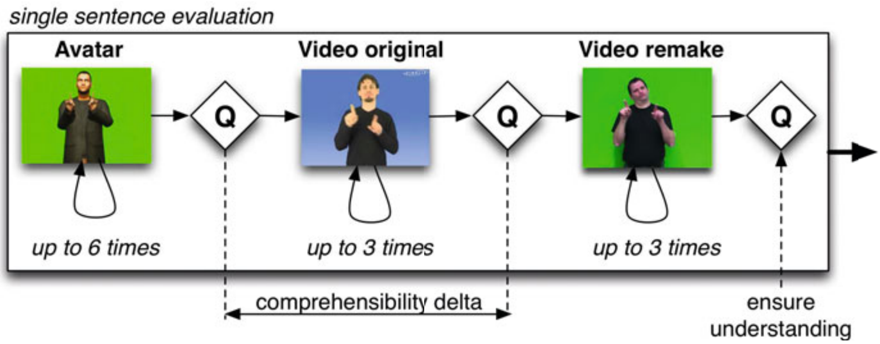

Fig. 3. Evaluation procedure for a single utterance. It was important to ensure understanding to prepare the following utterance test.

supervisors from outside our institute. Since deaf people may have difficulties understanding written commands, the briefing was done in sign language and the questionnaires included pictograms (e.g. smiley signs) for clarification.

To get the subject accustomed to the way our avatar performs sign language we showed three very basic clips ("My name is M-A-X", "My sign name is <sign>" and "My work is interesting") without further testing. Such a *warm-up phase* is quite common [27]. Then, we proceeded with the *evaluation phase* where each of the 11 utterances was displayed in the following scheme (depicted in Fig. 3): First, we showed the avatar version A which could be viewed up to 6 times. Second, we showed the original video V_{org} which could be viewed up to 3 times. Third, we showed the overarticulated remake V_{re} which could be viewed up to 3 times. After each of the three screenings the subject was asked to sign what s/he understood from the respective clip. After the three videos, we showed the video remake once more, this time with text subtitles[6], to make sure that this utterance was understood before proceeding with the next one.

5.2 Analysis and Results

According to [11] the participants' subjective impression of their understanding is not a good indicator of actual understanding. Therefore, we used two complementary methods for measuring comprehensibility. First, as an objective measure, we took the glosses of each utterance and tried to see which ones were repeated by the subject when asked to repeat the content of the utterance. The rate of understanding can be computed by dividing the number of repeated glosses by the total number of glosses. However, this can be misleading if subjects are able to recall unconnected parts of the utterance while not understanding the core meaning. Therefore, we asked our deaf experts to give a subjective estimation of how well the subject had understood the utterance on a 7-point scale. We then took the average of the two experts for each utterance.

[6] Subtitles may help subjects understand signs performed very quickly or in a sloppy manner or are unknown because of regional differences.

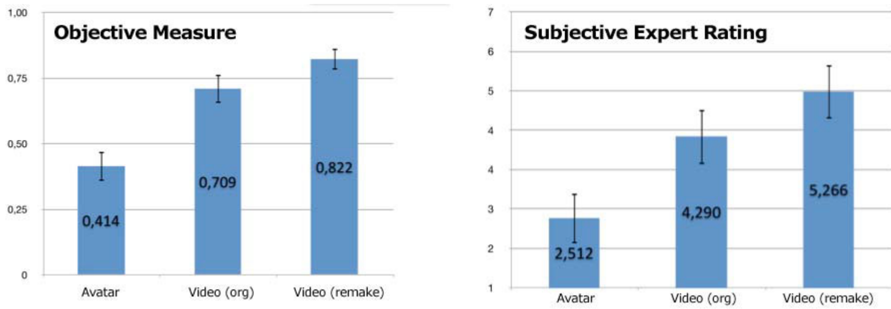

Fig. 4. Comprehensibility results of the objective measure and subjective expert rating

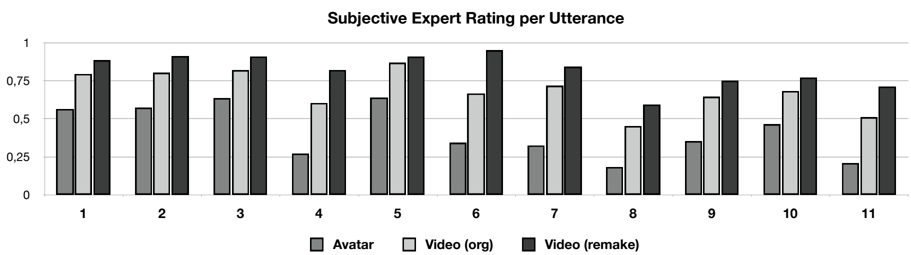

Fig. 5. Subjective expert ratings per utterance (in the order they were displayed)

Fig. 4 summarizes the results. The relative differences between the materials are similar in both measures. What is striking is that for the original video, absolute comprehensibility is only at 71% (objective) and 61% (subjective). Having comprehensibility scores for all three materials allows us to put the avatar score in relation to the others. If we put the avatar in relation to the original video we reach a comprehensibility of 58.4% (objective) and 58.6% (subjective). The harder comparison is that between avatar and remake with 50.4% (objective) and 47.7% (subjective).

Since, by design, we had to display utterances in a fixed order we examined potential ordering effects. One could assume that due to increasing context the understanding would increase. However, Fig. 5 indicates that understandability was quite varied, even for utterances in the later stages of the experiment.

5.3 Discussion

In our comprehensibility tests we take the comparison between avatar and original video to be our goal value. Here, our avatar reached 58.4% or 58.6% which is close to the evaluation results of state-of-the-art systems of around 62% in ViSiCAST [27,13]. Given a very short development time of 4 person months, we conclude that higher scores can be reached. We agree with [13] that a comprehensibility of 90% is possible, but only if a clear shift in research focus takes place in the field of sign language synthesis, towards *nonmanuals* and *prosody*. In terms of synthesis, this implies an extension of the gloss-based approach by adding gloss parameters and an utterance-level layer. In our current implementation nonmanuals are integrated only to a limited degree, partly because the utterance-level is missing. To see our study in the bigger picture we stress that our material of only 11 utterances was quite limited. While many other studies used similarly low quantities of around 10 utterances (e.g. [27,11]) for the future the same standards as in natural language processing should be reached (e.g. [8] used 12 stories of 48-80 signs each).

Regarding evaluation methods, we believe that our *delta evaluation* has two advantages over previous methods. First, it takes into account the limited comprehensibility of the original video material (in our case, this was as low as 71% / 61%) and thus, makes the comparison fairer and may inspire other researchers to dare the direct comparison with human signers. It allows to use more complex content and factors out dialect variation in sign languages that cause certain signs to be unknown in a participant's region. Second, setting our avatar into relation to human signers, we did not have to agree on any absolute measure of comprehension, e.g. that particular pieces of the utterance are more important than others. Defining such measures is work-intensive and subjective. By combining an objective measure (gloss counting) with a subjective expert evaluation we ensure that the understanding of the whole utterance is well captured. Due to our method's design the avatar is put to a slight disadvantage which means that the result represents a lower boundary: the avatar is *at least* as good as the measure indicates.

6 Conclusions

We presented the development and evaluation of a signing avatar, on the basis of an existing general-purpose avatar EMBR. We showed how a gloss-based approach with a tight cyclic animation development, in close cooperation with deaf experts, can lead to state-of-the-art performance for German sign language synthesis. We introduced an overarticulated video remake into the loop based on the working hypothesis that current avatar technology lacks the complexity of human multimodal signal generation. We also created a novel evaluation method we call *delta evaluation* where we compare avatar performance with human signers based on objective gloss counts and subjective expert opinions. This measure is a lower boundary of the real comprehensibility of the avatar. In the development process we identified nonmanual components and prosody as the most urgent issues for increasing comprehensibility significantly beyond 60% which we deem feasible. While theoretical work on nonmanual components and prosody exist (cf. [24]), the operationalization in avatars is scarce (see [8] for a notable exception).

Facial expression also needs more research, especially given the additional need of having the face generate visual interest so that listeners can fixate on it. Current research on the *uncanny valley* suggests that the face is of key importance for overcoming the acceptance problem of avatars [6]. Prerequisite for this is a consistent evaluation scheme like *delta evaluation*. This needs to be extended from the utterance level to the level of *whole text/discourse* understanding. We also stress that involvement of deaf people is crucial not only for defining use cases and for evaluation but, even more so, for creating animations and developing animation methods. Hence, we argue for a stronger *scientific* involvement of deaf individuals. This implies the development of better tools to allow animation building in the deaf community, e.g. with the help of novel interface technology [17].

Moreover, we conjecture that research on sign language synthesis will generate important insights for *coverbal gesture synthesis*. In "Kendon's continuum" sign language is one extreme pole with a fluid transition to coverbal gesture [21]. While speech is missing from sign language, it remains a highly multimodal problem that involves face, body, hands and arms. Making these cohere in natural orchestrated movements is a goal, both in speaking and non-speaking cultures.

Acknowledgements. This study was commissioned by the German Federal Ministry of Labour and Social Affairs (eGovernment Strategie Teilhabe). We would like to thank our two deaf DGS experts Peter Schaar and Iris König for their help in interviews, video recordings, DGS transcriptions and evaluation studies. Thanks to our reviewers whose comments were crucial for the revision. Part of this research has been carried out within the framework of the Excellence Cluster Multimodal Computing and Interaction (MMCI), sponsored by the German Research Foundation (DFG).

References

1. Elliott, R., Glauert, J.R.W., Kennaway, J.R., Marshall, I., Safar, E.: Linguistic modelling and language-processing technologies for avatar-based sign language presentation. Univers. Access Inf. Soc. 6, 375–391 (2008)
2. Filhol, M.: Zebedee: a lexical description model for sign language synthesis. Tech. Rep. 2009-08, LIMSI (2009)
3. Hanke, T.: iLex - a tool for sign language lexicography and corpus analysis. In: Proceedings of the 3rd International Conference on Language Resources and Evaluation, pp. 923–926 (2002)
4. Heloir, A., Kipp, M.: Realtime animation of interactive agents: Specification and realization. Journal of Applied Artificial Intelligence 24(6), 510–529 (2010)
5. Heloir, A., Kipp, M., Gibet, S., Courty, N.: Specifying and evaluating data-driven style transformation for gesturing embodied agents. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 215–222. Springer, Heidelberg (2008)
6. Hodgins, J., Jörg, S., O'Sullivan, C., Park, S.I., Mahler, M.: The saliency of anomalies in animated human characters. ACM Trans. Appl. Percept. 7, 22:1–22:14 (2010)
7. Holt, J.A.: Demographic, Stanford achievement test - 8th edition for deaf and hard of hearing students: Reading comprehension subgroup results. Amer. Annals Deaf. 138, 172–175 (1993)

8. Huenerfauth, M.: A linguistically motivated model for speed and pausing in animations of american sign language. ACM Trans. Access. Comput. 2, 9:1–9:31 (2009)
9. Huenerfauth, M., Hanson, V.L.: Sign language in the interface: Access for deaf signers. In: Stephanidis, C. (ed.) The Universal Access Handbook. CRC Press, Boca Raton (2009)
10. Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J.: Evaluating american sign language generation through the participation of native ASL signers. In: Proc. of the 9th International ACM Conference on Computers and Accessibility (ASSETS), pp. 211–218. ACM, New York (2007)
11. Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J.: Evaluating american sign language generation by native ASL signers. ACM Transactions on Access Computing 1(1), 1–27 (2008)
12. Johnston, T.: Australian Sign Language (Auslan): An introduction to sign language linguistics. Cambridge University Press, Cambridge (2007)
13. Kennaway, J.R., Glauert, J.R.W., Zwitserlood, I.: Providing signed content on the internet by synthesized animation. ACM Transactions on Computer-Human Interaction (TOCHI) 14(3), 15–29 (2007)
14. Kipp, M.: Anvil: The video annotation research tool. In: Durand, J., Gut, U., Kristofferson, G. (eds.) Handbook of Corpus Phonology. Oxford University Press, Oxford (to appear, 2011)
15. Kipp, M., Heloir, A., Schröder, M., Gebhard, P.: Realizing multimodal behavior: Closing the gap between behavior planning and embodied agent presentation. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 57–63. Springer, Heidelberg (2010)
16. Kipp, M., Neff, M., Albrecht, I.: An Annotation Scheme for Conversational Gestures: How to economically capture timing and form. Journal on Language Resources and Evaluation - Special Issue on Multimodal Corpora 41(3-4), 325–339 (2007)
17. Kipp, M., Nguyen, Q.: Multitouch Puppetry: Creating coordinated 3D motion for an articulated arms. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (2010)
18. Kipp, M., Nguyen, Q., Heloir, A., Matthes, S.: Assessing the deaf user perspective on sign language avatars. In: Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS). ACM Press, New York (2011)
19. Kita, S., van Gijn, I., van der Hulst, H.: Movement phases in signs and co-speech gestures, and their transcription by human coders. In: Wachsmuth, I., Fröhlich, M. (eds.) GW 1997. LNCS (LNAI), vol. 1371, pp. 23–35. Springer, Heidelberg (1998)
20. Liddell, S.K., Johnson, R.E.: American sign language: The phonological base. Sign Language Studies 64, 195–277 (1989)
21. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago (1992)
22. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture Modeling and Animation Based on a Probabilistic Recreation of Speaker Style. ACM Transactions on Graphics 27(1), 1–24 (2008)
23. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: An interactive expressive ECA systems. In: Proc. of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pp. 1399–1400 (2009)
24. Pfau, R., Quer, J.: Nonmanuals: their prosodic and grammatical roles. In: Brentari, D. (ed.) Sign languages (Cambridge Language Surveys), pp. 381–402. Cambridge University Press, Cambridge (2010)

25. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J.: HamNoSys Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guides, Signum (1989)
26. Schröder, M., Trouvain, J.: The german text-to-speech synthesis system mary: A tool for research, development and teaching. International Journal of Speech Technology 6, 365–377 (2003)
27. Sheard, M., Schoot, S., Zwitserlood, I., Verlinden, M., Weber, I.: Evaluation reports 1 and 2 of the EU project essential sign language information on government networks, Deliverable D6.2 (March 2004)
28. Stokoe, W.C.: Sign language structure: An outline of the visual communication system of the American deaf. Studies in linguistics, Occasional papers 8 (1960)
29. Swisher, V., Christie, K., Miller, S.: The reception of signs in peripheral vision by deaf persons. Sign Language Studies 63, 99–125 (1989)
30. Thiebaux, M., Marshall, A., Marsella, S., Kallman, M.: SmartBody: Behavior realization for embodied conversational agents. In: Proc. of the 7th Int. Conf. on Autonomous Agents and Multiagent Systems, AAMAS (2008)
31. Wolfe, R., McDonald, J., Davidson, M.J., Frank, C.: Using an animation-based technology to support reading curricula for deaf elementary schoolchildren. In: The 22nd Annual International Technology & Persons with Disabilities Conference (2007)