Signing Avatars: a Feasibility Study

Alexis Heloir, Quan Nguyen, Michael Kipp
DFKI
Embodied Agents Research Group
Saarbrücken, Germany
firstname.lastname@dfki.de

ABSTRACT

Signing avatars, by offering a flexible and cost-effective alternative to written content can make the Internet more accesible to Deaf people. In order to be adopted, signing avatars should produce intelligible animations and the technology should be well accepted by Deaf users. In this paper, we firstly show how the acceptance of signing avatar technology can be assessed using two well known methods: focus groups and online studies. Secondly, we present a gloss-driven signing avatar based on an existing solution that we improved by taking into account Deaf users' criticism and sign language experts' feedback. We finally address the open problem of designing standard evaluation methods for assessing signing avatars' comprehensibility by comparing avatars with human signers. We show that the sign language animations produced with our system reach a comprehensibility level that is comparable to the levels achieved by state of the art systems.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*language generation, machine translation*; K.4.2 [**Computers and Society**]: Social Issues—*assistive technologies for persons with disabilities*

General Terms

Acceptance, Experimentation, Measurement

Keywords

German Sign Language, Sign Language Synthesis, Accessibility Technology for Deaf people

1. INTRODUCTION

An increased use of video-recorded human signers can be observed on websites publishing information for Deaf people. However, video recordings imply considerable production cost, their content cannot be modified after production, and they cannot be anonymized. In contrast, when using signing avatars, i.e. virtual characters that perform sign language, one can change appearance (gender, clothes, lighting), they are inherently anonymous and the production of new content is potentially easy and cost-effective (no studio setup, no expert performer required, may even be created collaboratively) [10]. Most importantly, avatar animations can be dynamic, i.e. they can be computed and adjusted on-the-fly, allowing for the rendering of dynamic content (e.g. inserting locations, dates, times ...) and interactive behavior (question answering). In this paper, we focus on avatar technology that allows for this flexibility.

In order to be adopted by the Deaf, signing avatar technology should satisfy two conditions: the produced sign language animation should be intelligible and the technology should be well accepted by Deaf communities. While research on signing avatar has been carried out for two decades, an effort to clarify its general acceptance is, to the best of our knowledge, still missing. Acceptance implies identifying potential negative sentiments or fears concerning this technology. Ultimately, not only the assessment of acceptance but also the question of how to increase acceptance must be addressed.

Avatar based sign language generation is still considered an unsolved problem. However, the signing avatar community is small and lacks the budget to create the same international networks that have fostered spoken language research. It is therefore desirable to identify the most critical points of improvements early so that the research community can optimize its effort.

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 for German Sign Language (DGS[1]) and the acceptance in the German Deaf community.

This paper summarizes the feasability study which has been presented in more detail in two recent publications [16, 18]. Next Section presents the most relevant related work, we then present how we collected insights about opinions, potential negative feelings or fears but also recommendations, criticism, desired applications and hopes raised by signing avatar technology using focus groups and online questionnaires. In a 4th section, we present how the collected criticism and a tight collaboration with two native Deaf experts drove the improvement of both the animation

[1]Deutsche Gebärdensprache

SLTAT 2011, 23 October 2011, Dundee, UK

and appearance of an existing general-purpose character animation framework. In Section 5, we present an incremental approach that led to a gloss-based authoring workflow. In Section 6, we address the open problem of designing standard evaluation methods for assessing signing avatars comprehensibility by comparing avatars with human signers. A last Section concludes this paper.

2. BACKGROUND

This section briefly presents the most relevant work related to these three challenges. See [16, 18] for a more detailed discussion. In this study, we address three main issues:

- How can we assess the acceptability of signing avatar technology and collect meaningful information from the Deaf in the specific context of Deaf people?
- How can we, in a limited time, adapt an existing general purpose animation framework to generate comprehensible sign language utterances?
- How can we design a standard evaluation method for assessing the comprehensibility of the generated sign language utterances?

A well known method used to elicit people's perceptions and attitudes about any particular product or concept early on in the design process are *focus groups*. A focus group is a guided discussion with 3-10 participants led by a trained moderator where a preset agenda guides the discussion [22]. The goal of focus groups is to collect *in-depth* information about what issues participants consider important, what preferences they have and how they prioritize these. Focus groups with the Deaf were pioneered in 1999 in five groups with an experienced hearing moderator and a sign language interpreter/assistant [1]. Although all existing studies stressed the importance of establishing a deaf friendly environment when conducting the focus groups [1, 2, 3, 19, 8, 19], none of these studies were conducted by a native Deaf moderator.

Existing signing avatars frameworks can be categorized into two groups: The ones following articulatory approaches and the ones following concatenative approaches [10]. 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 [4, 13], transitioning from a concatenative to an articulatory approach. However, these system are not freely available. other more recent systems [12, 23, 9, 5, 24] are at the stage of a research prototype and are not yet mature and integrated enough to be used outside their original labs.

Finally, assessing the comprehensibility of the sign language output produced by an avatar is not straightforward and no agreed-on methodology exists. The most relevant work dealing with assessing the comprehensibility of sign language relied on questionnaires [13, 23], subjective assessment of comprehensibility [12] and multiple choice tests [11]. Avatar's signing has been compared to written english or signed english [12]. We instead suggest to use the comprehensibility of a Deaf signer as the control condition.

3. IMPROVING ACCEPTANCE

Identifying the potentially negative sentiments or fears concerning signing agents is critical in our context since, for historical reasons, Deaf may be skeptic toward any technology developed by the hearing. Therefore, not only the assessment of acceptance but also the question of how to increase it must be addressed. To answer these questions, we therefore combined in-depth discussion that is possible in focus groups with the quantitative strengths of online studies.

Figure 1: The left image shows the setup of the second focus group for 5 participants, the moderator and his assistant. A screenshot from the synchronized and arranged video for the analysis is shown on the right.

3.1 Focus groups

We conducted two focus groups, G1 and G2, with 3 and 5 participants each. Each group took about four hours. Participants should be native signers and should consider themselves members of the Deaf community.

During the focus groups we used different media to stimulate discussion: a video projector was used to project videos of existing avatars and still images. A whiteboard was used to stick flash cards with keywords on it that could be used for voting. Throughout the session, written text was avoided. All sessions were videotaped for later analysis (all subjects signed an agreement to grant us scientific usage of the material). Each focus group was structured in cycles of information–discussion–voting. The project was introduced by a sign language video and two initial questions: "Do you think avatars are useful?" and "Do you think Deaf people would use avatars?". Following this question, we showed a selection of recent signing avatar videos (Fig. 2). The avatars featured in these videos were leveraging different technologies and were fulfilling different goals, from research prototype (Max from the eSign project) to completely handcrafted animation (invitation to the 2007 Deaf World Congress: Deaf-World). Participants discussed and criticized the avatar. Keywords were taken and ordered according to their subjective importance. We then showed images depicting several suggestions for applications, depicted on a photo montage. In a round of voting, we collected the preferences for the most interesting applications. A similar round dedicated to online applications followed. Finally, participants were asked the first two initial questions once again. Additionally, we asked the participants whether they thought that the government should invest money in this technology.

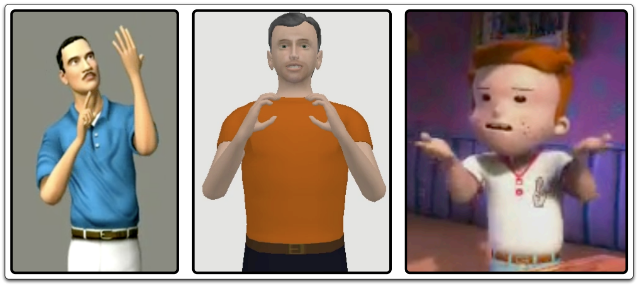

Figure 2: Three of the presented existing avatars (left to right): The Forest (ASL, created by VCom3D), Max (DGS, created by Univ. East Anglia) and DeafWorld (International Sign, commissioned by the World Federation of the Deaf).

3.2 Online study

We set up an accessible online study to quantify several results suggested by the focus group. This study mostly shared the structure of the focus group. For accessibility we provided DGS video explanations for all questions. Replies were collected on 5-point scales, visually enhanced by smileys and color-coding. In total, 330 people completed the questionnaire. There were more deaf (85 %) than hard of hearing (2%) or hearing (13%) participants. see [18] for further details on participants and results.

3.3 Results

The focus group interviews turned out to be an excellent method to elicit criticism, constructive suggestions and opinions of the Deaf participants. Especially for a topic like avatars, where participants might not have a clear idea of the opportunities associated with this technology, the focus group interviews allowed them to develop their criticism and suggestions throughout the session.

Application scenarios, as discussed in the focus groups, are mainly situated in the area of one-way communication situations. The participants can neither envision avatars in dialogic settings nor for very complex or emotional content.

Potential applications that emerged during the discussion were: (Online) translation services for simple sentences, static announcements (job offers, news letters, election campaigns) and static texts (legal texts, manuals), information usually communicated via speakers (train station, airport), daily news and news feeds, lexicon and dictionaries, museum guides. The online study (Fig. 3) showed a much more diverse picture with more entertainment and leisure time applications. For internet applications, the top applications were educational (17%), social network websites (16%) and public administration pages (11%).

In the focus groups, it was extremely important for all participants that avatars should not be seen as a replacement for human interpreters and that every Deaf should always have the choice between the two. This is reflected in the online study where 25% of mentioned effects concerned job cuts for interpreters or for Deaf people. Another concern was the danger that using an avatar may lower the motivation for the Deaf to properly learn reading/writing. Maybe not surprisingly, online study participants had much more con-

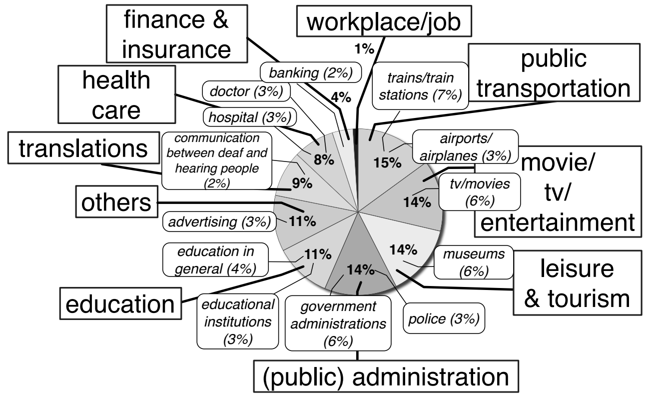

Figure 3: General applications (online study)

cern about technical feasibility (20%) and reliability (22%) of avatars since they had not discussed potential scenarios in depth. See Fig. 4 for the mentioned possible risks mentioned in the online study.

However, advantages in the use of avatars are seen. Focus group participants found most important that avatars are *available anytime*, while interpreters are often hard to find. *Personalization* is possible, e.g. regarding appearance, speed, or perhaps even language output (sign language vs. sign supported spoken language). Avatars also allow for *anonymity* in the internet (e.g. for discussions on controversial topics).

Finally, The general attitude towards avatars, as asked for in the beginning and at the end of the focus groups, was overall positive and increased throughout the sessions. In the online studies, this effect even reached high statistical significance. This underlines the potential of involving the Deaf community, not only for general assessments but also for increasing acceptance.

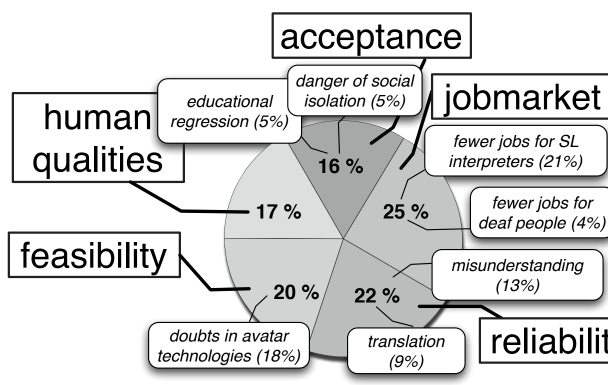

Figure 4: Fears of potential negative impacts in these areas (online study)

4. IMPROVING THE ANIMATIONS TOOLS AND WORKFLOW

In this section we describe hoy we extented a general purpose character control framework to produce gloss-based sign language animations. We firstly present the criticism

Important avatar aspects
Facial expression (7)
Natural movement (5)
Mouthing (4)
Emotions (4)
Body motion/posture (4)
Appearance (3)
Synchronisation of sign and mouthing (3)
Charisma (2)
Comprehensibility (2)

Table 1: Voting on most important avatar aspects

collected in the previous section and the feedback obtained from our Deaf experts. Secondly, we show how these insights drove the improvement of our exisitng agent framework and animation production workflows up to state-of-the-art comprehensibility levels.

4.1 Criticism from the two studies

The feedback of both the focus groups and the online study clearly shows that much improvement in the performance of sign language avatars is still needed. In the focus groups the criticism was not so much focussed on a single aspect than rather on the general appearance that was mostly described as stiff, emotionless and unnatural. Most of the criticism targeted nonmanual features, mainly facial expression and mouth patterns, but also movements of the head, shoulders and torso.

Regarding facial expressions, missing variation in eyebrow, eyelid, as well as eye movement were specifically mentioned. Permanent eye contact was regarded as unnatural and causing discomfort. The absence of mouth patterns, especially mouthings (i.e. mouth patterns derived from the spoken language), seemed to be one of the most disturbing factors for the participants since this is an important element of DGS. Also cheeks, teeth and tongue were said to be needed as a crucial element for understanding certain mouthings. For the torso, movements like hunching and twisting were mentioned to improve naturalness, as well as clear sideway rotations (e.g. for marking role shifts). In general, the participants wished for more smooth and relaxed movements of all parts of the upper body.

The wish for naturalness and emotions also show that the general appearance should not be underestimated. On the contrary, manual components (i.e. the hands) were not in the participants' focus. The most positive votes in the focus groups as well as the online study were given to a fully handmade animation. This underlines the gap between avatar approaches which can be automated and handmade animations. Table 1 summarizes the most important aspects elicited by the focus groups. These aspects served as guidelines for improving our animation methods.

4.2 Improving animation methods

The insights gathered from both the focus groups and the online study served as guidelines for improving EMBR [6, 15], the free and open character animation engine that we initially developed for coverbal gesture generation.

The EMBR character animation engine offers a high degree of control over the animation through the EMBRScript animation language. Arbitrary animation sequences can be specified and edited without having to program. Because the EMBR animation system has grown out of research on coverbal gesture production [20, 17, 7], it 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. To the existing set of 12 handshapes that is sufficient for coverbal gesture, we added 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. 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 of the OpenMARY speech synthesis system. Note that mouthing implies a number of questions in terms of selection (which word to mouth), timing (when to onset), duration and phoneme selection (how much of the word to mouth, often the first part is enough). 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. 5 shows some of the posing capabilities of the extended EMBR.

4.3 Improving the avatar's appearance

The feedback we collected from the focus groups and the online study provided us with guidelines for crafting a new agent that would elicit a more positive response from the deaf users. The design of this new agent was conducted in collaboration with a professional 3D artist from the school of design and art in Saarbruecken[2] and matched the following requirements collected in the focus groups: agent should look sympathetic and not dominant, he should look young but not eccentric, he should have realistic proportions, his appearance should be stylized and cartoony rather than realistic, he should have an expressive face and be capable of performing, specific facial expression which are used in sign language like, puffing or squeezing cheeks.

The result of this collaboration is depicted in Fig. 6. Hand drawings and intermediate still renderings we showed during the conception phase received very positive feedback from all the deaf participants who saw it. However, for technical reasons, this new agent was not assessed while performing signed utterance.

5. WORKFLOW

Regarding the animation creation workflow, our first attempt was based on a direct recreation of sentences extracted from e-Learning videos. Sign language sentences authoring was conducted using the *BehaviorBuilder* authoring tool [15].

5.1 Learning from failure

Our first pilot animation with a single EMBRScript animation was *not comprehensible* by our deaf assistant – not a single sign. Our initial attempt failed for a number of reasons, some on the level of a single sign, some on the sentence

[2]Hochschule der Bildenden Künste Saar

Figure 5: Based on the EMBR character animation system, we created a signing avatar to explore the technical feasibility and develop evaluation methods.

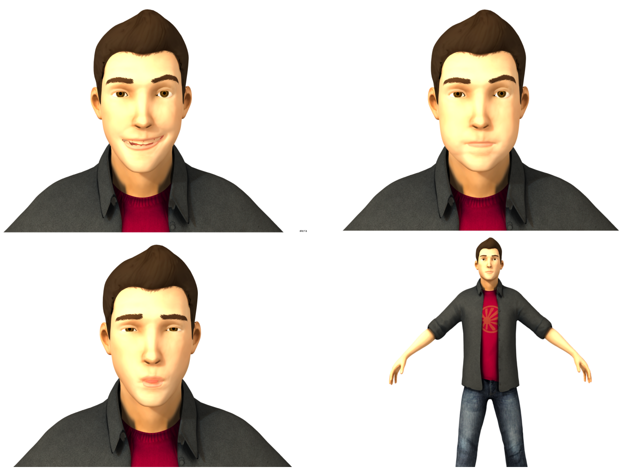

Figure 6: Our new agent Max displaying a neutral expression (a) and two expressions that are used to express size in DGS: puffed cheecks (b) is used to describe large objects, pinched cheecks (c) is used to describe thin objects. Last picture is a full-length portrait of our agent (d).

level. On the sign level, our animation expert could not always figure out which were the most salient sign features – the ones that required the most care and attention when reproduced in our animation framework. On the sentence level, careful observation and thoughtful discussions with the sign language expert made us realize that the sentences in the original videos were not suited as a base material for guiding our animations: many signs were significantly influenced by coarticulation, handshapes were not always completely achieved and many crucial non-manual features operating at a prosodic level were so subtle that they could hardly be recognized in the video because of the insufficient recording framerate (25Hz) and the deterioration caused by the video compression.

5.2 Over-ar-ti-cu-la-tion

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 native user of DGS with the following instructions: make single signs as clear as possible, include clear mouthing, separate signs cleanly from each other while maintaining overall fluidity.

We created a database of single gloss animations based on 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* [6], i.e. a collection of generalized poses (including IK constraints, morph targets and predefined skeletal configurations), which 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 k-pose-sequence) and the sequencing of glosses to a complete utterance. Fig. 7 shows the revised tool that allows the interactive creation of single poses, pose sequences (glosses) and gloss sequences. We used the OpenMARY[3] text-to-speech synthesis system to generate viseme animations which were assigned the same start time as the corresponding gloss.

This gloss-based approach is a simplification that does not take into account the possibility of variations and flections (e.g. for directed verbs like GIVE or SHOW). Moreover, glosses can obviously not contain animation information concerning the utterance level, for instance concerning information structure. However, we consider the approach a useful point of departure that must be extended using e.g. parameterized glosses and an added layer for utterance-level information.

6. COMPREHENSIBILITY EVALUATION

Assessing the comprehensibility of the signing produced by our avatar is a necessary and challenging task. Necessary because most of the experts working in the field are not native signers. Challenging because the theoretical problem of defining understanding is still largely open. It is already known that the subjective rating of understanding by the participant him/herself turns out to be highly unreliable

[3]http://mary.dfki.de

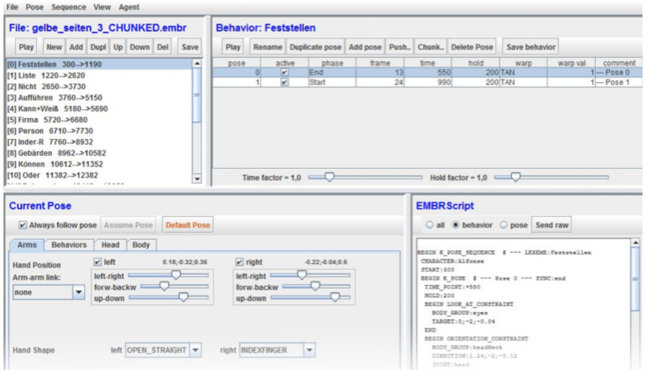

Figure 7: Screenshot of the BehaviorBuilder tool which allows to create animation on three levels: single pose (bottom left), single gloss (top right) and gloss sequence (top left). The result is a declarative script in the EMBRScript language (bottom right).

[12]. A more general challenge is to define a control condition, i.e. what is the avatar's signing compared against? [11] suggested *Signed English* (SE) as a control condition. Since Signed English and Sign Language are two distinct languages, the former is sometimes even harder to understand than the latter. We instead 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).

As material we used a corpus of 11 utterances from two e-learning videos. 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.

In the evaluation phase, 11 signed utterances were displayed in the following scheme: First, we showed the avatar version which could be viewed up to 6 times. Second, we showed the original video which could be viewed up to 3 times. Third, we showed the overarticulated remake 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 (Fig. 8), we showed the video remake once more, this time with text subtitles[4], to make sure that this utterance was understood before proceeding with the next one.

6.1 Analysis and results

According to [12] 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 asked our deaf experts 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 deafexperts 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.

Fig. 9 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).

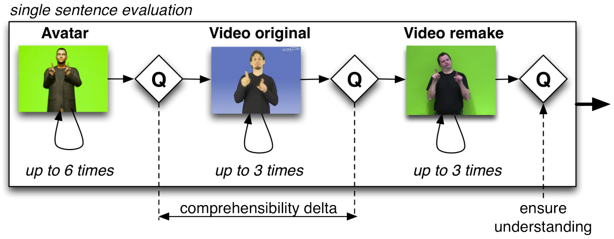

Figure 8: Evaluation procedure for a single utterance. It was important to ensure understanding to prepare the following utterance test.

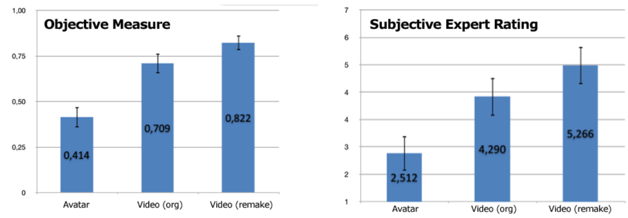

Figure 9: Comprehensibility results of the objective measure and subjective expert rating.

7. CONCLUSION

To sum up, this paper wraps up he most significant elements of a feasibility study dedicated to signing avatars. This study covers three main aspects: assessing the acceptance of signing avatars for the Deaf community, extending an existing general purpose character control framework in order to produce intelligible sign language utterances and developing new standard methods for assessing the comprehensibility of generated sign language utterances.

By conducting both focus groups and online studies in a pure sign language environment, we collected valuable feedback on existing avatars and ideas on application scenarios. A significant increase in positive opinion showed that both methods help increase the acceptance in the community.

We then showed how a close cooperation with experts can bring a general purpose character animation system to the challenging level required by the production of intelligible

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

signed language animations. We introduced an overarticulated video remake into our workflow based on the working hypothesis that current avatar technology lacks the complexity of human multimodal signal generation. We also created a novel evaluation method where we compare avatar performance with human signers based on objective gloss counts and subjective experts opinions. 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. [21]), the operationalization in avatars is scarce (see [9] for a notable exception).

We found that nonmanual components were found to be at least as important as manual ones. This indicates that research needs to make a major shift toward new challenges in the nonmanual area. In terms of applications, deaf subjects favored non-interactive, simple scenarios where avatars give information (train station, museums) or help in educational contexts (sign language lexicon, exam questions). However, also many other small everyday scenarios which may be made easier with an avatar were identified.

For the future we hope to conduct further focus group interviews on more specific topics, with better avatar materials or interactive mockup scenarios [23]. An important question is how to combine different media (video, text, avatars) so as to reach a maximum of comprehensibility and comfort for people with different degrees of reading skills. The ultimate question in the avatar domain is, however, how nonmanual components can be automatically integrated into existing systems and how that improves comprehensibiltiy which will in turn affect overall acceptance.

8. ACKNOWLEDGMENTS

We want to thank our excellent deaf moderators Iris König and Peter Schaar. We also thank Thomas Hanke (Univ. Hamburg) for valuable preparatory discussions as well as Horst Ebbinghaus (HU Berlin) for his explanations of Deaf culture. We finally thank Janosch Obenauer who dedicated his 3D modeling and character design expertise to our new agent. This research was funded by Federal German Ministry of Labour and Social Affairs. Parts of this research have been carried out within the framework of the Excellence Cluster Multimodal Computing and Interaction (MMCI), sponsored by the German Research Foundation (DFG).

9. REFERENCES

[1] G. I. Balch and D. M. Mertens. Focus group design and group dynamics: lessons from deaf and hard of hearing participants. *The American Journal of Evaluation*, 20(2):265 – 277, 1999.

[2] A. Cavender, R. E. Ladner, and E. A. Riskin. Mobileasl:: intelligibility of sign language video as constrained by mobile phone technology. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, Assets '06, pages 71–78, New York, NY, USA, 2006. ACM.

[3] A. C. Cavender, J. P. Bigham, and R. E. Ladner. Classinfocus: Enabling improved visual attention strategies for deaf and hard of hearing students. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, Assets '09, pages 67–74, New York, NY, USA, 2009. ACM.

[4] R. Elliott, J. R. W. Glauert, J. R. Kennaway, I. Marshall, and E. Safar. Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Univers. Access Inf. Soc.*, 6:375–391, January 2008.

[5] M. Filhol. Zebedee: a lexical description model for sign language synthesis. Technical Report 2009-08, LIMSI, 2009.

[6] A. Heloir and M. Kipp. Realtime animation of interactive agents: Specification and realization. *Journal of Applied Artificial Intelligence*, 24(6):510–529, 2010.

[7] A. Heloir, M. Kipp, S. Gibet, and N. Courty. Specifying and evaluating data-driven style transformation for gesturing embodied agents. In *Proceedings of the 8th International Conference on Intelligent Virtual Agents (IVA-08)*, LNAI 5208, pages 215–222. Springer, 2008.

[8] F. W.-l. Ho-Ching, J. Mankoff, and J. A. Landay. Can you see what i hear?: the design and evaluation of a peripheral sound display for the deaf. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '03, pages 161–168, New York, NY, USA, 2003. ACM.

[9] M. Huenerfauth. A linguistically motivated model for speed and pausing in animations of american sign language. *ACM Trans. Access. Comput.*, 2:9:1–9:31, June 2009.

[10] M. Huenerfauth and V. L. Hanson. Sign language in the interface: Access for deaf signers. In C. Stephanidis, editor, *The Universal Access Handbook*. CRC Press, 2009.

[11] M. Huenerfauth, L. Zhao, E. Gu, and J. Allbeck. 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)*, pages 211–218, New York, NY, USA, 2007. ACM.

[12] M. Huenerfauth, L. Zhao, E. Gu, and J. Allbeck. Evaluating american sign language generation by native ASL signers. *ACM Transactions on Access Computing*, 1(1):1–27, May 2008.

[13] J. R. Kennaway, J. R. W. Glauert, and I. Zwitserlood. Providing signed content on the internet by synthesized animation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(3):15–29, 2007.

[14] M. Kipp. Anvil: The video annotation research tool. In J. Durand, U. Gut, and G. Kristofferson, editors, *Handbook of Corpus Phonology*. Oxford University Press, 2011. to appear.

[15] M. Kipp, A. Heloir, P. Gebhard, and M. Schröder. Realizing multimodal behavior: Closing the gap between behavior planning and embodied agent presentation. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents (IVA-10)*. Springer, 2010.

[16] M. Kipp, A. Heloir, and Q. Nguyen. Sign language avatars: Animation and comprehensibility. In

Proceedings of the 11th International Conference on Intelligent Virtual Agents (IVA), 2011. submitted.

[17] M. Kipp, M. Neff, and I. Albrecht. 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, December 2007.

[18] M. Kipp, Q. Nguyen, A. Heloir, and S. Mates. Assessing the deaf user perspective on sign language avatars. In *The 13th International ACM SIGACCESS Conference on Computers and Accessibility*, 2011.

[19] T. Matthews, J. Fong, and J. Mankoff. Visualizing non-speech sounds for the deaf. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, Assets '05, pages 52–59, New York, NY, USA, 2005. ACM.

[20] M. Neff, M. Kipp, I. Albrecht, and H.-P. Seidel. Gesture Modeling and Animation Based on a Probabilistic Recreation of Speaker Style. *ACM Transactions on Graphics*, 27(1):1–24, March 2008.

[21] R. Pfau and J. Quer. Nonmanuals: their prosodic and grammatical roles. In D. Brentari, editor, *Sign languages (Cambridge Language Surveys)*, pages 381–402. Cambridge University Press, 2010.

[22] H. Sharp, Y. Rogers, and J. Preece. *Interaction Design: Beyond Human-Computer Interaction*. Wiley, second edition, March 2007.

[23] M. Sheard, S. Schoot, I. Zwitserlood, M. Verlinden, and I. Weber. Evaluation reports 1 and 2 of the EU project essential sign language information on government networks, Deliverable D6.2, March 2004.

[24] R. Wolfe, J. McDonald, M. J. Davidson, and C. Frank. Using an animation-based technology to support reading curricula for deaf elementary schoolchildren. In *The 22nd Annual International Technology & Persons with Disabilities Conference*, 2007.