

2015-10

## Emotional Facial Expressions in Synthesised Sign Language Avatars: a Manual Evaluation.


Robert G Smith

*Technological University Dublin, robert.smith@tudublin.ie*

Brian Nolan

*Technological University Dublin, brian.nolan@tudublin.ie*

Follow this and additional works at: <https://arrow.tudublin.ie/itbinfooth>

 Part of the [Applied Linguistics Commons](#), [Computational Linguistics Commons](#), [Computer Sciences Commons](#), [Language Description and Documentation Commons](#), and the [Other Linguistics Commons](#)

### Recommended Citation

Smith, R.G. and Nolan, B. (2016). Emotional facial expressions in synthesised sign language avatars: a manual evaluation. *Universal Access in the Information Society*, 15(4), 567-576. DOI:10.1007/s10209-015-0410-7

This Article is brought to you for free and open access by the Computational Functional Linguistics at ARROW@TU Dublin. It has been accepted for inclusion in Other Resources by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#)  
Funder: ITB

# Emotional facial expressions in synthesised sign language avatars: a manual evaluation

Robert G. Smith<sup>1</sup> · Brian Nolan<sup>1</sup>

Published online: 23 May 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** This research explores and evaluates the contribution that facial expressions might have regarding improved comprehension and acceptability in sign language avatars. Focusing specifically on Irish sign language (ISL), the Deaf (the uppercase “D” in the word “Deaf” indicates Deaf as a culture as opposed to “deaf” as a medical condition) community’s responsiveness to sign language avatars is examined. The hypothesis of this is as follows: augmenting an existing avatar with the seven widely accepted universal emotions identified by Ekman (Basic emotions: handbook of cognition and emotion. Wiley, London, 2005) to achieve underlying facial expressions will make that avatar more human like and improve usability and understandability for the ISL user. Using human evaluation methods (Huenerfauth et al. in *Trans Access Comput (ACM)* 1:1, 2008), an augmented set of avatar utterances is compared against a baseline set, focusing on two key areas: comprehension and naturalness of facial configuration. The approach to the evaluation including the choice of ISL participants, interview environment and evaluation methodology is then outlined. The evaluation results reveal that in a comprehension test there was little difference between the baseline avatars and those augmented with emotional facial expression. It was also found that the avatars are lacking various linguistic attributes.

**Keywords** User-centred design · Deaf · Sign language synthesis · Emotion · Natural variance · Avatar · SiGML · HamNoSys · Accessibility · Disability · HCI

## 1 Introduction

ISL is the indigenous language of the Deaf community in Ireland, standing apart from English and Irish. There are approximately 5000 native users of ISL in the Republic of Ireland [3], while it is estimated that some 50,000 non-Deaf people also know and use the language to a greater or lesser extent [4]. Unlike spoken language, signed languages have multiple articulators designated: manual features (MF) that are the hands/arms and non-manual features (NMF), which include everything else. Emotion and prosody are expressed in SL primarily through NMFs [3], which are widely accepted to carry up to 70 % of a sign’s meaning,<sup>1</sup> and this, therefore, makes emotion a significant factor in the credibility and acceptance of an avatar. The average reading age of Deaf school leavers is comparable to that of an 8- to 9-year-old hearing child [5]. Thus, there is a requirement for communication materials in a sign language format, yet the costly production of sign language video using actors means that these materials are limited. Synthesised sign language avatars are a cost-effective solution to this requirement. This paper outlines how an existing synthesised avatar framework is currently being used to evaluate comprehension levels of signing avatars amongst a portion of the Irish Deaf community. Particularly, the study investigates the effect of adding emotional facial expressions (EFEs) and the advantages, if any, of a human-like avatar versus a caricature-like avatar.

---

✉ Robert G. Smith  
robert.smith@itb.ie

Brian Nolan  
brian.nolan@itb.ie

<sup>1</sup> School of informatics and engineering, Institute of Technology Blanchardstown, Blanchardstown Road North, Dublin 15, Ireland

<sup>1</sup> This statistic is primarily based on anecdotal evidence gathered during discussions with members of the Deaf community.

## 2 Evaluation framework

In order to conduct any form of evaluation, one must first have an avatar in place along with some synthesised output. With this in mind, a subset from the well-established Signs of Ireland (SOI) corpus [6] and the JASigning [7] synthesised sign language avatar system were chosen as the instruments to work with. These technologies are discussed further in the subsequent sections.

## 3 Elicitation

The development of a new corpus is not a straightforward process, particularly with regard to the elicitation of data. Common difficulties include time limitations, attracting participants, authenticity of the data collected, not to mention confidentiality and other ethical issues. For these reasons, the building of a corpus was never within the scope of this project. Currently there are only two corpora with ISL content: the aforementioned Signs of Ireland (SOI) corpus [6] and the patient–receptionist dialogue corpus [8]. The latter of these two corpora has been fully transcribed with HamNoSys and outputs using the JASigning platform. Using this, a corpus would save much time with the transcription process allowing it to be circumvented entirely. The patient–receptionist dialogue corpus is very much focused on the domain of patient–receptionist dialogue. This fact, plus the fact that the dialogue is staged, makes it well suited to its purpose: the machine translation of sign languages with a small dataset.

On the other hand, the SOI corpus is well established and, as one of the largest digitally annotated signed language corpora in Europe, it gives a rich selection of utterances with emotional facial expressions (EFEs). The primary purpose of the corpus is to record ISL as it is currently used in Ireland. As a result, the subjects of the corpus were encouraged to relax and sign naturally. Corpus data, in which the signer is relaxed and using his/her natural sign, may be the best material to impartially evaluate the comprehension of a sign language avatar.

Through a simple keyword search, the story *A Scare in Belfast* was selected from the SOI corpus and identified as having a high level of EFE content. A manual inspection of the data confirmed that all seven emotions were present. Each EFE was manually identified and annotated using the ELAN software in which the SOI corpus was initially constructed. Five segments of the story containing a high concentration and variety of emotional content were chosen as the best candidates for the evaluation.

The SOI corpus contains natural, authentic sign language usage. However, as is true of all systems, errors may occur. For example, during the elicitation process:

**Table 1** Frequency of which each EFE appears

EFE	Frequency
Happy	18
Disgust	15
Anger	10
Fear	10
Contempt	8
Surprise	4
Sad	2

- The content may be recorded with an incorrect or lesser-known sign.
- In the transcription process, the transcriber may incorrectly identify a sign.
- Further down the pipeline, the software may display a sign incorrectly.

In an effort to avoid such errors, content was selected from the well-established SOI corpus. The content from this corpus, being ‘natural’ conversational signing produced by native ISL users, is difficult to challenge in its authenticity. One area in which the creators of the corpus would not mind being challenged is the accuracy of the ISL. There are many linguistic ‘errors’ throughout this corpus as a result of grammatical ‘mistakes’ or the introduction of a non-ISL sign. However, these ‘errors’ are representative of how ISL users currently use the language and are exactly what the SOI corpus was designed to record. By using the SOI corpus, some content that may indeed have some linguistic ‘errors’ was gathered though more importantly, it has an authenticity that cannot be challenged. In an effort to avoid errors with regard to the HamNoSys transcription, only one experienced transcriber was involved in transcribing the content and only a small set of 154 utterances was transcribed so that time could be allocated to rechecking transcriptions. A breakdown of the utterances by emotion is provided in Table 1. To avoid software errors, the JASigning framework was used, which is currently the state-of-the-art tool for sign language synthesis. It has inherited many of the limitations of HamNoSys, and some rendering bugs have yet to be resolved. Overall it was found to be a useful and proficient tool with a very useful modular structure.

## 4 Baseline system

The Virtual Humans group at the University of East Anglia (UEA), Norwich, are leaders in the field of synthesised SL avatars; this accomplishment is a result of their JASigning framework [7, 9]. Initially conceived during the ViSiCAST project and further developed during the eSIGN and subsequent projects, the framework, the driving force behind



**Fig. 1** Avatars Luna and Anna

avatars such as Anna (see Fig. 1), was developed with a modular structure such that researchers not associated with the initial project can easily pick up the technology and build it.

Figure 2 gives a high-level overview of the JASigning framework. The framework can receive input in the form of HamNoSys. The Hamburg Notation System (Fig. 3) or HamNoSys, pronounced: “Ham.No.Sys”, is one of the few well-established transcription systems and developed by the Institute for German Sign Language and Deaf communication at the University of Hamburg for all SLs [10]. HamNoSys is a phonetic notation system purpose-built for use by linguists in their detailed analytical representation of signs and sign phrases as opposed to a writing system for SLs.

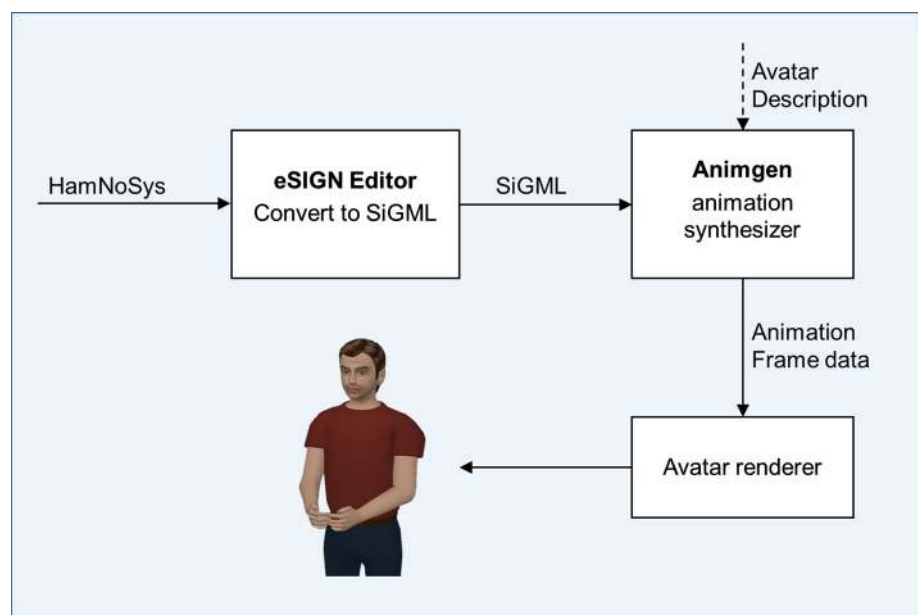
A transcriber may represent a signed utterance at the phonetic level using HamNoSys. In this work, all transcriptions were carried out by the author. The HamNoSys

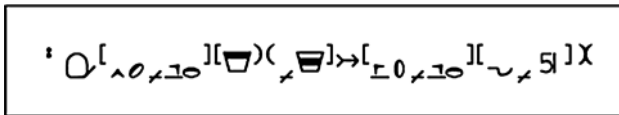
is then represented in the computer readable markup language known as SiGML (Signing Gesture Markup Language) [11] (See Fig. 4). SiGML defines a set of XML tags for each iconic symbol in HamNoSys. The eSIGN Editor tool does this automatically and has the ability to output SiGML at this point or send it to the animation synthesiser, AnimGen. AnimGen enriches the SiGML data with the avatar geometry data such as vertex coordinates and rotation values. These combined data are fed into the avatar-rendering engine which will produce a 3D avatar in real time.

#### 4.1 Improving the baseline system

Many of the limitations of the eSign Editor can be overcome using the existing framework. What makes JASigning popular amongst researchers is the modular structure. The framework has been developed so that a vast amount of fine-tuning or basic changes may take place without having to delve into a labyrinth of code. The framework takes input from a number of external XML and property files, which can be altered to affect facial morphs and movement speeds and to change avatar and a whole lot more. For this evaluation, it was possible to create seven new facial morphs using the ARToolkit [11]. The new morphs contain the facial configuration and movement for each of the seven emotions outlined by Ekman [1]. These are as follows: happiness, sadness, anger, disgust, contempt, fear and surprise. Using the ARToolkit, it is possible to export the files required to create and run an avatar locally using the SiGML service player, which, by default, pulls down the avatar data from a server based in UEA. For

**Fig. 2** JASigning framework (dataflow)





**Fig. 3** HamNoSys transcription for the sign “Stop”

```

<hns_sign gloss="BEGIN">
  <hamnosys_nonmanual>
    <hnm_mouthpicture picture="BE:"/>
    <hnm_head tag="TL"/>
    <hnm_extramovement tag="X69"/>
  </hamnosys_nonmanual>
  <hamnosys_manual>
    .
  </hamnosys_manual>
</hns_sign>
  
```

**Fig. 4** SiGML notation including ‘extra movement’ tag

these new morphs to take effect, some XML files need to be updated and the property file needs to be pointed to the newly created, local avatar. After identifying, visually, which signed utterances required the addition of EFEs, the final step was to add the new markup to each file. By manually adding the EFEs markup to each SiGML file, it was possible to circumvent the limitations of the eSign Editor, therefore enriching the existing output while using the existing framework. Figure 3 illustrates how EFE is added using the ‘hnm\_extramovement’ tag, where ‘X69’ represents the EFE ‘disgust’.

## 5 Evaluation

The multichannel visual nature of the sign language avatars as well as the requirement to identify a level of understandability made an automatic evaluation unfeasible, leaving a manual evaluation as the only viable option. A manual evaluation was undertaken with fifteen sign language users over a 2-day period on site at the newly developed Deaf Village of Ireland (DVI). The evaluation was designed such that all participants are native ISL users and a demographic balance was achieved. Barriers such as different levels of technical knowledge and pre-formed opinion of the technology would be identified early in the interview. Some barriers, like communication, for example, were overcome with the support of a certified ISL interpreter.

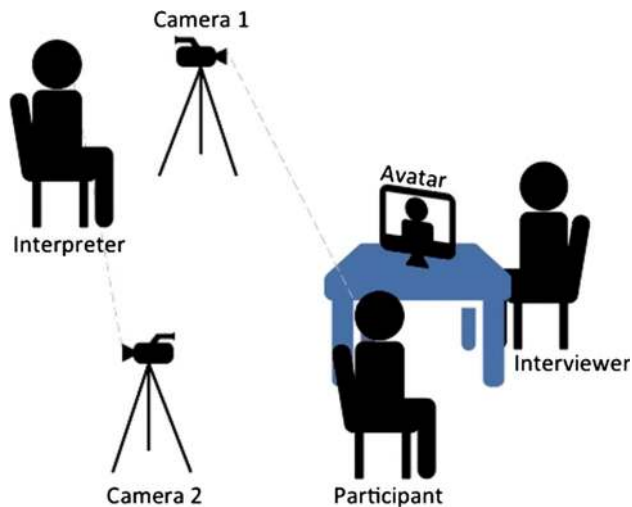
All of the five story segments selected were recreated as closely as possible to the original corpus entry using the JASigning platform described in Sect. 4, resulting in a set of digital avatar animated videos varying in duration from

**Table 2** Avatars used

	EFE encoding	Realism
Anna baseline	No	Human looking
Luna baseline	No	Caricature
AnnaE	Yes	Human looking
LunaE	Yes	Caricature

9 to 73 s. Each of the five story segments was present with one of four different avatars: (a.) Anna, a human looking avatar with baseline encoding, (b.) Luna, a caricature avatar, again with baseline encoding and both (c.) AnnaE and (d.) LunaE enriched with EFEs (see Table 2). This resulted in a total of twenty digital avatar animated videos, hereafter referred to simply as ‘videos’ or ‘avatar videos’. Each participant was presented the videos in a different order, the sequence of which was derived using a Latin square model in an effort to avoid learning. To further this effort and to lessen the interview duration, no participant saw all five story segments. The longest video was always shown in isolation or with one other to prevent fatigue in the participant. After watching each video, the participants were asked a number of comprehension questions as well as being asked to score their own comprehension of the video content on a scale of zero to five. During a trial run of the evaluation, it became obvious that some context was required and each video would need to be watched a second time; therefore, the same set of questions were asked after both viewings in a bid to track the level of comprehension after each pass.

The recruitment of voluntary participants was challenging given the closed nature of the Deaf community. Nevertheless, thanks to the efforts of the Irish Deaf Society (IDS) a total of fifteen participants took part. Evaluations, each 30 minutes in duration, took place over a 2-day period. Participants were asked a series of questions in an interview scenario. A digital video camera designated ‘camera 1’ captured footage of each interviewee as he/she watched the avatar videos and responded to the interviewer’s questions. A second camera, ‘camera 2’, filmed the ISL interpreter as he/she interpreted the conversation between participant and the interviewer. The room layout is illustrated in Fig. 5. The format of the interview stayed consistent throughout. Participants initially answered a set of establishing questions consisting of demographic information as well as some exploratory questions designed to establish their level of exposure and acceptance towards new technologies with particular focus on signing avatars. The participants had their first glimpse at the avatars in phase two. In this phase, each participant was asked to watch an avatar video and then answer some comprehension questions based on that video. The video



**Fig. 5** Evaluation room layout

was viewed a second time, and the same set of comprehension questions was asked again. This process was repeated for each video in a given video set as designated by the Latin square model. The final phase of the interview, phase three, was designed to allow the participants direct feedback regarding each avatar. The focus was primarily on the participants' acceptance/non-acceptance of the avatars, what use they might see for them in the future and how their own views may have changed since seeing the avatars in person.

## 6 Results

Demographically, a broad range of participants took part in the evaluation. All fifteen participants were aged between 19 and 60, with 60 % of those falling into the 31 to 40 age bracket. There were a comparatively even number of males to females with female participation slightly lower at 40 %. As the evaluation took place in Dublin, it is not surprising that 67 % of participants were from the province of Leinster. Munster was the only province with no representation as representatives of the other two provinces, Ulster (30 %) and Connaught (13 %), took part. Of all the participants, 93 % listed ISL as their first language with 87 % attending a Deaf-only school as a child. On a scale of zero to five, all participants ranked themselves either four or five for ISL competency, 87 % ranking themselves a five. 27 % of participants studied ISL at third level.

### 6.1 General findings

During the first phase of the interview, before having been shown the avatars, 40 % of all participants declared that they had never been exposed to signing avatar technology

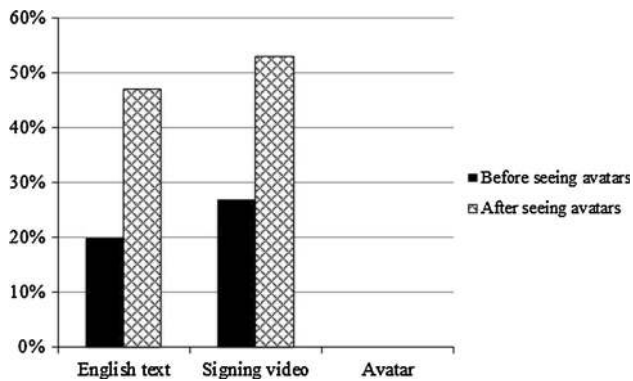
before. The remainder indicated only limited exposure, with only 7 % having had hands-on experience of the technology. Surprisingly, 20 % of participants indicated no interest in 3D graphics, including 3D animated movies. When asked whether difficulties might arise when introducing avatar technology to the Deaf community, 67 % of participants said there would be some difficulties. The majority of these cited: the lack of facial expression and robotic-like movement as the primary factors in this. All participants indicated a preference for a human signer. 33 % of participants fear that signing avatars will replace sign language interpreters in the future, and 60 % indicated a willingness to use this technology if it improves to an acceptable point.

73 % of participants declared themselves as having a general interest in new gadgets and technologies, identifying smartphones and tablets as their most used gadgets. When asked whether they prefer web content to be word-based or signed video,<sup>2</sup> 53 % said they would prefer content in both formats, 27 % would prefer signing video only, and the remaining 20 % would prefer English text. 33 % of the participants stated that they often have problems reading English text on the web. Participants stated that this was a common issue on websites with a lot of jargon or advanced English.

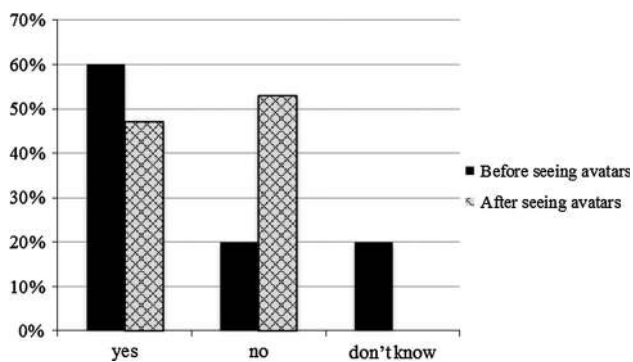
In phase three of the interview, after watching the avatar videos, participants were asked which medium is preferable for web content. 53 % of participants listed signing video as their first choice for web content and the remaining 47% listed written English as their first choice (Fig. 6). It is noteworthy that not one participant selected a signing avatar as their first choice for web content. Yet, 27 % did choose avatars as their second choice, and 73 % chose avatars as their third choice. When asked directly whether they would use a signing avatar video, 47 % said they would if the avatar was of a high enough quality. This is a 13 % decrease from the 60 % acceptance rate recorded in the first phase of the interview (see Fig. 7). The fact that 90 % of participants said that the avatars' movements do not look natural is a definite factor in this. Frequently, participants stated that the avatars looked "stiff" and "robotic" and required a "lot of effort to read". When asked whether the avatars had been easy to understand, 50 % said "no", 10 % said "yes", and 40 % said "sometimes".

As to whether the participants preferred a caricature avatar (Luna) or a more human-like option (Anna), 40 % preferred Luna, 50 % preferred Anna, and 10 % said they had no preference either way (Fig. 8). Generally, participants commented that Anna would be a better choice of

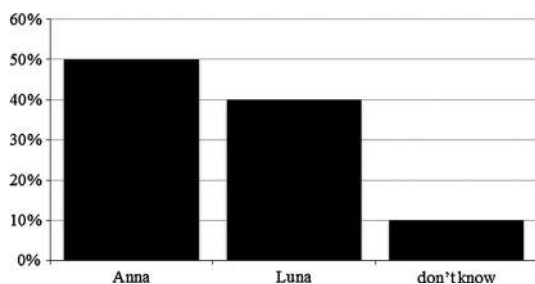
<sup>2</sup> A pre-recorded video of a 'real person' using sign language to provide an alternative to text on the web.



**Fig. 6** Participants' preference of medium (text, video and avatar)



**Fig. 7** Would you use a signing avatar video?



**Fig. 8** Avatar preference

avatar for formal content, whereas Luna would be best suited to content for children. A number of participants mentioned that Luna's longer fingers worked well and Anna's face is better suited to deliver facial expression. Luna's larger eyes received a mixed reaction; some felt they made the avatar more engaging while other participants considered them too big, while one participant mentioned that they were alien like. In an effort to quantify these comments, each positive comment was assigned a weight of '1' and each negative comment was assigned a weight of '-1'. These weightings were recorded in a matrix for each of the most commonly remarked upon attributes. A summary of this matrix is displayed in Table 3.

**Table 3** Attributes—Anna versus Luna

Attribute	Anna	Luna
Emotion	-3	-3
Facial movement	-1	-3
Eyes (engaging)	0	2
Eyes (size)	0	-4
NMF amount	-3	-3
Fingers/hands/arms	0	2
Body movement	-1	0
Naturalness	-4	-8
Presence	-2	0
Content	-3	-4
Clear signing	-1	0
Finger spelling	-4	-4
Singing space	-3	-3
Timing/flow	-1	-3
Clothes/hair/colours	0	-1
Suitable for adults	2	0
Suitable for kids	0	2
Total	-24	-30

It is clear that many of the attributes and characteristics that were strongly disliked such as emotion, naturalness, NMF amount, finger spelling and signing space were related to linguistic clarity and linguistic performance of the avatar. This would suggest that the avatars perform poorly with the more fundamental linguistic aspects of ISL. For the most part, the attributes that scored greater than or equal to zero are more aesthetic in nature and may lend themselves more to personnel taste.

Again, in phase three, participants were asked: Whether the technology was improved, where could this technology be used in the future? 80 % of participants would like to see the technology used to translate web content, 47 % said it may be a valuable teaching aid or suit a classroom environment, 43 % believe it suitable for television signing, and only 17 % think that it could be a suitable replacement for live interpreters in a sensitive setting. Other uses suggested include: social networking, a VOIP alternative, console gaming and video relay interpreting (see Fig. 9).

## 7 Comprehension

Results indicate that participants, when directly asked, underrated their own comprehension on each avatar video shown. Figure 10 shows that, on average, participants' self-score across all avatars is at 46 %. This is considerably lower than the score achieved in the comprehension exercise 60 %. In the case of avatars that had been enriched with EFEs, the self-applied score was 14 % lower. At the

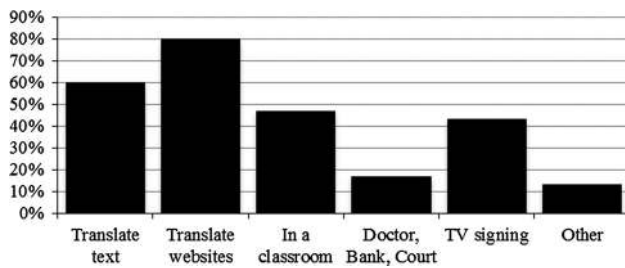


Fig. 9 Possible use for avatar technology

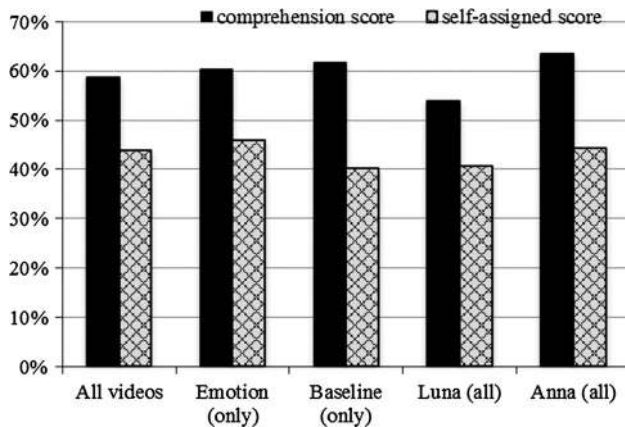


Fig. 10 Comprehension score versus self-assigned score

other extreme, in the case of the avatar Anna (with and without EFEs), the score was 44 % lower. This indicates that the participants’ perceived comprehension is substantially lower than their actual comprehension, which may be one reason for the low uptake of this technology amongst the Deaf community.

The most surprising result was the difference in comprehension score between baseline avatars and those augmented with EFEs. The results indicate that participants understood 62 % of the content delivered through the baseline avatars, yet when EFE was added the comprehension level fell to 60 % (Fig. 10). This would seem to indicate that instead of improving comprehension, the addition of EFE had a negative effect albeit marginal.

A further breakdown of the results in Fig. 11 gives a clearer picture as to how each of the four avatars performed. The AnnaE avatar recorded a higher comprehension score than the LunaE avatar scoring, 64 and 54 %, respectively. Anna also scored higher with the baseline encoding, scoring 4 % higher than Luna with 63 and 59 %, respectively. Again, the gulf between the EFE and baseline avatars may be noted.

These results demonstrate that the addition of EFEs for comprehension was more successful with the ‘human looking’ avatar than with the caricature avatar. In addition,

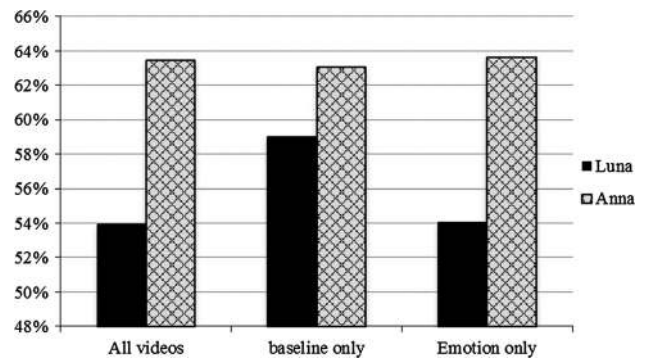


Fig. 11 Average comprehension score by avatar

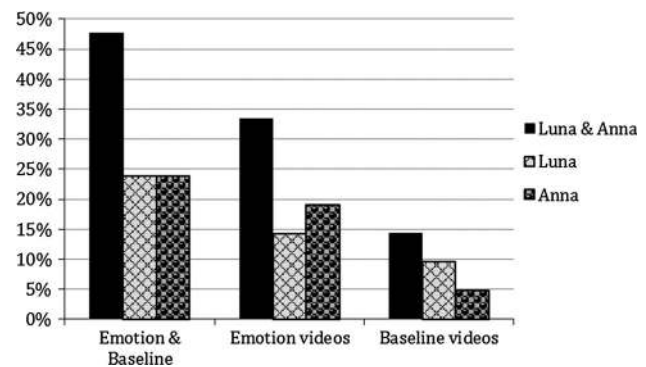


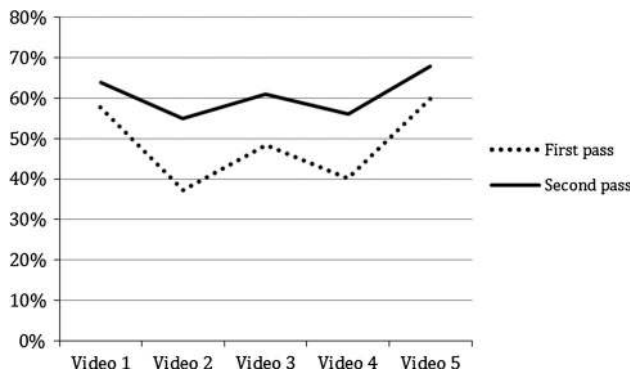
Fig. 12 Did you see emotion?

these results also confirm that regardless of EFEs, Anna was the easier avatar to comprehend.

After watching each avatar video, participants were asked whether they had seen emotion. Figure 12 shows that participants recognised emotion in 48 % of all videos. 33 % of the videos in which emotion was identified; EFEs were added to the baseline coding. Emotion was also identified in 14 % of videos with no additional EFEs. This may be due to participants incorrectly identifying basic facial movement as an attempt at EFEs. In addition to this, participants’ remarks indicate that Luna’s permanent smile was a cause of some confusion.

As stated earlier in this paper, each participant was asked to watch each video twice. For example, after watching a video once, a participant would be asked a series of comprehension question and then directly afterwards the participant was asked to view the video a second time and asked the very same set of questions again. It is acknowledged at this point that the score for the second pass is skewed by a degree of learning. It must also be noted, however, that due to inexperience with the technology, participants struggled to capture any information from the first viewing of each video as previously experienced during a trial evaluation.





**Fig. 13** Average comprehension score—first and second viewing

Figure 13 illustrates a comparison between the average comprehension score achieved based on each video for the first and second viewings, including EFE and baseline scores for all avatars. It is clear that comprehension scores are higher after the second viewing of each video. The difference between the score for the first and second pass ranges from 6 % for video 1 up to 18 % for video 2. The most possible cause of this is the video content. Video 2 contains finger spelling, place names, role shift and classifiers; although all of the videos contain these to some extent, video 2 has a higher concentration. This also accounts for the fact that video 2 has the lowest average comprehension score in both the first and second pass. The second trough in the graph represents a lower comprehension score for video 4. At 73 s and 77 utterances, video 4 is the longest video in the set. It also contains much of the same difficult content as video 2. Videos 1 and 5 are two of the shortest videos in the set and contain little of the difficult content described for video 2 and 5.

## 8 Conclusion

As with all studies, the findings outlined in the previous section must be considered in context to other, similar works. One similar study is that of Schnepf et al. [12]. The findings outlined in this paper neither affirm or disaffirm the findings of Schnepf. Although the comprehension and acceptance results do differ greatly between the two studies, the source utterances are too dissimilar to draw a direct comparison between the two studies. One may draw the conclusion from both studies that simple direct utterances, such as those used by Schnepf, are more suited to signing synthesis than the more complex examples of ‘natural discourse’ used in this study. Further investigation is required to confirm or deny this hypothesis. Also worth noting is a feasibility study conducted by Kipp et al. [13]. Unlike the study outlined in this paper, which primarily

investigates comprehension of signing avatar utterances, Kipp’s study was primarily aimed at obtaining opinions about signing avatars as well as new uses for avatar technology going forward. Kipp argues that “future research needs to focus on non-manual aspects and prosody to reach the comprehensibility levels of human signers”. A conclusion that corresponds to the findings of this paper is discussed in this section. It should be highlighted at this point that it is often difficult to draw direct comparisons between competing signing avatar frameworks. These frameworks are based on multiple, often dissimilar, layers of technology, an example of which is illustrated in Fig. 2. It may be the case that, due to the animation synthesiser or rendering engine, one avatar will sign more smoothly than another. It is also possible that the lower level transcription, elicitation or markup structure affects the final performance of the avatar. The multiple points of error equate to a difficulty to directly compare studies involving different signing avatar frameworks. This is particularly the case when focusing on one attribute such as EFEs.

The primary focus of the evaluation was to ascertain whether or not the addition of emotional facial configuration increased the understandability of a signed utterance. The results presented here would indicate that this is not the case. In fact, Fig. 10 shows that the addition of EFEs made very little impact with the score for the baseline avatars and the EFEs augmented avatar almost identical, overall having a marginally negative effect of  $-2\%$ . These findings are inconsistent with those of the aforementioned Schnepf et al. [12]. This is most likely due to the elicitation methodologies used, resulting in 2 distinct levels of linguistic and prosodic complexity of utterances presented to the participants of each study.

Also evident from the results is the higher comprehension levels achieved with the avatar Anna. Anna was designed to be as close to human looking as possible while using lower levels of 3D data for speedy rendering. This result could have a significant impact on future development of sign language avatars and their facial configuration. Commonly, participants commented that Anna looked quite the serious avatar and that Luna may be better suited for children. It was also suggested that a repertoire of avatars be available for various tasks. Such a repertoire would have a place for both Anna and Luna. The fact remains, however, that regardless of preference, participants understood Anna better than Luna. Participants’ remarks and the results highlighted in Figs. 8 and 11 are enlightening as to a possible reason for this: the EFEs are more easily identified in the AnnaE avatar. The difference in participants own perception of emotion recognition between Luna and Anna is marginal at 5 % (Fig. 12); however, when the relatively high false positive of the baseline Luna avatar (10 %) is considered, one can surmise that the

participants, at least 42 % of the time, falsely identify emotion in Luna. This is most likely due to the avatars' perpetual smile (see Fig. 1).

Figure 13 indicates a comprehension score of between 55 and 68 % (or an average of 61 %) on the second viewing of the videos and an average of 49 % on the initial viewings. The most common use suggested for this technology was the translation of websites (Fig. 9); in that instance, given the level of control provided to the user for video on the web, the score achieved after the second viewings is relevant. For practically every other purpose, again see Fig. 9, the scores achieved after the first viewing are of the utmost importance. These figures are encouraging but show that there is much work yet to be done before the various Deaf communities can use these avatars widely. As to why the mean comprehension level is low, particularly on the first viewing, Table 3 highlights a number of attributes of linguistic importance that scored badly amongst participants. One must surmise that these linguistics attributes are directly linked to the participants' comprehension and indeed the perceived comprehension scores reported in Fig. 10. Although the average comprehension scores indicate only a minor effect of EFEs, Table 3 indicates that attributes such as emotion and NMF are desired by the Deaf community and, furthermore, are required to improve comprehension.

It can be seen in Fig. 6 that, predictably, the majority of participants preferred signing video with a real person for web content. It was surprising, however, to see that there was an almost even split in those that chose English text and signing video. This revelation would seem to contradict most of the literature available on the level of the Deaf community's literacy skills [5]. This may, in part, be a result of the relatively young demographic: 89 % are less than 50 years of age, and 67 % are less than 40. Another contributing factor to this revelation may be the 73 % interested in new technologies reported in Sect. 6.1. It is reasonable to infer that daily use of mobile devices such as smartphones and tablet computers for casual web browsing, SMS and email would result in more exposure to the written word and therefore a higher level of literacy.

It is interesting to see (in Fig. 7) that, despite 60 % of participants indicating a willingness to use this technology before seeing the avatar videos, only 47 % held that view after viewing the avatar videos with the caveat of increased performance. This indicates that the avatar quality presented was below the standard that was anticipated by the participants. This is compounded by low perceived comprehension score (Fig. 10) in addition to the results illustrated in Fig. 6, in which no participant chose avatar video as their first choice of web content and only 27, % chose it as their second choice. All hope is not lost; however, hearts may be lightened by the 47 % willingness to engage with

the technology as well as the 20 % of participants who answered "Don't know" and, of course, the willingness of participants to elect some potential uses for the technology in the future (Fig. 9).

Qualitative feedback suggests the avatars are an applicable technology that has not yet evolved to a point for mainstream use. Common remarks include "robotic", "unnatural", "stiff" and, as mentioned by one participant, "avatary". This feedback alongside the statistic that 90 % did not think the avatars looked natural demonstrates that there is still a lot of work to be done with regard to the avatars movement. Feedback relating to the speed and timing of signs illustrates a need for work in this area, in particular an appropriate synchronization of manual feature and non-manual feature and timing at the sign level, particularly for finger spelling. Finally, feedback regarding facial movement and emotional expression indicates that there is still quite a long way in this case also. Although some change in facial configuration may be applied at the texture-map and polygon morph levels, an improvement in the naturalness of movement and timing has a huge effect on facial movement also and perhaps these are a more suitable place to begin making changes.

## 9 Future work

Further work is necessary in order to achieve a usable, comprehensible avatar with particular focus on the linguistic attributes that fared badly in Table 3. A further investigation would be beneficial to identify why these attributes fared badly and how best to deliver a solution that will not only address these attributes but, by proxy, increase the comprehension level also.

**Acknowledgments** This research is supported by The School of Informatics and Engineering at the Institute of Technology Blanchardstown, Dublin. The authors would like to acknowledge the support of colleagues at UEA and the efforts of the Irish Deaf Society, particularly Susan Whelan and the participants who freely offered their time for this evaluation.

## References

- Ekman, P.: *Basic Emotions: Handbook of Cognition and Emotion*. Wiley, West Sussex, England (2005)
- Huenerfauth, M., Zhao, L., Gu, E., Allbeck, J.: Evaluation of american sign language generation by native ASL signers. *Trans. Access Comput.* 1, 1 (2008)
- Matthews, P.: *The Irish Deaf Community, Survey Report, History of Education, Language and Culture*, vol. 1. The Linguistics Institute of Ireland, Dublin (1996)
- Leeson, L.: *Aspects of Verb Valency in Irish Sign Language*. Unpublished Doctoral Dissertation, Centre for Language and Communication Studies, TCD, Dublin (2001)

5. Conrad, R.: *The Deaf School Child: Language and Cognitive Function*. HarperCollins Publishers, London (1979)
6. Leeson, L., Nolan, B.: Digital deployment of the signs of ireland corpus in elearning. In *LREC, 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora.*, pp. 112–121. Marrakech (2008)
7. Kennaway, R.: Experience with and requirements for a gesture description language for synthetic animation. In *Gesture-Based Communication in Human–Computer Interaction*, pp. 300–311. Springer, Berlin (2003)
8. Morrissey, S., Somers, H., Smith, R.G., Gilchrist, S., Dandapat, S.: Building a sign language corpus for use in machine translation. In *LREC, 4th Workshop on Representation and Processing of Sign Languages Corpora and Sign Language Technologies*, vol. 172–177, Valletta, Malta. European Language Resources Association (2010)
9. Smith, R.G., Somers, H., Morrissey, S.: HCI for the deaf community: developing human-like avatars for sign language synthesis. In *4th Irish Human Computer Interaction Conference (IHCI)*. Dublin, Dublin City University (2010)
10. Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J.: Hamnosys version 2.0: Hamburg notation system for sign languages—an introductory guide. *International Studies on Sign Language and Communication of the Deaf* 7, 225–231 (1989)
11. Elliott, R., Glauert, J., Jennings, V., Kennaway, R.: An overview of the SiGML notation and SiGMLSigning software system. In *LREC, Workshop on the Representation and Processing of Sign Languages.*, pp. 98–104 (2004)
12. Schnepf, J., Wolfe, R., McDonald, J.: Synthetic corpora: a synergy of linguistics and computer animation. In *proceedings of 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, Valetta, Malta (2010)
13. Kipp, M., Heloir, A., Nguyen, Q.: Sign language avatars: animation and comprehensibility. In *Intelligent Virtual Agents*, pp. 113–126. Springer (2011)