

2013

Emotional Facial Expressions in Synthesised Sign Language Avatars: A Manual Evaluation

Robert G. Smith Mr

Institute of Technology Blanchardstown, robert.smith@tudublin.ie

Brian Nolan

Institute of Technology, Blanchardstown, brian.nolan@tudublin.ie

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



Part of the [Linguistics Commons](#)

Recommended Citation

Smith, Robert G. Mr and Nolan, Brian (2013) "Emotional Facial Expressions in Synthesised Sign Language Avatars: A Manual Evaluation," *The ITB Journal*: Vol. 14: Iss. 1, Article 2.

doi:10.21427/D7XF1F

Available at: <https://arrow.tudublin.ie/itbj/vol14/iss1/2>

This Article is brought to you for free and open access by the Ceased publication at ARROW@TU Dublin. It has been accepted for inclusion in The ITB Journal by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](#)

Emotional Facial Expressions in Synthesised Sign Language Avatars: A Manual Evaluation

Robert Smith, Brian Nolan

Department of Informatics, Institute of Technology Blanchardstown, Dublin, Ireland.

{robert.smith, [brian.nolan](mailto:brian.nolan@itb.ie)}@itb.ie

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), we examine the Deaf¹ community's responsiveness to sign language avatars. The hypothesis of this is: Augmenting an existing avatar with the 7 widely accepted universal emotions identified by Ekman [1] 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 [2] we compare an augmented set of avatar utterances against a baseline set, focusing on 2 key areas: comprehension and naturalness of facial configuration. We outline our approach to the evaluation including our choice of ISL participants, interview environment and evaluation methodology. The evaluation results reveal that in a comprehension test there was little difference between the baseline avatars and those augmented with emotional facial expression also we found that the avatars are lacking various linguistic attributes.

Keywords

User-centered 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 5,000 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) which are the hands/arms and non-manual features (NMF), 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 meaning² 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 8-9 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 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

¹ The uppercase "D" in the word "Deaf", indicates Deaf as a culture as opposed to a medical condition.

² Anecdotal evidence based on interaction with the Deaf community [13]

expressions (EFEs) and the advantages, if any, of a human-like avatar versus a caricature-like avatar.

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 regards 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 2 corpora with ISL content: the aforementioned Signs of Ireland (SOI) corpus [6] and the patient–receptionist dialogue corpus [8]. The later of these 2 corpora has been fully transcribed with HamNoSys and outputs using the JASigning platform. Using this 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.

The SOI corpus on the other hand is well established and as one of the largest digitally annotated signed language corpus 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 7 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 contained a high concentration and variety of emotional content were therefore 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 the content may be recorded with an incorrect or lesser-known sign, in the transcription process the transcriber may incorrectly identify a sign or further down the pipeline the software may display a sign incorrectly. In an effort to avoid such errors we selected content from the, well-established, SOI corpus. The content from this corpus being ‘natural’ signing 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 content we have gained some content that may indeed have some linguistic ‘errors’ but, more importantly, it has an authenticity that cannot be challenged. In an effort to avoid errors with regards 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 3-1. To avoid software errors we used the JASigning framework, 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 we found it a useful and proficient tool with a very useful modular structure.

Table 3-1 The frequency of which each EFE appears

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



Figure 1 Avatars Luna and Anna

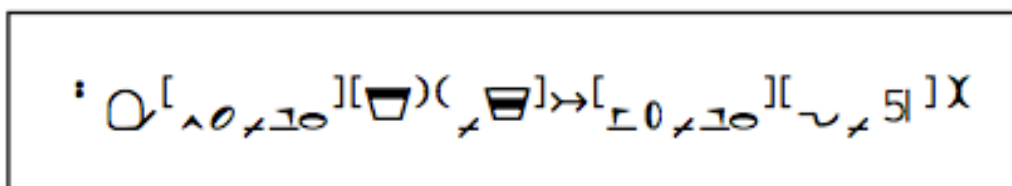


Figure 2 HamNoSys transcription for the sign "Stop"

```

<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>

```

Figure 3 SiGML notation including 'extra movement' tag

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 avatars such as Anna (Figure 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 4 gives a high level overview of the JASigning framework. The framework can receive input in the form of HamNoSys. The Hamburg Notation System (Figure 2) or HamNoSys 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 as 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 1st author. The HamNoSys is then represented in the computer readable markup language known as SiGML (Signing Gesture Mark-up Language) [11] (See Figure 3). 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. This combined data is 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, movement speeds, 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 being: 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 these new morphs to take effect, some XML files need to be updated and the properties 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. 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 15 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 5 story segments selected were recreated as closely as possible to the original using the JASigning platform described in section 4, resulting in a set of digital videos varying in duration from 9 seconds to 73 seconds. Each of the 5 story segments was present with 1 of 4 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 5-1). This resulted in a total of 20 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 5 videos. The longest video was always shown in isolation or with 1 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 0-5. 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 15 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 interviewers 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 Figure 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 2. In this phase each participant was asked to watch an avatar video and then answer some comprehension questions based on that video. The video 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 participants’ video-set as designated by the Latin square model. The final phase of the interview, phase 3, was designed to allow the participants direct feedback regarding each avatar. Focusing 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.

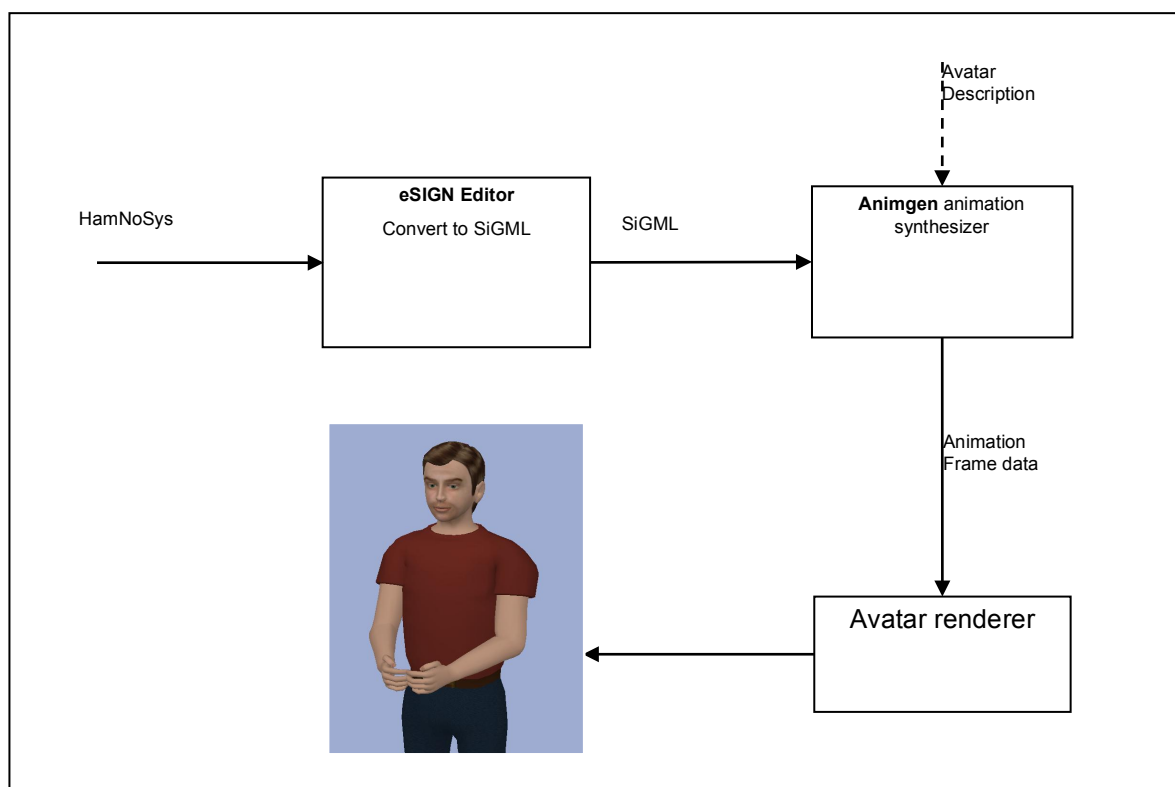


Figure 4 the JASigning framework

Table 5-1 Avatars used

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

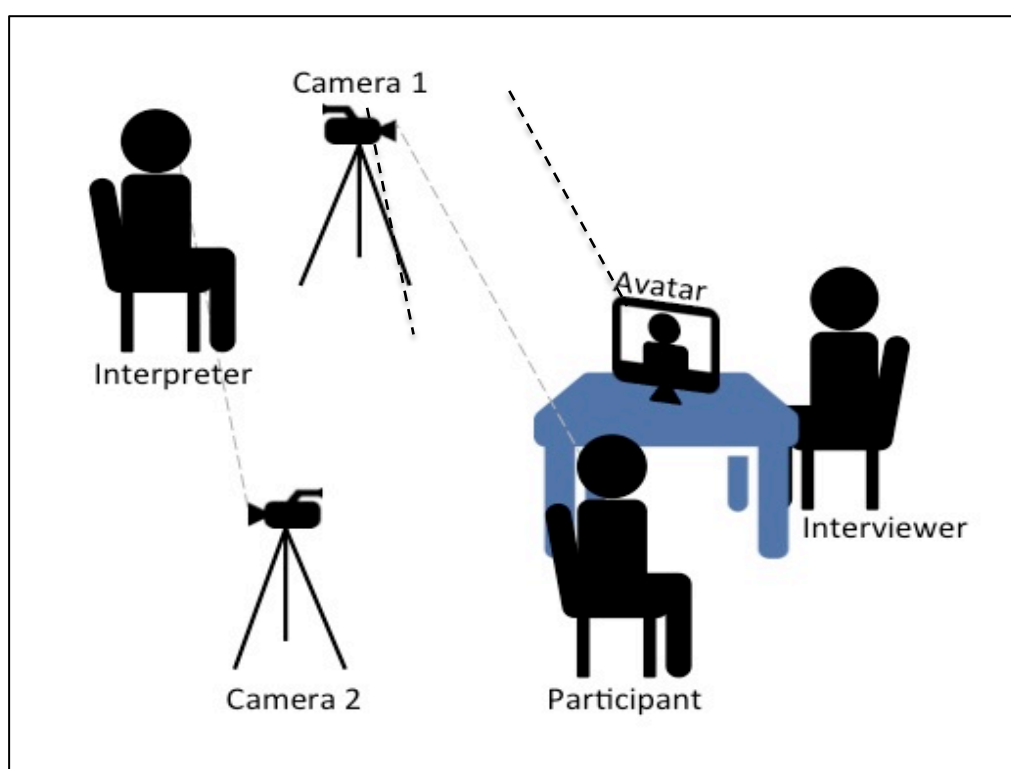


Figure 5 Evaluation room layout

6. RESULTS

Demographically, a broad range of participants took part in the evaluation. All 15 participants were aged between 19 and 60, with 60% of those falling into the 31 to 40 age bracket. There was 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 2 provinces: Ulster 30% and Connaught 13% took part. 93% of the participants listed ISL as their first language with 87% attending a Deaf-only school as a child. On a scale of 0 to 5, all participants ranked themselves either 4

or 5 for ISL competency, 87% ranking themselves a 5. 27% of participants studied ISL at 3rd level.

6.1 General findings

During the 1st phase of the interview, before participants had 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 if 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 if they prefer web content to be word-based or signed video³, 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 3 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 (Figure 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 if 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 Figure 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”, “robotic” and “required a lot of effort to read”. When asked if the avatars had been easy to understand, 50% said “no”, 10% said “yes” and 40% said “sometimes”.

³ A pre-recorded video of a ‘real person’ using sign language to provide an alternative to text on the web.

Table 6-1 Attributes - Anna vs. 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

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 (Figure 8). Generally, participants commented that Anna would be a better choice of 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, 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 6-1. It is clear that many of the attributes and characteristics that were strongly disliked such as emotion, naturalness, NMF amount, fingerspelling 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 ≥ 0 are more aesthetic in nature and may lend themselves more to personnel taste.

Again, in phase 3, participants were asked: if 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 Figure 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-scoring across all avatars, 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 other extreme, in the case of the avatar Anna (with and without EFEs), the score was 44% lower. This indicates that the participant's 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% (Figure 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 Figure 11 gives a clearer picture as to how each of the four avatars performed. AnnaE recorded a higher comprehension score than LunaE scoring 64% and 54% respectively. Anna also scored higher with the baseline encoding, scoring 4% higher than Luna with 63% and 59% respectively. Again we note the gulf between the EFE and baseline avatars.

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

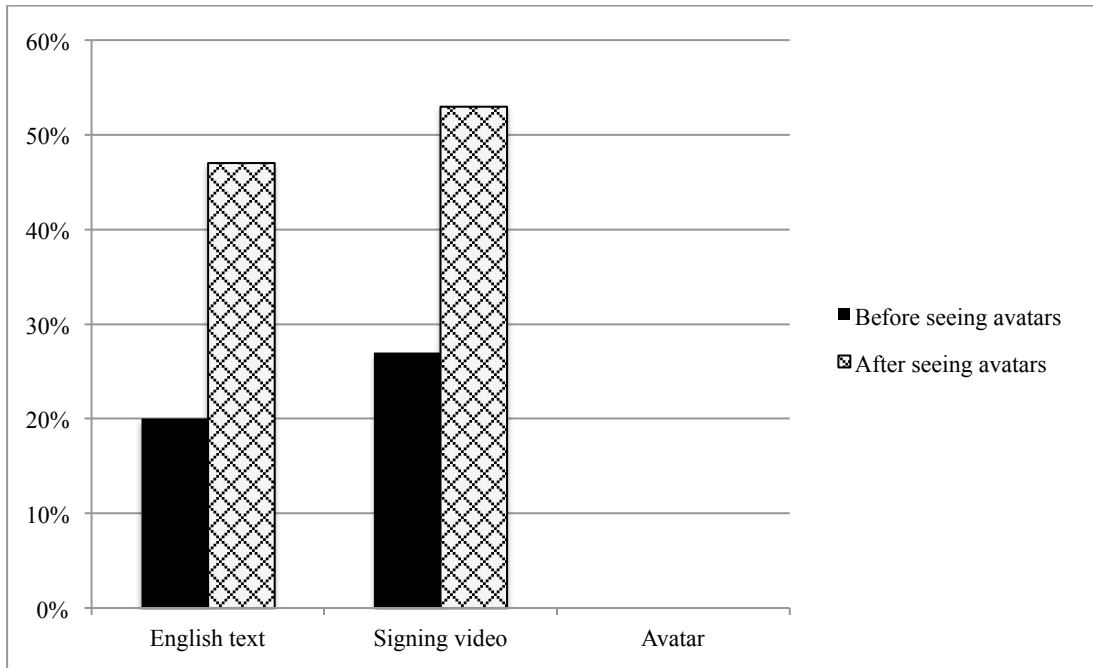


Figure 6 Participants' preference (text, video, and avatar)

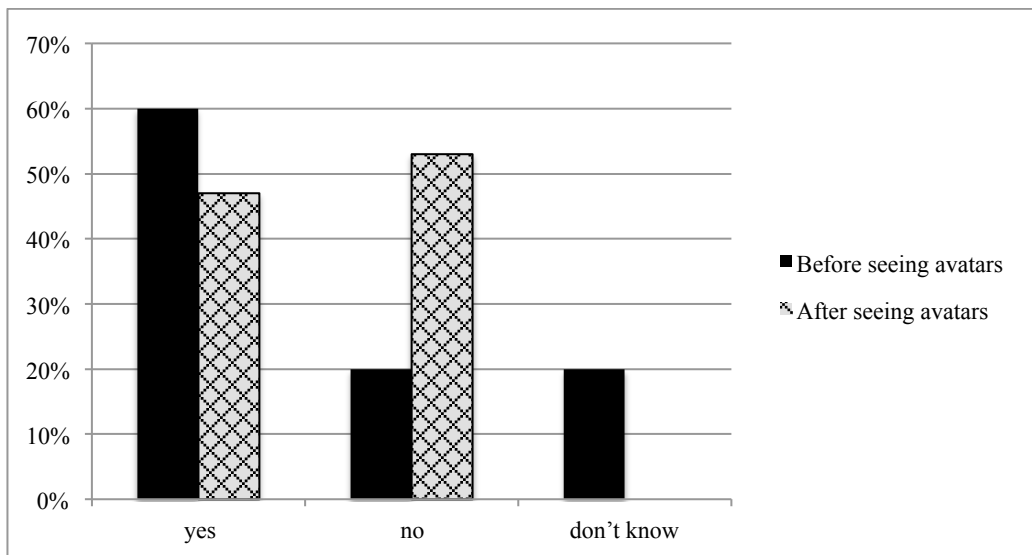


Figure 7 Would you use a signing avatar video?

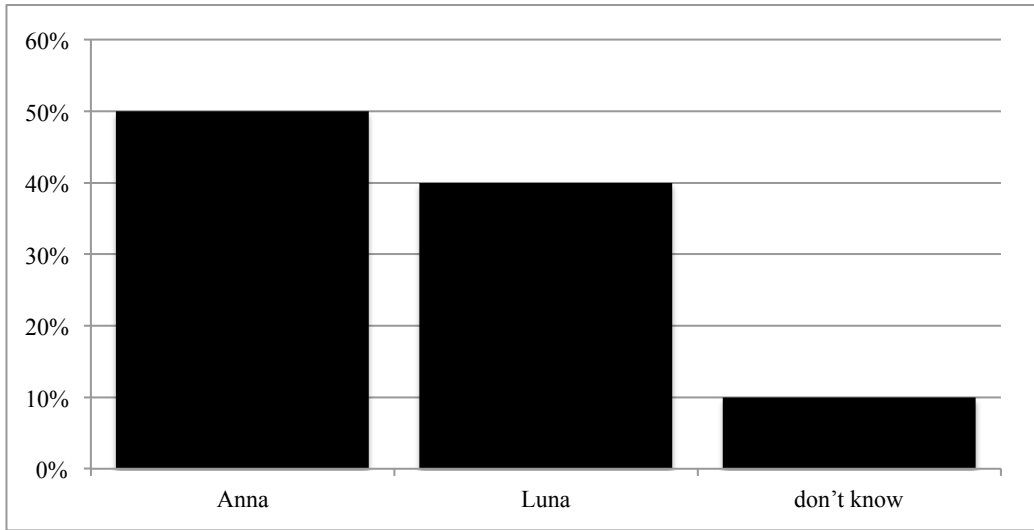


Figure 8 Avatar preference

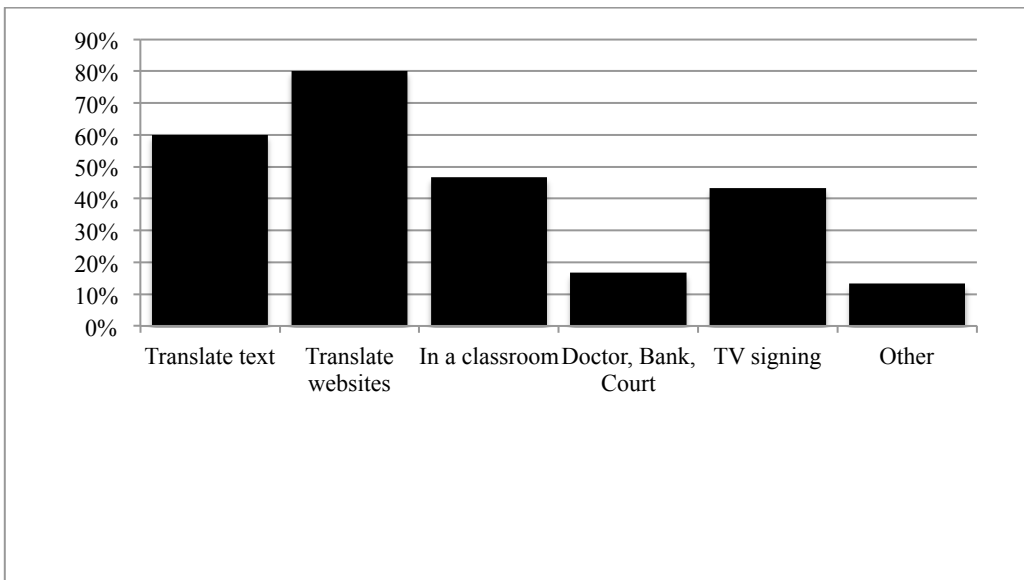


Figure 9 Possible use for avatar technology

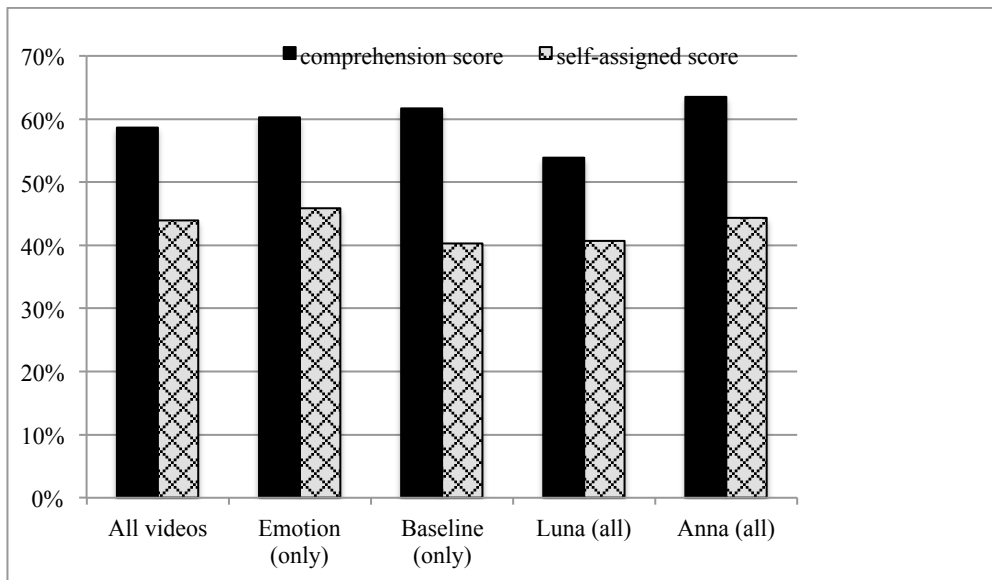


Figure 10 Comprehension score vs. self-assigned score

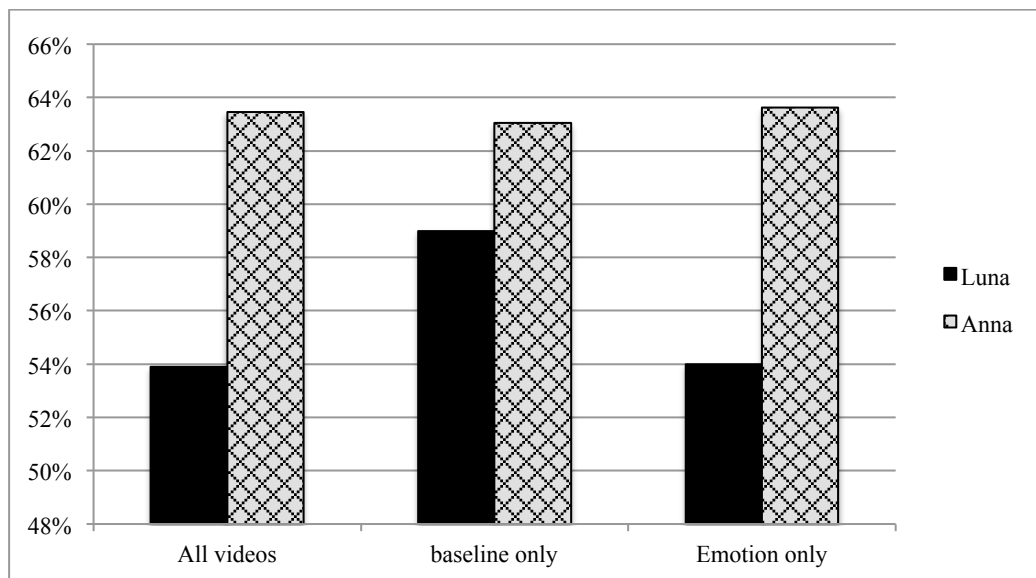


Figure 11 Average comprehension score by Avatar

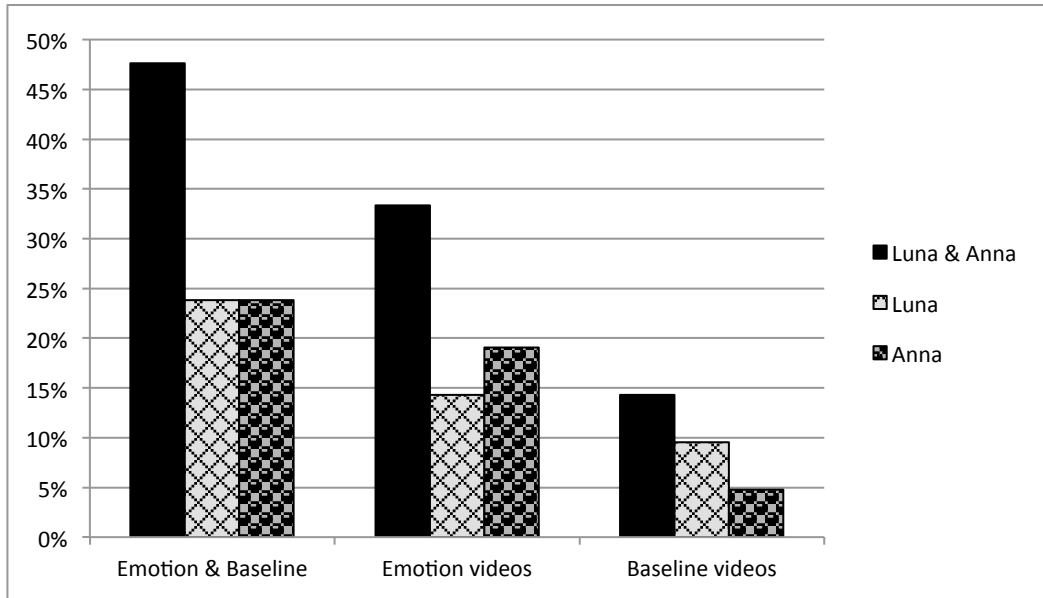


Figure 12 Did you see emotion?

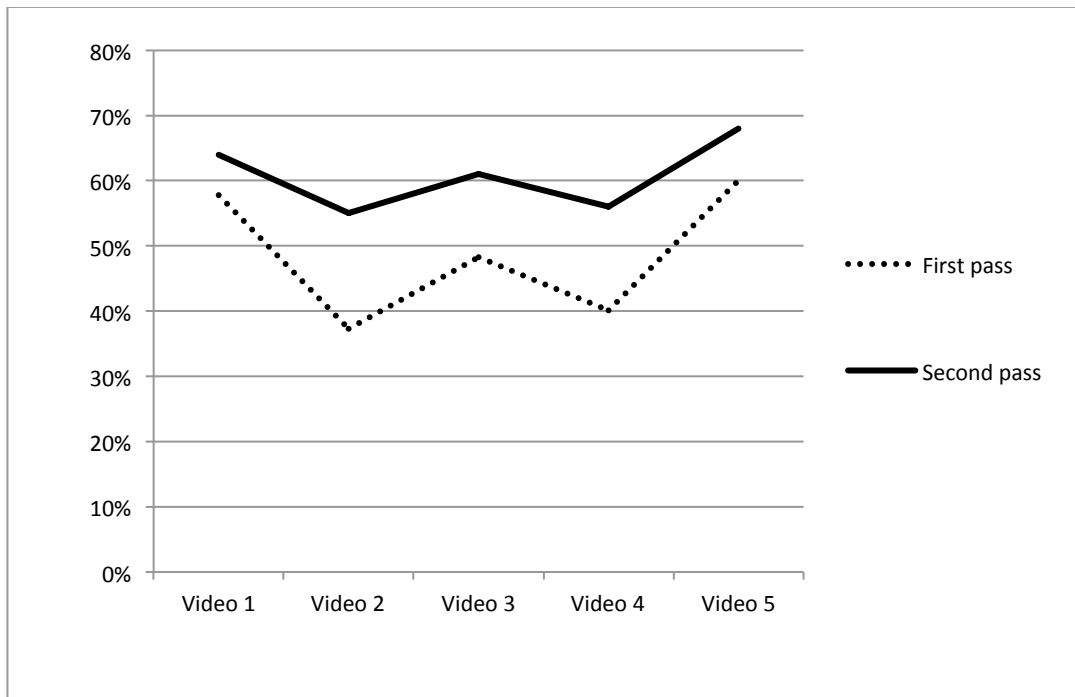


Figure 13 Average comprehension score - 1st and 2nd viewing

After watching each avatar video, participants were asked if 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 e.g. After watching a video once, a participant would be asked a series of comprehension question 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.

Figure 13 illustrates a comparison between the average comprehension score achieved based on each video for the first and second viewings, this include 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. We believe the 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 seconds 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

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, Figure 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%.

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. Participant's remarks and the results highlighted in Figure 8 & Figure 11 enlighten us 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% (Figure 12) but when we also consider the relatively high false positive of the baseline Luna avatar (10%) we 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 Figure 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 (Figure 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 Figure 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 6-1 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 Figure 10. Although the average comprehension scores indicate only a minor effect of EFEs,

Table 6-1 indicates that attributes such as emotion and NMF are desired by the Deaf community and furthermore, are required to improve comprehension.

We saw in Figure 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 much 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 section 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 Figure 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 (Figure 10) in addition to the results in Figure 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 (Figure 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 in one case a participant coined the new phrase “it looks 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 regards 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 here 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 have a huge effect on facial movement also and perhaps these are a more suitable place to begin making changes.

9. FUTURE WORK

Much work must be done to achieve a usable, comprehensible avatar with particular focus on the linguistic attributes that fared badly in

Table 6-1. 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, also increase the comprehension level.

10. ACKNOWLEDGMENTS

This research is supported by: The School of Informatics and Engineering at the Institute of Technology Blanchardstown. We would like to acknowledge the support of our 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.

11. REFERENCES

- [1] P. Ekman, "Basic Emotion," in *Handbook of cognition and emotion*, UK, Wiley, 1999, pp. 45-60.
- [2] M. Huenerfauth, L. Zhao, E. Gu and J. Allbeck, "Evaluation of American Sign Language Generation by Native ASL Signers," *Transactions on Access Computing*, vol. Vol. 1, no. 1, p. Article 3, 2008.
- [3] P. Matthews, *The Irish Deaf Community, Survey Report, History of Education, Language and Culture*, vol. Vol 1, Dublin: The Linguistics Institute of Ireland, 1996.
- [4] L. Leeson, "Aspects of Verb Valency in Irish Sign Language," *Unpublished Doctoral Dissertation, Centre for Language and Communication Studies, TCD, Dublin*, 2001.
- [5] R. Conrad, *The Deaf School Child: Language and Cognitive Function*, London: HarperCollins Publishers, 1979.
- [6] L. Leeson and B. Nolan, "Digital Deployment of the Signs of Ireland Corpus in Elearning," in *3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, Marrakech, 2008.
- [7] R. Kennaway, "Experience with and requirements for a gesture description language for synthetic animation," in *A. Camurri and G. Volpe (eds) Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop*, Genova, 2003.
- [8] S. Morrissey, H. Somers, R. Smith, S. Gilchrist and S. Dandapat, "Building a Sign Language corpus for use in Machine Translation," in *4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, LREC, Valletta, Malta*, 2010.
- [9] R. Smith, H. Somers and S. Morrissey, "HCI for the Deaf community: Developing human-like avatars for sign language synthesis," in *Proceedings of the 4th Irish Human Computer Interaction Conference (IHCI)*, Dublin, 2010.
- [10] S. Prillwitz, *HamNoSys Version 2.0: Hamburg notation system for sign languages: An introductory guide*, Signum, 1989.
- [11] V. Jennings, R. Elliott, R. Kennaway and J. Glauert, "Requirements For A Signing Avatar," in *LREC, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 2010.
- [12] R. Elliott, J. Glauert, V. Jennings and R. Kennaway, "An overview of the SiGML notation and SiGMLSigning software system," in *Workshop on the Representation and Processing of Signed Languages, 4th International Conference on Language Resources and Evaluation*, 2004.