HCI for the Deaf community: Developing human-like avatars for sign language synthesis

Article · October 2010

Source: OAI

3 authors:

Robert Smith
Technological University Dublin - Blanchardstown Campus

Sara Morrissey
Munster Technological University

Harold Somers

HCI for the Deaf community: Developing human-like avatars for sign language synthesis

Robert Smith, Sara Morrissey, Harold Somers
Centre for Next Generation Localisation
School of Computing
Dublin City University
Glasnevin, Dublin 9, Ireland.
{rsmith, smorri, hsomers}@computing.dcu.ie

ABSTRACT

With ever increasing computing power and advances in 3D animation technologies it is no surprise that 3D avatars for sign language (SL) generation are advancing too. Traditionally these avatars have been driven by somewhat expensive and inflexible motion capture technologies and perhaps this is the reason avatars do not feature in all but a few user interfaces (UIs). SL synthesis is a competing technology that is less costly, more versatile and may prove to be the answer to the current lack of access for the Deaf in HCI. This paper outlines the current state of the art in SL synthesis for HCI and how we propose to advance this by improving avatar quality and realism with a view to ameliorating communication and computer interaction for the Deaf community as part of a wider localisation project.

General Terms

Design, Human Factors, Languages.

Keywords

User-centered design, Deaf, Sign Language synthesis, Emotion, Prosody, Natural variance, Involuntary movement, 3D Avatar, SiGML, HamNoSys

1. INTRODUCTION

Interfaces are typically designed with the "average" user in mind. Special consideration is sometimes given to designing for disabled users but typically only a minimal amount of accessibility options are implemented. A good example of this is with current operating systems, which for the Deaf user allow visual warnings to appear when the system makes a sound. This functionality is useful to the Deaf user but unfortunately, it does not resolve the main issue facing Deaf users, namely that of literacy. In countries where SLs are not legally recognised and where children are still expected to learn via Oralism[1] ("oral education"), the average reading age of Deaf school leavers is comparable to that of an 8-9 year old hearing child [1][9][10].

The first language of the Irish Deaf Community is Irish Sign language (ISL). ISL is an indigenous language standing apart from English and Irish. There are approximately 5,000 native users of ISL in the Republic of Ireland [11], while it is estimated that some 50,000 non-Deaf people also know and use the language to a greater or lesser extent [7].

The average human interacts with a computer in their native language. The potential of synthesised SL avatars means the same level of access could be provided for Deaf users allowing them to interact in their first and preferred language.

If we compare[2] the task performance of an Irish Deaf person to that of an Irish hearing person, when both are introduced to a new interface containing written English text, we can assume that the Deaf person will have a heavier cognitive workload. In fact the workload will be similar to that of hearing person whose first language is not English. This is because both people are confronted with the same challenge, they both must translate the English text to their first language before they process the information contained within the text. In spoken languages this problem is easily solved by offering the text in many languages and with machine translation tools such as Google Translate. Because SL does not have a written form this presents a unique challenge in both synthesis and translation.

Achieving a truly universal design, i.e. software that can be used by all, equally is a difficult but worthy goal. The introduction of synthesised SL avatars into interface designs during the early

[1] Oralism is the education of Deaf students restricting SL within the classroom and using methods such as lip reading, speech, the process of watching mouth movements, and mastering breathing.

[2] This comparison is theoretical and based on anecdotal evidence from the Deaf community.

stages of development will progress this goal with regards to one minority user group, the Deaf community.

The World Federation of the Deaf (WFD), an international non-governmental organisation, approximates that it represents 70 million Deaf people worldwide [17]. This is equivalent to 1% of the world population or, in other terms, is larger than the population of the UK and Ireland combined. This is a large user base to be overlooked by global software development companies. Deaf Human Computer Interaction (DHCI) has the potential to be a lucrative field in the future.

Access to everyday services for the Deaf population of Ireland is non-existent in most cases. Considering that the ratio of qualified interpreters to Deaf people in Ireland equates to 250 profoundly Deaf people for every 1 fully accredited interpreter, low access levels are not surprising [8]. DHCI can help improve access for the Deaf by providing, as a starting point, accessible off-the-shelf software, websites and information kiosks.

Naqvi [13] discusses several artificial forms of digitally representing SLs which have emerged over recent times. SLs are visual-gestural languages, which do not have a written form. This makes translation difficult and limits the mediums in which the translation can be produced. Despite this obstacle, some webpages have their content translated into the appropriate SL. The most common medium for SL over the internet is streaming videos (of real people/humans) signing. This process is expensive, inefficient and non-transferable with difficulties in reproduction. Minor alterations to a webpage would mean whole videos must be reshot, reedited and reposted to the website. In choosing a medium for SL output we must consider these factors and how they influence the cost of developing that representation.

In contrast to video production, synthesised SL avatars offer a cheap and effective solution to the lack of SL representation in HCI. Other advantages include a low bandwidth requirement, ease of reproduction and the option to translate too many SLs simultaneously.

The remainder of this paper is organised as follows: In section 2, we discuss synthesised SL avatars in more detail and highlight some existing synthesised SL avatar systems. Section 3 briefly presents the baseline SL synthesis system we are working with. In section 4 we outline our proposed amendments to the baseline for human-like avatar synthesis and section 5 we describe how we plan to achieve this by introducing realistic human features, Section 6 concludes the paper.

2. HCI WITH SYNTHESISED SL AVATARS

Arguably the most interesting aspect of synthesising SL and gesture is how we “drive” the animated avatar. In this section we outline some of the many projects that use a variety of methods to drive their avatar. Most have been developed as proof of concept or pilot models for the purpose of research and have never entered a commercial development stage. For this reason none are acceptable to the Deaf community as replacements for human interpreters.

SL avatar synthesis is a young area of research but it is not new. Many projects have attempted to use synthesised avatars in a variety of scenarios.

The ViSiCAST project was a 3-year, EU-funded project involving a collaborative approach by a number of institutions to improve access to services and facilities for the Deaf by means of virtual signing technology. This project used motion capture technology to develop a pilot project called “TESSA” [2], a SL avatar translation tool, tested in UK post offices aimed at providing limited access for Deaf customers.

A more recent project, also for 3 years and EU funded, the eSign project [4] succeeded ViSiCAST. eSign built upon the technology already developed in the ViSiCAST project by introducing synthesised SL. eSign was responsible for the later versions of SiGML, which we will discuss in section 3. eSign moved away from motion capture technology to fully synthesised signing which allowed the project to bring SL to eGovernment websites in a small scale pilot project [4].

Weather forecasting systems are a sensible choice for SL synthesis. The domain is small and phrases used are quite predictable. In a separate proof-of-concept project, Grieve-Smith [5] used Albuquerque weather forecasts because they presented less variability than in many other cities, thus the domain was smaller still. The emphasis of this research was on Machine Translation (MT) for SL and therefore little attention was paid to the realism of the avatar which was generated using keyframe interpolators in Virtual Reality Modeling Language (VRML).

The previously discussed eSign project developed a plug-in for Internet Explorer that would display weather forecast summaries in a number of SLs including British Sign Language, Sign Language of the Netherlands, and German Sign Language [4].

The Greek project, Vsigns [14]developed a virtual human avatar tool which was used to teach SL to interested parties. They used VRML and Mpeg-4 body animation to animate the avatar in real-time and as a result the avatars take the form of caricatures more than human-like representations.

VCom3D [16], a commercial entity, provides exclusively designed avatars for scenario-based learning and language and culture learning. These tools are used by a number of clients including the US Department of Defense. During an evaluation carried out by VCom3D [6] on comprehension of kindergarten to 12 year old Deaf and Hard-of-Hearing students, an increase of 17% to 67% was observed when shifting from text-only to text accompanied by SL, when using Vcom3D’s “SigningAvatar” technology.

3. BASELINE SYSTEM

The Virtual Humans group at the University of East Anglia (UEA), Norwich, lead the field of synthesised SL avatars through their involvement in innovative projects such as ViSiCAST and eSign. Both projects were based on the SiGMLSigning processing pipeline shown in figure 1.

The pipeline receives input in the form of HamNoSys (Hamburg Notation System). HamNoSys is among few well-established transcription systems, developed by the Institute for German Sign Language and Deaf communication at the University of Hamburg for all SLs [15]. HamNoSys is a phonetic notation system purpose-built for use by linguists in their detailed analytical representation of signs and sign phrases rather than as a writing system for SLs. It consists of approximately 200 iconic symbols representing such parameters as hand shape, hand orientation, location and movement. The notation is essentially phonemic, so the transcriptions are very precise, but on the other hand also very long and cumbersome to decipher.

Closely associated with HamNoSys is SiGML (Signing Gesture Mark-up Language) [3], a form of XML developed by the Virtual Humans group (UEA). SiGML defines a set of XML tags for each iconic symbol in HamNoSys. SiGML files are represented as plain text which means the files are small allowing for rapid transmission, and efficient MT.

ISL is articulated by manual features (MFs) and Non-Manual features (NMFs). MFs are meaningful units of SL that are articulated primarily by the hands and arms. NMFs are meaningful units of SL not articulated by the hands such as facial movements, head tilts and body tilting. The fact that HamNoSys is limited with regards to how it represents NMFs means SiGML is restricted too.

The eSignEditor tool can be used to create HamNoSys transcriptions which are converted to SiGML and sent 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.

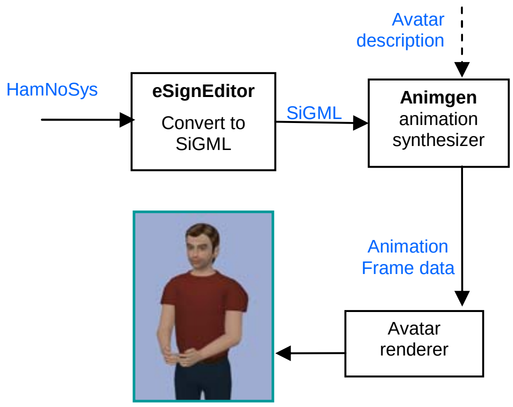

Figure 1. UEA SiGMLSigning processing pipeline [4]

4. IMPROVING THE BASELINE SYSTEM

MFs carry only 30% of a SL utterance's meaning, leaving the remaining 70% of the meaning to be expressed through the NMFs[3]. The disregarding of NMFs means some important linguistic information is missing and consequently the avatar can produce incorrect or confusing SL. This tendency to exclude NMFs has made the current state of the art in SL synthesis comparable to the robotic and artificial nature of early speech synthesis output and consequently there has been little uptake of the technology. Our proposed enhancements to the existing system may result in a higher uptake of the technology.

We propose to collaborate with the team at UEA in the development of their system. By giving to the baseline system more linguistic data in the form of greater support for NMFs, we improve the communication medium between the system and the Deaf user, ultimately making the SL easier to understand. We mean to focus on the following, which all relate to NMFs:

- Prosody
- Natural variance
- Involuntary movement

Emotion and prosody are expressed in SL through both NMFs and in the intensity of the MF [11]. A good example of this is how commands are used in ISL. In speech we can intensify the command by using volume, tone of voice and stress patterns. The ISL for the English word "NOW" will vary depending on how the word is used. If the word is used in a relaxed context, the sign is carried out in a relaxed manner, however, if the word is used in a command context such as "NOW!" or "NOW!!!!", then there is more intensity in the NMF and the pace and force of delivery of the MF is increased. Another example is the effect of missing NMFs for a question in ISL. If the eyebrows are not raised or frowned to indicate the question, then it is not always obvious that a question is being asked. Without the appropriate NMFs vital linguistic information is missing possibly leading to a mis-translation/misinterpretation.

We aim to simulate natural variance in signs such that the same sign is not repeated in exactly the same manner each time and to make the avatar less symmetrical when conducting two-handed signs. We will define a specification for the non-linguistic attributes of the avatar such as weight shift and involuntary movements.

In order to produce a functional SL synthesis system for DHCI, it is our intention to amalgamate our research on SL MT with our synthesis module to facilitate automatic English to SL translation.

5. CONCLUSION

As new forms of communication and HCI establish their role in society, it is important that we take measures to ensure that the technology is readily available to address barriers to interaction for the Deaf in the future.

SL synthesis will be a key technology in providing access to the Deaf community in their chosen language of SL.

Through our research we expect to advance the state of the art in human-like avatar synthesis and in the process advance HCI for the Deaf community, thus, promoting the research area such that synthesised SL is more easily understood and accepted amongst the Deaf community.

6. ACKNOWLEDGMENTS

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (http://www.cngl.ie) in the School of Computer Science at Dublin City University. We would also like to acknowledge the generous assistance of our colleagues at UEA, Norwich.

[3] Anecdotal evidence from conversations with ISL tutors and our ISL animation evaluators [12]

7. REFERENCES

[1] Conrad, R. (1979).The Deaf School Child. Language and Cognitive Function. London: Harper and Row, U.K.

[2] Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., Abbott, S. (2002). TESSA, a system to aid communication with deaf people. In Proceedings of the Fifth International ACM SIGCAPH Conference on Assistive Technologies, Edinburgh, U.K., pages 205–212.

[3] Elliott, R., Glauert, J.R.W., Jennings, V. and Kennaway, J.R. (2004). An overview of the SiGML notation and SiGMLSigning software system. In Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal, pages 98–104.

[4] Glauert, J. R. W., Kennaway, J. R., Elliott, R. and Theobald, B. (2004). Virtual Human Signing as Expressive Animation. In Proceedings of the AISB 2004 Convention on Motion, Emotion and Cognition, Leeds, U.K, pages 98-106.

[5] Grieve-Smith, A.B. (1999). English to American Sign Language Machine Translation of Weather Reports. In Second High Desert Student Conference in Linguistics, Albuquerque, New Mexico, [no page numbers].

[6] Hurdich, J. (2008). Utilizing Lifelike, 3D Animated SigningAvatar Characters for the Instruction of K-12 Deaf Learners. In Proceedings of the Exploring Instructional and Access Technologies Symposium National Technical Institute for the Deaf, New York, USA. [no page numbers].

[7] Leeson, L. (2001). Aspects of Verb Valency in Irish Sign Language. PhD Dissertation, Centre for Language and Communication Studies, Trinity College, Dublin, Ireland.

[8] Leeson, L. (2003). Sign Language Interpreters: Agents of Social Change in Ireland. In M. Cronin and C. Ó Cuilleanàin (Eds.), Languages of Ireland, Dublin, Ireland: Four Courts Press, pages 148–164.

[9] Leeson, L. (2006). Signed Languages in Education in Europe – a Preliminary Exploration. Preliminary Study. Strasbourg, France: Language Policy Unit, Council of Europe.

[10] Leeson, L. (2007). Seeing is Learning. A Review of Education for Deaf and Hard of Hearing People in Ireland. Report submitted to the Irish National Council for Special Education, Trim, Co. Meath, Ireland.

[11] Matthews, P. (1996). The Irish Deaf Community, Survey Report, History of Education, Language and Culture, Volume 1, Dublin, Ireland: The Linguistics Institute of Ireland.

[12] Morrissey, S. (2008). Data-Driven Machine Translation for Sign Languages. PhD Thesis, Dublin City University, Dublin, Ireland.

[13] Naqvi, S. (2007). End-User Involvement in Assistive Technology Design for the Deaf: Are Artificial Forms of Sign Language Meeting the Needs of the Target Audience? In Proceedings of the Conference and Workshop on Assistive Technologies for People with Vision & Hearing Impairments (CVHI), Granada, Spain, [no page numbers].

[14] Papadogiorgaki, M., Grammalidis, N., Tzovaras, D. and Strintzis, M.G. (2005). Text-To-Sign Language Synthesis Tool. In Proceedings of the 13th European Signal Processing Conference (EUSIPCO2005), Antalya, Turkey, [no page numbers].

[15] Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. (1989). HamNoSys. Version 2.0: Hamburg Notation System for Sign Languages. An introductory guide. Hamburg: Signum.

[16] VCom3D inc http://www.vcom3d.com/ last viewed 12-05-2010

[17] World Federation of the Deaf (WFD) http://www.wfdeaf.org/ last viewed 12-05-2010.