


2014-09-01

The Role of Emotional and Facial Expression in Synthesised Sign Language Avatars

Robert G Smith

Technological University Dublin, robert.smith@tudublin.ie

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

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

Recommended Citation

Smith, R.G. (2014) The role of emotional and facial expression in synthesised sign language avatars, Master Thesis, Technological University Dublin.

This Theses, Masters 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, aisling.coyne@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#)

THE ROLE OF EMOTION AND FACIAL EXPRESSION IN SYNTHESISED SIGN LANGUAGE AVATARS

Robert Smith

A dissertation submitted in fulfillment of the requirements for the award of

Master of Science (M.Sc.) in Computing



Institute of Technology Blanchardstown

Computational and Functional Linguistics Group

Irish Sign Language Special Interest Group

Supervisor: Dr. Brian Nolan

March 2014

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of M.Sc. is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

Student No.:

Date:

Contents

Abstract.....	8
Acknowledgments	9
1 Introduction.....	10
2 Literature Review	14
2.1 Signed languages and the Deaf community.....	14
2.2 Linguistics of Signed Language	15
2.2.1 Linguistic levels.....	15
2.2.2 Signing space.....	17
2.2.3 Classifiers and Placement.....	18
2.2.4 MFs and NMFs	19
2.3 Representing Sign Language	20
2.3.1 Transcription for Sign Language	21
2.3.2 Mark-up Languages	26
2.4 Animation	31
2.4.1 Animation paradigms.....	32
2.4.2 Human emotion.....	34
2.4.3 3D character development	39
2.4.4 3D rendering systems.....	42
2.5 Synthesised signing systems.....	43
2.6 Motion Capture and Digital Video	46
2.6.1 Digital Video	46
2.6.2 Motion Capture.....	47
3 Methodology.....	49
3.1 Corpus.....	49
3.1.1 Patient–receptionist dialogue corpus	49
3.1.2 Signs of Ireland corpus.....	50
3.1.3 Choosing a subset.....	51
3.1.4 Corpus transcription	53
3.2 The JASigning platform	67
3.2.1 JASigning real-time avatar generation	68
3.2.2 ARP toolkit.....	75

3.2.3 Mark-up for new morph set	77
3.3 Evaluation.....	79
3.3.1 Manual evaluation methodology.....	79
4 Findings	82
4.1 General findings.....	82
4.2 Comprehension results.....	86
5 Discussion.....	91
5.1 Implementation Discussion	91
5.2 Problems with evaluation	92
6 Conclusions	94
6.1 Future work.....	97
Bibliography.....	98
Appendix A – Avatar EFes	105
Appendix B – HamNoSys Symbol set.....	106
Appendix C – ARP Toolkit workflow	115
Appendix D – Motion capture rig.....	116
Appendix E – interview questions	119

List of Figures

Figure 2-1 Linguistic levels (Bungeroth, 2002).....	16
Figure 2-2 Signing space - front and side view.....	18
Figure 2-3 Classifier - person falling (ASLU, 2013)	19
Figure 2-4 A passage from the children's story "Goldilocks", transcribed	22
Figure 2-5 A passage from Goldilocks, notated with.....	23
Figure 2-6 A passage from Goldilocks, notated with.....	25
Figure 2-7 Szczepankowski's gestographic notation PSL sign "TO WRITE" (Fabian & Francik, 2001)....	25
Figure 2-8 ASL-phabet writing system – ASL notations.....	26
Figure 2-9 markup hierarchy	27
Figure 2-10 the evolution of SiGML	28
Figure 2-11 The ISL sign "I" represented as HamNoSys, H-SiGML and G-SiGML	29
Figure 2-12 the representation in SWML of the sign for "idea" in 'LIBRAS' the Brazilian Sign Language (Costa & Dimuro, 2001).....	30
Figure 2-13 the famous half-filled flour sack (Thomas & Johnston, 1995).....	34
Figure 2-14 Plutchik's three-dimensional model describes the relations among emotion concepts.....	38
Figure 2-15 Three levels of subdivision in a polygon mesh (Velho, 2001)	39
Figure 2-16 armature structure with polygonal mesh (unity3D.com, 2013).....	40
Figure 2-17 the avatar Anna and the various texture maps used in her creation.....	41
Figure 2-18 Signing Avatars from 6 different projects	45
Figure 2-19 Motion capture setup (Optitrack, 2013)	47
Figure 3-1 ELAN with emotion annotations.....	52
Figure 3-2 eSIGN Editor - maintenance options.....	54
Figure 3-3 eSIGN Editor - signs database	54
Figure 3-4 eSIGN Editor - new sign window	55
Figure 3-5 eSign Editor – HamNoSys input panel.....	56
Figure 3-6 eSIGN Editor - mouth picture panel & mouth gesture panel	57
Figure 3-7 eSIGN Editor - untitled document.....	58
Figure 3-8 eSIGN Editor - new sign utterance window.....	59
Figure 3-9 eSIGN Editor - new sign transcription window	59
Figure 3-10 eSIGN Editor - working environment	60
Figure 3-11 eSIGN Editor - data flow	61
Figure 3-12 eSIGN Editor - Facial Expressions	62
Figure 3-13 HamNoSys - general structure.	64
Figure 3-14 HamNoSys - Hand Shapes	65
Figure 3-15 HamNoSys - Extended Finger Direction (EFD).....	65
Figure 3-16 UEAs JASigning data flow	68
Figure 3-17 hns-SiGML for DGS (German Sign Language) sign "Internet"	69

Figure 3-18 G-SiGML for DGS (German Sign Language) sign "Internet":	70
Figure 3-19 config file partial contents	71
Figure 3-20 Attack, Hold, Release envelope using NMFs	72
Figure 3-21 Extract from a nonmanuals.xml file with emotion face configuration	72
Figure 3-22 ARP Toolkit data flow	75
Figure 3-23 ARP Toolkit - vertex weights.....	76
Figure 3-24 ARP Toolkit - morph groups.....	77
Figure 3-25 nonmanuals.xml "sad" element	78
Figure 3-26 SiGML notation for the ISL sign "Begin" - with emotion tag.....	78
Figure 3-27 Avatars: Luna & Anna	80
Figure 3-28 Evaluation room layout	81
Figure 4-1 Participants' preference (text, video, and avatar)	83
Figure 4-2 Would you use a signing avatar video?	84
Figure 4-3 Avatar preference	87
Figure 4-4 Possible use for avatar technology	87
Figure 4-5 Comprehension score vs. self-assigned score	88
Figure 4-6 Average comprehension score by Avatar	88
Figure 4-7 Did you see emotion?	88
Figure 4-8 Average comprehension score - 1 st and 2 nd viewing	89

List of Tables

Table 2-1 principals of animation	32
Table 2-2 Ekman's 7 Universal emotions demonstrated by actor Tim Roth	36
Table 3-1 the frequency of which each EFE appears	53
Table 3-2 HamNoSys symbol classification (examples)	63
Table 3-3 nonmanuals.xml timing codes	73
Table 3-4 Avatars used	80
Table 4-1 Attributes - Anna vs. Luna	85

Abstract

This thesis explores the role that underlying emotional facial expressions might have in regards to understandability in sign language avatars. Focusing specifically on Irish Sign Language (ISL), we examine the Deaf¹ community's requirement for a visual-gestural language as well as some linguistic attributes of ISL which we consider fundamental to this research. Unlike spoken language, visual-gestural languages such as ISL have no standard written representation. Given this, we compare current methods of written representation for signed languages as we consider: which, if any, is the most suitable transcription method for the medical receptionist dialogue corpus². A growing body of work is emerging from the field of sign language avatar synthesis. These works are now at a point where they can benefit greatly from introducing methods currently used in the field of humanoid animation and, more specifically, the application of morphs to represent facial expression.

The hypothesis underpinning this research is: augmenting an existing avatar (eSIGN) with various combinations of the 7 widely accepted universal emotions identified by Ekman (1999) to deliver underlying facial expressions, will make that avatar more human-like. This research accepts as true that this is a factor in improving usability and understandability for ISL users. Using human evaluation methods (Huenerfauth, et al., 2008) the research compares an augmented set of avatar utterances against a baseline set with regards to 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. Remarkably, the results of this manual evaluation show that there was very little difference between the comprehension scores of the baseline avatars and those augmented with EFEs. However, after comparing the comprehension results for the synthetic human avatar "Anna" against the caricature type avatar "Luna", the synthetic human avatar Anna was the clear winner. The qualitative feedback allowed us an insight into why comprehension scores were not higher in each avatar and we feel that this feedback will be invaluable to the research community in the future development of sign language avatars. Other questions asked in the evaluation focused on sign language avatar technology in a more general manner. Significantly, participant feedback in regard to these questions indicates a raise in the level of literacy amongst Deaf adults as a result of mobile technology.

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

² A multimodal, multimedia, parallel corpus outlined in (Morrissey, et al., 2010).

Acknowledgments

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

This body of work would not have been possible without the infinite patience and unwavering support of my wife Gemma and my children Andrew aged 6 and Ben aged 2. You have my gratitude always.

I would like to express my gratitude also, to my supervisor Brian Nolan for the useful comments, remarks and engagement through the learning process of this master thesis.

Finally, a big thank you to my sister Barbara Smith and friend Peter Alexander who offered their time to help me eradicate those nasty grammatical and spelling errors that so often attach themselves to these works.

Chapter 1

1 Introduction

“The single biggest problem in communication is the illusion that it has taken place.”

- George Bernard Shaw

“If you talk to a man in a language he understands, that goes to his head.

If you talk to him in his language, that goes to his heart.”

- Nelson Mandela

The quotes above serve to illustrate the importance of effective communication and more importantly the language in which that communication takes place. Languages differ across the globe and although there are many people who speak many languages, there are more that speak only their native tongue. Regardless to whether or not one may be multilingual, it is fair to say that the majority of people prefer to speak in their own language. Unfortunately, however, a large portion of the world’s communications take place only in one or more of the world’s most common languages. For example the United Nations have adopted 6 of the world’s most common languages to be their ‘official languages’. These are Arabic, Chinese, English, French, Russian and Spanish (U.N, 2013). According to the United States’ CIA world fact book, 71% of the world’s population do not use any of the more popular languages as their first language (C.I.A, 2009). As a consequence, those 39% whose first language is more widely used have the benefit of being ‘information rich’. Let us consider for a moment printed and electronic publications. These publications, in the majority of cases, exist for some commercial benefit. Whether that benefit translates into direct sales, advertisement revenue or support information is not relevant to this work. What is relevant is the motivation: profit. That is not to say that all publications are profit driven. The Rosetta foundation are a non-profit organisation who strive to plug the information gap that is so visible in many minority languages. The Rosetta foundation believe that ‘information poverty’ is endemic across minority languages and results, in no small way, in poverty as well as poor healthcare, education and justice systems (Rosetta, 2013).

Clearly, there are larger issues caused by communication barriers for those who use minority languages and in this, signed languages are no exception. Technologies such as subtitling and video may be leveraged to address some of these communication barriers, particularly for signed languages. However,

the very nature of these technologies limit the scenarios and the mediums in which they may be implemented. These limitations provoked the development of sign language avatars, which are perhaps a more flexible and applicable technology for signed languages. The development of sign language avatars began a little over 10 years ago. A short time span in any field of research but it should come as little surprise, when one considers that research into the representation of sign languages only began in the 1960s, after William Stokoe began a linguistic study of what would later be named American Sign Language (ASL) (Stokoe, 1960). The Deaf community around the world are yet to embrace sign language avatars as a communication tool. Undoubtedly, this is due to the somewhat robotic nature of the current state of the art. This is in itself, a symptom of the short time period dedicated to researching and developing this technology.

The hypothesis on which this thesis is based is: The comprehensibility and acceptability of synthesised sign language avatars can be improved upon with the provision of a more human-like avatar than that which is provided by the current state of the art. In this case ‘human-like’ applies primarily to the avatars appearance, the level of emotion simulated and to a lesser extent the fluidity of movement. This hypothesis has prompted 3 research questions:

RQ1. What is the current uptake of signing avatar technology within the Irish Deaf community? And what are the factors that have influenced this?

Before launching into the development of a signing avatar platform, it is important to understand the Deaf community’s mood towards assistive technologies. Particularly, it is prudent to investigate the relative acceptance or non-acceptance of signing avatar technologies. This would include a look at the technologies involved in existing signing synthesis systems and the underpinning written and computer based forms used to represent sign languages.

RQ2. Does the presence of simulated emotional data improve avatar comprehension or acceptance? How may we best evaluate: to what extent Emotional Facial Expressions (EFEs) are significant in signing avatars?

According to Cochran & Claspell (1987), “An emotion is a communication”. In the book ‘Communicating Emotion’ (Planalp, 1999), Planalp uses this quotation to emphasise the importance of emotion in human communication. The presence of an emotion does not carry any linguistic meaning but that is not to say that it does not carry any meaning at all. A statement made in anger can communicate something very different than the same statement if it was communicated with a mocking or happier tone. The presence of an emotion or, indeed, the absence of emotion in a communication carries meaning. It is for this reason that, of all the attributes that could make an avatar more ‘human-like’, this thesis will concentrate primarily on emotion. Emotion is a relatively abstract term, it covers everything from the cognitive process involved in feeling an emotion to how that emotion is outwardly

projected. Consequently it is important to narrow our focus onto a specific aspect of emotion. Therefore this work will concentrate primarily on EFEs and their effect on avatar acceptance and comprehension. For the purpose of this thesis, EFEs are defined as facial configurations that are representative of the outward projection of an emotion.

Emotion and prosody are expressed in sign language primarily through Non-Manual Features (NMFs) (Matthews, 1996), which are widely accepted to carry up to 70% of a sign's meaning and this therefore makes emotion a significant factor in the credibility and acceptance of an avatar. NMFs include movements of the head, mouth, shoulders and eyebrows amongst others. Unlike spoken language, signed languages have multiple articulators and every NMF is considered an articulator. Also considered articulators are the 'Manual Features' (MF), which include the hands and arms. The linguistic properties of signed language are discussed in more detail in section 2.2.

RQ3. To what extent is a signing avatar output understood? And how may we best evaluate the comprehension level of a signing avatar performance?

Deaf users may like or dislike the idea of avatar technology. They may simply like or dislike an avatar's appearance, all of which effect the acceptance levels of avatar technology. The third research question relates primarily to comprehension. How well do sign language users actually understand an avatars performance? We will answer how this level of comprehension may best be captured and what is the significance of this in relation to acceptance levels.

To answer our research questions we will work with the Signs of Ireland corpus (SOI), a well-established corpus that uses ISL data with English glosses. Also we will work with the HamNoSys notation system for representing signed languages which will drive a pre-existing avatar framework developed in the University of East Anglia (UEA). Data from the SOI corpus will be transcribed with HamNoSys and then performed by the UEA synthesised sign language avatar. The synthesised sign language performances will be generated based on the current state of the UEA avatar software and then again with an augmented avatar state that will include the 7 universal facial expressions identified by Paul Ekman (1999). In this instance, the HamNoSys transcriptions and the augmentations to the UEA system will be carried out by the author. A manual evaluation will establish if the addition of EFEs can increase the level of comprehension. In addition to this, in an effort to identify acceptable avatar styles, 2 distinctly different avatars will be used to perform the same data. The first avatar is designed to be human-like and the second is more caricature-like in nature.

The evaluation techniques used in this study are loosely based on those used by Huenerfauth, et al. (2008) in their evaluation of American Sign Language generation by native ASL signers as well as those used in Kipp, et al. (2011).

We expect that the findings of the evaluation will uncover some interesting indications of the Irish Deaf community's attitude towards signing avatars and technology in general. An indication of the effects of mobile technology on Irish Deaf adults is expected although this is not our primary objective. The primary objective of the evaluation is to gain some insight into the Deaf community's attitude to avatar technologies. The evaluation will answer the 3 research questions outlined above, supplying quantitative and qualitative data on Deaf acceptance and comprehension of signing avatar technology as well as preference (or lack thereof) for avatars that contain EFE data. It is hoped that the qualitative data collected can be used to identify key failings of the technology as well as isolating priority areas for improvement. This qualitative data is expected to be re-enforced by the quantitative data in many cases.

Chapter 2 of this thesis provides an overview of sign language in the community, some signed languages linguistic aspects that are relevant to this work and how signed languages are currently recorded and represented. Additionally, in chapter 2 there is a discussion around the various technologies required for sign language avatar synthesis and how they may have been utilised in previous projects.

Chapter 3 outlines the methodology used including the selecting and implementation of the corpus materials, avatar signing software and the methodology behind a manual evaluation.

In **Chapter 4** we discuss the findings of the evaluation. This chapter includes some general findings as well as some statistical representation of the comprehension results.

Chapter 5 discusses the problems encountered before, during and after the evaluation process.

Chapter 6 concludes this thesis and outlines some possible future work.

The research presented in this thesis was accepted in the proceedings of 2 peer-reviewed conferences. A short paper based on a summary of the main evaluation findings was accepted to the well-established ASSETS conference (Smith & Nolan, 2013a) and a long paper was accepted to the smaller but more specialised symposium on Sign Language Translation and Avatar Technology (Smith & Nolan, 2013b).

2 Literature Review

2.1 Signed languages and the Deaf community

The first language of the Deaf Community in Ireland is Irish Sign Language (ISL). 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 (Matthews, 1996), while it is estimated that some 50,000 non-Deaf people also know and use the language to a greater or lesser extent (Leeson, 2001).

A common misconception among the hearing community is that written text is an adequate means of providing access to information for the Deaf. Unfortunately this is not the case. In countries where signed languages are not legally recognised and where children are still expected to learn via Oralism³, the average reading age of Deaf school leavers is comparable to that of an 8-9 year old hearing child (Irish National Rehabilitation Board, 1991; Conrad, 1979). For this reason, information in a written format is not adequate. The Deaf community would prefer to access information through their first language.

Unbeknown to many in the hearing community is the fact that Deaf people across the world speak different languages. In Ireland the Deaf community communicate using Irish Sign Language, the British Deaf community use British Sign Language (BSL), Americans use American Sign Language (ASL) and so on. Each of these languages has their own complex linguistic structure, with their own syntax and morphology, making each sign language as rich a natural language as any spoken equivalent. The World Federation of the Deaf (WFD), an international non-governmental organisation, approximates that it represents 70 million Deaf people worldwide (WDF, 2013). This is equivalent to 1% of the world population or, to put that figure into context is larger than the population of the UK and Ireland combined. Like the wider community, Deaf people in every developed country face similar difficulties accessing services, but then, that is only a symptom of the communication divide between the Deaf and hearing communities.

For the Deaf population of Ireland, access to everyday services is problematic in the majority of cases. Considering that the ratio of qualified interpreters to Deaf people in Ireland equates to 250:1. That's 250 profoundly Deaf people for every 1 fully accredited interpreter. When we consider this ratio low

³ Oralism is the education of Deaf students which restricts signed language use within the classroom instead using methods such as lip reading, speech, the process of watching mouth movements, and mastering breathing.

access levels are not surprising (Leeson, 2003). The result of this and low literacy levels, compounded by other sociolinguistic difficulties, is an obvious communication barrier between the Deaf community and their hearing counterparts, not only when conversing directly, but in other mediums also. It is plain to see the everyday forms of media that hearing people take for granted are not accessible or are accessible to a lesser extent to Deaf people. These include written communications such as electronic mail, websites, notice boards or advertisements. Also Radio and Television/video without subtitles is completely inaccessible.

Technology can help improve access for the Deaf by providing, as a starting point, accessible off-the-shelf software, websites and information kiosks as well as plugins for social media platforms and a plethora of other applications. Naqvi (2007) discusses several artificial forms used to digitally represent signed languages which have emerged over recent times. Despite this obstacle, some webpages offer signed language content. The most common medium for signed language over the internet is streaming videos (of real people signing) which for the purpose of this document are referred to as ‘signing videos’. The process of creating signing videos is expensive and inefficient with difficulties in reproduction. Minor alterations to a webpage might mean whole videos must be re-shot, re-edited and reposted to the website, often with continuity problems.

2.2 Linguistics of Signed Language

To understand why certain conventions are used when representing a signed language in written form or indeed through an avatar, one must first understand some of the basic linguistic principals of signed languages. Irish Sign Language is an indigenous language, a natural language that has evolved over generations, constantly adapting to and incorporating new signs and meaning. However, as with spoken languages, most signed languages have the same traits. It is those traits that we discuss in this chapter. For example; one of the more notable differences between signed languages and spoken languages is the articulators used. Signed languages use the hands, eyes, shoulders, head and some parts of the face as articulators (Matthews, 1996) whereas spoken languages use only the vocal tract. Hearing people often use the same body parts to indicate emotion or intonation but for signed languages the various articulator’s carry linguistic meaning.

2.2.1 Linguistic levels

Every language, whether spoken or signed, has various levels of linguistic detail based on the size of the chunk in question from sentence level to the phonetic level. Figure 2-1 illustrates the correlation between spoken language and signed language at each level.

Phonetics is the study of ‘phones’ or the sequence of sounds that make up a word. Phones are the smallest unit of speech usually categorised as a vowel sound or consonant sounds. In signed languages it is difficult to identify what constitutes a phone. Liddell (1993) identified segments of movement and

hold as a signed language's equivalent to a phone, however, these segments may also be considered phonological.

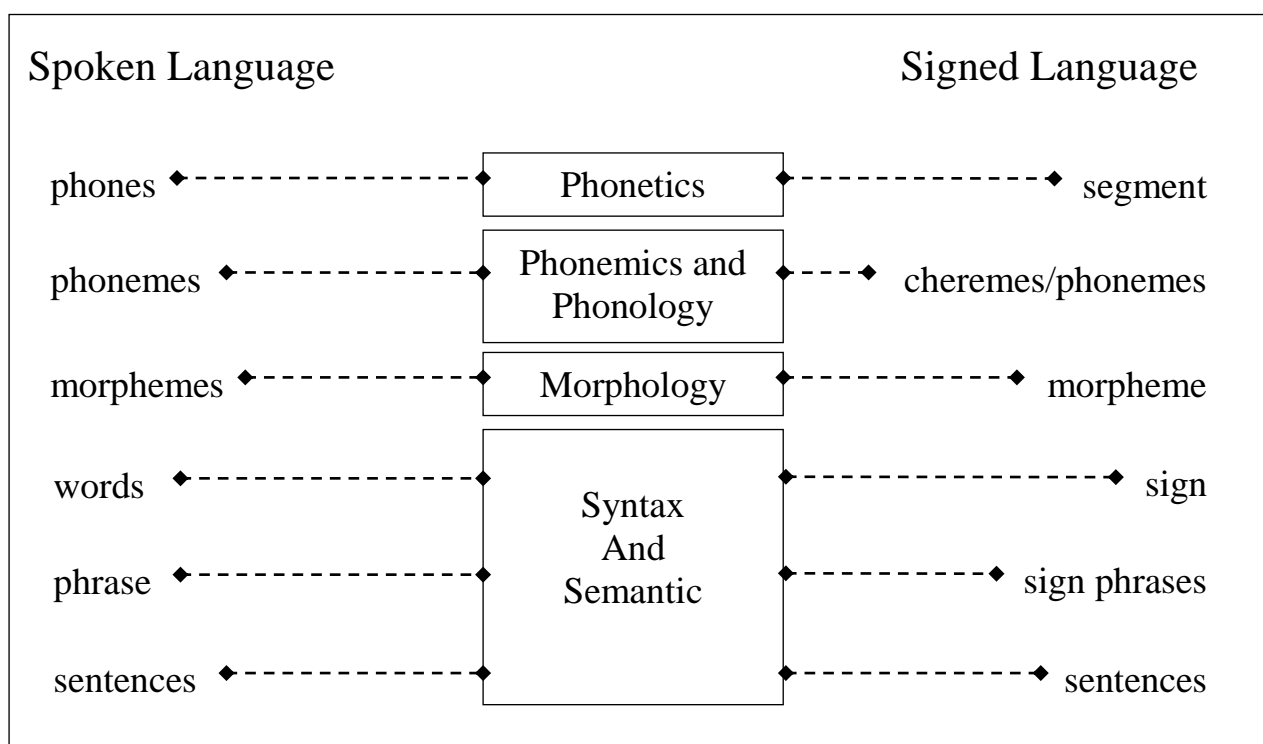


Figure 2-1 Linguistic levels (Bungeroth, 2002)

Most sign language notation systems record signs at the phoneme level. The Phonological units of signed languages were identified by Stokoe in his paper on signed language structure (Stokoe, 1960). Stokoe identifies 'Cheremes' as analogous to spoken language phonemes. The two terms are interchangeable in the publications of the past decade. In Stokoe's view a chereme is a phonological component of either hand shape, location or movement. Since Stokoe's ground-breaking research, hand orientation as well as various NMFs such as facial expression and mouthing's have been recognised as phonemic components. In Spoken language, a phoneme is the abstract application of phones to a language i.e. a phone exists but is that applied to a given word in a given dialect? The close relationship between phones and phonemes makes it difficult to categorise various sign language components as one or the other. This is the motivation behind Stokoe's term 'chereme'. Due to the limited articulation of spoken language only one phone is produced at a time which creates a linear sequence of phonemes i.e. one meaningful sound after another. When put together, these phonemes create a word or phrase which represents some semantic meaning. Signed languages on the other hand have many articulators which can be used simultaneously creating a multichannel, nonlinear sequence of phonemes. A sign must comprise of all of the phonemes in a given sequence to represent the accurate semantic meaning.

Morphemes are larger units that together make up a word. A simple example is the word *walking* where *walk* is the ‘stem’ and *ing* is the ‘suffix’. There are 2 morphemes in the word *walking* and both add meaning. If the suffix *ing* was to be changed to *er* then the sense of the word would change from the action of *walking* to the subject, the *walker*. These morphemes are represented in signed language by the quickening or slowing down of a sign, repetition, and circular movements.

The linguistic principals employed at a syntactic level are virtually the same for signed language and spoken language. Every language categorises words into groups or ‘word classes’ based on their function, such as nouns, adjectives, adverbs and verbs. The words from these word classes are used to create phrases and sentences according to that language’s syntactic rules. It is the syntactic structure of a phrase that implies some semantic meaning. In the case of signed languages, signs are used instead of words.

2.2.2 Signing space

Signing space refers to the area around the body in which a sign is performed. The primary signing space is at the front of the body and extends from the head down to the waist and 12 inches to the left and right. The most natural signing space takes place directly in front of the torso such that the ‘listener’ can view the face and hands with little effort. This is illustrated in Figure 2-2.

Distance from the body is important and often implies linguistic meaning or intonation. Signs are naturally performed close to the body with the elbows bent. It is not very often that the arm is fully extended positioning the hand far from the body. This usually occurs when using classifiers, placement or when performing a gesture. The same is true when the hand moves to a position behind the body. It is often the case that the hand or fingers touch the body during a sign. This may be to identify a part of the body but there are many signs in ISL that touch the body with no semantic meaning relating to that body part e.g. “BROTHER” and “SISTER”.

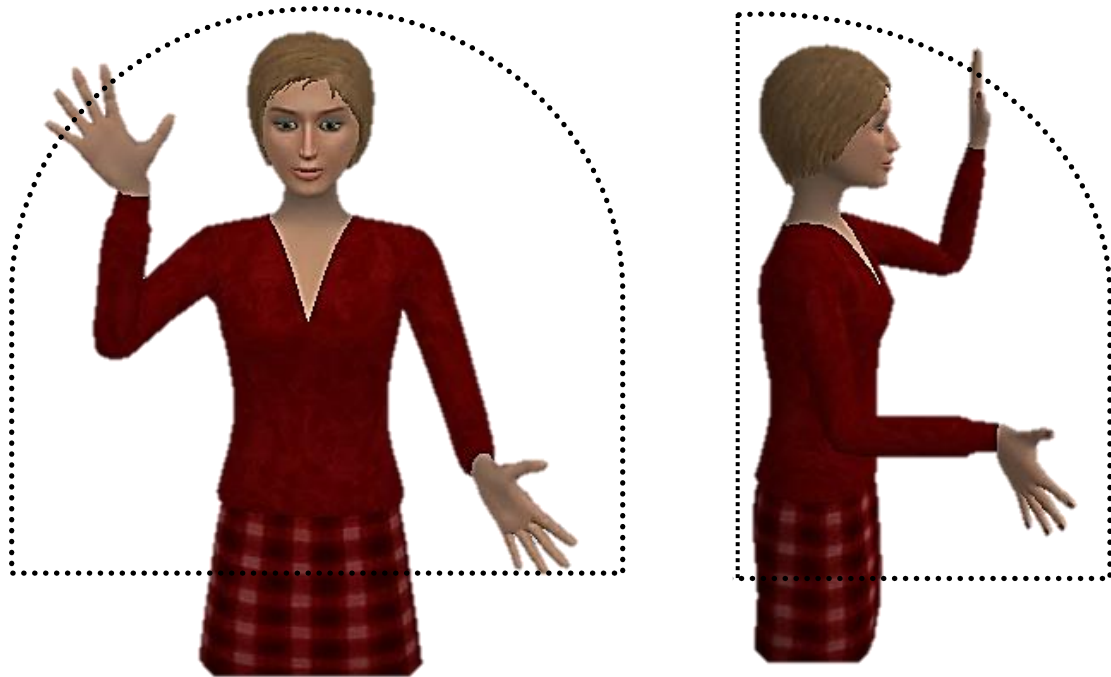


Figure 2-2 Signing space - front and side view

2.2.3 Classifiers and Placement

Classifiers are used in signed language to signify movement, location, and appearance. They, very efficiently, deliver information to the ‘listener’ using ‘mime’ type gestures. For example a gesture in which a one-handed catch is mimed would be considered an ‘Instrument classifier’. ‘Semantic Classifiers’ can be used to indicate a person slipping and falling as seen in Figure 2-3. To do this, the signer would identify the person who is the subject of the sign and proceed to replace that person with a handshape. In this case the handshape for the dominant hand would be that of a closed hand with the index and ring fingers straightened fully. Much like an inverted letter “V” (ISL sign). These fingers signify the person’s legs. The tips of these fingers are placed on the palm of the non-dominant hand which is opened flat with the palm facing upward representing the ground. The signer then mimes as if the subject’s feet slip out and the subject falls flat onto the ground (palm). Other handshapes are used to represent such objects as vehicles, buildings and animals. These handshapes vary depending on the quantity of objects.

Semantic and instrumental classifiers are used here to exemplify the purpose of classifiers. Many other classifier types exist for various purposes, an explanation of these is outside the scope of this thesis.

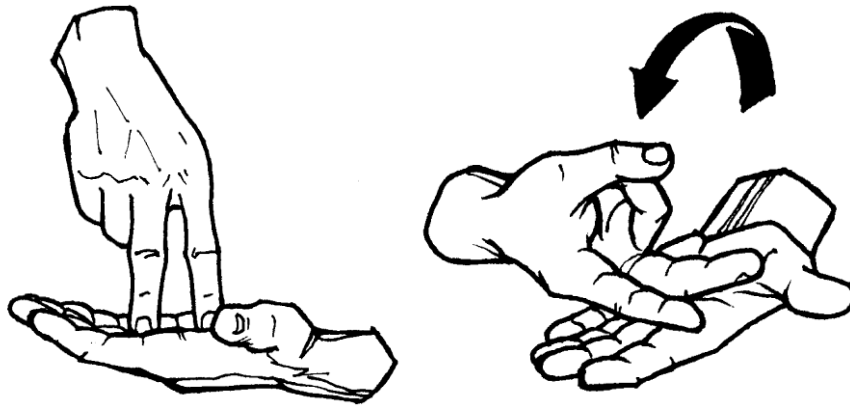


Figure 2-3 Classifier - person falling (ASLU, 2013)

Placement uses classifiers to represent one or more objects in the signing space which can be referred to, by the signer, throughout a conversation. An example might be the location of a book on a shelf or the location of a house relative to a road. In the latter example, the house can be small or large in comparison to the road. There may be other objects between the house and the road and the house may be quite close to the road or set quite far back. All of this information is relayed to the ‘listener’ through some very effectual placement signing. Interestingly, it is quicker and easier to describe a scene through placement than spoken English.

2.2.4 MFs and NMFs

Earlier in this chapter we spoke of the many articulators in signed language. These articulators are split into 2 categories: manual signs and non-manual signs also referred to, in publications, as manual features (MFs) and non-manual features (NMFs). During this document the latter naming convention is used.

MFs include hand motion and shapes made by the arm, hand and finger articulators. These are often mistaken by those who are unfamiliar with sign language as the only articulators. In actual fact, a large amount of linguistic and prosodic meaning is expressed primarily through NMFs (Matthews, 1996), which are widely accepted to carry up to 70% of a sign's meaning. NMFs consist of facial expressions, head/shoulder motion and body posture. Without NMFs it is difficult to recognise a sign's meaning, in fact, it is not possible to distinguish between a sign and a negation of that sign. For example, the MF for the ISL sign “GO” is the same as the ISL sign “DON’T GO”. The difference in the signs being only in the NMF. In this case turning the head from side to side in a ‘no’ gesture carries the linguistic meaning that negates the MF.

According to O'Baoill & Matthews (2000), the functions performed by NMFs are:

1. to show degrees of emotion, including non-emotion
2. to denote intensification/modulation
3. to distinguish declarative/interrogative sentences
4. to denote negation
5. to define topic/comment structure
6. to indicate conditional clauses
7. to show sarcasm

Currently, state of the art signing avatars have reached an acceptable level of proficiency in performing MFs but do not adequately represent NMFs. A justifiable explanation for this is the complex and often subtle nature of NMFs. This fact is compounded by the number of research fields that intersect in the synthesis of NMFs such as 3D animation, sign language linguistics, anatomy and the psychology of emotion and body language.

2.3 Representing Sign Language

The visual-gestural nature of signed language limits the mediums in which signed languages can be recorded and makes it particularly difficult to represent them in a written form. With the introduction of mobile devices and increased bandwidth now available to everyone, it has become easier for sign language users to record sign and communicate with each other across large distances. Video has been used for many years to record signed language. It is reasonable to assume that the volume of video footage with signing content has vastly increased to coincide with the high sales levels of smartphones and other mobile devices (Gartner, 2013). It is also reasonable to assume that by the very nature of these personal mobile devices, the content recorded is of a personal nature and/or not of professional quality. These materials will undoubtedly become useful in some circles but amongst professional bodies, only video of a professional quality with the appropriate content is used in publications, be they web based or otherwise. Generating such content has a high cost in monetary terms as well as being labour intensive and time consuming to create. The result of which, is the low access to information in signed language that we see today.

As a result, the Deaf community turn to other methods of recording their languages such as transcription and avatar synthesis. There are various reasons why one may wish to record sign language in a transcribed format. Those interested in the linguistic analysis of signed languages are interested in various linguistically rich aspects of the language whereas, those interested in recording the language such that it may be reproduced at a later date, may be interested in aspects other than those of a linguistic nature. Those interested in representing a signed language through avatar synthesis require an initial transcription, a computer readable format that may include detail not required in the other 2 scenarios outlined.

2.3.1 Transcription for Sign Language

To date, synthesised signed language systems have used transcription methods that had been designed primarily for linguistic study and therefore lack the explicit detail required for synthesis. We mentioned in section 2.2.1 that signed language makes use of multiple communication channels composed of Manual and non-manual features. Spoken language can be considered linear if we consider it in written text format, which is common practice when computational processing is to be performed. A single channel linear format is easily inputted into a computer system but a multi-channel format presents a unique problem. If this is the case, one might ask: why not record a signed language sentence in written English? This would require some on-the-spot interpreting and serves to highlight the fact that we are recording one language with another language. This would undoubtedly result in a loss of information.

A number of transcription systems have been developed for the purpose of recording sign language. However, a standard or universal system has never been adopted that could represent signed language in a written format. This has resulted in the existence of many. Transcription systems are primarily used in signed language linguistics research. With the exception of the Sutton SignWriting system, which we will discuss later in this chapter, it is not common for Deaf people in general to make use of a transcription system unless they are involved in such research. In this section we provide a brief introduction to a number of these systems, however, the SignWriting, HamNoSys and Stokoe notation systems have gained most ground for various research tasks. In addition, to understand why a transcription system might be lacking in a particular area, we must understand the motivation behind its conception.

2.3.1.1 Stokoe notation system

Stokoe Notation (Stokoe, 1960; Stokoe, et al., 1976) first emerged in 1960 and was the first notation system developed for sign language. William Stokoe, a linguist, developed the self-named notation system in a successful attempt to prove that American Sign Language was in fact, a real language. Stokoe showed that ‘Hand shape’ is a parameter in signed language and through further linguistic analysis particularly with minimal pairs⁴, Stokoe also identified ‘Location’ and ‘Movement’ as parameters. In doing so he also showed that, for signed languages, the hands could replace the tongue as the primary articulator.

Unlike the phonemic level script used for the Roman alphabet, Stokoe notation has a feature level script; a script made up of symbols representing one level lower than the phonemic level i.e. the feature level (Sampson, 1985). In the sense that the script delimits symbols at a word level, it is also logographic. An example of the Stokoe notation is provided in Figure 2-4.

⁴ A minimal pair is two words that differ in only one segment. E.g. rode and rude differ in only the second segment. The two words require the articulator, the tongue, to be in different locations.

Due to the motivation behind the creation of Stokoe Notation it has been criticised as not being technical enough for linguistic research of sign languages (Wilbur, 1987). The notation lacks the ability to categorise important prosodic information such as stress, rhythm and intonation. Stokoe also failed to address non-manual components, stating that it would be much more feasible to do so after the analysis of the basic aspects of ASL had been completed (Stokoe, 1960). Since Stokoe first introduced his notation system, it has been altered and enhanced to fit various research agendas. Arguably, the most noteworthy is the BSL Dictionary Notation which included additional information about finger and palm orientation, body contact information and the hands relationship to each other (Brien, 1992).

The taxonomy of Stokoe notation permits only a finite number of symbols and combinations of those symbols. For example Stokoe identified 19 symbols for handshape, however, the number of shapes a human can make with one hand is, in theory, infinite. Subsequent research has identified a higher number of handshapes for ASL, Liddell and Johnson, for example, identified 150 handshapes (Liddell & Johnson, 1989). A schematic approach, discussed in the following sections, allowing a handshape to be ‘assembled’ as required, would offer a level of flexibility not currently available with the Stokoe notation. Theoretically, a schematic approach permits any conceivable handshape.

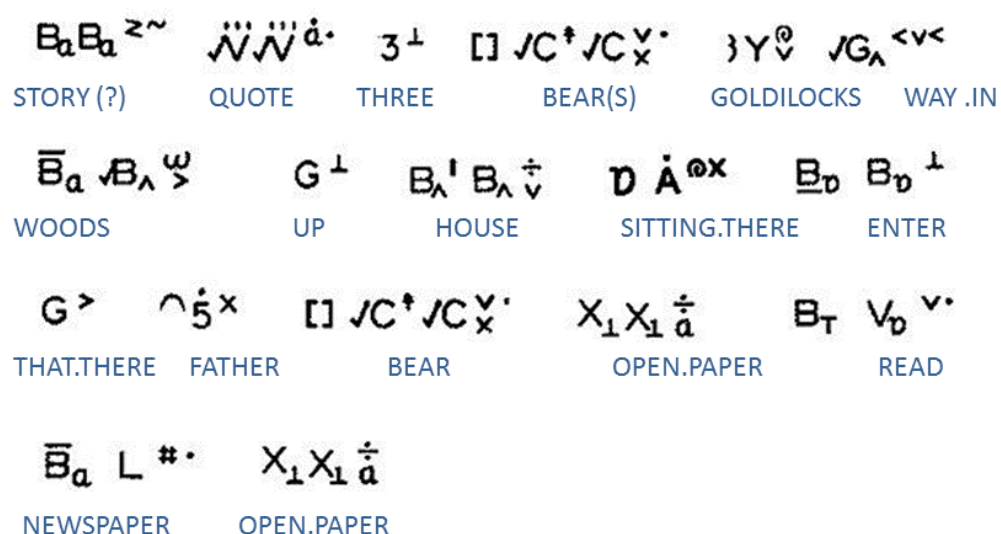


Figure 2-4 A passage from the children's story "Goldilocks", transcribed with Stokoe including English gloss (Martin, 2007)

2.3.1.2 Sutton SignWriting

Interestingly, the Sutton SignWriting notation system was not developed by a linguist but by a dancer and movement annotator. In 1974, Valerie Sutton adapted her Sutton DanceWriting notation to transcribe Danish Sign Language (DTS). The goal being, not to capture any linguistic data, but rather, to simply record the movements of DTS. The SignWriting notation is part of the bigger Sutton movement writing system (also named the international movement writing alphabet or IMWA) which included the aforementioned Sutton notation systems as well as SportsWriting and MimeWriting. The system is so diverse it can also be used to transcribe movement for craftsmen, skateboarders and animations.

In the book, ‘lessons in sign writing’ (2002), Sutton states that her notation system had been adopted by 27 different countries to represent 27 different sign languages. The SignWriting website (SignWriting, 2013a) currently identifies 42 countries using SignWriting. Although these sign languages may use different orthographies, they use the same featural alphabetic script (Jespersen, 1889) comprised of symbols that show phonetic details. This is possible because Sutton SignWriting simply records the movement of a sign. Although by recording the movement the notation also records the linguistic detail, capturing the linguistic detail was never Sutton’s goal. Sutton SignWriting cannot be considered an alphabetic script because it does not include letters like those found in the Roman alphabet. It could be considered Logographic in that it delimits ‘words’ but this is not strictly true either. It is, in fact, a ‘non-alphabetic’ script or an analphabetic script. During his work on phonetic transcription, Jespersen referred to an ultra-alphabetic system of writing which does not symbolize the sounds of characters but the elements of the sound. Jespersen termed this writing system “analphabetic” (Jespersen, 1889). Although there are no sounds in signed language and Jespersen did not consider signed language when he identified what constitutes an analphabetic writing system, the element of the sound as described using a script of this type is analogous to the featural level script used for signed languages. See Figure 2-5 for an example of the Sutton SignWriting notation system.

Today sign writing is a popular medium for recording signed languages in a written form. Quite a lot of literature for reading and writing in signed language is available from the online SignWriting library (SignWriting, 2013c), including popular children’s stories. It is upheld as an easy to learn notation and is popular in ASL. The non-linear format makes it challenging to process SignWriting computationally; however, later in this chapter we discuss the Sign Writing Mark-up Language which has enabled the development of many purpose built tools for SignWriting (SignWriting, 2013b).

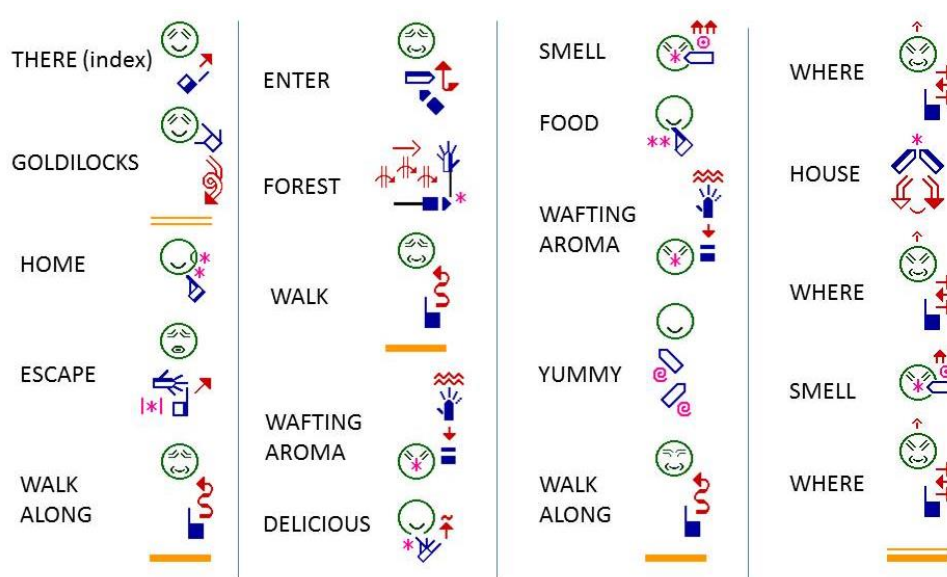



Figure 2-5 A passage from Goldilocks, notated with Sutton SignWriting including English gloss (Sutton, 1999)
Reads from top to bottom

2.3.1.3 Hamburg Notation System

The Hamburg Notation System – HamNoSys, was developed as a linguistic tool at the University of Hamburg in 1984 (Prillwitz, 1987; Prillwitz, et al., 1989). The initial objective of HamNoSys was to represent signed languages in all countries using a set of glyphs/symbols specifically designed for the task. With this in mind, HamNoSys was built on the phonemic structure of signed languages as identified by Stokoe. Also like Stokoe, HamNoSys is transcribed linearly and has a strict ordering of symbols. Sampson (1985) would describe HamNoSys as a “featural” script indicating that a number of features in the script, when combined, can represent a phoneme. Features are represented by symbols in the HamNoSys notation and often carry no meaning in isolation (Figure 2-6). SignWriting and Stokoe may also be considered featural under Sampson’s definition and all 3 notation systems identified so far may be identified as having an alphabetic script (Jespersen, 1889).

Exemplified by its own font (DGS-Corpus, 2013), HamNoSys has been successfully integrated into various software tools for editing and synthesising sign language (this is further discussed in section 3.1.4). Through this integration, the research team at the University of Hamburg were able to identify a number of shortfalls in the notation and subsequently released HamNoSys v4.0 in an effort to address them. Amongst other minor changes, Version 4.0 included more functionality for NMFs, choosing a syntactic pattern which allows the transcriber to replace the hand articulator with another part of the body and then apply various actions to the selected articulator e.g. ‘eyebrows up’ (). The additional NMF functionality, although welcome, is limited in the articulators available as well as the actions which may be applied to them. It was always intended for HamNoSys to support facial expressions by using a placeholder for the face and then describe facial expressions. This did not happen due to the limitations of the single-byte font (Schmaling & Hanke, 2001).

2.3.1.5 ASL-phabet

Earlier we mentioned that Stokoe's notations system has, over time, been altered and enhanced to fit various research agendas. The ASL-phabet is one of these. ASL-phabet, or the ASL Alphabet, is a writing system designed by Supalla, et al. (2001) for American Sign Language (ASL). Unlike the other notation systems mentioned, this notation system was designed primarily to teach signing students literacy skills (Supalla & Blackburn, 2003). The ASL-phabet is based on the SignFont (Lebourque & Gibet, 1999) and Stokoe notations, and consequently is a phonemic script. The number of characters in this notation system is less than those from which it is derived causing some ambiguity. For example, Stokoe has 24 letters encoding types of movement while ASL-phabet has just 5. Like the Stokoe notation, the ASL-phabet does not encode facial expressions or mouthing, and so is perhaps not sufficient for extended text. The ASL-phabet website, developed by the Canadian cultural society of the Deaf (2012) utilizes Supalla's notation to Deaf children learning English. Figure 2-8 illustrates the ASL-phabet writing system.

CAT..... \forall \subset \mathcal{L} \mathcal{H}
CHASE..... \cap \cap \cup \mathcal{L} \mathcal{H}
DOG..... ∂ \exists \approx \mathcal{H}

Figure 2-8 ASL-phabet writing system – ASL notations

2.3.2 Mark-up Languages

A number of markup languages have been developed for some of the more iconic notation systems. Each of these define a hierarchy of elements that represent the various parts of the human body or symbols of a notation system. This makes it possible for computer software to process input which may otherwise be unreadable. The hierarchical structure of a markup language offers a flexible and light weight solution that can be exploited to represent the multichannel structure of signed languages. Most markup languages are easily processed with modern programming languages and require little bandwidth when transmitting. As a result, the markup languages described below are not explicitly tied into any one project but rather, may be used for any project.

2.3.2.1 XML

XML, the Extensible Markup Language, is a restricted form of SGML (Standard Generalized Markup Language), created to structure, store, and transport information (Bray, et al., 2008). XML content is stored in a hierarchy of “elements” which are not predefined and therefore allow for great flexibility. XML was designed so that it is easily readable by humans and computers alike. The textual nature of an XML file allows for minimal storage and transition requirements.

The markup languages described in this section are all based on the XML specification (Bray, et al., 2008) (see Figure 2-9), however a number of projects have represented avatar data without specifying a new markup language, instead they use standard XML. Examples of these include the GesSyCa and THETOS projects (Gibet, et al., 2001; Suszczańska, et al., 2005), the Auslan tuition system (Yeates, et al., 2002) and the DIVA framework (Braffort, et al., 2008). The choice to use standard XML is most likely due to the fact that: 1) There are many open source resources available which are compatible with XML and 2) Unsurprisingly, there is no standard mark-up language for sign language avatars, undoubtedly a consequence of the absence of a standard notation system.

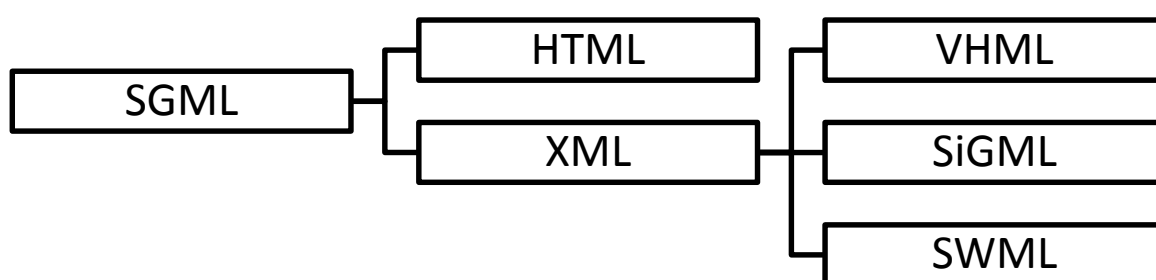


Figure 2-9 markup hierarchy

2.3.2.2 VHML

VHML, the Virtual Human Mark-up Language (Marriott & Stallo, 2002), is intended to facilitate the natural and realistic interaction of a virtual talking head or virtual human with a user through a web page or in an application. VHML addresses the movements required for a virtual human/avatar to converse. In doing this, much of the data in a VHML file contains the subject matter of a spoken conversation in textual form as well as data required to render the avatar. VHML consists of many ‘sub-languages’ that explicitly address a given aspect of a virtual human such as the face or body movements. This allows for more detail to be included at every level but as the literature would seem to show, has never been applied to model avatars for sign language.

2.3.2.3 SiGML

SiGML, the Signing Gesture Mark-up Language was developed during the ViSiCAST project as a way to represent the HamNoSys (version 4) notation system so that it could be employed by computational systems to represent a virtual human signer (Elliott, et al., 2001; Glauert, 2002; Bangham, et al., 2000). SiGML is based on the XML specification but has a more rigid structure in that, only predefined elements may be used. MFs are identified by elements analogous to HamNoSys symbols and although HamNoSys does represent NMFs at some level, it is not feasible to allow the transcriber to include the level of detail required to describe the intricate movements used to perform a NMF. Such transcriptions

would be verbose and difficult to use. For this reason SiGML represents NMFs not through HamNoSys, but through a series of encodings that identify predefined movements (Elliott, et al., 2004). How MFs and NMFs are defined in SiGML and applied through the use of the eSIGN Editor is explained in section 3.1.4.1.

Two types of SiGML were originally produced (Elliott, et al., 2001). The first form “HamNoSysML” was developed to closely represent HamNoSys. Elements exist in HamNoSysML that directly correspond to each HamNoSys symbol. Later HamNoSysML was renamed hns-SiGML (Elliott, et al., 2004) and in a progress report on extending the SiGML notation Glauert refers to it as H-SiGML (Glauert & Elliott, 2011).

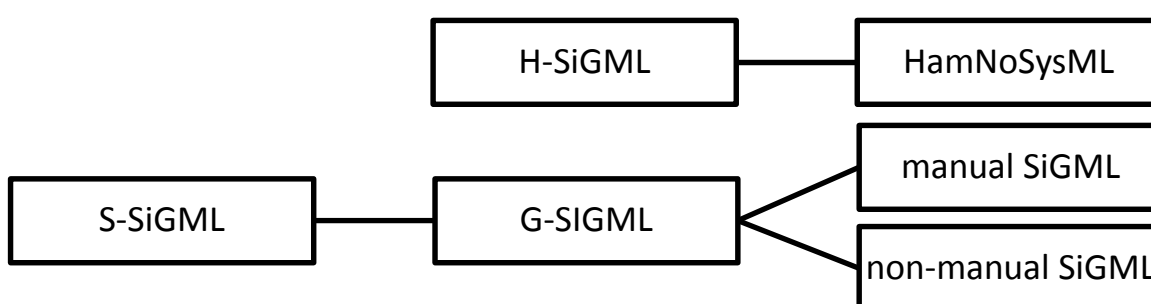


Figure 2-10 the evolution of SiGML

H-SiGML exists simply to represent HamNoSys and to represent a sign in a somewhat human readable format. The H-SiGML Markup for a given ‘signed unit’ can be lengthy and does not contain the necessary information to drive the synthesised avatar. In order to be readable by the synthesis engine ‘AnimGen’, H-SiGML must first be converted to G-SiGML (Elliott, et al., 2004). G-SiGML or “Gestural” SiGML is often referred to simply as “SiGML”. The initial SiGML definition (Elliott, et al., 2001) consists of manual SiGML and non-manual SiGML, both have dedicated DTD files, the entities of which appear together in a single SiGML file. The evolution of SiGML is illustrated in Figure 2-10 and a comparison of H-SiGML and G-SiGML is demonstrated in Figure 2-11 using a variation of the ISL sign for the English word “I” as it was used in SOI corpus.

SiGML was extended to “Segmental” SiGML or S-SiGML recently in an effort to increase precision and flexibility. The paper by Glauert & Elliott (2011) explains how “explicit control over timings” and greater control over direction is achieved through the influence of Johnson and Liddell’s segmental framework (Johnson & Liddell, 2011). The variety of extensions go a long way to addressing the shortcomings of G-SiGML but it is not clear exactly how far they go. Very few implementations of S-SiGML exist. The author knows only of one such implementation: the integration of animated avatars to the corpus annotation tool ‘ELAN’ (Elliott, et al., 2010). In that particular implementation Elliott exploited the new temporal control ability of S-SiGML to match the playback timings of a traditional video that displays a real human signer.

After the ViSiCAST project SiGML was implemented in various other projects outlined by Kennaway, et al. (2007), San-Segundo, et al. (2012), and Morrissey, et al. (2010).

Figure 2-11 illustrates the word “I” as transcribed with HamNoSys, H-SiGML and G-SiGML



HamNoSys

```

<sigml>
  <hns_sign gloss="I VS_4">
    <hamnosys_nonmanual>
      <hnm_mouthpicture picture="m"/>
      <hnm_head tag="NB"/>
    </hamnosys_nonmanual>
    <hamnosys_manual>
      <hamfinger2/>
      <hamthumbacrossmod/>
      <hambetween/>
      <hamfinger2/>
      <hamthumbacrossmod/>
      <hamfingerstraightmod/>
      <hamindexfinger/>
      <hamextfingeril/>
      <hambetween/>
      <hamextfingerl1/>
      <hampalmr/>
      <hamshoulders/>
      <hamlrat/>
      <hamclose/>
      <hammovei/>
      <hamsmallmod/>
    </hamnosys_manual>
  </hns_sign>
</sigml>

```

H-SiGML

```

<?xml version="1.0" encoding="UTF-8"?>
<sigml>
  <hamgestural_sign gloss="I VS_4">
    <sign_nonmanual>
      <mouththing_tier>
        <mouth_picture picture="m"/>
      </mouththing_tier>
    </sign_nonmanual>
    <sign_manual>
      <handconfig handshape="finger2" second_handshape="finger2"
        second_mainbend="bent" second_thumbpos="across" specialfingers="2"
        thumbpos="across"/>
      <handconfig extfidir="il" second_extfidir="l1"/>
      <handconfig palmor="r"/>
      <location_bodyarm contact="close" location="shoulders" side="right_at"/>
      <directedmotion direction="i" size="small"/>
    </sign_manual>
  </hamgestural_sign>
</sigml>

```

G-SiGML

Figure 2-11 The ISL sign "I" represented as HamNoSys, H-SiGML and G-SiGML

2.3.2.4 SWML

SWML, the SignWriting Mark-up Language is an XML-based format developed for the storing and processing of SignWriting texts and dictionaries (Costa & Dimuro, 2001; Papadogiorgaki, et al., 2006; SignWriting, 2013b). An example of SWML is available in Figure 2-12. Much like H-SiGML; elements in SWML are analogous to the symbols in SignWriting. This function of SWML has been useful in its primary application which is to display SignWriting in a digital format. To achieve this goal it is not enough to represent SignWriting in a markup format. The SWML is rendered as a gif image for each sign. This is made possible by the use of a SVG reference file that assigns vector graphics for each element/symbol until a complete sign is produced. Scalable Vector Graphics (SVG) is an open standard, XML-based vector image format for two-dimensional graphics that is packaged with most modern internet browsing software.

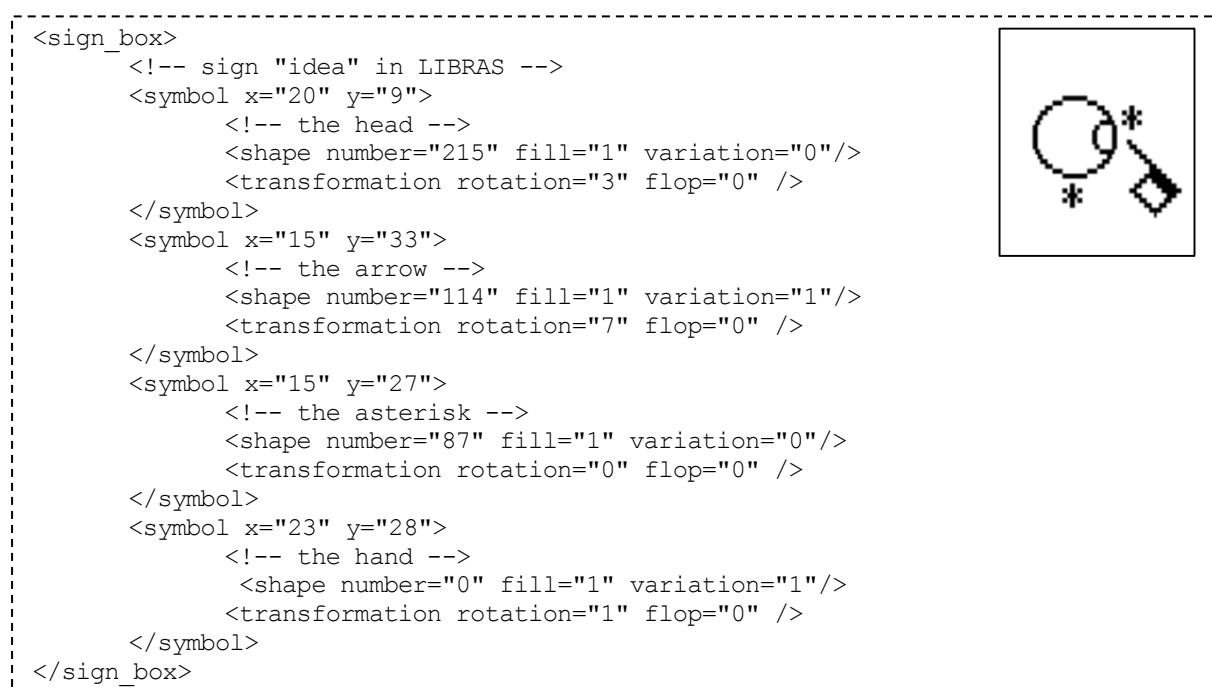


Figure 2-12 the representation in SWML of the sign for “idea” in ‘LIBRAS’ the Brazilian Sign Language (Costa & Dimuro, 2001)

Papadogiorgaki (2006) used SWML in the Greek sign language project Vsigns to exemplify a markup form of SignWriting. This was the form of the initial input for the Vsigns system which then converted the SWML signs into Body Animation Parameters (BAPs) which are used in the MPEG-4 standard. These BAPs could be used to animate a VRML avatar (Virtual Reality Mark-up Language avatar). VRML is an XML-based, ISO standard file format for real-time rendering of animation across the web which has since been superseded by X3D and not to be confused with VHML discussed earlier in this section.

Used in this way the V-Sign project followed a similar framework as the ViSiCAST project in that they both used a markup representation of a particular notation system which was later converted to a

markup more suited to animating an avatar.

2.4 Animation

As technology has advanced over the past number of decades, so too has the techniques involved in animation. Far from the day when Walt Disney first sat down with a pencil to animate ‘Alice in Wonderland’, we now live in a generation that has technologies that far exceed what Disney had access to. In the early 1920s, Disney’s animation team produced only one image every 3 weeks and sound did not appear in animations until 1928 (Thomas & Johnston, 1995). Of course various forms of animation existed before this, dating back to twenty five thousand years ago when our ancestors depicted animals on cave walls. Walt Disney, as well as all other animators up until the 1970s used ‘traditional methods’ of animation which, for the most part, consisted of drawing, by hand, each frame of an animated film. This would have required a substantial amount of time as images were shown at a rate of 12 pictures per second or 720 per minute.

Naturally, as animation techniques progressed, ways to save time were devised such as the use of transparent paper on which various aspects of a frame were drawn. This method allowed for the re-use of these aspects, for example: a background could be used in many scenes, effectively eliminating the requirement for an animator to re-draw that background each time it was required. This method increased consistency as well as reducing man hours.

It was the advent of key frame animation software that most significantly changed the animation landscape, ultimately resulting in the animated feature films we see today. Now recognised as ‘the fathers of computer animation technology’ (National-Film-Board-of-Canada, 1996), Nestor Burtnyk and Marceli Wein, while working at the National Research Council in Canada developed the first system capable of key frame animation (Burtnyk & Wein, 1971). This technology was used in the making of the 2D computer animated film “Metadata” (Metadata, 1971) and later in the making of “Hunger” (Hunger, 1973) which achieved honours at the Cannes film festival, the international film awards and was nomination for an Oscar.

There are now a plethora of key frame animation tools available on the market. The tools are relatively simple to use and can produce a feature animation in a fraction of the time it would take using traditional methods and at a fraction of the cost. However, these are only tools, talented animators are still required to produce quality films since the same animation paradigms apply now as applied in the time of Walt Disney. Only the medium has changed.

2.4.1 Animation paradigms

To appreciate the difficulties involved in the creation of signing avatars we must first understand the basic paradigms of animation and how they may apply to character animation.

2.4.1.1 Fundamental principles of animation

Frank Thomas and Ollie Johnson explain the principals of animation in the very comprehensive book the “The Illusion of life” (Thomas & Johnston, 1995). Originally published in 1981, the principals outlined in this book are used as much today as they were then. Below (Table 2-1) we outline those 12 principals, the description of each identifies how they directly affect avatar synthesis. Figure 2-13 is an illustration from Thomas and Johnson’s book that demonstrate how these principals can be employed to bring to life something as mundane as a sack of flower.

Table 2-1 principals of animation

Animation principal	Description
1 Squash and Stretch	Gives the illusion of weight and volume. Muscles stretch and contract, faces contort, footballs squash as they bounce. Squash and stretch give the illusion that a character is alive.
2 Anticipation	The viewer must be made aware what a character is going to do next. This can be simply a facial expression or a raising of the shoulders.
3 Staging	A pose or action that communicates the attitude, mood, reaction of a character. When an action is staged, it is unmistakably clear what is being communicated.
4 Straight through and pose to pose animation	Planned verses unplanned. Straight through scenes are unplanned and are usually more creative. Pose to pose scenes are planned, the animator draws the main frames and his/her assistant draws the “in-betweens”
5 Follow through and overlapping action	A character does not simply stop at the end of an action. Instead the characters momentum carries appendages, loose skin, clothing a little further before they bounce back. All parts of the body might not stop at the same time.

6	Slow-in and Slow-out	The majority of frame changes takes place at the beginning and the end of a movement. For example a runner's stride is the same or similar through a running sequence but the starting and stopping of that run are different and require more frames to animate.
7	Arcs	Most living creatures move in arcs. Movements are rarely straight left to right or up and down. Instead they tend to use circular movements.
8	Secondary action	Secondary movement may be used to support the main action. If a character is sad he may wipe away a tear also. This help to emphasise the primary action 'feeling sad'.
9	Timing	A characters personality can be determined through the timing of his/her movements. Timing can indicate whether a character is nervous or relaxed, excited or lethargic.
10	Exaggeration	Exaggerating a characters actions, facial expressions and emotions leads to a more 'realistic' animation. In this sense, 'realism' does not refer to a close replica of real world movement but a sense of authenticity from the viewer point of view.
11	Solid Drawing	Characters would need to be redrawn in many positions from many angles. The better one can draw the better an animator they can be. Disney studios asked all animators to achieve weight, depth and balance in every drawing.
12	Appeal	A character needs to have appeal in order to draw and keep the viewers' attention. This must involves some quality of charm such as a pleasing design, simplicity, communication or magnetism.

Table 2-1 offers only a short description of the 12 animation principles, however it is plain to see how these principles may be applied to avatar synthesis. Very few of the 12 principles should be excluded, for example, the users of a signing avatar will not show interest unless there is ‘Appeal’ and the avatar is drawn well or ‘Solid’. The principles tell us that, the way to show emotion in an avatar is not to emulate a real signer but to ‘exaggerate’ their movements perhaps with ‘timing’ and ‘secondary actions’. These 12 principles have been used for many years in some of the most iconic animated shorts and feature films ever created. It would be ill-advised to simply disregard them from synthesised avatar animation.

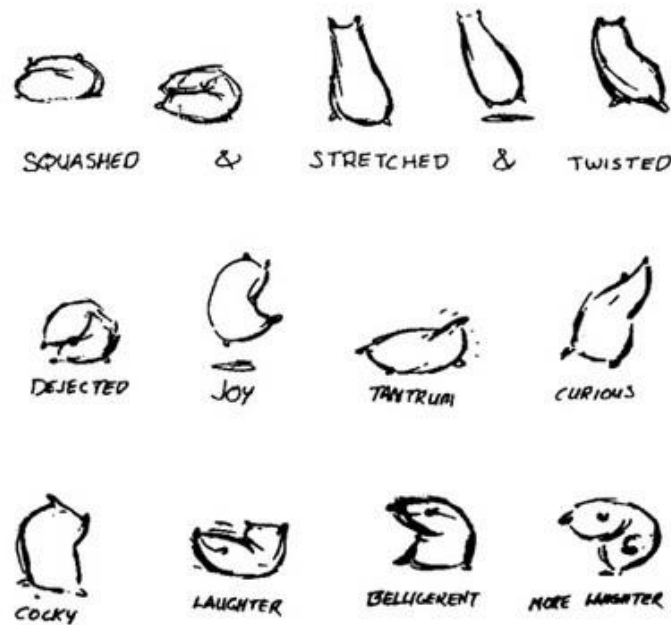


Figure 2-13 the famous half-filled flour sack (Thomas & Johnston, 1995)

2.4.2 Human emotion

In order to create an engaging avatar, it is not only important to understand basic animation principals but also the role and significance of facial expression. Facial expressions can be viewed as cues to an underlying emotion. An outward projection of that emotion which, when synthesised, must be unmistakable. It is therefore important to explore what defines emotion and their outward projection that we refer to as emotional facial expressions (EFEs).

Humans and several primates have the ability to recognise various facial expressions. This allows us to not only recognise familiar faces, but also to identify a particular mood or feeling. Face-sensitive neurons in the brain have been identified that carry out this very task (Malim & Birch, 1998). However, studies have yet to agree on a method by which we can define even basic emotion. This is because, at least in part, human emotion belongs to many fields of research: biology, anthropology, sociology, physiology, neuroscience, facial recognition and animation, to name but a few. Charles Darwin (1872) theorised that; a humans basic emotions, like that of an animal, are primitive or

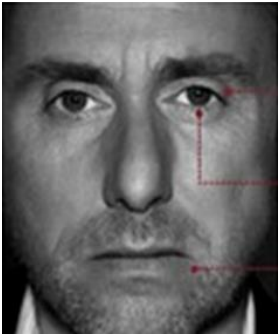



biological in nature, a consequence of the evolutionary process. Others (Allport, 1920; Peiper, 1963; Averill, 1983) consider emotion as either wholly or primarily sociological, a learned experience which humans gain from their cultural surroundings. Emotions such as guilt and jealousy are primary examples of emotions that are cultivated by interacting within society.

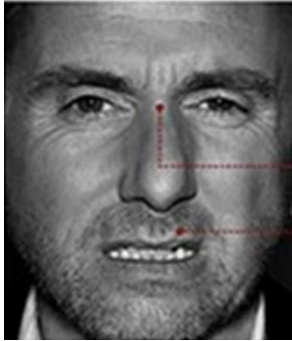
Ekman (1971) originally agreed with Darwin's theory, at least in part. He spent 40 years investigating the concept of basic emotion and universal facial expressions across many cultures. Ekman's experiments, which support Darwin's theory, were based on multi-culture groups, some, like the group in the southeast highlands of New Guinea, were totally isolated from popular culture. Although many anthropologists (Lutz & White, 1986) argue that emotions are not universal. Ekman's literature initially identified 6 universal facial expressions: disgust, sadness, happiness, fear, anger and surprise (see Table 2-2 Ekman's 7 Universal emotions demonstrated by actor Tim Roth. This was supported later by Plutchik (1980) who expanded Ekman's list by adding 2 more emotions: anticipation and trust. Plutchik's list of basic emotional concepts identifies 8 emotions, 4 emotions with 4 direct opposites which cannot be experienced at the same time (Figure 2-14). Later Ekman revised his original set of basic emotion to include contempt which became the 7th universal facial expression, see Table 2-2(Ekman, 1999).

Whether a human face can express any more than these seven emotions is a matter of some debate. Ekman (1999) indicates there could be specific facial expressions for other emotions such as contentment, excitement, pride, relief, guilt, and shame. As they have yet to be outlined in any literature to date, this thesis will focus primarily on Ekman's 7 universal emotions.

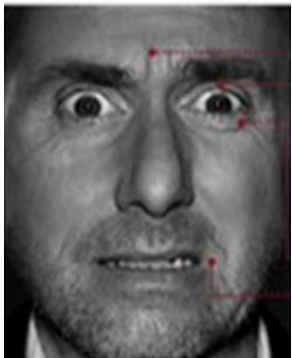
Ekman's work was made famous outside of his field of research when he became a consultant on the American TV series "Lie to me". The show premiered on the Fox network on January 21, 2009 and followed the exploits of a character played by Tim Roth who could identify a lie simply by facial expressions. The premise of this TV show is based on Ekman's actual research on micro-expressions which can be used to identify deception (Ekman, 2003).

Table 2-2 Ekman's 7 Universal emotions demonstrated by actor Tim Roth

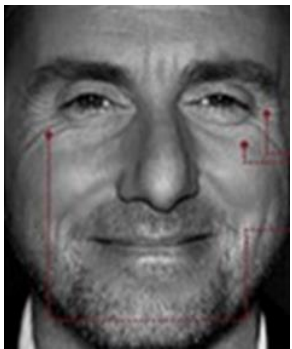
<p>SADNESS</p> 	<ul style="list-style-type: none"> - the eyelids droop - the inner corners of the brows rise (<i>in extreme sadness, the brows draw together</i>) - the corners of the lips pull down - the lower lip may push up in a pout
<p>SURPRISE</p> 	<ul style="list-style-type: none"> - the upper eyelids and brows rise, - the jaw drops open
<p>ANGER</p> 	<ul style="list-style-type: none"> - the lower and upper eyelids tighten - the brows lower and draw together (<i>Intense anger raises the upper eyelids as well</i>) - the jaw thrusts forward - the lips press together - the lower lip may push up a little
<p>CONTEMPT</p> 	<ul style="list-style-type: none"> - appears on just one side of the face - one half of the upper lip tightens upward

DISGUST

- the nose wrinkles
- the upper lip rises while the lower lip protrudes

FEAR

- the eyes widen
- the upper lids rise, as in surprise
- the brows draw together
- the lips stretch horizontally

HAPPINESS

- the corners of the mouth lift in a smile
- the eyelids tighten,
- the cheeks rise
- the outside corners of the brows pull down

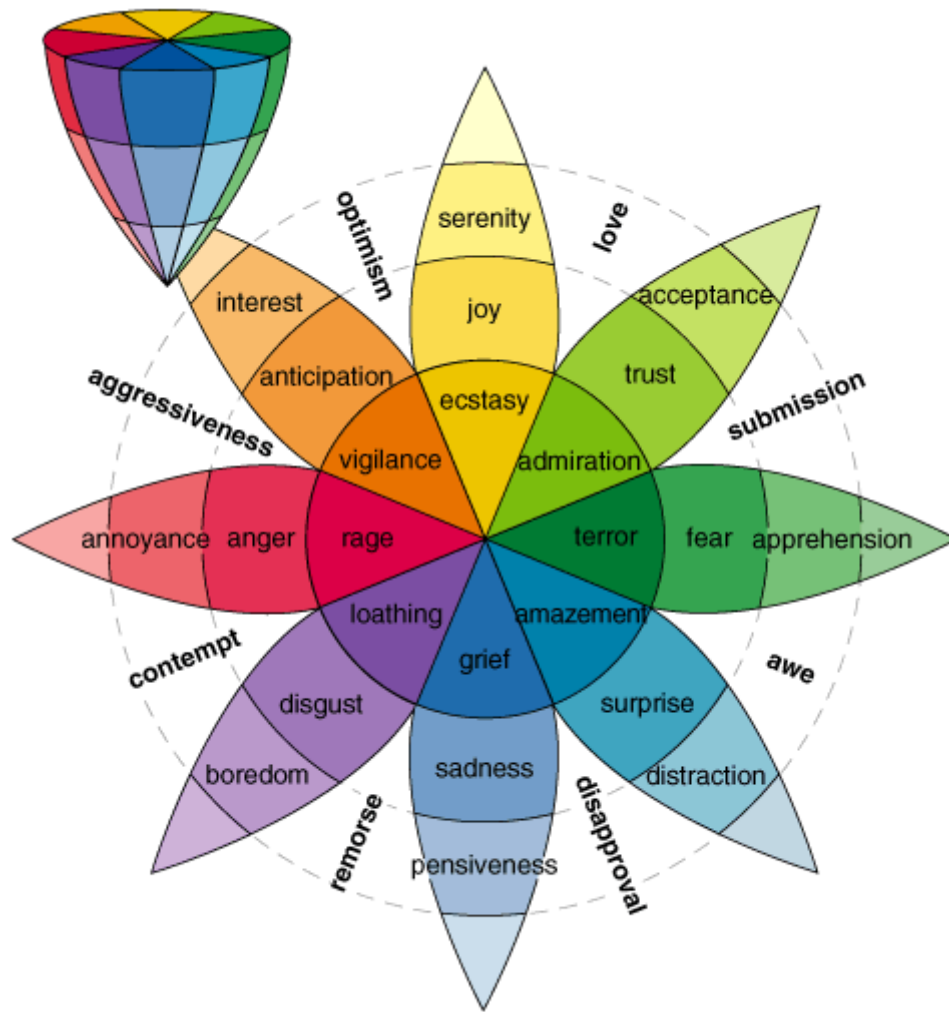


Figure 2-14 Plutchik's three-dimensional model describes the relations among emotion concepts

2.4.2.1 Facial Measurement

For the purpose of synthesising a humans movements, particularly facial expressions, it is not enough to simply identify facial expressions for the 7 universal emotions. There must be a scientifically appropriate way to identify and measure these movements. Simply, pulling faces in the mirror is not enough. With this in mind, the Facial Action Coding System (FACS): a technique for the measurement of facial movement was published by Ekman and Friesen in 1978. This paper outlines a system that may be used to identify facial muscles and related facial movement. Using this system one could pin-point, at an anatomical level, muscles that are used to express each emotion using “observable components” called action units (AUs) (Ekman & Friesen, 1978). In 2002 a FACS manual was released on CD-ROM which describes how to identify and use these AUs (Ekman, et al., 2002). Supplementary to this a database called FACSaid was created to map various facial movements to a series of AUs. This database may be used to identify AUs for a given emotion or vice versa. FACSaid is an acronym for ‘Facial Action Coding System Affect Interpretation Dictionary’. According to this dictionary, the facial expression for the emotion ‘Fear’ uses the action units 1 + 2 + 4 + 5 + 20 + 26 (Hager, 2003).

2.4.3 3D character development

When considering 3D character development for the first time, it is good to picture, in the mind's eye, a physical analogy such as a 'Papier Mâché' sculpture. In its most basic form, the process of developing a papier mâché sculpture involves placing a layer of paper and adhesive over a rigid structure. The rigid structure can be constructed of many different materials, a common material being a thin flexible wire mesh such as chicken wire. The flexible mesh may be manipulated into any shape the sculptor desires. Considering the subject matter of this thesis, it would be appropriate to take the human head as an example. During the creation of this human head, the sculptor, depending on the level of detail they wish to portray, may use a 3 step approach: 1) create the head shape with wire mesh, 2) cover the mesh with paper and adhesive and 3) paint the head, including eyes, lips, and other features. This process is very similar to that used in the creation of a virtual 3D human head and in both cases the 3 step approach may be extended, depending on the level of detail and functionality required. Extra detail, such as hair for example, may require an extra step.

The first step in creating a 3D character or any 3D object for that matter, is to create a wire mesh which is moulded to the desired shape. In a virtual environment this is done by creating a geometric representation for the desired shape using a series of polygons which together make a 'polygon mesh'(Figure 2-15). The point at which these polygons intersect is called a vertex and it is actually these points that are recorded. Similar to mesh modelling with papier mâché, polygonal modelling is a process in which one can create a complex 3D model of an object using polygons. Smaller polygons usually means a more detailed mesh which would result in a more computational expensive rendering process.

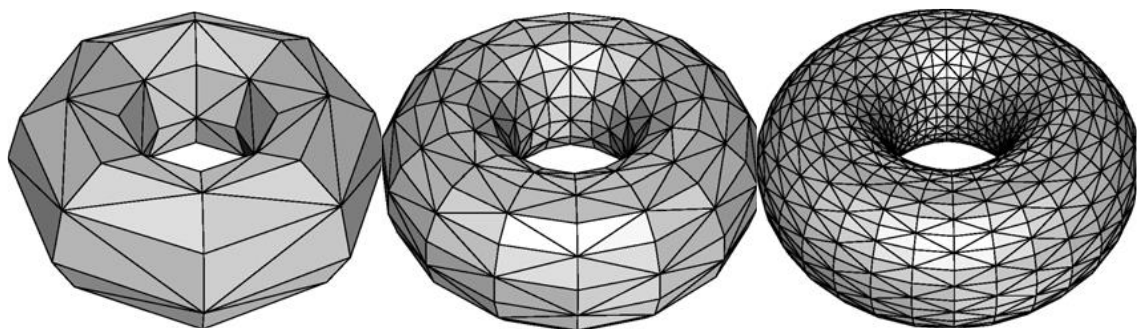


Figure 2-15 Three levels of subdivision in a polygon mesh (Velho, 2001)

Once a mesh has been created, movement may be added. This is made possible through the use of a skeletal-like structure called an armature. The armature is mapped to points on the polygonal mesh such that if a bone is moved, the corresponding section of the mesh is moved also. This process is known as rigging. Bones in an armature are conventionally connected by a hierarchical chain, in this regard the hip bone is connected to the upper leg, the upper leg is connected to the lower leg, the lower leg to the foot and so on. At this point the mesh can be positioned by moving or rotating the armature bones, which may be restricted at points to stop joints rotating on an 'unnatural' axis. Positioning a 3D model by means of its armature structure can be quite cumbersome as each bone must be moved separately.

To simplify this process, Inverse Kinematics (IK) is often used. IK allows the animator to position the last element in the IK chain and allow local controls to determine the rotation and translation of joints up the control chain's hierarchy. For example, if the foot bone was to be moved, the lower leg bone will move with it as will the upper leg and so on up the chain as required.

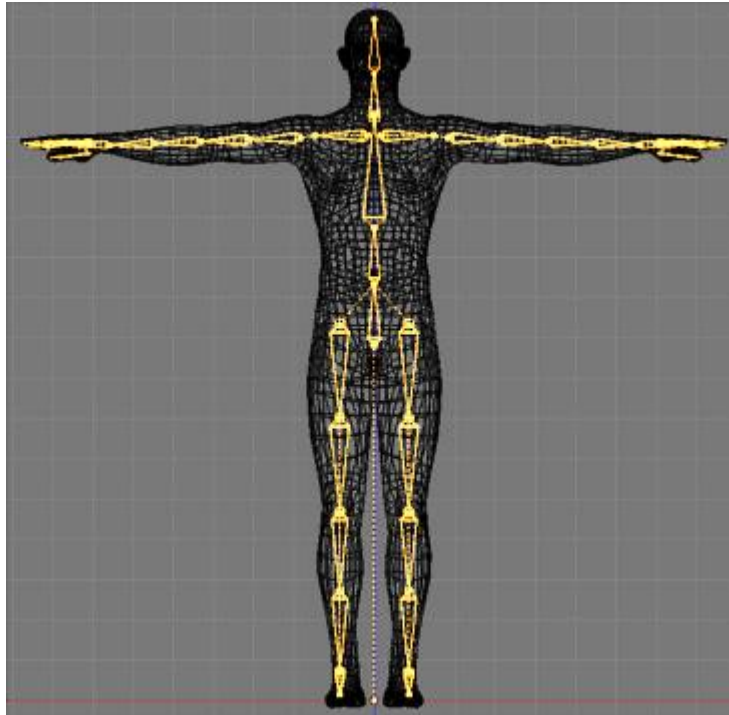


Figure 2-16 armature structure with polygonal mesh (unity3D.com, 2013)

Morph target animation is often used alongside armature animation for movements that are too small or too cumbersome to rig. In reality, the animator deforms the appropriate polygons on the mesh as required on a frame by frame basis. What the viewer sees, however, is character movements such as the cheeks filling with air or an eyebrow raise. This topic is discussed in more detail in section 3.2.2.

Texture mapping is the process of wrapping 2D images around a multidimensional object to define the characteristics of that object's surface. This process is analogous to the second and third steps in the papier mâché example given earlier: wrapping paper and adhesive to a mesh and painting features. The texture map adds a suitable appearance and feel to the object surface. That may be to give an object a textured look such as metal, glass, wood or skin. With texture mapping a simple rectangular cuboid can be transformed into a skyscraper with a simple image. Texture maps can be created from digital photographs or created by hand using image editing software. 3D animation packages have the ability to assign a coordinate value of the texture map image to a vertex in the mesh. RGB values are then altered, depending on surface depth. In the case of 3D characters, texture maps are used to simulate skin, hair and clothing. Clever use of lighting and camera angles will help give a 3D scene depth and highlight focal points. Bump maps and Normal maps, which would go beyond the simple 3 step papier mâché example, take advantage of the scenes lighting to simulate bumps and wrinkles on a 3D object. A bump map, a type of texture map may be used to identify minor areas of shadow or light, ideal for simulating

pores, acne, scars or wrinkles on skin. Other texture maps do exist such as reflection maps, opacity maps and specular maps which simulate reflection, transparency and brightness respectively.



Figure 2-17 the avatar Anna and the various texture maps used in her creation

2.4.3.1 Character animation software

The creation of any 3D object must be done in a specialised virtual environment which includes the tools required for each step of the process. There are many editing tools available for 3D modelling and animation. Most have been developed for creating a whole 3D world such as 3DS Max, Blender3D and Maya but others have been created specifically for animating virtual humans. Popular packages for this include Poser and Daz3D. The game and movie industries often develop characters in a package such as Poser only to import them at a later stage into a broader 3D modelling/animation package such as 3DS Max. This allows for a high level of complex detail to be added to a character with relative ease while other members of the team model the landscape for a scene. Often the tools with more advanced features (content libraries and user friendly interfaces) are commercial products and are an expensive investment for a short term research project. One of the very few exceptions to this rule is Blender, a free open source 3D content creation suite, available under the GNU general public license. With regard to tools specifically developed for animating virtual humans, it is also possible to download a free version of Daz3D.

2.4.4 3D rendering systems

Rendering software is required to process the avatar and scene data created in a 3D modelling or animation package. Put simply, a renderer will take a series of values and transform them, on screen into a graphical representation of those values. For example an internet browser will render simple HTML as a well-designed website. At a lower level, graphics software and hardware work together to render a string of 1s and 0s into the graphical display seen on a computer screen. Rendering is simply the process of taking computer readable data and presenting it in a human readable format. This includes text, video and graphics (2D and 3D).

This process is applied to rendering 3D data. 3D scenes are represented as numerical data in the form of geometric values, light values, texture, camera angles and more. Rendering software allows these values to be transformed into a 3D animated scene by utilising a computers CPU, graphics card and other hardware. A high level of complexity in a 3D scene requires more data to be processed at a higher speed. This is computationally expensive and requires more of the computer's hardware resources for the rendering process.

Given a particular scene the renderer calculates individual pixel values for each frame in that scene. If each frame takes 15 seconds to render: A 60 second animation would take approximately 7.5 hours to render (60 seconds x 30 frames per second = 1,800 frames x 15 seconds per frame = 7.5 hours). The latest 3D feature film will have been rendered using a cluster of high power machines but even with this, rendering is far from instant and animators must plan production well as scenes are often rendered, changed and rendered again.

In the case of signing avatar synthesis, avatars will most likely be rendered on a home computer, without high spec hardware to increase the rendering speed. In this instance it is important to reduce the amount of data that needs to be passed to the render. This includes reducing the number of polygons in the avatars mesh as well as reducing the complexity of the armature and the texture maps. In doing this the animator must walk a fine line between the acceptable level of detail in the avatar, required by the user, and the performance requirements of the average home computer.

High level rendering tools come packaged in all 3D modelling and animation packages. These save time and money but do not offer much flexibility. For that reason many projects such as ViSiCAST chose to develop their own high level tools which allow, amongst other things, real-time rendering. These tools must make use of the API library used to access a machines graphics hardware. The 2 most common API packages are OpenGL and Direct3D. Both can be accessed through a host of programming languages such as C, C++ and Java. The current incarnation of the ViSiCAST project utilizes JOGL to access OpenGL. JOGL is required to allow Java to interface with the OpenGL libraries. Direct3D is one of the DirectX APIs developed by Microsoft. Direct3D and OpenGL offer a very similar level of access and performance with graphical hardware. The primary difference being that of licensing. Direct3D is

licensed by Microsoft whereas OpenGL is open source and therefore works well with Java through JOGL.

2.5 Synthesised signing systems

In contrast to motion capture and video production, synthesised signed language avatars offer a more cost effective and efficient solution to the lack of signed language representation in human computer interaction. Other advantages include a low bandwidth requirement, ease of reproduction and the option to simulate many signed languages simultaneously. This is made possible by the idea that the avatars movements can be completely synthesised from some form of textual input. This may be direct input from a text based file or a transcriber, equally, it may be output from another source such as a machine translation system.

Sign language avatar synthesis is not a large research field in respect to the number of researchers in it. The relative newness of the field has led to a landscape of many ideas not yet brought to fruition. Current and past projects attempt to use avatars in a variety of scenarios, for example some projects such as the Albuquerque weather forecast system (Grieve-Smith, 2002; Grieve-Smith, 1999), focused on various aspects of machine translation for sign language (SLMT), for this reason, little attention was paid to the realism and usability of the avatar which was generated using keyframe interpolators in Virtual Reality Modelling Language (VRML). Another example of such a project is the Greek project, Vsigns (Papadogiorgaki, et al., 2005), which developed a virtual human avatar learning tool which was used to teach signed language to interested parties over the web. 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 rather than human-like representations.

Research carried out by Kipp, et al. (2011), indicates that the visual appeal of an avatar is an important attribute for Deaf users. Not only this but also the facts that linguistic information is lost if fingers and facial movement are not synthesised or they are synthesised badly. With this in mind many researchers have aimed to create avatars of the highest practical photo-realistic quality. A project from University of west Bohemia, Pilsen have developed their avatar with this in mind. The 'Czech sign speech synthesiser' (Krňoul, et al., 2008; Krňoul & Železný, 2007; Chaloupka & Chaloupka, 2009) is a phrase-based translation system for Czech-to-CSE. Underlining their desire to achieve a photo-realistic avatar, they have created a proprietary rendering engine with C++ and OpenGL which, as the literature would seem to indicate, surpasses any other projects for photo-realism in the face.

Another venture that puts the primary focus on the avatar quality is VCom3D (VCom3D, 2007). Unlike the other projects mentioned here VCom3D is a commercial entity. Their website describes the company as a leader in providing multi-cultural, context sensitive, virtual communicator characters for enhanced learning. The company has made much headway in developing and patenting many tools to provide this service. Their tools are used by a number of clients including academic institutes and the

US department of defence. With their Vcommunicator tool, they have realized a superior quality avatar which may be configured by means of an editing suite so that users can create their own animations. The suite provides the user with a timeline interface to drag and drop avatar movements. A number of generic avatars come packaged in the software or a bespoke avatar can be provided at additional cost. This application is aimed directly at corporate and government entities such that they may personalise their VCom3D e-learning experience.

VCom3D have also developed a number of avatar based applications that endeavour to address the communication needs of the Deaf community including some e-learning tools specifically aimed at the needs of Deaf and hearing people. One such tool is the Sign Smith Studio which may be used to construct signed phrases and sentences which will then be performed by an avatar. Like the Vcommunicator this suite provides the user with a timeline interface, in this package the user may choose a signed utterance from a pre-defined lexicon of ASL signs and add them to the timeline to construct a phrase or sentence. New signs may be added or existing signs can be edited to include various NMFs.

During a comprehension evaluation carried out by VCom3D (Hurdich, 2008) on Deaf and Hard-of-Hearing students from kindergarten age to 12 year old, an increase of 17% to 67% was observed when shifting from text-only to text accompanied by signed language, when using Vcom3Ds SigningAvatar technology.

There is one project or, more accurately, a series of projects that has attempted to address a range of communication issues facing the Deaf community across a number of signed languages. The initial ViSiCAST (Bangham, et al., 2000) and eSIGN projects resulted in an avatar synthesis system that could be driven by the output of a machine translation module. The current state of the project can be seen in the EU funded Dicta-Sign project which was established to leverage existing web 2.0 technologies such that wikis, social networking sites and blogs can be accessed through signed language. To get a full picture of this project we must first appreciate the projects history: The ViSiCAST project was a 3-year, EU-funded project involving a collaborative approach by 3 institutions 1) The Virtual humans group at the UEA Norwich, 2) IDGS - University of Hamburg and 3) Televirtual. The goal of the project was 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 (Cox, et al., 2002), a signed language avatar translation tool, tested in UK post offices aimed at providing limited access for Deaf customers. Critically, the software did not translate to British Sign Language (BSL), instead it translated to Sign Supported English (SSE) which follows the grammatical structure of English rather than any signed language.

The lessons learned in the ViSiCAST project were built upon in the eSIGN project. Also, a 3 year EU funded project built upon the technology already developed in the ViSiCAST project by introducing

synthesised signed language. Up to this point all of the signing content was based on motion capture technology. eSIGN was responsible for the later versions of SiGML, which is discussed at length in section 2.3.2.3. eSIGN moved away from motion capture technology to fully synthesised signing which allowed for more flexibility with the signing content which the project utilised in its ultimate goal: to bring multiple signed languages to eGovernment websites (Glauert, et al., 2004). These language include BSL-English, DGS-German and NGL-Dutch. Ultimately, the modular based SiGMLSigning framework (now named JASigning) was developed during both the ViSiCAST and eSIGNs projects. The Virtual human team at UEA endeavour to keep progressing the system. During the Dicta-Sign project the SiGML notation system was extended to include temporal controls (see section 2.3.2.3) and the framework was rewritten in the Java programming language for cross platform compatibility.

None of the systems mentioned here sufficiently synthesise NMFs. This is unfortunate because, as previously mentioned in this thesis, a large portion of a signs linguistic meaning is communicated through the NMF channel. To be reasonable, many of the systems mentioned here have progressed in their representations of NMFs but none have fully resolved the issue. Another problem with these systems, particularly with the VCom3D Sign Smith software is the inability to accurately represent sign language specific anomalies such as spatially inflected verbs (Huenerfauth & Lu, 2010). Figure 2-18 illustrates avatars from various projects including ViSiCAST, eSIGN and VCom3D

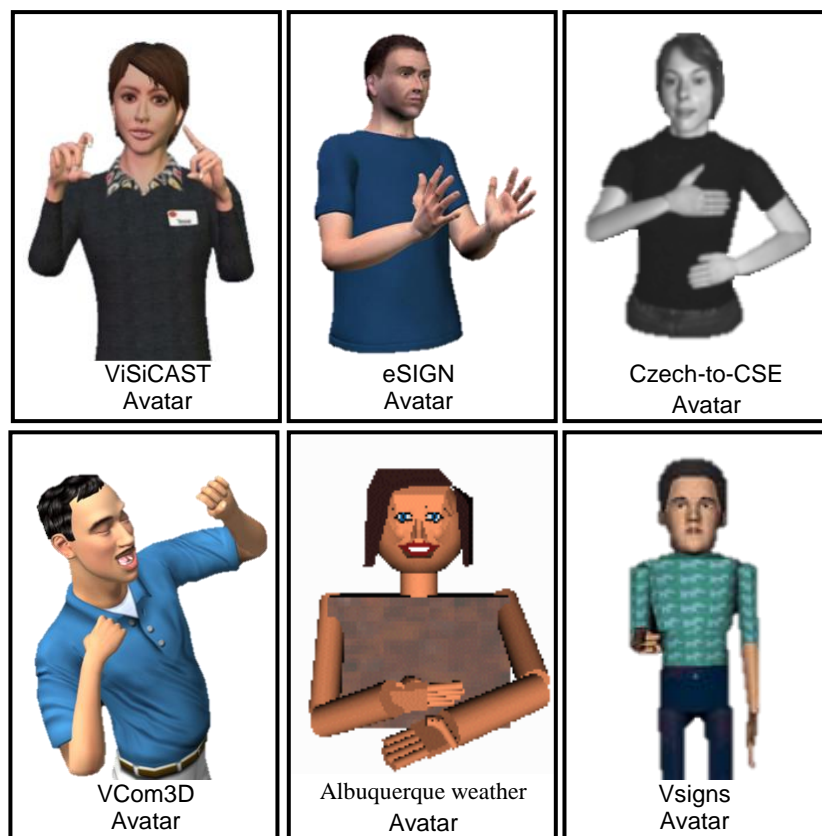


Figure 2-18 Signing Avatars from 6 different projects

2.6 Motion Capture and Digital Video

Besides synthesising data for signing avatars there are 2 other popular methods used to communicate sign language in digital format: 1) digital video and 2) motion capture driven avatars. These methods should be considered separate to written forms of signed language as they attempt to record and playback signed language in its natural form i.e. a signed performance.

2.6.1 Digital Video

Today, digital video is the most common format for presenting signed language. Digital video or more specifically a signing video as we refer to it in this document, consists of a real person signing in the foreground of the scene. Typically only the signer's head and torso are visible. The signer's body below the waist is not recorded in the frame, allowing for a greater focus on the arms, hands and face. These signing videos can be seen across many mediums including the web, DVD, and CD-ROM.

Video is undoubtedly the most accurate form in which to represent signed content, as every aspect of the signer and the signing content are recorded exactly as they were originally performed. If this is the case, why bother with avatars at all? As advantageous as video is, it is not without its limitations. True, it is the most accurate way to record signed languages but it is also the most problematic to produce, store, transmit, update and concatenate. Video production is costly, not just in monetary terms but also with respect to time and resources. Updating signing video content, on a web page for example, requires more production and therefore more time and an additional budget. Such costs are not usually associated with the corresponding textual content. All video files require a large amount of physical memory, which impacts not only on storage requirements but also transmission speeds. In the instance of video on the web, a fast network connection is required at the user end and as well as a large bandwidth allowance on the hosting network. The more popular the content, the more costly the supporting network infrastructure will become. Finally, video clips will not blend well together unless they were designed to do so. By this we mean: footage captured for an information DVD for example, can be easily edited with transitions to breakup subject matter or transition between camera angles etc. In this example consistency is important: the actor/presenter would need to wear the same clothing and there would be strict control over props, lighting and backgrounds. It would be unacceptable to source footage from a number of such sources and produce a 'mish-mash' of presenters, costumes, backgrounds etc. in order to structure a new sentence from existing content. Even if all the footage was of the same presenter in the same clothes there is still the issue of seamlessly blending the clips together. In the television industry this is most commonly referred to as seamless splicing (Cheng, et al., 2004). Seamless splicing is possible when clips have minor differences and so long as extra time is available at the beginning and end of each clip in which the splicing may occur.

In many cases, the costs involved in the provision of signing video content, with regards to time, money and recourses are too high. Many potential providers often fail to see any benefit in making their information accessible to the Deaf community. The result is that, in a world where information should be at everybody's fingertips, a deepening divide is forming between the quantities of information available in popular languages compared to that which is available in minority languages, including signed languages. This is the driving force behind signing avatar research.

2.6.2 Motion Capture

One may consider motion capture (mocap) as an advanced form of rotoscoping, a technique that was used by Disney in the creation of Snow White in 1937. The process involves tracing real-world actors frame-by-frame and using those drawings as a basis for an animated movie. Much like rotoscoping, mocap uses real-world actors as a basis for animated content. However, in the case of mocap, the process is much less painstaking due to the way in which the initial data is recorded. Instead of tracing each frame by hand, information from sensors positioned on key points of the actor's body is captured and used to move a character in a 3D virtual space. These days the sensors are not actually sensors but most commonly take the form of small lights or coloured dots. The movement of which is recorded by an array of 2-96 specialised cameras. The creators of the Optitrack infrared camera system (Optitrack, 2013), recommend a minimum of 6-8 cameras to affectively capture the movements of one person. Typically, the cameras are setup around the subject in close proximity as is the case in Figure 2-19. State of the art mocap technologies include a face rig that can capture even the smallest amount of facial movement. One such rig was used in the filming of the 2012 movie "The Hobbit: An Unexpected Journey" (see Appendix D – Motion capture rig).

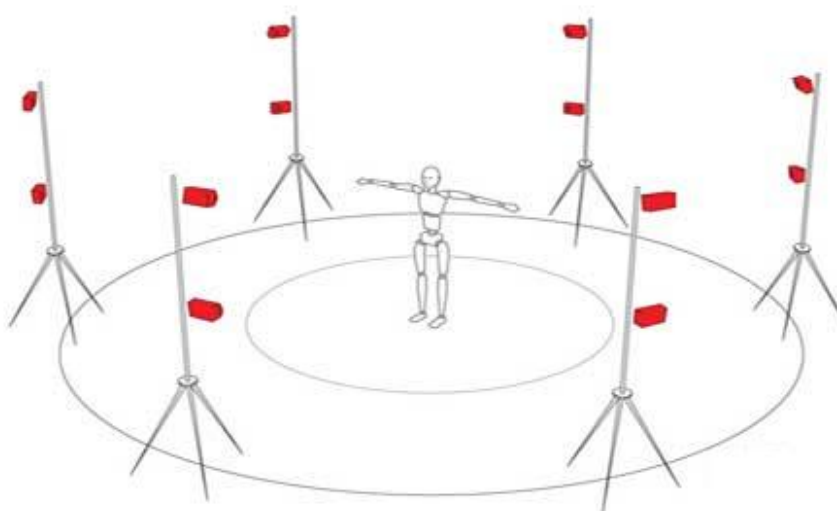


Figure 2-19 Motion capture setup (Optitrack, 2013)

Motion capture was used in projects such as ViSiCAST but was quickly abandoned due to the costs and inflexibility involved. This technology does solve many of the problems inherent in digital video, for example, the footage requires drastically less physical storage and is more easily transmitted due to the numerical or textual format in which it is recoded. The issue of seamless splicing no longer exists as 'in-between' frames may now be synthesised to seamlessly transition between one clip to the next. Motion capture does fall down, however in two areas: the production process is often more costly than video and, like video, the prospective library of pre-recorded signs is finite. If a sign is not in the library, it cannot be used. Of course, one may produce more signs and add them to the library, although this is expensive. Even after the initial capital expense, equipment must be maintained, studios must be made available and often, external expertise must be sourced.

State of the art motion capture does have the ability to capture realistic human movement and facial expression. In this regard, synthesised sign language does not yet compete. Often avatars driven by synthesised input are unnatural, almost robotic in their movements. Their key advantage over motion capture being that of flexibility. Fully synthesised avatars do not have a finite library of possible signs. Instead any conceivable sign is possible, theoretically at least. That does not mean that motion capture has no place in the creation of signing avatars. It is the author's belief that motion capture has a role to play in the recording of facial expressions, which may become 'base morphs', that may later be tweaked by a synthetic system.

3 Methodology

An evaluation was conducted to ascertain whether or not it is possible to improve the understandability of a synthesised Sign Language avatar through the introduction of emotion. We further evaluate the Deaf community's disposition towards signing avatar technology as well as their preference between avatars designed to resemble humans and caricature styled avatars. In order to conduct these evaluations we must first have an avatar in place along with some synthesised output. With this in mind we have chosen to use a subset from the well-established Signs of Ireland (SOI) corpus (Leeson & Nolan, 2008) and the JASigning (Kennaway, 2003) synthesised sign language avatar system as our platform. We discuss these technologies in the subsequent sections.

3.1 Corpus

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. An existing corpus must be used. Currently there are only 2 corpora with ISL content: the Signs of Ireland (SOI) corpus (Leeson & Nolan, 2008; Leeson, et al., 2006) and the patient–receptionist dialogue corpus (Morrissey, et al., 2010).

3.1.1 Patient–receptionist dialogue corpus

The patient–receptionist dialogue corpus is a multimedia parallel corpus specifically developed for the purposes of English to ISL machine translation (MT). Designed to facilitate a very specific subdomain of the healthcare domain, this corpus focuses on appointment scheduling dialogue between the medical secretary and the patient. Due to confidentiality and other ethical issues the original corpus data (audio recordings) were elicited through a role-play exercise involving native English speakers and a GP's receptionist. Consisting of approximately 350 dialogue turns or an average of 3,000 words, the corpus is small in regard to original content. However the corpus also consists of six different parallel modalities. Considering the time and expertise it would take to convert the original data into a new modality, not least with regards to HamNoSys transcription, this makes the corpus all the more relevant.

The six modalities are:

- audio recordings of the original material
- written English transcription
- ISL video recordings
- HamNoSys transcriptions
- SiGML notations
- Bangla text (for additional testing on MT for minority languages)

Conveniently the corpus has already been fully transcribed with HamNoSys and SiGML making it suitable to output using the JASigning platform. Using this corpus would save much time allowing the transcription process to be circumvented entirely. The patient–receptionist dialogue corpus has a very much focused domain. 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. Further reading about the planning and construction of this corpus is available in a publication by Morrissey, et al. (2010).

3.1.2 Signs of Ireland corpus

The SOI corpus is well established and is one of the largest digitally annotated signed language corpora in Europe. Developed as part of the ‘Languages of Ireland’ programme at the School of Linguistic, Speech and Communication Sciences, TCD, it gives a rich selection of utterances with EFEs. The primary purpose of the corpus is to record ISL as it is currently used in Ireland. During the course of this programme the researchers sourced data from 40 different signers. In eliciting the dataset researchers carefully selected signers that demographically represented a snapshot of the Irish Deaf community at the time of elicitation. Criteria used included: age (18 – 65), geographical location (5 locations across Ireland), and fluency in ISL (ISL must be their first language acquired before the age of 6). Care was taken to include a balance of male and female participants and also to avoid including those who had a formal education in sign language linguistics in order to avoid signers who would endeavour to deliver “correct” or “pure” ISL. All participants were asked to choose their own anecdote as well as narrate a set of stories that have been used widely in signed language research. Data for the corpus was collected, primarily, by members of the Deaf community and as a result, the subjects of the corpus were free to relax and sign naturally. This relaxed and natural sign may be the best material to impartially evaluate the comprehension of a sign language avatar. In this instance, the ISL captured is an accurate representation of live signing used by Deaf people in Ireland today as opposed to grammatically correct ISL. Over the years, as with any language, ISL has changed and will continue to do so. There is no real motivation behind this change, it simply occurs as a result of, amongst others, influential popular signs, family signs, the influence of various dialects, the adoption of signs from other languages or new signs

for new technologies. The use of grammatically correct ISL content for this evaluation discounts all of the aforementioned anomalies and therefore does not represent live signing used by Deaf people in Ireland today.

3.1.2.1 SOI corpus annotation

The terms annotation and transcription are often confused. According to Johnston (2008), Linguistic annotations are used to identify units of a language. For example, they may identify some phonological, morphological or syntactic meaning within an utterance. Sign language corpora often use English glosses to identify lexical items. The SOI corpus have done much the same but include 12 more tiers of annotation to include: Mouthing, Dominant Hand, Non-dominant hand, Eyebrows, Eye aperture, Eye-gaze, Head movement, Body movement, Iconic information, Point of view, Translation and a final tier for Notes. Transcription is defined by Johnston (2008), as a dedicated script or graphic representation of an oral/signed language which can be used to accurately record and ultimately reproduce a signed/oral utterance. HamNoSys is an example of transcription for signed language.

As mentioned in the previous section, the SOI corpus is the largest digitally annotated signed language corpus in Europe. The corpus was annotated using the digitally based EUDICO Linguistic Annotator (ELAN) developed at the Max-Planck-Institute in Nijmegen (Hellwig, et al., 2012). Many linguistic annotation tools exist⁵ but ELAN is a tool specifically designed for multi-level annotation of video and/or audio, making it ideally suited for the annotation of signed languages. For that reason, ELAN is used widely across the signed language linguistic research community (Johnston, 2008; Bungeroth, et al., 2008).

3.1.3 Choosing a subset

Using ELAN's search facility, we searched the SOI corpus for utterances which exhibited any of the 7 universal emotions outlined earlier in this document. 63 utterances were discovered. This figure is not the sum total of utterances that contain EFEs within the corpus, but rather, an indication of where to begin looking. A matrix was devised to indicate how frequently each EFE appears in each story narration, to further narrow down the number of possible utterances for the subset. The matrix highlighted 33 different story narrations of which 14 featured only 1 of the EFEs, 10 narrations featured 2 EFEs, 7 featured 3 EFEs and only 2 featured 4 or more EFEs. These were: "Fergus D (Dublin) - Frog Story" and "Lianne (Dublin) - A Scare in Belfast". Interestingly, the "Frog Story" by "Fergus D" is a narrative that all participants were asked to sign, whereas the "Scare in Belfast" by "Lianne" is a unique narrative chosen by the participant herself. A manual investigation showed that all 7 emotions were present in the story narration "A Scare in Belfast". Furthermore, during the investigation, 5 segments of the story were identified as having a high concentration and variety of emotional content and were

⁵ http://annotation.exmaralda.org/index.php/Linguistic_Annotation

therefore the best candidates for the evaluation.

To begin the process of identifying each EFE we added a 14th layer to the corpus such that we could annotate EFEs. After that, each facial configuration relating to one of the 7 universal emotions was annotated. Some of these annotations are suprasegmental as the beginning and end of each utterance was not considered during annotation. Instead, each annotation begins at the start of an EFE and ends when the EFE is completed. In the instance where a facial configuration exists but does not correspond to one of the universal emotions, the annotation designation “undefined” was applied. Figure 3-1 Shows a screen grab of the ELAN interface as it is being used to annotate on the “emotion” tier (Note: only 2 tiers are selected as “visible” in this image).

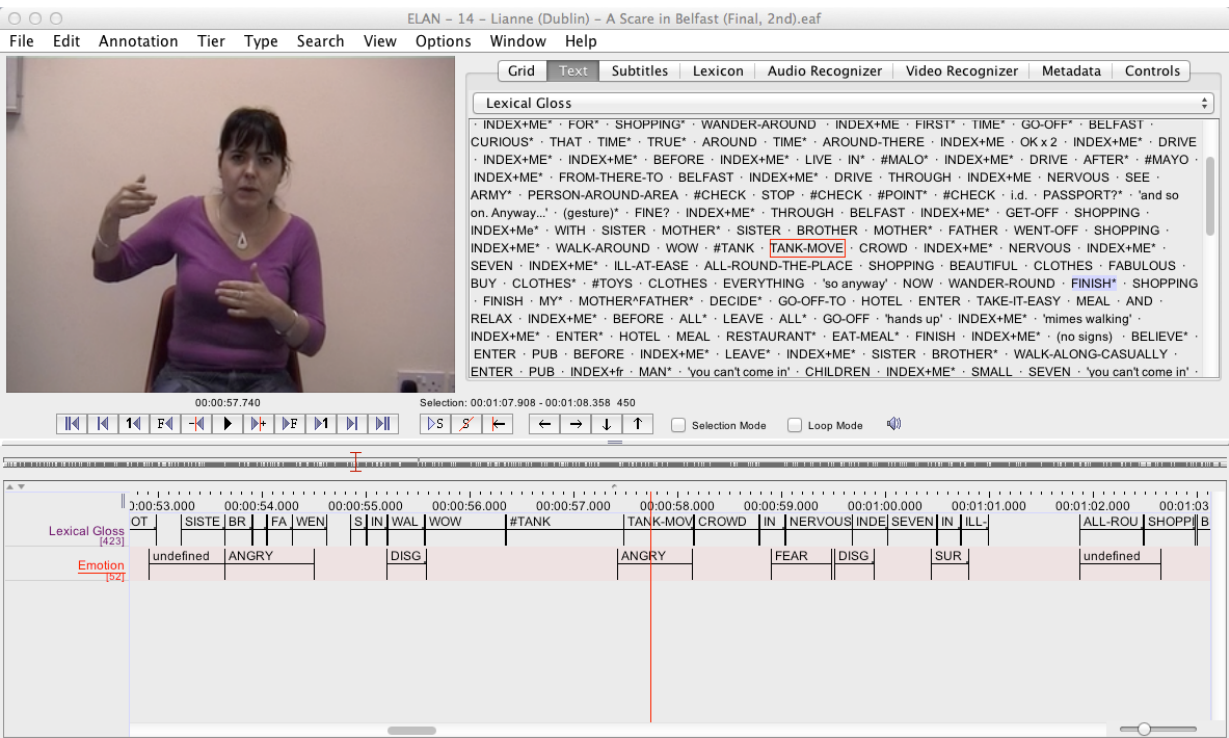


Figure 3-1 ELAN with emotion annotations

Across the 5 story segments a total of 154 utterances are present. 67 of these utterances have been annotated where recognisable EFE is present. The frequency of which each EFE appears is listed in Table 3-1. ‘Happy’ was the most common facial expression, appearing in 18 utterances. On the opposite end of the scale, ‘Sad’ appears only 2 times. Ideally the sample set would carry an equal quantity of each emotion, in this instance, however, we are simply trying to ascertain if any of the 7 universal emotions can improve comprehension. For that task, the above-mentioned set of 67 utterances with EFE was deemed suitable.

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

3.1.4 Corpus transcription

Although ELAN does support the HamNoSys font, it does not fully support HamNoSys input. HamNoSys can be typed using the keyboard, a cumbersome task, or simply pasted into the appropriate space on the timeline. Neither of these options are conducive to an efficient working environment. As a result, the transcription of the corpus could not be done using the ELAN tool. Instead we look to the eSign Editor tool developed as part of the JASigning framework which we explain in more detail later in this document.

3.1.4.1 eSIGN Editor

The eSign Editor is a simple tool designed specifically to allow a user input HamNoSys and output SiGML. The tool allows transcriptions to be stored in a searchable database that comes preloaded with BSL signs. As this is a simple tool, the database takes the form of a text file called “Import1” and can be replaced with more appropriate data if required, as was the case for this project.

Preparing a Sign Language Database

We created an empty Import1 file and used the “import signs” option on the maintenance menu (see Figure 3-2) to access our blank database. We found that the software was unreliable at writing to the database and therefore recommend using the “export signs” button, also on the maintenance menu, to regularly backup any work. The chief purpose of this option is not to backup work but to export your database such that it may be used elsewhere if desired.

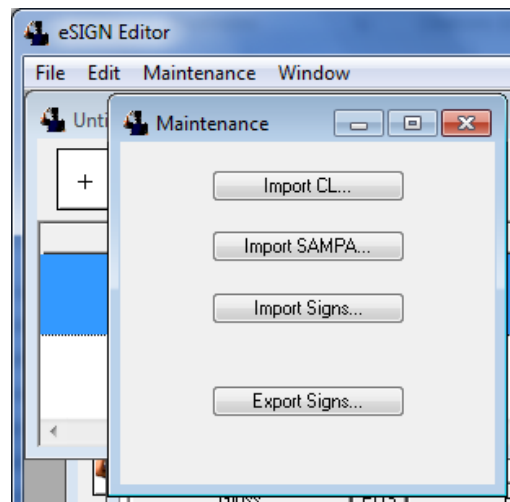


Figure 3-2 eSIGN Editor - maintenance options

The maintenance menu also contains an option to access the database signs directly through the interface (see Figure 3-3). This interface may be utilised to search for existing signs in the database, to edit a sign, to duplicate a sign or to create a new database entry. The database can be used to search for the sign gloss only. Figure 3-3 illustrates a search for the word “Go”. The search will present any database entries that have the letter combination “Go”.

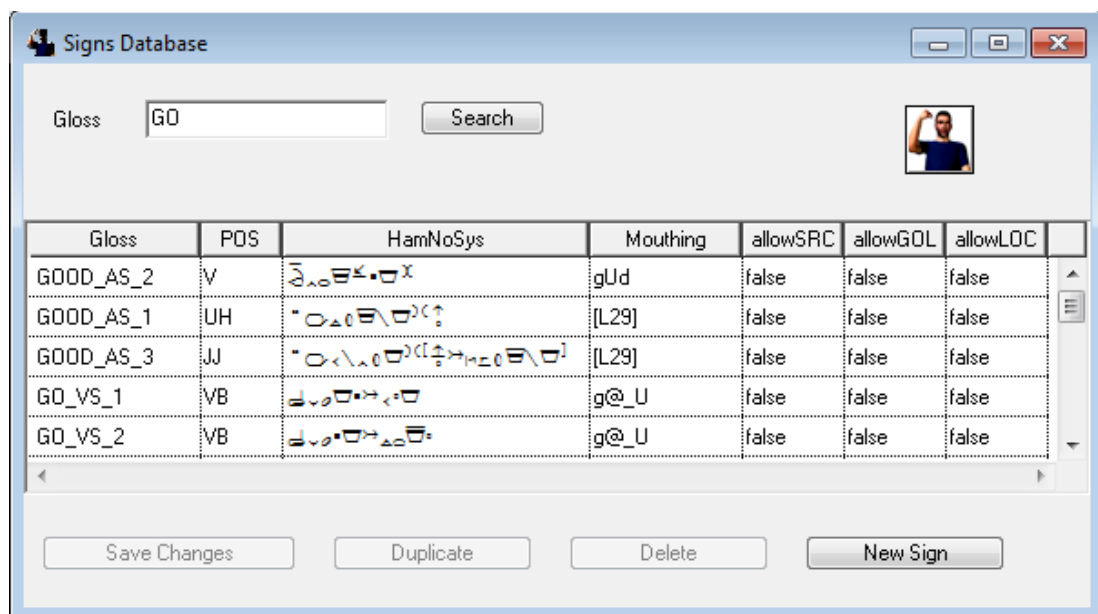


Figure 3-3 eSIGN Editor - signs database

Upon clicking the “new sign” button a new window appears (see Figure 3-4) with textboxes for sign gloss, Parts of Speech (POS), HamNoSys transcription and mouthing. To the right of the HamNoSys textbox is a button with the label “...”. This button is used to open the HamNoSys input panel (Figure 3-5). The same button to the right of the mouthing textbox is used to open the mouthing input panel (Figure 3-6). These are the primary methods of inputting HamNoSys and mouthing for a sign into the database although the interface will accept keyboard input and pasted input also. The new sign may then

be added to the database by clicking the “insert” button or if a number of signs are to be created, the “insert & new” button may be exploited such that the sign is saved and a new “new sign” window is opened.

Figure 3-4 eSIGN Editor - new sign window

HamNoSys Input

The user interface includes a panel of buttons with HamNoSys symbols as labels. Each button represents a HamNoSys symbol or part thereof. It is possible, by editing a simple text document, to change the layout and content of this panel. One may wish to do this to make their environment more comfortable or efficient. In the course of this project we added the ISL alphabet handshapes to the panel. In doing so, the most popular handshapes used in ISL were easily transcribed. This saved quite a bit of time because without the additions to the panel, these handshapes would have to be assembled multiple times and the more complex handshapes often require some experimentation to get them right. Figure 3-5 shows the interface panel used for this project. On the top of the panel, there are a number of tabs which allow the HamNoSys symbols for handshape, orientation, location and movement to be separated and therefore easier to find quickly. Each tab contains a similar set of buttons representative of that particular domain. The tab open in this example is for “Hsh” or “Handshape” and contains all of the HamNoSys symbols relating to handshape. On the panel in Figure 3-5, there are 3 clusters of buttons. The top row of the first cluster contains thumb positions and varying degrees figure of curvatures. The second row contains symbols used to identify each finger and various locations on each finger such as fingertip and knuckle bone. The second cluster of buttons is a little more complex. The first column contains 9 distinct handshapes and the buttons to the right of each handshape contain variations on each. The third cluster has 26 buttons, each represents a handshape used in a letter of the ISL alphabet. The

buttons in this cluster are ordered alphabetically A-Z so that each handshape is easy to find.



Figure 3-5 eSign Editor – HamNoSys input panel

Mouthing

As HamNoSys has limited control of the complex mouth movement required for mouthing or mouth gestures, the eSIGN Editor uses an alternative method. The mouthing panel, which is accessed when creating or editing a sign (Figure 3-4), is used to create 2 types of mouth movement: mouth pictures and mouth gestures.

The simplest of these are mouth gestures. These are, in essence, predefined mouthing gestures that follow their own timeline in that they have a fixed duration. These include such gestures as smacking of the lips, inflating the cheeks and pushing out the Jaw. Mouth gestures are chosen from a submenu on the left of the mouth gesture panel above which is a dropdown menu that can be used to select gestures specific to the Teeth, Jaw, Lips, Cheeks or Tongue. Each mouth gesture is allocated a code and when one is chosen from the submenu the code appears in the textbox at the bottom of the panel. In Figure 3-6 the code is “[L02]”. This code is then used as part of the database entry to identify the mouthing for a particular sign. To the right of the submenu, taking up 2/3 of the panel is a video that provides examples for each of the mouth gestures.

The mouth picture panel has a vastly different interface than that of the mouth gesture panel (Figure 3-6). However, regardless of the interface, the final product from both is a code to identify the mouthing

for a particular sign. The primary difference is how that code is derived. As mentioned in the previous paragraph, mouth gestures are fixed movements identified by a unique code. Mouth pictures, on the other hand, are more flexible in nature and use SAMPA computer readable, pronunciation encoding. Gibbon et al. (1997) defines SAMPA as a language independent system for phonemic transcription and annotation. Each symbol in the SAMPA system represents a phoneme. For example, the word “car” is represented as “kA:” where the uppercase “A” represents a strong “aaa” sound and the colon identifies an elongated sound. Each character in the SAMPA system that represents a mouth shape has a corresponding ‘picture’ or morph target which illustrates that mouth shape. Rendering software is deployed to concatenate these ‘pictures’ seamlessly such that the user sees only a smooth video. Virtually any combination of SAMPA is accepted by the system, therefore, any combination of mouth pictures is possible which makes mouth pictures a more flexible solution than mouth gestures, although these are still required. Another reason why mouth pictures are flexible is that they follow the main timeline in that they start at the beginning of a manual sign and finish at the end of that sign. They are dynamic enough to speed up or slow down as required to fill the signing space.

The mouth picture panel (Figure 3-6) can accept 3 types of input: Standard Orthography, SAMPA or International Phonetic Alphabet (IPA). The user can type in an English word such as “Car” into the standard orthography textbox and then click the button with the down arrow icon “↓” which is located directly under the Standard orthography textbox. This button will convert the standard orthography to SAMPA and IPA and fill the appropriate textboxes. Alternatively, the user, after typing the word “car”, may click the button labelled with a question mark “?”. This button will search a pronunciation database for words with those letters and their corresponding SAMPA encoding. When a SAMPA code is selected, the down arrow has been pressed and the code appears in the bottommost textbox, the user may click “Select” to finalise the selection.

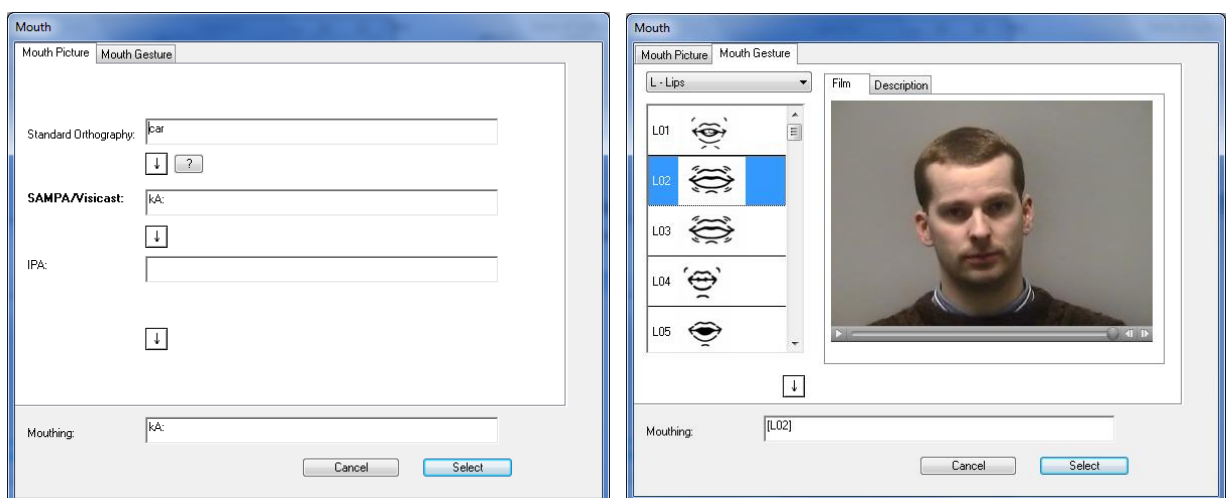


Figure 3-6 eSIGN Editor - mouth picture panel & mouth gesture panel

Building Phrase Sized Chunks

When the sign database has been populated, it is possible to build sign phrases or sentences in order for signs to be exported in larger than ‘utterance sized’ chunks as required. When the eSIGN Editor is initially opened, a new “untitled” document automatically opens in the window. If the user wishes to open a new or existing document, they use the file menu as is the convention in many software packages (see Figure 3-7). To add a new utterance to the new document, the user must click the button with the plus “+” icon. This adds a new field to the document, highlighted with blue in Figure 3-7.

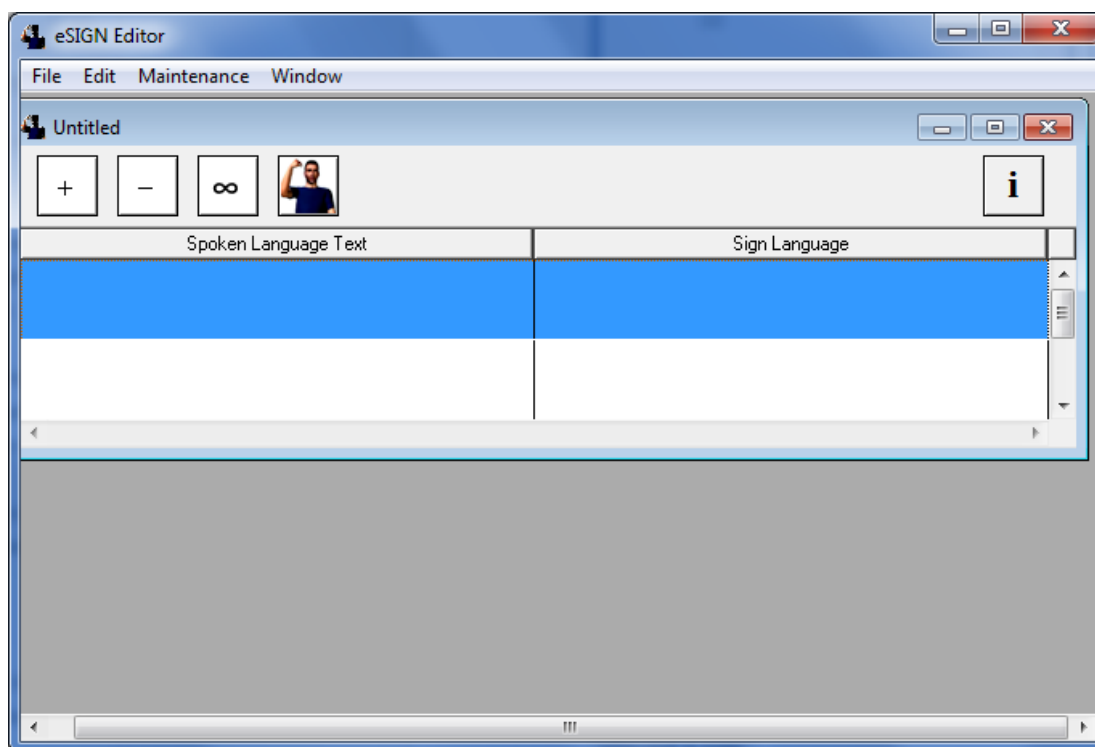


Figure 3-7 eSIGN Editor - untitled document

When the users clicks the column on the right hand side, a new window appears to add signed language transcription (see Figure 3-8). The spoken language text or gloss may be added in the topmost textbox of the new sign utterance window (see Figure 3-8). A number of signed utterances may be added to this window. As before, clicking the “+” button will open a new window which will allow the user to add a transcription (see Figure 3-9). If the database is populated, the user may simply type the gloss name and click search. When the required transcription is selected from the results list, the user can click “select” to confirm the selection. Alternatively, the user may click the “Form” tab and build a new HamNoSys transcription. This is not recommended due to the fact that the transcription is not stored in the database and therefore cannot be re-used when forming additional phrase sized chunks later on.

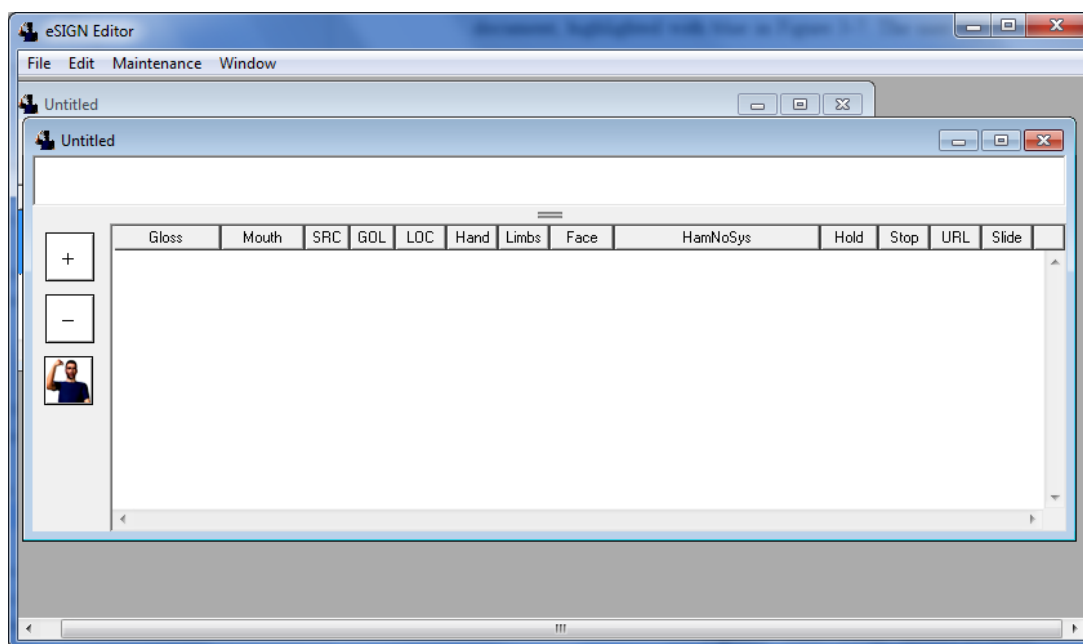


Figure 3-8 eSIGN Editor - new sign utterance window

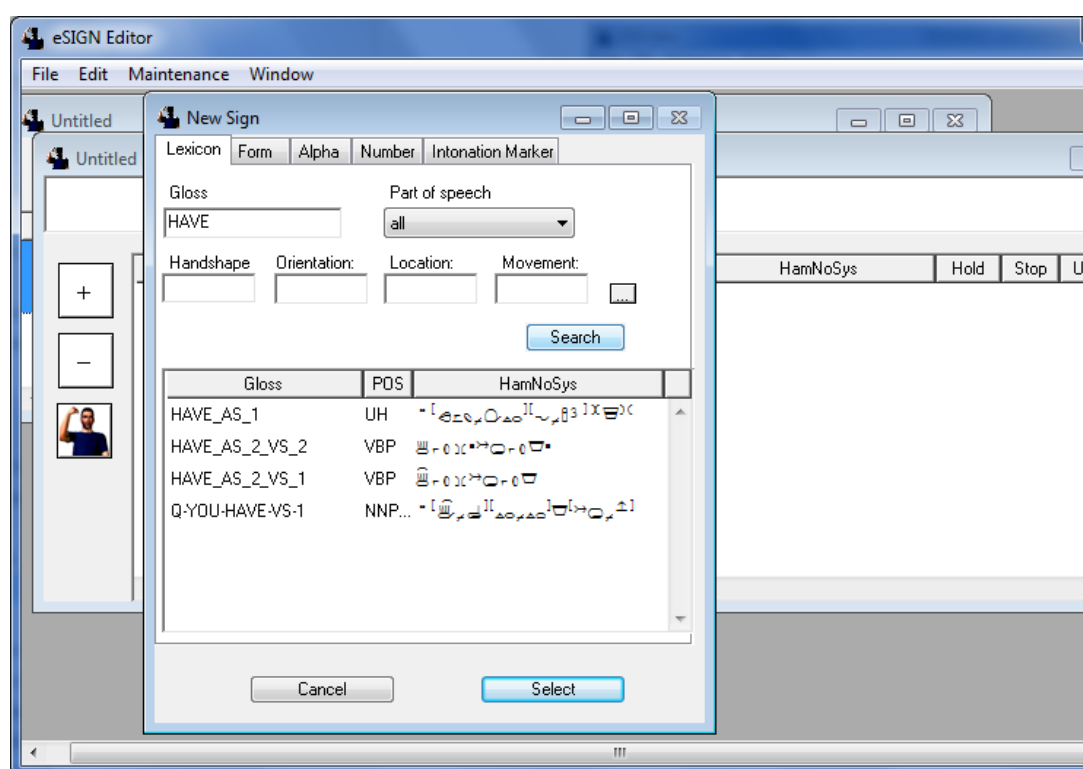


Figure 3-9 eSIGN Editor - new sign transcription window

Export to Avatar

The majority of the windows and panels in the eSIGN Editor have a button with an avatar icon, for example, this button can be seen to the left of Figure 3-9. This button is used to send the signing data to the SiGML service player. The SiGML service player is another piece of software developed as part of

the JASigning framework and must be installed and running for the avatar button to work. This is one of the most advantageous features of the eSIGN Editor as it allows the user to test his/her transcriptions instantaneously. The SiGML service player is one way to output data from the eSIGN Editor, in fact we used screen recording software to capture this output. It is also possible to export a SiGML file using the “Export SiGML” option on the file menu. Figure 3-10 illustrates how the typical work environment might look while transcribing with the eSIGN Editor. It is often necessary to have a number of windows open simultaneously within the eSIGN Editor window as well as a visible SiGML service player window in order to check transcriptions.

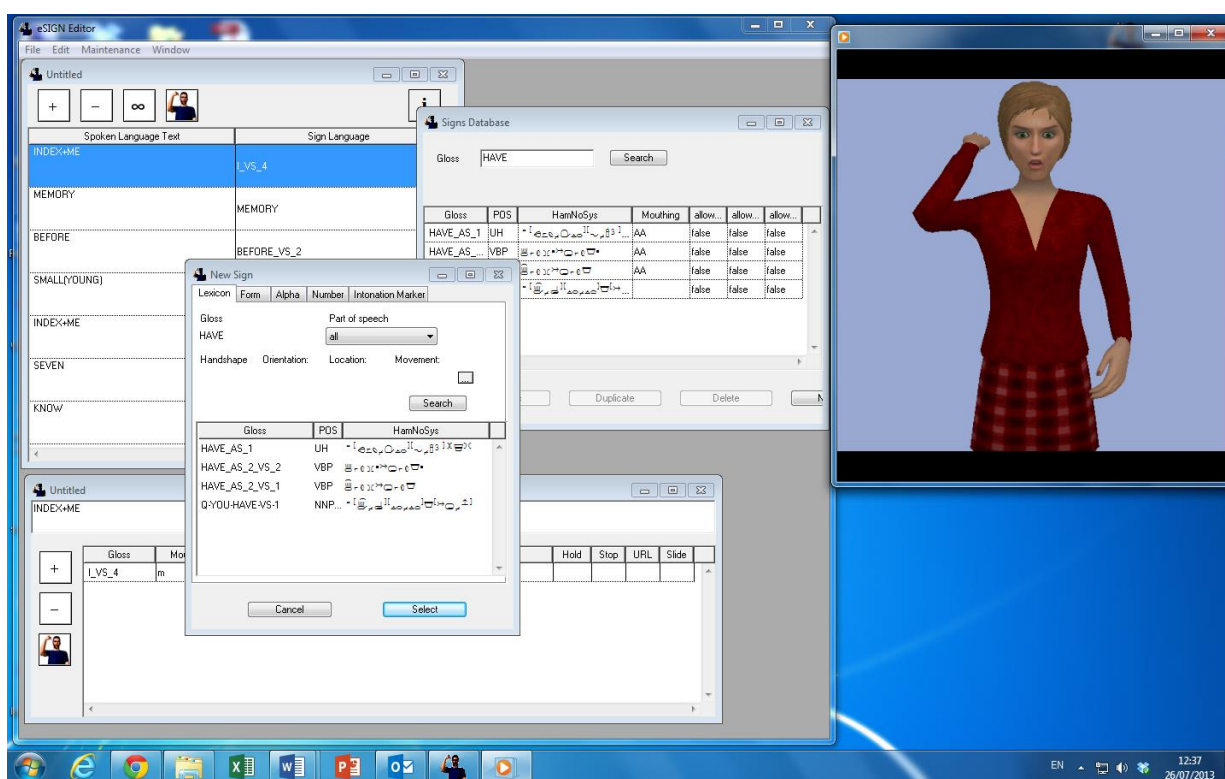


Figure 3-10 eSIGN Editor - working environment

eSIGN Editor and NMFs

Earlier in this section we discussed how mouth pictures are represented by SAMPA codes and mouth gestures are represented by a proprietary code arrangement. The eSIGN Editor uses a similar proprietary coding system to represent various other NMFs such as head, body, shoulder and facial movement. These codes may be manually entered into the editor in a similar manner as HamNoSys and SAMPA code (see Figure 3-11).

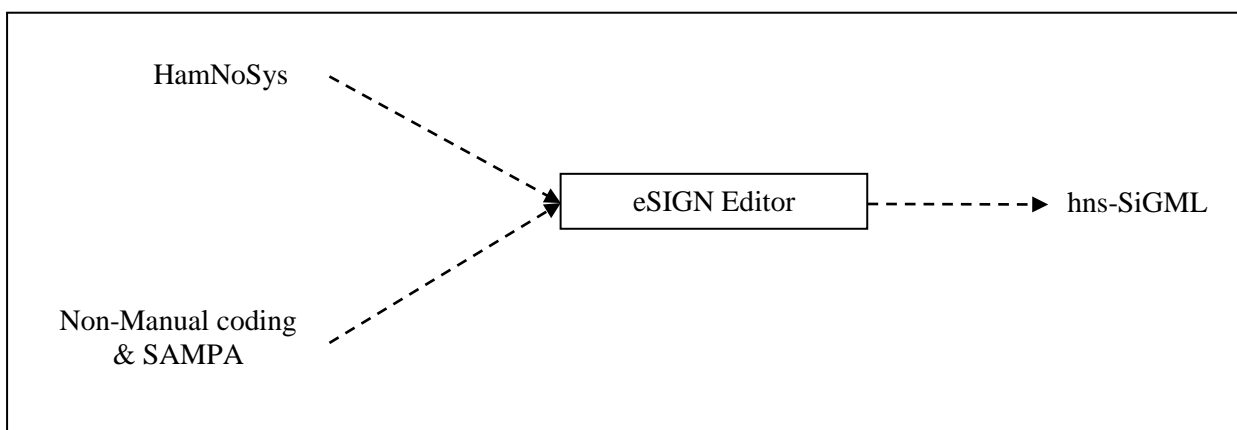


Figure 3-11 eSIGN Editor - data flow

The column headings in the ‘new utterance’ window, illustrated in Figure 3-8, includes a column named ‘Limbs’ and another named ‘Face’. When a new utterance has been created the corresponding cell under these column headings can be clicked to add head, body, shoulder and facial movement. Figure 3-12 illustrates what happens when the cell in the ‘Face’ column is clicked. A new window appears with various drop down menus for eye gaze, eye brows, eye lids and nose. The figure shows that the code value for eye brows “RR” has been selected. A description for each code is provided in the dropdown menu. In this case, “RR” will raise the right eyebrow. The code “AD” in the eye gaze textbox refers to “towards addressee”.

Each of these dropdown menus have a number of options available with positions or movements of the same nature as those just mentioned. When a cell under the column heading ‘Limbs’ is clicked a similar window appears, this time with the option to move the body, shoulders and head. Body movements include tilting and rotating, shoulder movements include hunch, raise & shrug and the head dropdown menu includes such movements as turn, tilt and push.

During the process of building the corpus with this tool it became clear that, although many NMF options are available, there is not enough control over these options. The ability to change NMFs at a micro level is required, however, fine tuning of movements is not currently possible as each NMF code relates to a pre-defined morph or armature movement. Additionally, it is not possible to apply more than one NMF for each menu, even at the SiGML level. For example, it is possible to turn the head to the right or tilt it to the right but it is not possible to turn and tilt the head at the same time. Such functionality would make the eSIGN Editor a far more powerful tool for NMFs. As it is, such limitations greatly affect the quality of the NMFs in a signed utterance.

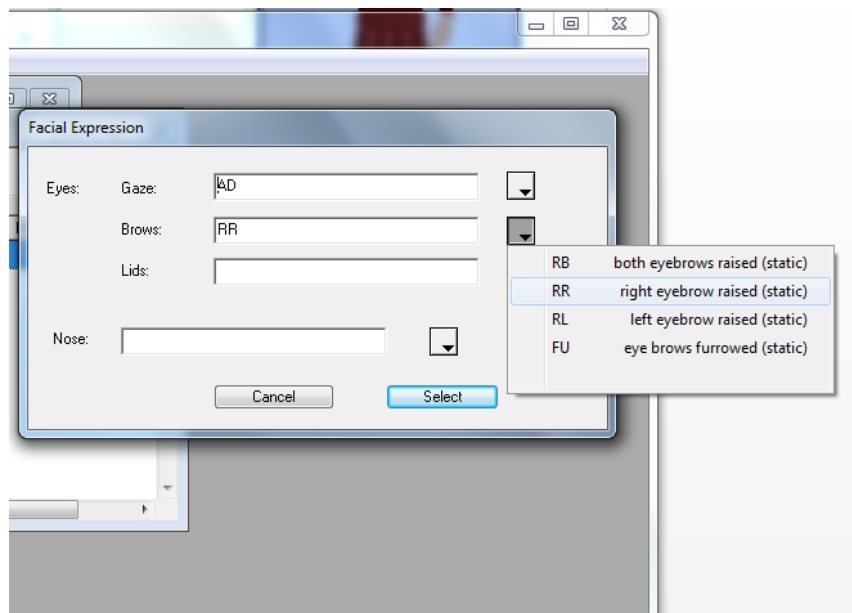


Figure 3-12 eSIGN Editor - Facial Expressions

When the appropriate codes are applied to a sign transcription, such as “RR” and “AD” in Figure 3-12, a SiGML file, such as the one below can be exported. Note the codes for mouth picture, eye gaze and eyebrows and included within the <hamnosys_nonmanual> tags.

```
<sigml>
  <hns_sign gloss="GO">
    <hamnosys_nonmanual>
      <hnm_mouthpicture picture="g@_U"/>
      <hnm_eyegaze tag="AD"/>
      <hnm_eyebrows tag="RR"/>
    </hamnosys_nonmanual>
    <hamnosys_manual>
      .
      .
      .
    </hamnosys_manual>
  </hns_sign>
</sigml>
```

A more comprehensive guide on the use of the eSIGN Editor was developed as part of the eSIGN project (Hanke & Popescu, 2003).

3.1.4.2 HamNoSys Annotation

3.1.4.2.1 HamNoSys General Structure

The learning curve for HamNoSys is a steep one which gradually plateaus once the transcriber becomes accustomed to the symbol set and general structure. HamNoSys, also discussed in section 2.3.1.3 of this document, has an antalphabetic script with an inventory of approximately 200 symbols. The symbols in this script can be categorised primarily using Stokoe's (1960) initial phonemic structure for sign languages, these being, 'Hand Shape', 'Hand Position/Configuration', 'Location', and 'Movement'. A range of other symbols exist to represent such elements as sentence punctuation, repetition, sequence and 'handedness' which identifies whether a sign is one or two handed and if the hands should move in a symmetrical or asymmetrical pattern. Table 3-2 illustrates only some of the symbols for each 'category' or 'classification' while Figure 3-13 demonstrates the structure of a HamNoSys transcription.

Table 3-2 HamNoSys symbol classification (examples)

HamNoSys Symbols	Classification
⋮ ⋯ ⋮̄ •	Handedness
○ ʞ ʞ̂ ʞ̃ ʞ̄ ʞ̅	Hand Shape
^ ʌ ʔ ʔ̂ ʔ̃ ʔ̄ ʔ̅	Extended Finger Direction
○ ʔ ʔ̂ ʔ̃ ʔ̄ ʔ̅	Palm Orientation
○ ~ ∞ ? ʔ̂ ʔ̃ ʔ̄	Location (used for NMFs also)
→ ʔ̂ ʔ̃ ʔ̄ ʔ̅ ʔ̆ ʔ̇ ʔ̈ ʔ̉	Movement (used for NMFs also)

HamNoSys symbols must appear in a strict sequence in order for the transcription to be valid. This sequence is made more stringent when using the eSIGN Editor due to various parsing rules. This section discusses HamNoSys as it is interpreted by the eSIGN Editor software as opposed to more traditional transcriptions which may be interpreted by a human. There are categories of symbols that, due to their optional nature, will not cause an adverse effect to a transcription if they are excluded. On the other hand there are categories that will cause a transcription to fail. Figure 3-13 illustrates the general structure of a HamNoSys transcription. In this example the 'Hand Shape', 'Hand Position', 'Location', and 'Movement' are each identified using coloured labels. All of these categories make up a

transcriptions ‘Initial Configuration’ and with the exception of ‘Location’ all of these categories must be represented in all HamNoSys transcriptions. Symbols that represent symmetry, NMFs and location are not required for every sign. This has been highlighted in the diagram with a broken line which surrounds the applicable symbols. If these symbols are omitted from a transcription the default values for each are assumed.

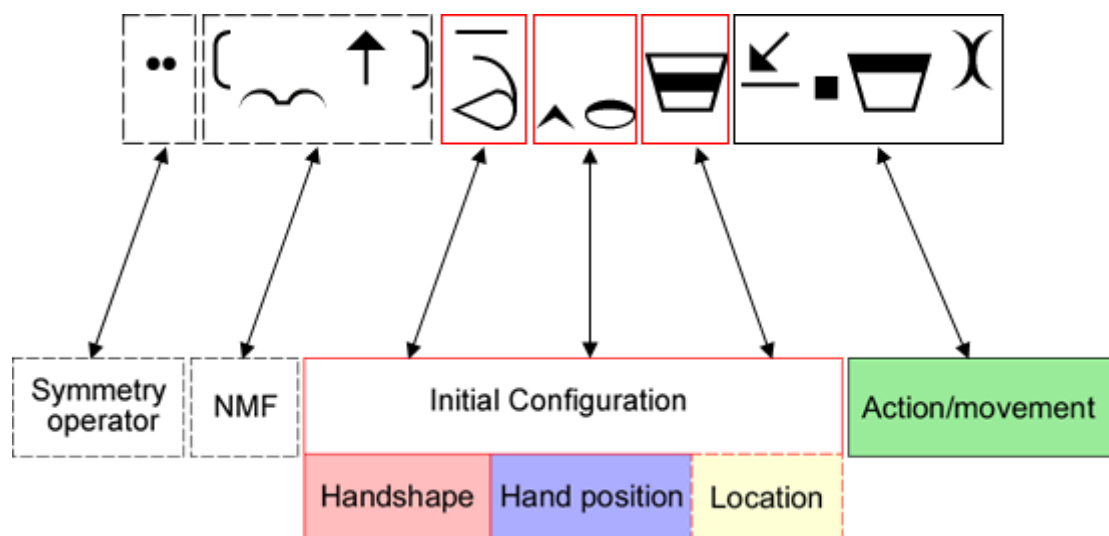


Figure 3-13 HamNoSys - general structure.
The components in the boxes with the broken border are optional

Hand Shape

Like ‘Location’, ‘Hand Position’ and ‘Movement’, ‘Hand Shape’ is a parameter of a signed utterance and no signed utterance is complete without all of these parameters. The SiGML service player will simply not play without all of this information. HamNoSys has 12 basic hand shapes, all of which may be altered through various additional thumb and finger configurations. Figure 3-13 illustrates a transcription using the “ $\overline{\text{D}}$ ” handshape. This handshape is derived from the basic symbol “ D ” with the variant “ $\overline{\quad}$ ”. The basic symbol produces the round "okay" gesture with the thumb and forefinger. The ‘straight line’ variant tells the fingers to be straight and as a result flattens the rounded fingers into a “pinch” gesture (see Figure 3-14).



Figure 3-14 HamNoSys - Hand Shapes

Hand Position

The ‘Hand Position’ category may be further broken down into ‘Extended Finger Direction’ (EFD) and the ‘Palm Orientation’. EFD represents the orientation of the wrist which is identifiable by the direction that the index finger would point if it was extended into a pointing gesture (see Figure 3-15). EFDs symbols resemble small arrow heads pointing in various direction on the X, Y and Z axis’s of the 3D plane. The transcription illustrated in Figure 3-13 uses the EFD “^” which identifies an upward direction. This symbol would also be used to transcribe the EFD in Figure 3-15. Further examples of EFD may be seen in Table 3-2.

The palm orientation, as the name suggests, represents the orientation in which the palm of the hand is facing. Palm orientation is represented by an ellipse shape with only half shaded in black. The shaded half of the ellipse represents the palm and the non-shaded half represents the back of the hand. In Figure 3-13 the palm is facing the sky whereas in Figure 3-15 the palm is facing left so the “0” symbol would be used.

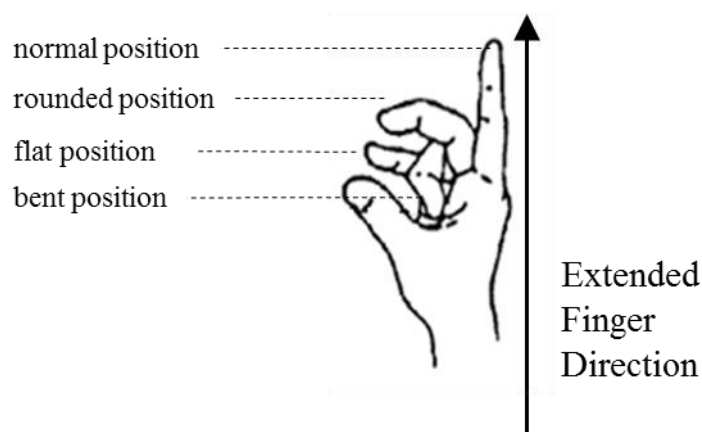























Figure 3-15 HamNoSys - Extended Finger Direction (EFD)

Location

The ‘Location’ parameter may be omitted from any transcription. In doing so, the sign is performed in the default location which is just out from of the signer’s chest. Choosing to omit a location parameter can lead to difficulties in describing more complex signs. When transcribing a location, HamNoSys has quite a comprehensive set of symbols for manual signs. These include but are not limited to the shoulder line “”, breast line “”, belly line “”, head “” and the eyes “”. The full set of the HamNoSys symbols can be seen in Appendix B – HamNoSys Symbol set.

A number of modifications may be used with the location symbols including ‘distance’ symbols which may be used to indicate if a sign is touching the body “”, close to the body “”, far to the left or right of the specified location “” or just slightly to the left or right “”. Again these examples are not exhaustive of the full symbol set and the full set of the HamNoSys symbols as well as a diagram which illustrates ‘distance’ can be seen in Appendix A – The number of body locations identifiable through HamNoSys is greatly increased by use of the ‘between’ symbol “”. By utilising this symbol a transcriber can describe a location as between the shoulder line and the left of the breast line as follows: “  ”. The between symbol may also be used with handshape and hand position.

Movement

Movement may take the form of an action symbol, an interpolation of a new position or a combination of both as is the case in Figure 3-13. In this diagram, the symbol “” represents a movement from the right towards the left of the body. This is an action symbol. The symbol combination “  ” refers to a position touching the left side of the shoulder line. In this example the movement and end position are both explicitly transcribed. It is also possible to use the action symbol in isolation where it is not required to specify the final position of the hand. Equally, it is possible to specify the final position only and the software will interpolate the animation frames between the start and end position. Such transcriptions require the use of the ‘replace’ symbol i.e. “   ”. The ‘replace’ symbol tends to apply a faster duration to a sign than the action symbols whereas the action symbols often allow only a small degree of accuracy when positioning. A combination of both will often overcome these restrictions.

NMF

HamNoSys does not explicitly support NMFs in that it lacks the necessary detail to describe small movements, particularly small facial movements. There are no symbols for opening and closing and not

all parts of the face have corresponding HamNoSys symbols. That said, it is possible to transcribe some NMFs using the existing symbol inventory. An acceptable method of doing this is to replace the default articulator (the hand) with a non-manual articulator such as the head or shoulders and then apply the appropriate action. For example, a shrug of the right shoulder would look like this: $\left[\square \cdot \uparrow \right]$ and a nodding of the head would look like this: $\left[\bigcirc \rightleftharpoons \updownarrow \right]$. These partial transcriptions can then be added to the beginning of a full transcription as seen in Figure 3-13.

Although the transcription structure described in Figure 3-13 is valid, the JASigning platform does not currently support the non-manual component of this transcription. Instead it uses a mixture of the SAMPA encoding and proprietary encodings discussed in section 3.1.4.1.

Handedness

A sign may be one handed or two handed. During a two handed sign the right hand is identified as the dominant hand by default, however, it is possible to change this by using the non-dominant symbol “ \bigcirc ”. A symmetry symbol is used to identify a two handed sign. The symmetry symbol “ \cdot ” identifies a two handed sign where the non-dominant hand copies the dominant hand. This is useful if you want both hands pointing to the right. The symmetry symbol “ \cdot ” identifies a two handed sign where the non-dominant hand performs a mirror image of the parameters defined for the dominant hand. This is useful if you want both hands to point in opposite direction with no additional transcription. It is possible to apply different movements to each hand or to apply a stationary gesture to one hand while moving the other. Typically if there is a symmetry operator or non-manual component it will be before the handshape as seen in Figure 3-13. The details on how this and other HamNoSys features work in practice are available in a HamNoSys user manual edited by Smith (2013).

3.2 The JASigning platform

When we discuss the JASigning platform in this document, we refer to all of the software components therein including the eSIGN Editor and the ARP Toolkit. However, often the JASigning ‘system’ refers to the real-time aspects of the platform only. JASigning is, in fact, a synthetic virtual human signing system designed by the Virtual Humans group at the University of East Anglia (UEA). The software architecture supersedes the earlier SiGMLSigning system developed during the ViSiCAST and eSIGN projects. At the very heart of JASigning is the avatar independent, AnimGen engine, which is used to generate the signing data for the rendering software in real-time. The eSIGN Editor may be considered a component of this platform because it outputs the hns-SiGML that becomes the initial input in Figure 3-16, however the content generated in the eSIGN Editor, although it may be transmitted ‘instantaneously’, is produced offline by a transcriber and therefore is not part of the real-time dataflow

of the JASigning system. A large portion of the previous chapter is dedicated to discussing the eSIGN Editor. Figure 3-16 shows a high level data flow of the JASigning system, including all of the components that operate in real-time.

Another component of the JASigning framework, not explicitly illustrated in Figure 3-16 is the ARP toolkit which generates the files: ASD.xml, nonmanuals.xml, config.xml and AvatarDef.arp. Like the eSIGN Editor, this tool is used in an offline capacity and therefore is discussed in section 3.2.2.

3.2.1 JASigning real-time avatar generation

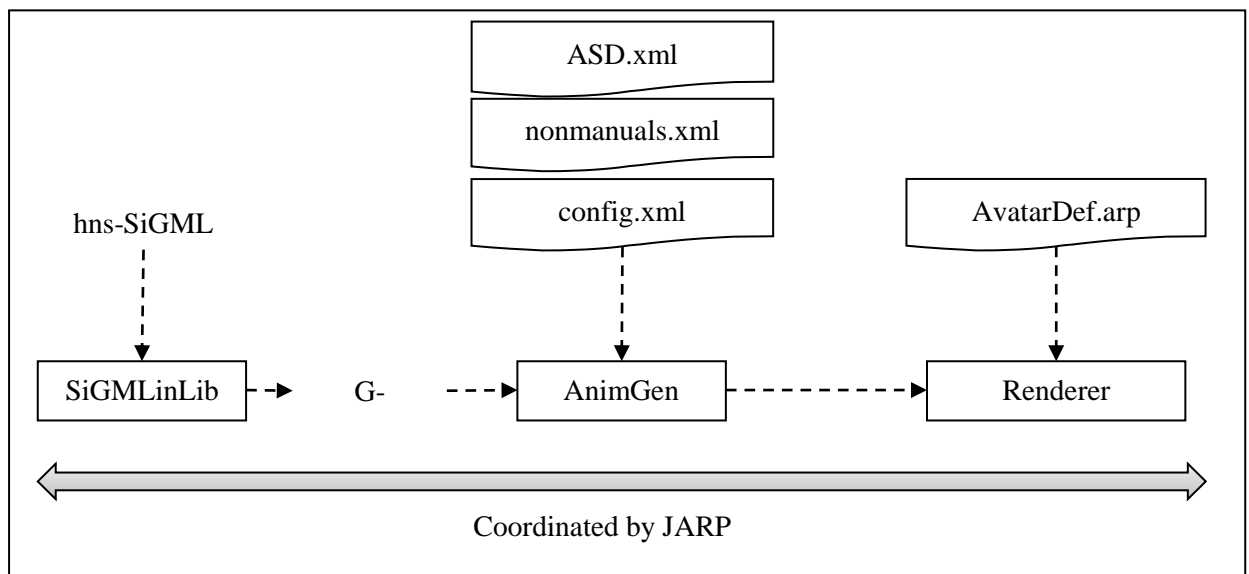


Figure 3-16 UEAs JASigning data flow

As mentioned in the previous section, Figure 3-16 illustrates a data flow as it would occur in the real-time generation of a signing avatar. The data flow consists of 3 components which are coordinated by JARP control software. These components are: SiGMLinLib, AnimGen and the Rendering software which are described in the subsequent sections.

3.2.1.1 SiGMLinLib

SiGMLinLib acts as a pre-processor for AnimGen. It is effectively a text processing tool used to tokenise and re-order hns-SiGML into a more explicit form that it can be understood by AnimGen. The re-ordered form is called gestural SiGML or G-SiGML (Elliott, et al., 2004). Figure 3-17 & Figure 3-18 are examples of the DGS (German Sign Language) sign for “internet” represented in hns-SiGML and G-SiGML respectively, as they appear in (Glauert & Elliott, 2011). It is clear from these examples that where hns-SiGML simply lists an element for each HamNoSys symbol, G-SiGML further categorises these into elements such as <handconfig> and <handconstellation> and further again with attributes such as ‘extfidir’, a shorthand for ‘extended finger direction’ and ‘palmor’, shorthand for ‘palm orientation’. It should also be clear that hns-SiGML is a more ‘human friendly’ form of markup and

would be preferable when writing the markup manually.

```
<?xml version="1.0" encoding="utf-8"?>
<sigml>
  <hns_sign gloss="INTERNET">
    <hamnosys_nonmanual>
      <hnm_mouthpicture
        picture="nEt"/>
      <hnm_head tag="NB"/>
      <hnm_eyebrows tag="FU"/>
    </hamnosys_nonmanual>
    <hamnosys_manual>
      <hamsymmlr/>
      <hampinchall/>
      <hamparbegin/>
      <hamextfingeru/>
      <hampalmd/>
      <hambetween/>
      <hampalmdl/>
      <hamplus/>
      ...
      <hamparend/>
      <hamparbegin/>
      ...
      <hamparend/>
      <hamtouch/>
      <hamshouldertop/>
      <hamclose/>
      <hamparbegin/>
      ...
      <hamplus/>
      <hamparbegin/>
      <hammovei/>
      <hamsmallmod/>
      <hamarcl/>
      <hamreplace/>
      <hamfinger2345/>
      <hamthumboutmod/>
      <hamextfingeru/>
      <hambetween/>
      <hamextfingerur/>
      <hampalml/>
      <hamparend/>
      <hamparend/>
    </hamnosys_manual>
  </hns_sign>
</sigml>
```

Figure 3-17 hns-SiGML for DGS (German Sign Language) sign "Internet"

```

<sigml>
  <hamgestural_sign gloss="INTERNET">
    <sign_nonmanual>
      <head_tier>
        <head_movement movement="NB"/>
      </head_tier>
      <facialexpr_tier>
        <eye_brows movement="FU"/>
      </facialexpr_tier>
      <mouthing_tier>
        <mouth_picture picture="nEt"/>
      </mouthing_tier>
    </sign_nonmanual>
    <sign_manual both_hands="true" lr_symm="true">
      <handconfig handshape="pinchall"/>
      <split_handconfig>
        <handconfig extfidir="u" palmor="d" second_palmor="dl"/>
        <handconfig extfidir="u" palmor="u" second_palmor="ur"/>
      </split_handconfig>
      <handconstellation contact="touch">
        <location_hand digits="3" location="tip"/>
        <location_hand digits="3" location="tip"/>
        <location_bodyarm contact="close" location="shouldertop"/>
      </handconstellation>
      <split_motion>
        <par_motion>
          <directedmotion curve="r" direction="o" size="small"/>
          <tgt_motion>
            <changeposture/>
            <handconfig handshape="finger2345" thumbpos="out"/>
            <handconfig extfidir="u" palmor="l" second_extfidir="ul"/>
          </tgt_motion>
        </par_motion>
        <par_motion>
          <directedmotion curve="l" direction="i" size="small"/>
          <tgt_motion>
            <changeposture/>
            <handconfig handshape="finger2345" thumbpos="out"/>
            <handconfig extfidir="u" palmor="l" second_extfidir="ur"/>
          </tgt_motion>
        </par_motion>
      </split_motion>
    </sign_manual>
  </hamgestural_sign>
</sigml>

```

Figure 3-18 G-SiGML for DGS (German Sign Language) sign "Internet":

3.2.1.2 AnimGen

AnimGen is at the core of the JASigning system. It is AnimGen that takes the avatar data and the signing data together to make a frame by frame stream of animation data for the renderer. At the point of entering AnimGen, the signing data contained within the G-SiGML file is ‘avatar independent’, which means the data can be applied to any avatar conforming to the strict AnimGen specifications. In contrast the avatar data contains geometrical data that could only be used by one avatar making it ‘avatar specific’. Much of this data is located inside the avatar definition file which feeds directly into the rendering software and not AnimGen. However, the 3 XML files that do feed into AnimGen are also avatar specific which means they rely on the accuracy of the data contained in the definition file. The files: ASD.xml, nonmanuals.xml and config.xml all provide additional information for the avatar and enrich the data contained in the avatar definition file.

Asd.xml

The Avatar Standard Description (ASD) file, asd.xml, defines the hierarchy of the bone names in the armature structure as well as a default/reference position. This data may then be used by AnimGen to calculate rotation values for joints. The ASD file also defines in the region of 380 reference points that may be used to determine locations in signing space (Jennings, et al., 2010).

Config.xml

The AnimGen configuration data file format, config.xml, defines values for timings, signing space, constraints, trajectories, hand shapes, constants, repetitions, and rest poses. For example each hand shape has a series of values assigned to each finger which identifies to what degree they are bent. These values are applied to the handshapes identified in the SiGML.

```
<handshapes>
  <fist specialbends="0000" ordinarybends="4440" extendedfingers="" class="fist" />
  <flat specialbends="0000" ordinarybends="0000" extendedfingers="2345" class="flat" />
  <finger2 specialbends="0000" ordinarybends="4440" extendedfingers="2" class="fist" />
  <finger23 specialbends="0000" ordinarybends="4440" extendedfingers="23" class="fist" />
  .
  .
  .
</handshapes>
```

Figure 3-19 config file partial contents

Nonmanuals.xml

The NonManuals file format (nonmanuals.xml) maps nonmanual morph targets to those in the main avatar definition file. It is important that the same names are used to identify morphs in both files. The nonmanuals.xml file also defines durations and trajectories (timings) for these nonmanuals.

Within the nonmanuals.xml file each NMFs duration is represented using an ‘Attack, Hold, Release’ envelope commonly seen in sound synthesis. On the attack time is the time it takes for the morph to reach its peak. When it has reached 100% it enters the ‘Sustain state’ for the specified duration. Once the specified duration is complete the morph enters the release state where the intensity of the morph decreases again. The various states of intensity and duration are given via the nonmanuals.xml file. These envelopes are commonly overlapped to show that one NMF may begin while another is sustained or released (See Figure 3-20)

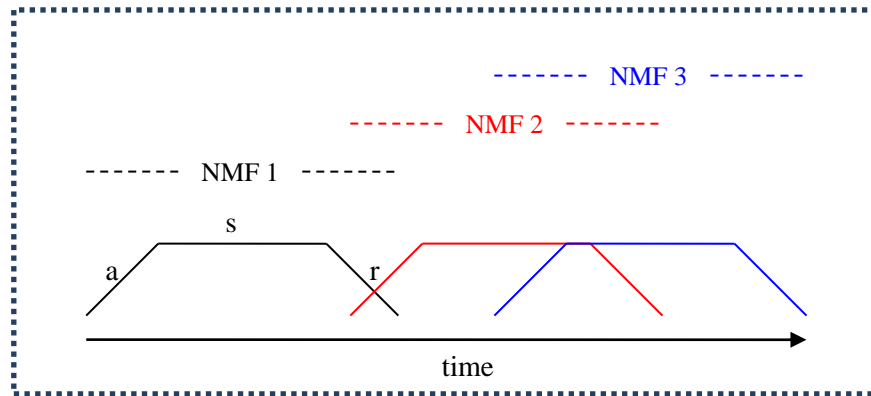


Figure 3-20 Attack, Hold, Release envelope using NMFs

```
<extra_movement sigmlName="X65">
  <morph name="Sad" amount="1.0" timing="x m t m s l x"/>
</extra_movement>

<extra_movement sigmlName="X66">
  <morph name="Happy" amount="1.0" timing="x m t m s l x"/>
</extra_movement>

<extra_movement sigmlName="X67">
  <morph name="Fear" amount="1.0" timing="x m t m s l x"/>
</extra_movement>
```

Figure 3-21 Extract from a nonmanuals.xml file with emotion face configuration

Figure 3-21, an extract from nonmanuals.xml, illustrates how the ‘Attack, Sustain, Release’ method is implemented in practice. Each of the ‘extra movement’ elements in the diagram represent an EFE, which are identifiable by their ‘morph name’ attribute: X65 = Sad, X66 = Happy and X67 = Fear. Each of the elements has an attribute called ‘amount’ which is set at 1.0 or full. This attribute refers to intensity and has a scale of 0 – 1. Morphs can be turned on or off but they may also have a value of somewhere in between. Each element also has an attribute for timing. Timing has 7 encoded values. Each of these values influences a different aspect of a morphs timing and each of these aspects may contain a character/code which refers to the values such as ‘fast’ or ‘slow’ (see Table 3-3).

Table 3-3 nonmanuals.xml timing codes

code	meaning
‘x’	anchored
‘e’	elastic
‘f’	fast
‘m’	medium speed
‘s’	slow
‘-‘	zero
‘t’	targeted
‘l’	lax

The example in Figure 3-21 shows that the 3 elements have the default encoded values for time which are ‘x m t m s l x’. The first value in this string, in this case ‘x’, indicates whether the morph is anchored to the start of the interval during which it is played. The second value, ‘m’, specifies the attack time. The third, ‘t’, represents the attack trajectory which is the manner in which the morph approaches the full amount. The fourth character, ‘m’, identifies the sustain time. The fifth, ‘s’, specifies the release time and finally the sixth, ‘l’, and seventh, ‘x’, characters identify the release trajectory and whether the morph is anchored to the end of the interval during which it is played. These values are simply default values and can be changed. For example, the sustain time can be shortened by increasing the time from its current value ‘m’ (medium speed) to ‘f’ (fast) (see Table 3-3).

The token that represents trajectory must be specified as targeted ‘t’ or lax ‘l’. Ordinarily the attack trajectory is targeted and the release trajectory is lax such that a smooth transition with the next sign can be achieved. The first and seventh encoded values will always be either ‘x’ or ‘e’, however, it is possible to omit these tokens entirely leaving only 5 tokens. In this instance, the anchor values would revert to their default values (Jennings, et al., 2010).

3.2.1.3 Renderer

The renderer has 2 inputs, the first being a stream of frame by frame, avatar specific animation data from AnimGen in the CAS XML format. This animation data includes numeric values that represent movement such as joint angles, position vectors and morph weight targets. The second input is the avatar definition file, avatarDef.arp, which is created offline and contains the data required to render a stationary avatar.

AvatarDef.arp

Much like the asd.xml, config.xml and nonmanuals.xml files, the main avatar definition file, avatardef.arp, is created using the ARP toolkit and is avatar specific. It contains basic information for a given avatar i.e. the data required for an avatar to perform typical animations such as a basic walk cycle or head nod. Unlike the aforementioned XML files, the definition file is a single binary file, which contains data such as: Vertex List, Texture Map, Armature, Mesh-to-Skeleton Attachment Data and Morph Targets (Jennings, et al., 2010). Unless further developing an avatar, there is no requirement to alter the definition file. However, to create new morph targets it is essential to generate a new avatardef.arp file using the ARP toolkit tool.

The JARP rendering software, utilizes the JOGL library to access the required OpenGL APIs such that the information in the avatardef.arp, and CAS files may be rendered into a 3D animated avatar at a rate of 25 frames per second.

3.2.1.4 Use-jarp

A developer may utilise the APIs provided by JARP and SiGMLinLib to develop their own interface in which they may display and control their avatar. However the use-jarp module includes various applets and applications which were designed to demonstrate some of the core features of the platform. During the course of this project, the SiGML service player was utilised to play avatar animations from SiGML input. This input was generated by the eSIGN Editor for the baseline data, however after the SiGML was directly edited to include new elements for emotion we used the SiGML service client application.

3.2.2 ARP toolkit

The ARP (Avatar Research Platform) is a set of proprietary tools developed by UEA in order to develop avatars that are compatible with the AnimGen engine. Many commercial tools already exist which were specifically designed to create 3D environments and 3D characters yet none of these tools produce the data required specifically for Deaf signing. The ARP toolkit generates this data as well as providing simple interfaces and automation for some of the arduous tasks such as armature rigging. This allows users with less technical knowledge to focus on the development of signing avatar and not some complex 3D design task.

The default avatars that are currently used in the JASigning system were developed, in the initial stages with the use of commercial tools. The data required for signing was added later in the process using the ARP toolkit. The commercial 3D character design tool ‘poser 5’ was employed to create the initial avatar meshes and morph targets. The 3D modelling software ‘3D Studio Max’ was then used to append these meshes with texture maps (designed with ‘Photoshop’) and armature structures. At this point the avatar is ready for standard animation but lacks the detail required for signing animation. The unique features of the ARP toolkit are exploited to create feature points, apply 2D to 3D mapping, create morph targets and create dynamic textures. Only at this point is the avatar ready for signing animation. The toolkit is ready to export the 4 above-mentioned files, these being ASD.xml, nonmanuals.xml, config.xml and avatarDef.arp (see Figure 3-22). These files are then used by AnimGen and the rendering software to animate the avatar in real-time.

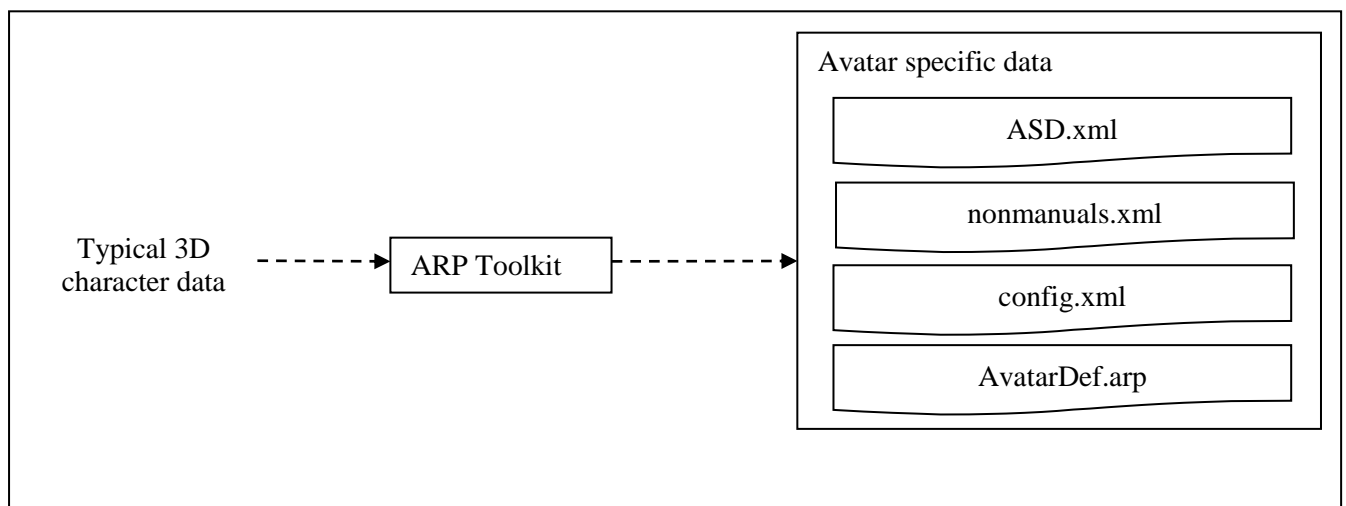


Figure 3-22 ARP Toolkit data flow

3.2.2.1 New Morph set

Only one tool from the ARP toolkit was used during the course of this project. The Morph toolkit allows the user to manipulate or ‘morph’ the vertices in the polygon mesh such that a movement can be created. A starting point, an end point, trajectory and duration exist for each of the vertices contained within any given morph. This data is referred to as a ‘morph target’. The Morph toolkit provides a graphical user interface to select individual (or groups of) vertices to which a ‘weight’ may be added. Varying degrees of movement may be added to these vertices on the X, Y or Z axis. The heavier the ‘weight’ value applied to a vertex, the greater the level of displacement that is applied. This is particularly useful when applying movement to the upper lip for example. The lip itself needs to move when the mouth is opened but so too does the facial area that surrounds the upper lip, but to a lesser degree. Figure 3-23 illustrates how these weightings appear in the GUI. The heavier weightings are the red vertices and the lightest weightings are yellow in colour with various shades of orange in-between.

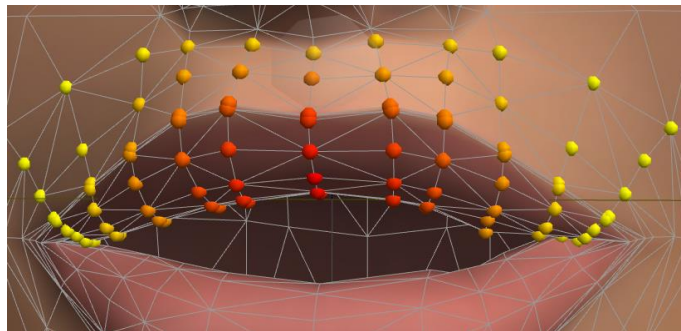


Figure 3-23 ARP Toolkit - vertex weights

When the desired end position and duration have been selected for all the required vertices a ‘morph primitive’ is formed. A morph primitive is a morph applied only to part of the mesh such as the upper lip, jaw, or the outer corner of one eye. They are usually very specific but it is possible to include a much larger portion of the mesh, though this is not common practice. Morph primitives can be grouped into ‘morph groups’ such that a collection of primitives can combine to create one movement. This is the process that was used to create morphs for each of the 7 universal emotions. Figure 3-24 illustrates the morph groups for the universal emotions.



Figure 3-24 ARP Toolkit - morph groups

When the avatar is ready for use, the user can export the required data (see Figure 3-22) from the ARP Toolkit which will contain a new avatarDef.arp file containing all of the new morph data. The new morphs must be manually added to the newly generated nonmanuals.xml file as illustrated in Figure 3-21. At this point the new morphs can be applied to the SiGML notation.

3.2.3 Mark-up for new morph set

Currently the existence of a nonmanual element within a SiGML notation simply informs the AnimGen software to turn a morph on. Nonmanual elements refer to the predefined elements in the nonmanual.xml file as described in section 3.2.1.2 which in turn refer to the morphs in avatarDef.arp.

As was the case with this project, newly created facial morphs may be added to a SiGML notation by referring to the “sigmlname” as it appears in nonmanuals.xml (see Figure 3-25). The new SiGML element must be placed in between the ‘hamnosys_nonmanuals’ tags with the same element name used in the nonmanuals.xml file using the “sigmlname” as a value of the “tag” attribute (see Figure 3-26). Elements may be added to signs notations as required.

```

<extra_movement sigmlName="X65">
  <morph name="Sad" amount="1.0" timing="x m t m s l x"/>
</extra_movement>

```

Figure 3-25 nonmanuals.xml "sad" element

```

<hns_sign gloss="BEGIN">
  <hamnosys_nonmanual>
    <hnm_mouthpicture picture="BE:"/>
    <hnm_head tag="TL"/>
    <hnm_extramovement tag="X65"/>
  </hamnosys_nonmanual>
  <hamnosys_manual>
    <hamsymmlr/>
    <hamfinger2345/>
    <hamextfingerol/>
    <hampalml/>
    <hamparbegin/>
    <hamwristback/>
    <hamplus/>
    <hampinkyside/>
    <hamparend/>
    <hamtouch/>
    <hamparbegin/>
    <hammoveil/>
    <hamsmallmod/>
    <hamplus/>
    <hamnomotion/>
    <hamparend/>
  </hamnosys_manual>

```

Figure 3-26 SiGML notation for the ISL sign "Begin"- with emotion tag

3.3 Evaluation

Automated computer software evaluation is not a new area, various established methods exist for specific domains. This is true of machine translation for example, however the multi-modal and visual nature of the sign language results in a significantly higher level of difficulty when automating the evaluation process than would be the case with a corresponding spoken language text evaluation. The problems involved in automatically evaluating sign language machine translation are comprehensively explored by Morrissey (2011) and Morrissey & Way (2013). Morrissey ultimately used a manual evaluation method as did Huenerfauth, et al. (2008) and Kipp, et al., (2011). Signed language avatars pose a unique question with regards to evaluation: What should be evaluated? The signed meaning, the signing accuracy and fluidity or the avatars aesthetics. It may be possible for an avatar to perform the appropriate movements, as instructed for a given sign but the movements may be incorrect in the signing space used or from a temporal point of view. Additionally it may also be the case that the instructions sent to the avatar rendering software may have been incorrect. With these difficulties in mind, how can we best evaluate the avatar? We may consider evaluating each utterance at the HamNoSys or SiGML level but that would only serve to test the quality of the transcriptions. Moreover, there is not enough HamNoSys or SiGML data available in ISL to produce a corpus from which a gold standard can be extracted. Without a gold standard there is nothing to test our transcriptions against. The only option left available is to evaluate the avatars performance i.e. the final output of the synthesis system. For this, the author knows of no automatic evaluation methodology. Computer vision technology would seem the most appropriate technology for this task. There is a possibility that this technology could be utilised to test positioning for manual and non-manual signs but the complexities of the movement required and the nature of ISL linguistics would escape the current abilities of this technology.

Our conclusion then, is that an automatic approach is not feasible for sign language avatar performance evaluation, leaving a manual evaluation as the only viable option.

3.3.1 Manual evaluation methodology

A manual evaluation was undertaken with 15 users of signed language 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 3.2, 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 (the 4 avatars are derived from an original 2: Luna and Anna (see Figure 3-27)): (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 3-4). 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.



Figure 3-27 Avatars: Luna & Anna

Table 3-4 Avatars used

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

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 interlocutors questions. A second camera, ‘camera 2’, filmed the ISL interpreter as he/she interpreted the conversation between participant and the interlocutor. The room layout is illustrated in Figure 3-28.

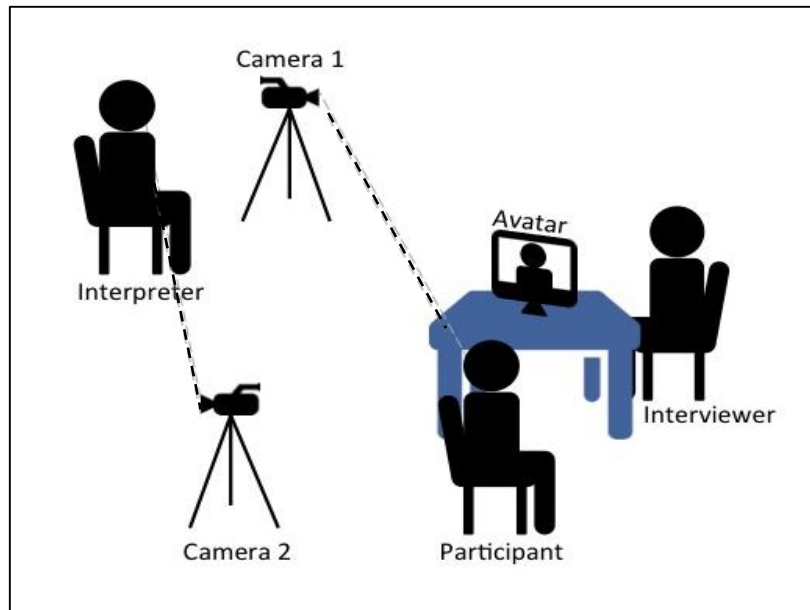


Figure 3-28 Evaluation room layout

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 a 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. A template of the questions from each phase of the interview is available in Appendix E – interview questions.

4 Findings

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.

4.1 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 signed 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 video⁶, 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.

⁶ A pre-recorded video of a 'real person' using sign language to provide an alternative to text on the web.

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 4-1). 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 4-2). 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".

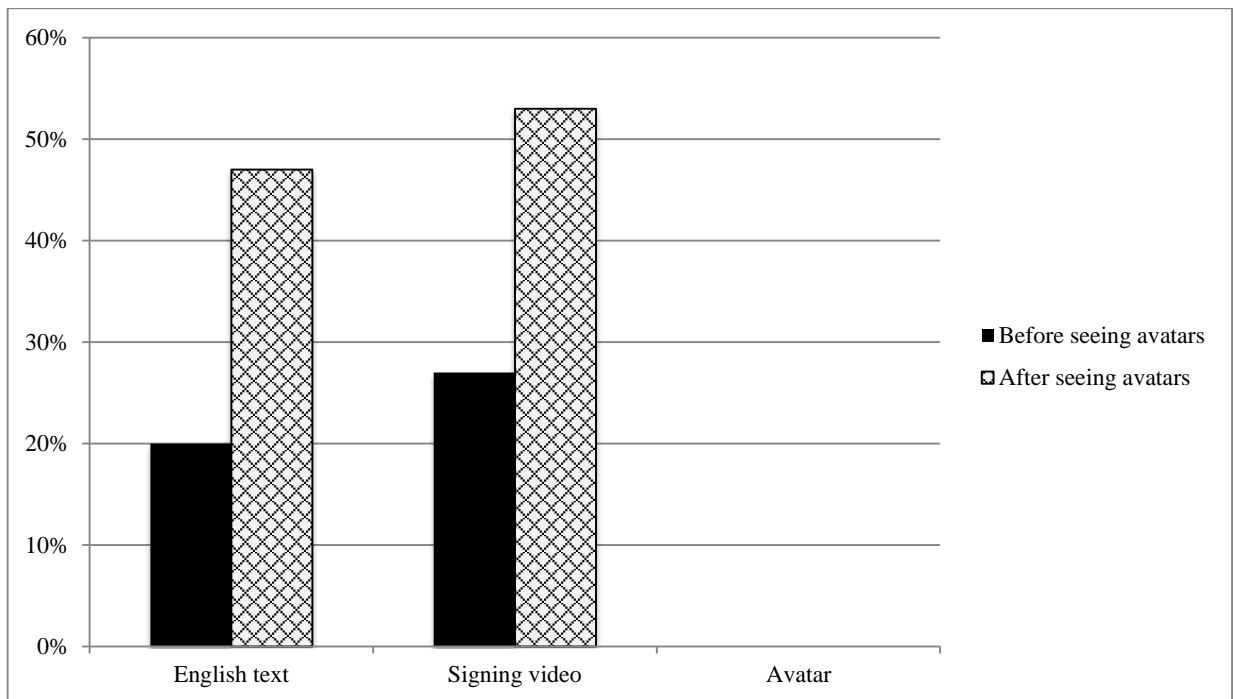


Figure 4-1 Participants' preference (text, video, and avatar)

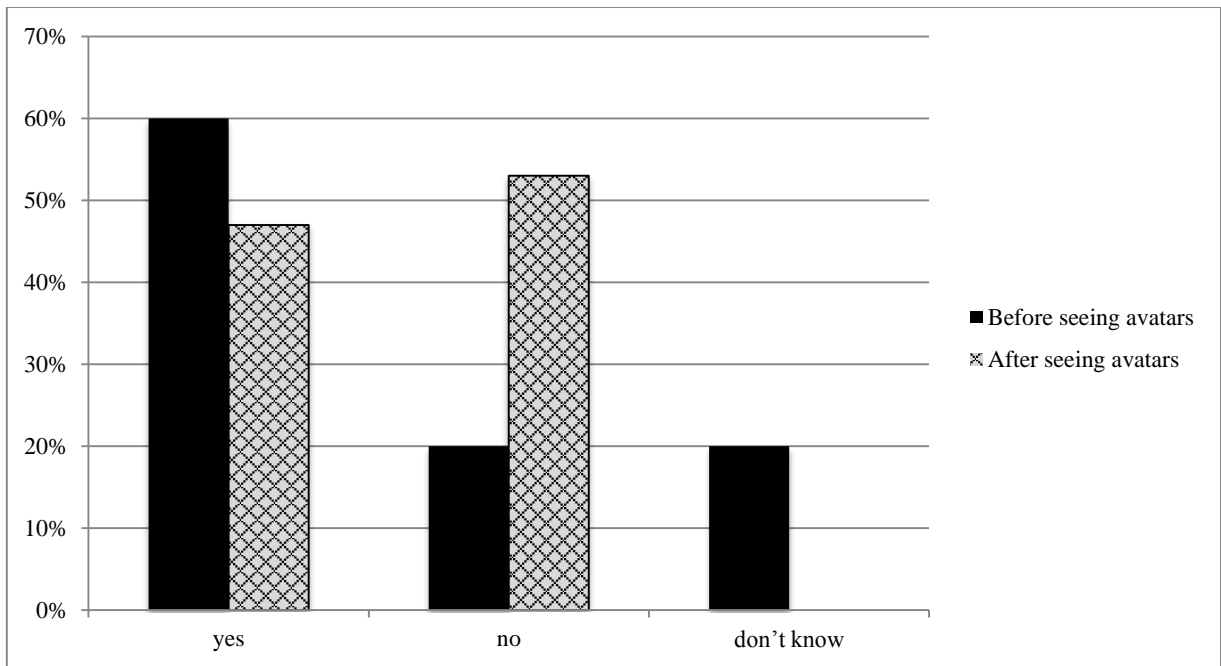


Figure 4-2 Would you use a signing avatar video?

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 4-3). 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 Figure 4-2. 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 4-4).

Table 4-1 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

4.2 Comprehension results

Results indicate that participants, when directly asked, underrated their own comprehension on each avatar video shown. Figure 4-5 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 4-5). 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 4-6 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.

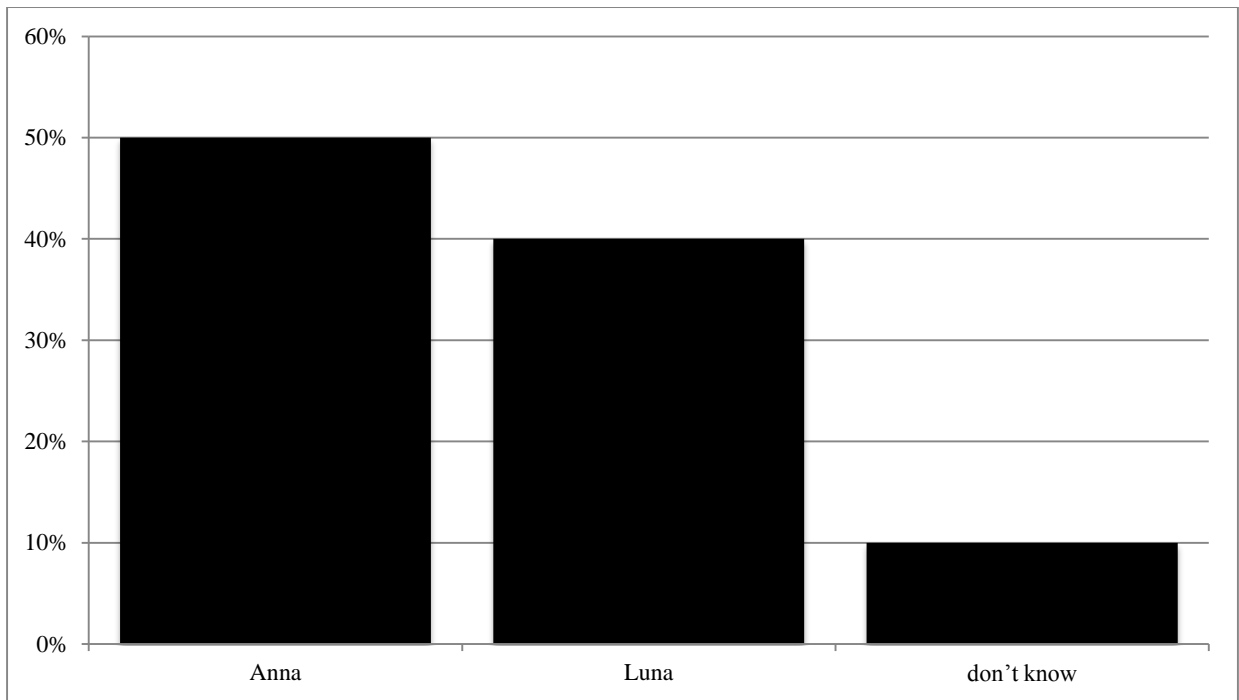


Figure 4-3 Avatar preference

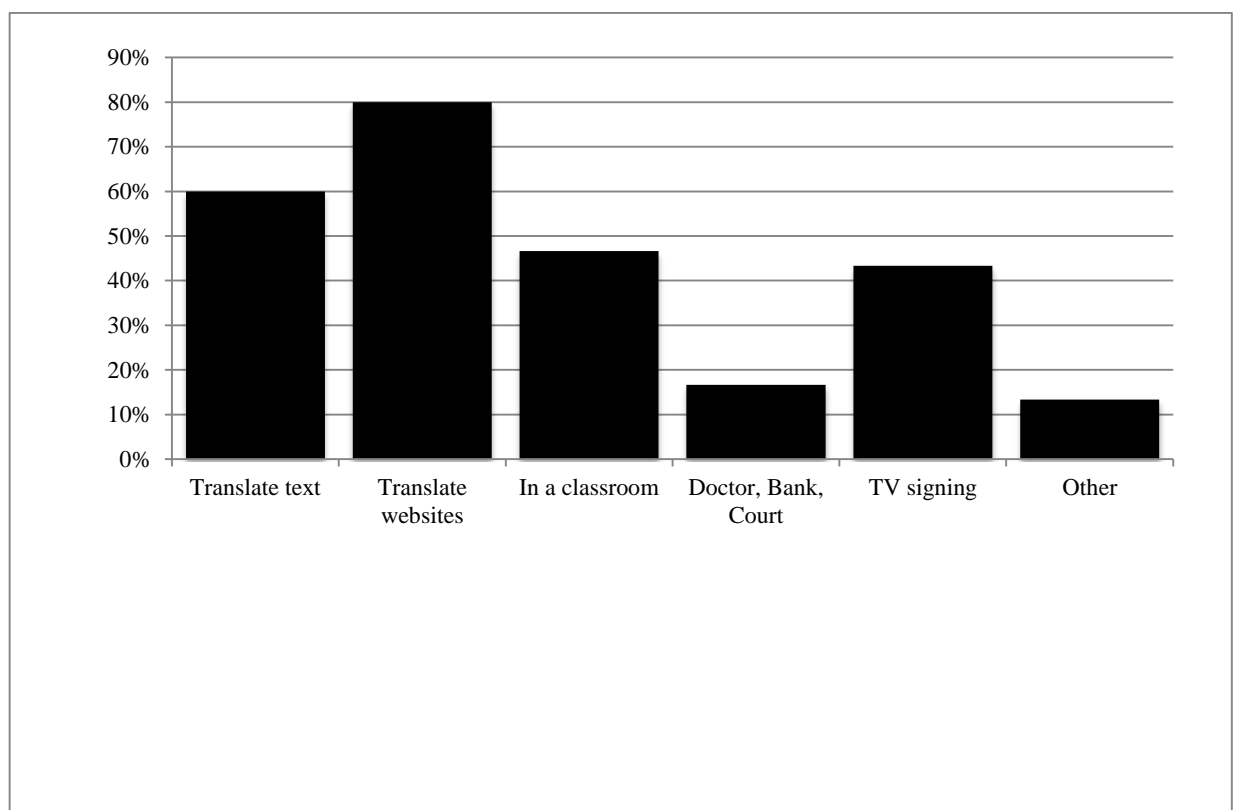


Figure 4-4 Possible use for avatar technology

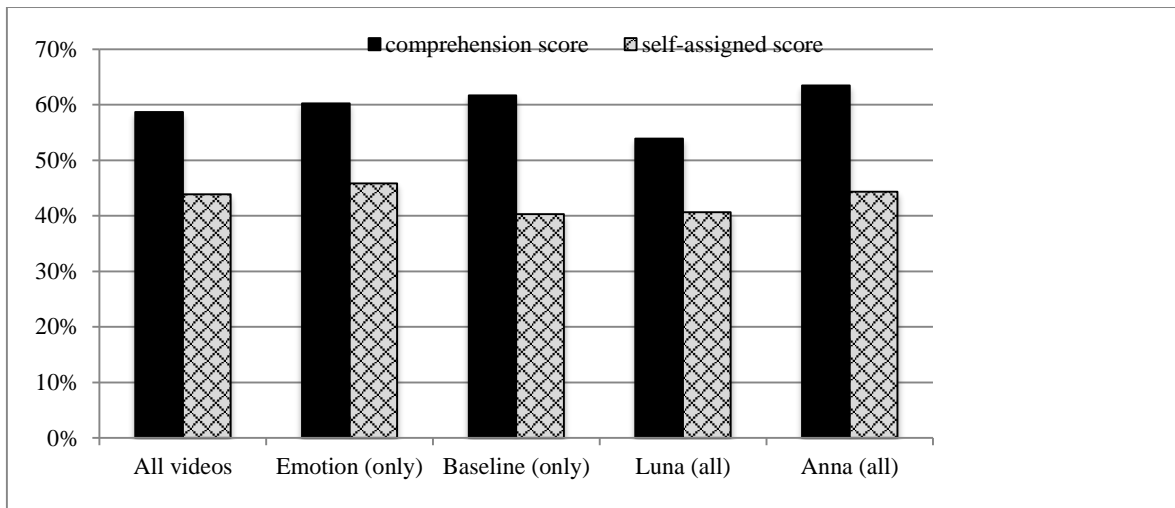


Figure 4-5 Comprehension score vs. self-assigned score

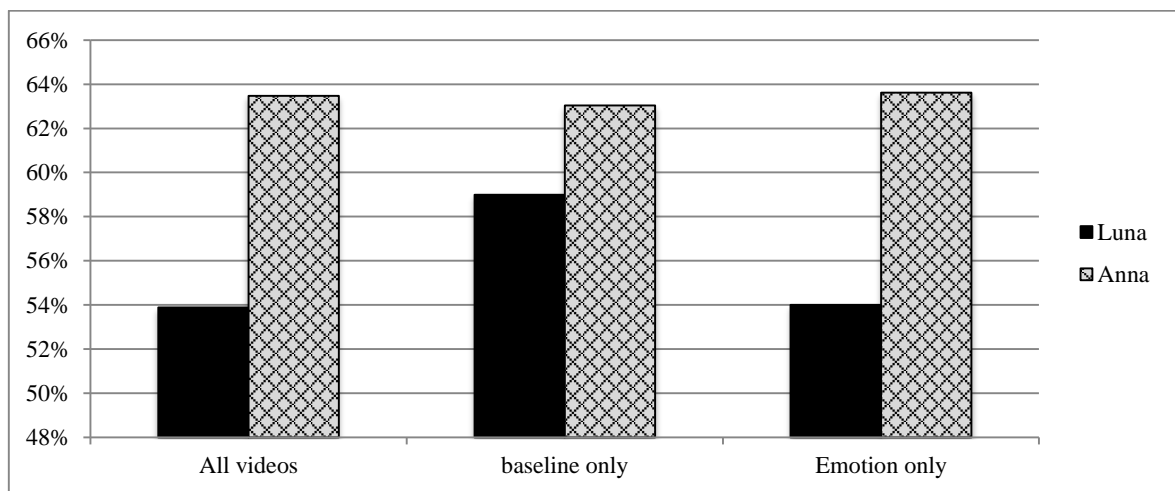


Figure 4-6 Average comprehension score by Avatar

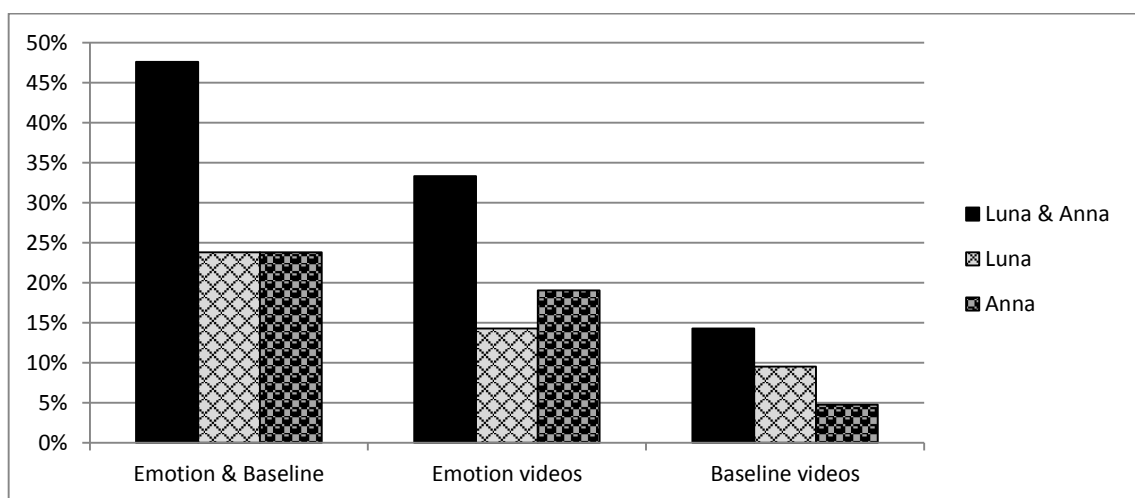


Figure 4-7 Did you see emotion?

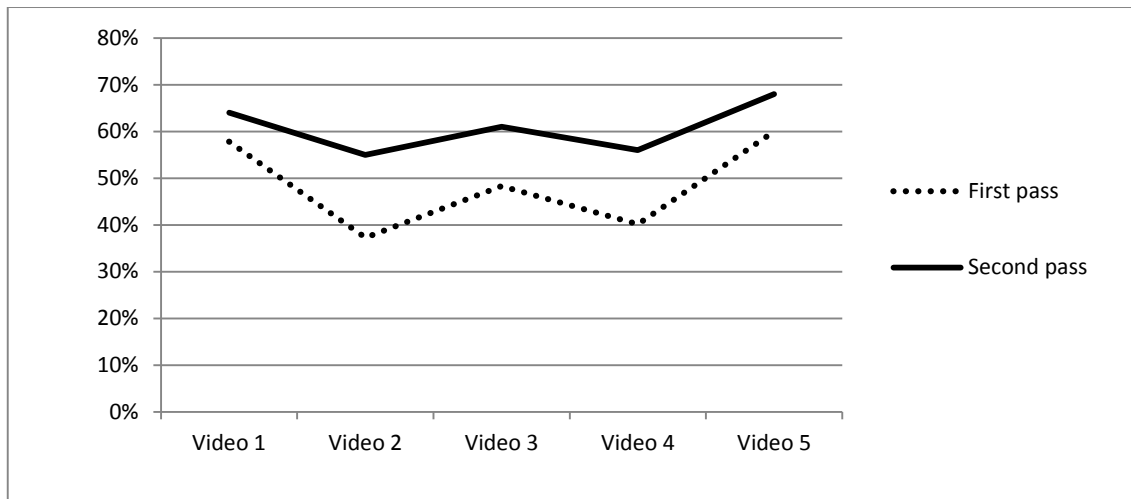


Figure 4-8 Average comprehension score - 1st and 2nd viewing

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.

After watching each avatar video, participants were asked if they had seen emotion. Figure 4-7 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 i.e. 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 4-8 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 videos 2 and 5.

5 Discussion

5.1 Implementation Discussion

The primary focus of the work was to ascertain whether or not the addition of emotional facial configuration increased the acceptability of signing avatars and understandability of a signed utterance. Ultimately, this data was discovered through a manual evaluation but a substantial portion of the work for this thesis was carried out prior to the evaluation in the form of researching relevant technologies, corpus transcription, understanding and implementing the UEA system, augmenting the UEA system with new morphs for the 7 universal emotions and synthesising the avatar output. In this, the author found that the learning curve for HamNoSys is steep at first and much time must be devoted to learning the rules and symbols before effective transcription can be attempted. The steep learning curve is compounded by the minimal documentation available for those who wish to learn HamNoSys. This was the motivation behind the creation of a new unpublished HamNoSys handbook (Smith, 2013). This handbook takes some of the existing minimalistic literature for HamNoSys and puts it all into one document. The document should be a useful recourse of those learning HamNoSys and also for those who just require a reference guide.

When utilising the UEA JASigning platform we again found that there was a steep learning curve. However, the problem around the documentation for this system was not that there was too little but that there was too much. There is a plethora of published papers and project documents available, most of which originate from the ViSiCAST and eSign projects. As expected, the technologies used in this system have moved on over the years and as a result, it can often be difficult to separate the old data from the new. A useful wiki (Elliott & Glauert, 2011) hosted by the virtual humans group in UEA has proven to be a huge asset in this regard. To manufacture materials that were usable by the JASigning platform the author was required to become familiar with proprietary software such as the eSign Editor and the ARP toolkit as well as acquiring an advanced understanding of SiGML and how JASigning processes it. Much like HamNoSys, this proved to be a time consuming process due to the volume of information involved. After using the platform and its features, our conclusion is that the modular nature of the system is its biggest advantage. Many of the software interfaces were developed as “in-house” tools, exemplar tools and often require some knowledge of programming and the system’s architecture to navigate. Although, in most cases, these interfaces can be circumvented by accessing the APIs or

markup directly. The level of access to this platform coupled with the modular structure means that a developer can create a new module to replace any existing module along the frameworks architecture. For example a new avatar may be developed or a new version of SiGMLInLib may be developed to take SWML input and translate it to G-SiGML.

5.2 Problems with evaluation

As with any project, a number of problems presented themselves during the course of this evaluation. Foremost of these was the challenge involved in attracting participants. Through the support of friends and past colleagues in the Irish Deaf Society we were successful in recruiting 15 suitable participants for the evaluation. In a community of 5000 sign language users that equates to a ratio of 3:1000 or 0.3% of the community. This level of participation is comparable to similar studies carried out by Huenerfauth, et al. (2008) and Kipp, et al. (2011) who achieved participation levels of 0.001% and 0.7% respectively. In the case of Kipp, et al. (2011), 330 of the 338 participants contributed to the evaluation through an online survey only.

It is often difficult to attract participants to get involved in a study such as this. Factors such as non-remuneration, demographic requirements and being outside of the closed Deaf community will no doubt have impacted on the number of people willing to participate. This is not a unique phenomenon however. Studies involving the Deaf community directly often struggle to attract participants (Leeson, et al., 2006; Huenerfauth, et al., 2008; Kipp, et al., 2011). In fact the problem is more general than that, research carried out with the assistance of any minority group will experience similar difficulties. With regards to the Irish Deaf community, there have been signs during the recruitment process for this study that the community is suffering from a type of ‘research participation fatigue’. Anecdotal evidence, from speaking with members of the Deaf community, would suggest that some Deaf people have been asked to partake in numerous studies, which have ultimately delivered no tangible benefit for them directly. Now these people feel that they have ‘done their part’ and it is up to others in their community to get involved. Symptoms of this fatigue include reluctance or downright refusal to take part in these studies.

Ideally, feedback from participants would have been multimodal, including formats such as survey, focus groups and interviews. However, a number of barriers presented themselves and it seemed that one-to-one interview sessions were the most appropriate approach. In retrospect, we still consider that interview was the correct choice. The 1-to-1 interview sessions were hugely beneficial in that they yielded much qualitative as well as quantitative data. The questions, although they did follow a strict script, could be reworded, avoided or asked in multiple ways if required. A huge amount of qualitative feedback was gained during the question-and-answer sessions simply because the participants ‘wanted’ to tell us their thoughts on various features of the avatar as they were performed. Gathering this

information would not have been possible through a survey. The result of the 1-to-1 scenario was that the participants were not influenced by others in the room which would be the case in a focus group scenario. Traditionally surveys require a high volume of participants. It was reasonable to accept that this high volume could not be achieved for two reasons. The literacy levels amongst Deaf adults, as highlighted in chapter 1, may result in some of the data being untrustworthy and therefore impracticable to use. The second reason is perhaps more compelling: Currently the Irish Deaf Society (IDS), who have unparalleled access to the Irish Deaf community, are conducting a survey across the whole Deaf community and have experienced much difficulty in attracting participants. The motivation behind that survey is to ascertain an accurate number of sign language users in Ireland. The IDS, as representatives of the Irish Deaf community, hope that the results of the survey will attract more resources to sign language and lead ultimately to ISL being recognised under Irish constitutional law (IDS, 2013). Clearly, the Irish Deaf community should be keener to take part in that particular survey than any other and yet a number of months into the survey, attracting participants has been difficult. For these reasons and because we didn't want to distract from the ongoing IDS survey, a survey component for our data collection was dismissed.

Choosing to collect data through interview helped avoid much of the bias that would be associated with a focus group scenario. None the less, a number of unavoidable biases have been introduced to the evaluation. All participants were aged between 19 and 60. Although the research set out to address this age group primarily, it does not accurately represent the whole Irish Deaf community. Similarly there were no participants to represent the Munster province. This level of sampling bias was unavoidable with the relatively small number of participants and the location in which the evaluation was carried out. Another bias was introduced by the use of an interpreter. This also was unavoidable due to the fact that the interlocutor was not a fluent ISL user. Finally, in an effort to minimise the Hawthorne effect⁷, the evaluation was conducted in the Irish Deaf village, in surroundings that were familiar to the participants and every effort was made to make each participant feel as comfortable as possible.

A control group such as signing video was considered to set a benchmark for comprehension exercises. However this was quickly dismissed as signing video would most certainly introduce learning by repetition which would undoubtedly bias the comprehension scores for each avatar. As it was, a Latin square model was employed to eliminate such bias amongst the avatar videos.

⁷ The Hawthorne effect or the observer effect: subjects improve or modify an aspect of their behaviour in response to the fact that they know that they are being studied.

6 Conclusions

Throughout this thesis we have discussed a number of relevant topics in an effort to communicate the work we have done and the motivations behind it. A discussion around the above topics was considered significant to addressing our primary aims which were:

- To ascertain whether or not the addition of emotional facial configuration increased the acceptability of signing avatars and understandability of a signed utterance.
- To establish a clear picture of the attitude towards signing avatar technology amongst the Irish Deaf community.
- To record, at what level, a fluent sign language ‘speaker’ can comprehend an avatars performance.

Ultimately these aims were set out to support our initial hypothesis: *The comprehensibility and acceptability of synthesised sign language avatars can be improved upon with the provision of a more human-like avatar than that which is provided by the current state of the art.* In this case ‘human-like’ applies primarily to the avatars appearance, the level of emotion simulated and to a lesser extent the fluidity of movement. From this hypothesis we derived 3 distinct research questions.

RQ1. What is the current uptake of signing avatar technology within the Irish Deaf community? And what are the factors that have influenced this?

RQ2. Does the presence of simulated emotional data improve avatar comprehension or acceptance? How may we best evaluate: to what extent EFEs are significant in signing avatars?

RQ3. To what extent is a signing avatar output understood? And how may we best evaluate the comprehension level of a signing avatar performance?

A set of 20 synthesised sign language avatar videos was created and used in a manual evaluation to address these questions. The source data used came from the SOI corpus, which was transcribed using HamNoSys and was synthesised using the UEA JASigning avatar system. The initial 5 videos created came from the baseline system for the avatar Anna. These videos contained limited facial movement and no EFEs. The second set of 5 videos were created with an augmented version of the JASigning system. These videos contained EFEs that were performed parallel to the limited facial movements of the original 5 videos. The Addition of EFEs was made possible through the creation of 7 new morph

targets. Each representing one of the 7 universal emotions and each created in the ARP toolkit. The new morph targets were then added to appropriate non-manual data files and then later to the appropriate elements in SiGML. This process was repeated for a second more caricature like avatar named Luna. The result being 20 videos: 10 baseline videos (5 for Anna and 5 for Luna) and 10 augmented EFE videos (5 for AnnaE and 5 for LunaE). A manual evaluation then took place which recorded and later compared comprehension scores for the videos in each set. Also included in this manual evaluation was an opportunity for feedback on avatar technology and technology in general. As a result some of the feedback received, although valuable, is not directly related to our original research questions.

The results presented in this thesis would indicate that, contrary to our initial hypothesis, the addition of EFEs did not increase the understandability of a signed utterance. In fact, Figure 4-5 shows that the addition of EFEs made very little impact with the score for the baseline avatars and the EFE augmented avatar being 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 4-3 & Figure 4-6 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 4-7) 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 3-27 and Appendix A – Avatar EFEs).

Figure 4-8 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 4-4), 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 4-4, 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, Figure 4-2 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 4-5. Although the average comprehension scores indicate only

a minor effect of EFEs, Figure 4-2 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 4-1 that, predictably, the majority of participants preferred signing video with a real person for web content. However, it was surprising 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 (Conrad, 1979). 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 that 73% were interested in new technologies as reported in section 4.1. 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. This is an important finding and could have implications across, not only the signing avatar field of research, but the broader field of sign language linguistics. This result merits further investigation.

It is interesting to see (in Figure 4-2) 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 the low perceived comprehension score (Figure 4-5) in addition to the results in Figure 4-1, in which no participant chose avatar video as their first choice of web content and only 27% chose it as their second choice. Despite this, 47% of participants indicated a 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 4-4).

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 to go 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.

6.1 Future work

There was ample room for advancement at each phase of this project. These developments range from small software improvements, to a new transcription frame work and a new methodology for synthesising NMFs. This section identifies some of those ‘improvements’ in more detail.

Much work must be done to achieve a usable, comprehensible avatar with particular focus on the linguistic attributes that fared badly in Figure 4-2.

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.

There is scope to develop a whole new transcription framework that can accurately describe, not just EFEs, but all NMFs such that the software can correctly synthesis the movement. This transcription may be used for other applications but should be specifically designed for synthesised sign language avatars and therefore various levels of NMF, for example, on a scale of 0-10. The new system would need to provide more explicit notation than existing systems while at the same time be simple (not laborious) to use. The transcription framework may allow for temporal overlapping of NMF to form larger units such as EFEs, and as such, any graphical user interface developed to support this framework should include commonly used groups of transcription which represents, for example, a particular EFE. A timeline interface would be hugely beneficial to aid temporal placement of MFs and NMFs. Such an interface should provide support for numerous suprasegmental parts.

The use of morphs was fundamental to this work. However, we found their uses quite limited. It would be worth investing some time into investigating the use of various bump maps which may be used to synthesise facial lines and wrinkles.

As discussed in this thesis, motion capture has many disadvantages for synthesised avatars. However it would be interesting to investigate the use of state of the art motion capture technologies to synthesis EFEs and NMFs in general. Focusing initially on facial movement, it may be possible to develop a catalogue of facial movements and configurations much like the catalogue of morphs that currently exist. The theory that underpins this exploration is that the motion capture technology would record a much more fluid and natural movement than the synthesised equivalent.

In an effort to make the JASigning system more accessible to the signing avatar research community it is worth opening the technology to other transcription languages as opposed to HamNoSys alone. For example the SiGMLInLib module can be altered to accept SignWriting via SWML. Many editors already exist that have the ability to generate SWML.

Bibliography

- Allport, F. H., 1920. Social Psychology. *Psychological Bulletin*, Volume 17, pp. 85-94.
- ASLU, 2013. *American Sign Language University*. [Online]
Available at: <http://lifeprint.com/index.htm>
[Accessed August 2013].
- Averill, J., 1983. Studies on anger and aggression: implications for theories of emotion. *American Psychologist*, Volume 38, pp. 1145-60.
- Bangham, J. A., Cox, S. J., Elliott, R. & Glauert, J. R. W., 2000. *An overview of ViSiCAST*. London, The Institution of Electrical Engineer.
- Bentele, S., 1999. [Online]
Available at: <http://www.signwriting.org/forums/linguistics/ling007.html>
[Accessed 07 03 2013].
- Braffort, A., Sansonnet, J.-P., Martin, J.-C. & Verrecchia, C., 2008. *A web-based framework for interactive deictic and sign language agents*. Paris, LIMSI-CNRS.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M. & Maler, E., 2008. *XML 1.0 specification*. [Online]
Available at: <http://www.w3.org/TR/REC-xml/>
[Accessed August 2013].
- Brien, D., 1992. *Dictionary of British sign language-English*. London: Faber & Faber.
- Bungeroth, J., 2002. *A Formal Description of Sign Language Using HPSG*, Karlsruhe: University of Stellenbosch.
- Bungeroth, J., Stein, D., Dreuw, P. & Ney, H., 2008. *The ATIS sign language corpus*. Marrakech, Morocco, LREC.
- Burtnyk, N. & Wein, M., 1971. Computer-generated key-frame animation. *Journal of the Society of Motion Picture and Television Engineers, SMPTE*, 3(80), pp. 149-153.
- C.I.A, 2009. *The world factbook*. [Online]
Available at: <https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html#135>
[Accessed Sept 2013].
- Cawley, C., 2010. *Avatar movie: behind the scenes*. [Online]
Available at: <http://avatarblog.typepad.com/avatar-blog/2010/05/behind-the-scenes-look-at-the-motion-capture-technology-used-in-avatar.html>
[Accessed August 2013].
- Chaloupka, J. & Chaloupka, Z., 2009. *Czech Artificial Computerized Talking Head George*. Springer Berlin Heidelberg, Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions, pp. 324-330.
- Cheng, G., Zhou, Y. & Zhao, Y., 2004. Frame accuracy seamless splicing of MPEG-2 video streams in compressed domain. *Consumer Electronics, IEEE Transactions on*, 50(1), pp. 268-273.
- Cochran, L. & Claspell, E., 1987. *The meaning of grief: A dramaturgical approach to understanding emotion*. 8 ed. Westport CT: Greenwood Press.
- Conrad, R., 1979. *The Deaf School Child: Language and Cognitive Function*. London: HarperCollins Publishers.
- Costa, A. C. D. R. & Dimuro, G. P., 2001. Supporting deaf sign languages in written form on the web. *The SignWriting Journal*, Volume 0, Article 1.

- Cox, S., Lincoln, M., Tryggvason, J. & Nakisa, M., 2002. *TESSA, a system to aid communication with deaf people*. In *Proceedings of the fifth international ACM conference on Assistive technologies*. Edinburgh, ACM, pp. 205-212.
- Darwin, C., 1872. *The expression of the emotions in man and animals*. New York: Appleton and Co.
- DGS-Corpus, 2013. *HamNoSys font*. [Online]
Available at: <http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys-97.html>
[Accessed August 2013].
- Ekman, P., 1999. Basic Emotion. In: *Handbook of cognition and emotion*. UK: Wiley, pp. 45-60.
- Ekman, P., 2003. Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1(1000), pp. 205-221.
- Ekman, P. & Friesen, W. V., 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), p. 124 – 129.
- Ekman, P. & Friesen, W. V., 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Press*.
- Ekman, P., Friesen, W. V. & Hager, J. C., 2002. *Facial Action Coding System: The Manual on CD ROM*, Salt Lake City: A Human Face.
- Elliott, R., Bueno, J., Kennaway, R. & Glauert, J., 2010. *Towards the integration of synthetic animation with avatars into corpus annotation tools*. Valletta, Malta, LREC.
- Elliott, R. & Glauert, J. R. W., 2011. *Virtual Humans Research Wiki*. [Online]
Available at: http://vh.cmp.uea.ac.uk/index.php/Main_Page
[Accessed Sept 2013].
- Elliott, R., Glauert, J. R. W., Jennings, V. & Kennaway, J. R., 2004. *An Overview of the SiGML Notation and SiGML Signing Software System*. *Sign Language Processing Satellite Workshop of the Fourth International Conference on Language Resources and Evaluation*. Lisbon, LREC, pp. 98-104.
- Elliott, R., Glauert, J. R. W. & Kennaway, J. R., 2004. *A framework for non-manual gestures in a synthetic signing system*. *2nd Cambridge Workshop on Universal Access and Assistive Technology*. Cambridge, CWUAAT, pp. 127-136.
- Elliott, R., Glauert, J. R. W., Kennaway, J. R. & Parsons, K. J., 2001. *ViSiCAST deliverable, D5-2: SiGML Definition*, Norwich: ViSiCAST.
- Fabian, P. & Francik, J., 2001. *Synthesis and presentation of the Polish sign language gestures*. Gabčíkovo, 1st International Conference on Applied Mathematics and Informatics at Universities, pp. 190-197.
- Francik, J. & Fabian, P., 2002. *Animating sign language in the real time*. Innsbruck, In 20th IASTED International Multi-Conference on Applied Informatics, pp. 276-281.
- Gartner, 2013. *Worldwide mobile phone sales*. [Online]
Available at: <http://www.gartner.com/newsroom/id/2573415>
[Accessed August 2013].
- Gibbon, D., Moore, R. K. & Winski, R., 1997. *Handbook of standards and resources for spoken language systems*. Berlin: Walter de Gruyter.
- Gibet, S., Lebourque, T. & Marteau, P., 2001. High-level Specification and Animation of Communicative Gestures. *Journal of Visual Languages & Computing*, 12(6), pp. 657-687.
- Gilsdorf, E., 2012. *Hobbit Week: A Conversation With Andy Serkis, Creator of Gollum*. [Online]
Available at: <http://www.wired.com/geekdad/2012/12/andy-serkis/>
[Accessed August 2013].

- Glauert, J., 2002. *ViSiCAST Milestone M5-8: SiGML Standardisation*, Norwich: ViSiCAST.
- Glauert, J. & Elliott, R., 2011. *Extending the SiGML Notation—a Progress Report*. Dundee, 2nd International Workshop on Sign Language Translation and Avatar Technology (SLTAT).
- Glauert, J. R. W., Kennaway, J. R., Elliott, R. & Theobald, B. J., 2004. *Virtual human signing as expressive animation*. Leeds, University of Leeds, pp. 98-106.
- Grieve-Smith, A. B., 1999. *English to American Sign Language machine translation of weather report*. Albuquerque, HDSL2, pp. 23-30.
- Grieve-Smith, A. B., 2002. SignSynth: A sign language synthesis application using Web3D and Perl. *Gesture and Sign Language in Human-Computer Interaction*, pp. 37-53.
- Hager, J. C., 2003. *FACSAID Demonstration*. [Online]
Available at: <http://www.face-and-emotion.com/dataface/facsaid/facsaid.jsp>
[Accessed August 2013].
- Hanke, T. & Popescu, H., 2003. *EDC-22124 ESIGN / 27960 Deliverable D2.3 :Intelligent Sign Editor*. [Online]
Available at: <http://robertsmith2012.files.wordpress.com/2012/10/esign-editor-d23.pdf>
[Accessed August 2013].
- Hellwig, B., Uytvank, D. v. & Hulsbosch, M., 2012. *ELAN-Linguistic Annotator, version 4.4.0*. Nijmegen, Max Planck Institute for Psycholinguistic.
- Huenerfauth, M. & Lu, P., 2010. *Modeling and synthesizing spatially inflected verbs for American sign language animations*. Orlando, 12th international ACM SIGACCESS conference on Computers and accessibility, pp. 99-106.
- Huenerfauth, M., Zhao, L., Gu, E. & Allbeck, J., 2008. Evaluation of American Sign Language Generation by Native ASL Signers. *Transactions on Access Computing*, Vol. 1(1), p. Article 3.
- Hunger*. 1973. [Film] Directed by Peter Foldès. Canada: National Film Board of Canada.
- Hurdich, J., 2008. *Utilizing Lifelike, 3D Animated Signing Avatar Characters for the Instruction of K-12 Deaf Learners*. New York, International Tech Symposium on Instructional Technology and Education of the Deaf, National Technical Institute of the Deaf, NTID.
- IDS, 2013. *Irish Deaf Society - ISL survey*. [Online]
Available at: <http://www.irishdeafociety.ie/irish-sign-language.html>
[Accessed March 2014].
- Irish National Rehabilitation Board, 1991. *Access to colleges: A survey of third level centres*. Dublin: National Rehabilitation Board, National Advisory Committee on Training and Employment.
- Jennings, V., Elliott, R., Kennaway, R. & Glauert, J., 2010. *Requirements For A Signing Avatar*. Malta, 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. LREC, pp. 133-136.
- Jespersen, O., 1889. *The articulations of speech sounds represented by means of alphabetic symbols*. Marburg: NG Elwert.
- Johnson, R. E. & Liddell, S. K., 2011. A segmental framework for representing signs phonetically. *Sign Language Studies*, 3(11), pp. 408-463.
- Johnston, T., 2008. *Creating a corpus of Auslan within an Australian National Corpus*. Sydney, HSCNet Workshop Proceedings: Designing the Australian National Corpus.
- Kennaway, J. R., Glauert, J. R. W. & Zwitterlood, I., 2007. Providing signed content on the Internet by synthesized animation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(14).
- Kennaway, R., 2003. *Experience with and requirements for a gesture description language for synthetic animation*. 5th International Gesture Workshop. Genova, Springer Verlag, p. 300–311.

- Kipp, M., Heloir, A. & Nguyen, Q., 2011. Sign language avatars: animation and comprehensibility. *Intelligent Virtual Agents*, pp. 113-126.
- Krňoul, Z., Kanis, J., Železný, M. & Müller, L., 2008. Czech text-to-sign speech synthesizer. *Machine Learning for Multimodal Interaction*, pp. 180-191.
- Krňoul, Z. & Železný, M., 2007. Translation and conversion for Czech sign speech synthesis. *Text, Speech and Dialogue*, pp. 524-531.
- Labov, W., 1972. Sociolinguistic Patterns. In: Vol 4 ed. Philadelphia: University of Pennsylvania press, p. 209.
- Lebourque, T. & Gibet, S., 1999. *A complete system for the specification and the generation of sign language gestures*. Berlin, Springer, pp. 227-238.
- Leeson, 2001. *Aspects of Verb Valency in Irish Sign Language*, Dublin: PhD dissertation, Trinity College Dublin.
- Leeson, L., 2003. *Sign Language Interpreters: Agents of Social Change in Ireland*. s.l., Languages of Ireland, Dublin, Ireland: Four Courts Press.
- Leeson, L. & Nolan, B., 2008. *Digital Deployment of the Signs of Ireland Corpus in Elearning*. Marrakech, 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora, p. 112-121.
- Leeson, L., Saeed, J., Macduff, A. & Byrne-Dunne, D., 2006. *Moving Heads and Moving Hands: Developing a Digital Corpus of Irish Sign Language*. Carlow, Ireland, Information Technology and Telecommunications, pp. 25-26.
- Liddell, S., 1993. Holds and Positions: Comparing Two Models of Segmentation in ASL. *Phonetics and Phonology Volume 3: Current Issues in ASL Phonology*, Volume 3, pp. 198-211.
- Liddell, S. K. & Johnson, R. E., 1989. *American sign language: The phonological base*. Washington. DC: Gallaudet University Press.
- Lutz, C. & White, G., 1986. The anthropology of emotions. *Annual Review of Anthropology*, 15(4), pp. 405-436.
- Malim, T. & Birch, A., 1998. *Intorductory Psycology*. New york: Palgrave MacMillan.
- Marriott, A. & Stallo, J., 2002. *VHML – Uncertainties and Problems, A discussion*. Bologna, AAMAS, Workshop on Embodied Conversational Agents.
- Martin, J., 2007. *SignWriting*. [Online]
Available at: <http://www.signwriting.org/forums/linguistics/ling006.html>
[Accessed August 2013].
- Matthews, P., 1996. *The Irish Deaf Community, Survey Report, History of Education, Language and Culture*. Dublin: The Linguistics Institute of Ireland.
- Metadata*. 1971. [Film] Directed by Peter Foldès. Canada: National Film Board of Canada.
- Morrissey, S., 2011. *Assessing three representation methods for sign language machine translation and evaluation*. Leuven, Belgium, EAMT 2011, pp. 137-144.
- Morrissey, S., Somers, H., Smith, R. & Gilchrist, S., 2010. *Building a Sign Language corpus for use in Machine Translation*. Valletta, Malta, European Language Resources Association, LREC, pp. 172-177.
- Morrissey, S. & Way, A., 2013. Manual labour: tackling machine translation for sign languages. *Machine Translation*, pp. 1-40.
- 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*. Granada, Assistive Technologies for People with Vision & Hearing Impairments (CVHI).

- National-Film-Board-of-Canada, 1996. *Retired NRC Scientists Burtnyk and Wein honoured as Fathers of Computer Animation Technology in Canada*. [Online]
Available at: http://www.ieee.ca/millennium/computer_animation/animation_honoured.html
[Accessed August 2013].
- O'Baoill, D. & Matthews, P. A., 2000. *The Irish Deaf Community (Volume 2): The structure of Irish Sign Language*. Dublin: The Linguistics Institute of Ireland.
- Optitrack, 2013. *mocap hardware and software setup*. [Online]
Available at: <http://www.naturalpoint.com/optitrack/support/faq/mocap.html>
[Accessed August 2013].
- Papadogiorgaki, M., Grammalidis, N., Makris, L. & Strintzis, M. G., 2006. *Gesture synthesis from sign language notation using MPEG-4 humanoid animation parameters and inverse kinematics*. Athens, 2nd IET International Conference on Intelligent Environments, pp. 151-160.
- Papadogiorgaki, M., Grammalidis, N., Tzovaras, D. & Strintzis, M. G., 2005. *Text-to-sign language synthesis tool*. Antalya, 13th European signal processing conference (EUSIPCO2005).
- Peiper, A., 1963. *Cerebral function in infancy and childhood*. New York: Plenum Pub Corp.
- Planalp, S., 1999. *Communicating emotion: Social, moral, and cultural processes*. Cambridge: Cambridge University Press.
- Plutchik, R., 1980. Emotion: Theory, research, and experience. *Theories of emotion*, Volume 1.
- Prillwitz, S., 1987. *HamNoSys: Hamburger Notation System for Sign Language; an Introduction*. Hamburg: Center for German Sign Language.
- Prillwitz, S., Leven, R., Zienert, H. & Hanke, T., 1989. HamNoSys Version 2.0: Hamburg notation system for sign languages: An introductory guide. *International Studies on Sign Language and the Communication of the Deaf*, Volume 5.
- Rosetta, 2013. *About the rossetta foundation*. [Online]
Available at: <http://www.therosettafoundation.org/about/what-we-do/>
[Accessed Sept 2013].
- Sampson, G., 1985. *Writing systems: A linguistic introduction*. Stanford: Stanford University Press.
- San-Segundo, R., Montero, J. M., Cordoba, R. & Sama, V., 2012. Design, Development and Field Evaluation of a Spanish into Sign Language Translation System. *Pattern Analysis and Applications*, 15(2), pp. 203-224.
- Schmaling, C. & Hanke, T., 2001. *ViSiCAST Deliverable D5-1: Interface Definitions*, Hamburg: ViSiCAST.
- SignWriting, 2013a. *SignWriting participants*. [Online]
Available at: <http://www.signwriting.org/about/who/>
[Accessed August 2013].
- SignWriting, 2013b. *SignWriting software*. [Online]
Available at: <http://www.signwriting.org/forums/software/>
[Accessed August 2013].
- SignWriting, 2013c. *SignWriting Library*. [Online]
Available at: <http://www.signwriting.org/library/>
[Accessed August 2013].
- Smith, R., 2013. *HamNoSys handbook*. [Online]
Available at: <http://robertsmith2012.files.wordpress.com/2012/10/hamnosys-user-guide-rs-draft-v3-0.pdf>
[Accessed Sept 2013].

- Smith, R. & Nolan, B., 2013a. *Manual Evaluation of synthesised Sign Language Avatars*. Bellevue, Washington, In proceedings of ASSETS 2013: SIGACCESS ACM.
- Smith, R. & Nolan, B., 2013b. *Emotional Facial Expressions in Synthesised Sign Language Avatars: A Manual Evaluation*. Chicago, In Proceedings of 3rd International Symposium on Sign Language Translation and Avatar Technology: Universal Access in the Information Society, Springer.
- Stokoe, W., 1960. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *Studies in linguistics: Occasional papers*.
- Stokoe, W. C., Casterline, D. C. & Croneberg, C. G., 1976. *A dictionary of American Sign Language on linguistic principles*. Silver Spring: Linstok Press.
- Supalla, S. & Blackburn, L., 2003. Learning How to Read and Bypassing Sound. *Odyssey*, 5(1), pp. 50-55.
- Supalla, S. J., Wix, T. R. & McKee, C., 2001. Print as a primary source of English for deaf learners. In: J. Nicol, ed. *One mind, two languages: Studies in bilingual language processing*. Oxford: Blackwell, pp. 177-190.
- Suszczańska, N., Romaniuk, J. & Szmaj, P., 2005. *Automatic analysis of elliptic sentences in the Thetos system*. Nancy, SemDial: 9th workshop on the semantics and pragmatics of dialogue.
- Suszczańska, N., Szmaj, P. & Francik, J., 2002. *Translating Polish texts into sign language in the TGT system*. Innsbruck, Austria, 20th IASTED Multiconference Applied Informatics.
- Sutton, V., 1999. *Learn to Read American Sign Language in SignWriting - Goldilocks & The Three Bears*. San Diego, California: The Deaf Action Committee For SignWriting.
- Sutton, V. J., 2002. *Lessons in sign writing (revised)*. San Diego, California: Center Sutton Movement Writing.
- The-Canadian-Cultural-Society-of-the-Deaf, 2012. *ASL-phabet*. [Online]
Available at: <http://aslphabet.com/home.html>
[Accessed August 2013].
- Thomas, F. & Johnston, O., 1995. *The illusion of life: Disney animation*. (Original work published 1981) ed. New York: Disney Edition.
- Time-Magazine, 2012. *Ground breaking mocap technology*. [Online]
Available at: <http://entertainment.time.com/2012/12/05/gollums-getup-how-the-hobbits-groundbreaking-technology-works/>
[Accessed August 2013].
- U.N, 2013. *UN official languages*. [Online]
Available at: <http://www.un.org/en/aboutun/languages.shtml>
[Accessed Sept 2013].
- unity3D.com, 2013. *Using Blender and Rigify*. [Online]
Available at: <https://docs.unity3d.com/Documentation/Manual/BlenderAndRigify.html>
[Accessed August 2013].
- VCom3D, 2007. *Welcome page*. [Online]
Available at: <http://www.vcom3d.com/>
[Accessed August 2013].
- Velho, L., 2001. Quasi 4–8 subdivision. *Computer Aided Geometric Design*, 4(18), pp. 345-357.
- WDF, 2013. *World Deaf Federation*. [Online]
Available at: <http://wfdeaf.org/human-rights/crpd/sign-language>
[Accessed August 2013].
- Wilbur, R. B., 1987. *American Sign Language: Linguistic and Applied Dimensions*. 2 ed. unknown: Little, Brown.

Yeates, S., Holden, E. J. & Owens, R., 2002. *Real-time 3d graphics for human modelling and teaching sign language*. Zakopane, Association for Image Processing, Warszawa, Poland, pp. 815-821.

Appendix A – Avatar EFEs

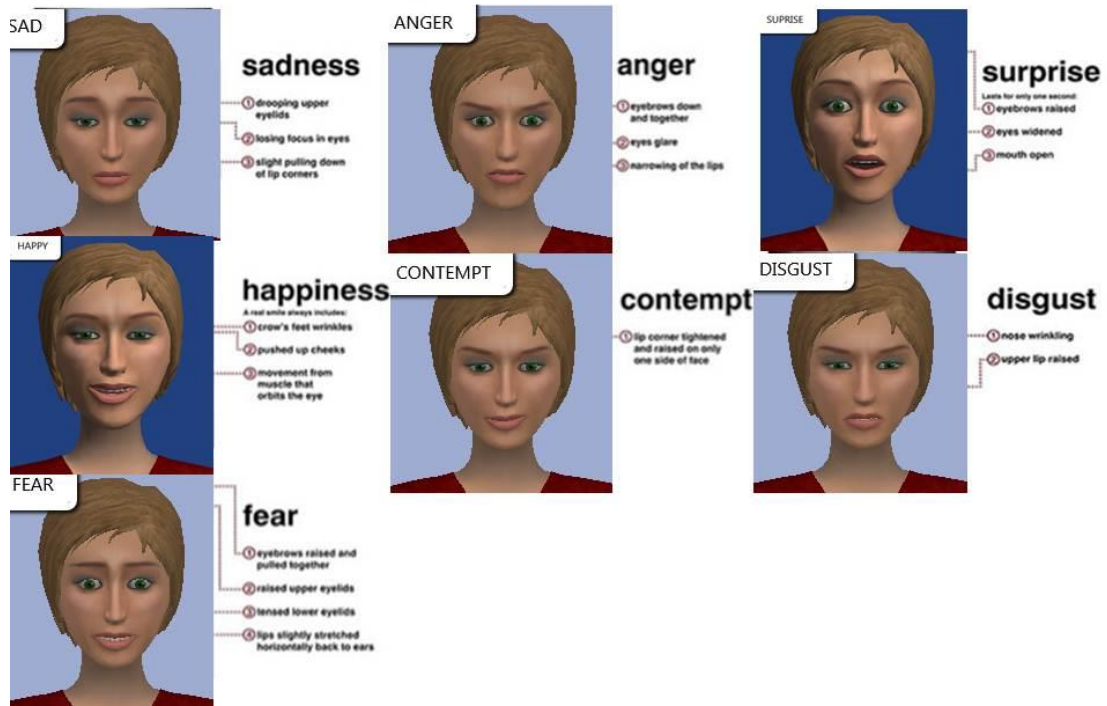


Figure A-1 EFEs for Anna



Figure A-2 EFEs for Luna
























Appendix B – HamNoSys Symbol set

Table B.1 HamNoSys Symbol set
























	UCS-1997	UCS	HNS	Octet	HNS-name
1	U+F009	U+0009		9	tab
2	U+F00A	U+000A		10	linefeed
3	U+F00C	U+000C		12	pagebreak
4	U+F00D	U+000D		13	return
5	U+FFFF	U+E0F1	<u>4.</u>	20	hamversion40
6	U+F020	U+0020		32	hamspace
7	U+F021	U+0021	!	33	hamexclaim
8	U+F022	U+003F	?	34	hamquery
9	U+F023	U+002E	.	35	hamfullstop
10	U+F024	U+002C	,	36	hamcomma
11	U+F025	U+E0E7	+	37	hamplus
12	U+F026	U+007C		38	hammetaalt
13	U+F027	U+E0B0	⌚	39	hamclocku
14	U+F028	U+E0B1	⌚	40	hamclockul
15	U+F029	U+E0B2	⌚	41	hamclockl
16	U+F02A	U+E0B3	⌚	42	hamclockdl
17	U+F02B	U+E0B4	⌚	43	hamclockd
18	U+F02C	U+E0B5	⌚	44	hamclockdr
19	U+F02D	U+E0B6	⌚	45	hamclockr
20	U+F02E	U+E0B7	⌚	46	hamclockur
21	U+F02F	U+E0B8	⊕	47	hamclockfull
22	U+F030	U+E0E8	:	48	hamsymmpar





















23	U+F031	U+E0E9	••	49	hamsymmlr
24	U+F032	U+E000	○	50	hamfist
25	U+F033	U+E001	◻	51	hamflathand
26	U+F034	U+E002	┐	52	hamfinger2
27	U+F035	U+E003	┌	53	hamfinger23
28	U+F036	U+E004	└	54	hamfinger23spread
29	U+F037	U+E005	┘	55	hamfinger2345
30	U+F038	U+E00C	,	56	hamthumboutmod
31	U+F039	U+E00D		57	hamthumbacrossmod
32	U+F03A	U+E006	◁	58	hampinch12
33	U+F03B	U+E007	◂	59	hampinchall
34	U+F03C	U+E008	◃	60	hampinch12open
35	U+F03D	U+E009	⋃	61	hamcee12
36	U+F03E	U+E00A	⊃	62	hamceeall
37	U+F03F	U+E00B	⊅	63	hamceeopen
38	U+F040	U+E00E		64	hamthumbopenmod
39	U+F041	U+E010		65	hamfingerstraightmod
40	U+F042	U+E011		66	hamfingerbendmod
41	U+F043	U+E012		67	hamfingerhookmod
42	U+F044	U+E0EA	◊	68	hamnondominant
43	U+FFFF	U+E013		69	hamdoublebent
44	U+FFFF	U+E014		70	hamdoublehooked
45	U+F048	U+E020	⤴	72	hamextfingeru
46	U+F049	U+E021	⤵	73	hamextfingerur
47	U+F04A	U+E022	⤶	74	hamextfingerr

48	U+F04B	U+E023	↙	75	hamextfingerdr
49	U+F04C	U+E024	↘	76	hamextfingerd
50	U+F04D	U+E025	↗	77	hamextfingerdl
51	U+F04E	U+E026	↖	78	hamextfingerl
52	U+F04F	U+E027	↗	79	hamextfingerul
53	U+F050	U+E028	↖	80	hamextfingerol
54	U+F051	U+E029	↗	81	hamextfingero
55	U+F052	U+E02A	↖	82	hamextfingeror
56	U+F053	U+E02B	↗	83	hamextfingeril
57	U+F054	U+E02C	↘	84	hamextfingeri
58	U+F055	U+E02D	↙	85	hamextfingerir
59	U+F056	U+E02E	↖	86	hamextfingerui
60	U+F057	U+E02F	↗	87	hamextfingerdi
61	U+F058	U+E030	↙	88	hamextfingerdo
62	U+F059	U+E031	↖	89	hamextfingeruo
63	U+FFFF	U+E048	?	93	hamearlobe
64	U+FFFF	U+E046	h	94	hamnostrils
65	U+FFFF	U+E050	h	95	hamshouldertop
66	U+F060	U+E038	o	96	hampalmu
67	U+F061	U+E039	o	97	hampalmur
68	U+F062	U+E03A	o	98	hampalmr
69	U+F063	U+E03B	o	99	hampalmdr
70	U+F064	U+E03C	o	100	hampalmd
71	U+F065	U+E03D	o	101	hampalmdl
72	U+F066	U+E03E	o	102	hampalmi

73	U+F067	U+E03F		103	hampalmul
74	U+F068	U+E0AA		104	hamreplace
75	U+F069	U+E0D4		105	hamarmextended
76	U+F06A	U+E0D5		106	hambehind
77	U+FFFF	U+E0EC		107	hametc
78	U+FFFF	U+E0ED		108	hamorirelative
79	U+FFFF	U+E04B		109	hamtongue
80	U+FFFF	U+E04C		110	hamteeth
81	U+F06F	U+E053		111	hamstomach
82	U+F070	U+E05F		112	hamneutralspace
83	U+F071	U+E040		113	hamhead
84	U+F072	U+E041		114	hamheadtop
85	U+F073	U+E042		115	hamforehead
86	U+F074	U+E043		116	hameyebrows
87	U+F075	U+E044		117	hameyes
88	U+F076	U+E045		118	hamnose
89	U+F077	U+E047		119	hamear
90	U+F078	U+E049		120	hamcheek
91	U+F079	U+E04A		121	hamlips
92	U+F07A	U+E04D		122	hamchin
93	U+F07B	U+E04E		123	hamunderchin
94	U+F07C	U+E04F		124	hamneck
95	U+F07D	U+E051		125	hamshoulders
96	U+F07E	U+E052		126	hamchest
97	U+F080	U+E054		128	hambelowstomach

98	U+F081	U+E058	◻	129	hamlrbeside
99	U+F082	U+E059	■	130	hamlrat
100	U+F083	U+E060	↵	131	hamupperarm
101	U+F084	U+E061	↵	132	hamelbow
102	U+F085	U+E062	↵	133	hamelbowinside
103	U+F086	U+E063	↵	134	hamlowerarm
104	U+F087	U+E064	↵	135	hamwristback
105	U+F088	U+E065	↵	136	hamwristpulse
106	U+F089	U+E066	↵	137	hamthumbball
107	U+F08A	U+E067	↵	138	hampalm
108	U+F08B	U+E068	↵	139	hamhandback
109	U+F08C	U+E070	1	140	hamthumb
110	U+F08D	U+E071	2	141	hamindexfinger
111	U+F08E	U+E072	3	142	hammiddlefinger
112	U+F08F	U+E073	4	143	hamringfinger
113	U+F090	U+E074	5	144	hampinky
114	U+F091	U+E069	1	145	hamthumbside
115	U+F092	U+E06A	5	146	hampinkyside
116	U+F093	U+E0E6	\	147	hambetween
117	U+F094	U+E075	↑	148	hamfingertip
118	U+F095	U+E076	↑	149	hamfingernail
119	U+F096	U+E077	●	150	hamfingerpad
120	U+F097	U+E078	●	151	hamfingermidjoint
121	U+F098	U+E079	↑	152	hamfingerbase
122	U+F099	U+E07A	▮	153	hamfingerside

123	U+F09A	U+E07C		154	hamwristtopulse
124	U+F09B	U+E07D		155	hamwristtoback
125	U+F09C	U+E07E		156	hamwristtothumb
126	U+F09D	U+E07F		157	hamwristtopinky
127	U+FFFF	U+E05A		158	hamcoreftag
128	U+F09F	U+E05B		159	hamcorefref
129	U+F0A0	U+E0AF		160	hamnomotion
130	U+F0A1	U+E080		161	hammoveu
131	U+F0A2	U+E081		162	hammoveur
132	U+F0A3	U+E082		163	hammover
133	U+F0A4	U+E083		164	hammovedr
134	U+F0A5	U+E084		165	hammoved
135	U+F0A6	U+E085		166	hammovedl
136	U+F0A7	U+E086		167	hammovei
137	U+F0A8	U+E087		168	hammoveul
138	U+F0A9	U+E088		169	hammoveol
139	U+F0AA	U+E089		170	hammoveo
140	U+F0AB	U+E08A		171	hammoveor
141	U+F0AC	U+E08B		172	hammoveil
142	U+F0AD	U+E08C		173	hammovei
143	U+F0AE	U+E08D		174	hammoveir
144	U+F0AF	U+E08E		175	hammoveui
145	U+F0B0	U+E08F		176	hammovedi
146	U+F0B1	U+E090		177	hammovedo
147	U+F0B2	U+E091		178	hammoveuo

148	U+F0B3	U+E0AD		179	hammovecross
149	U+F0B4	U+E0AE		180	hammoveX
150	U+F0B5	U+E0C6		181	hamsmallmod
151	U+F0B6	U+E0C7		182	hamlargemod
152	U+F0B7	U+E0B9		183	hamarcl
153	U+F0B8	U+E0BA		184	hamarcu
154	U+F0B9	U+E0BB		185	hamarcr
155	U+F0BA	U+E0BC		186	hamarcd
156	U+F0BB	U+E0BD		187	hamwavy
157	U+F0BC	U+E0BE		188	hamzigzag
158	U+F0BD	U+E0A4		189	hamfingerplay
159	U+F0BE	U+E0E2	[190	hamparbegin
160	U+F0BF	U+E0E3]	191	hamparend
161	U+F0C0	U+E092		192	hamcircleo
162	U+F0C1	U+E093		193	hamcirclei
163	U+F0C2	U+E094		194	hamcircled
164	U+F0C3	U+E095		195	hamcircleu
165	U+F0C4	U+E096		196	hamcirclel
166	U+F0C5	U+E097		197	hamcircler
167	U+F0C6	U+E0C4		198	hamincreasing
168	U+F0C7	U+E0C5		199	hamdecreasing
169	U+F0C8	U+E0D0)("	200	hamclose
170	U+F0C9	U+E0D1		201	hamtouch
171	U+F0CA	U+E0D2		202	haminterlock
172	U+F0CB	U+E0D3		203	hamcross

173	U+F0CC	U+E0C8	*	204	hamfast
174	U+F0CD	U+E0C9	—	205	hamslow
175	U+F0CE	U+E0CA	✱	206	hamtense
176	U+F0CF	U+E0CB	7	207	hamrest
177	U+F0D0	U+E0CC		208	hamhalt
178	U+F0D1	U+E0D8	+	209	hamrepeatfromstart
179	U+F0D2	U+E0D9	++	210	hamrepeatfromstartseveral
180	U+F0D3	U+E0DA	→	211	hamrepeatcontinue
181	U+F0D4	U+E0DB	→	212	hamrepeatcontinueseveral
182	U+F0D5	U+E0E0	(213	hamseqbegin
183	U+F0D6	U+E0E1)	214	hamseqend
184	U+F0D7	U+E0DD	~	215	hamalternatingmotion
185	U+F0D8	U+E0DC	↵	216	hamrepeatreverse
186	U+FFFF	U+E0D6	♣	217	hambrushing
187	U+FFFF	U+E0EB	Я	218	hamnonipsi
188	U+F0DC	U+E0C0	⊖	220	hamellipseh
189	U+F0DD	U+E0C1	∅	221	hamellipseur
190	U+F0DE	U+E0C2	∅	222	hamellipsev
191	U+F0DF	U+E0C3	∅	223	hamellipseul
192	U+F0E0	U+E0F0	U	224	hammime
193	U+FFFF	U+007B	{	225	hamaltbegin
194	U+FFFF	U+007D	}	226	hamaltend
195	U+FFFF	U+E0A5	↕	227	hamnodding
196	U+FFFF	U+E0A6	↔	228	hamswinging
197	U+FFFF	U+E0A7	↻	229	hamtwisting

198	U+FFFF	U+E0A8	ϣ	230	hamstircw
199	U+FFFF	U+E0A9	ϣ̇	231	hamstircw
200	U+FFFF	U+E0E4	⋖	236	hamfusionbegin
201	U+FFFF	U+E0E5	⋗	237	hamfusionend
202	U+F0F0	U+E098	⦿	240	hamcircleul
203	U+F0F1	U+E099	⦿̇	241	hamcircledr
204	U+F0F2	U+E09A	⦿̂	242	hamcircleur
205	U+F0F3	U+E09B	⦿̃	243	hamcircledl
206	U+F0F4	U+E09C	⦿̄	244	hamcircleol
207	U+F0F5	U+E09D	⦿̅	245	hamcircleir
208	U+F0F6	U+E09E	⦿̆	246	hamcircleor
209	U+F0F7	U+E09F	⦿̇	247	hamcircleil
210	U+F0F8	U+E0A0	⦿̈	248	hamcircleui
211	U+F0F9	U+E0A1	⦿̉	249	hamcircledo
212	U+F0FA	U+E0A2	⦿̊	250	hamcircleuo
213	U+F0FB	U+E0A3	⦿̋	251	hamcircledi
214	U+F0FE	U+00FE	þ	254	hamnbs

Appendix C – ARP Toolkit workflow

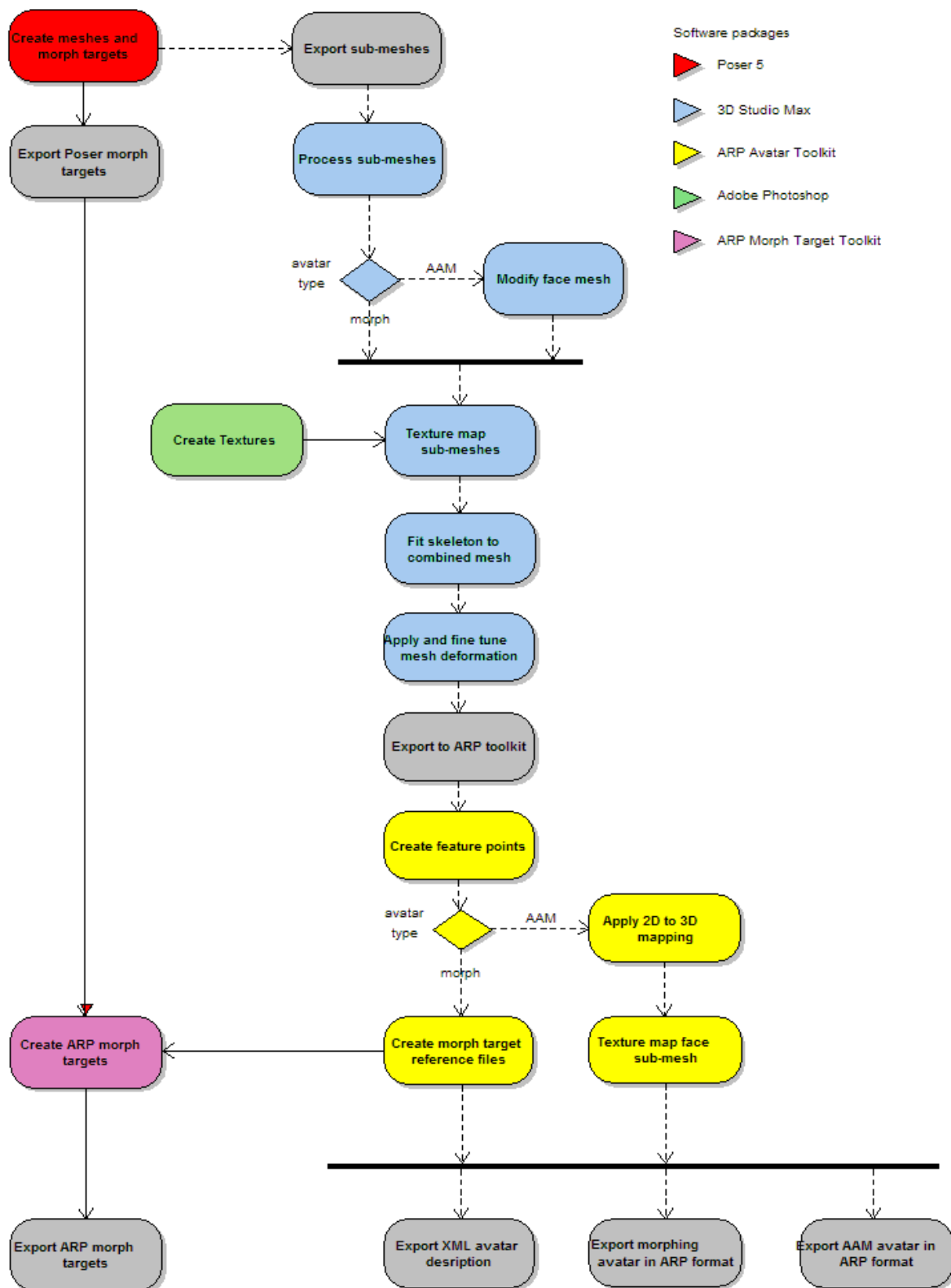


Figure C-1 ARP Toolkit workflow

Appendix D – Motion capture rig

Gollum's Getup How performance capture works

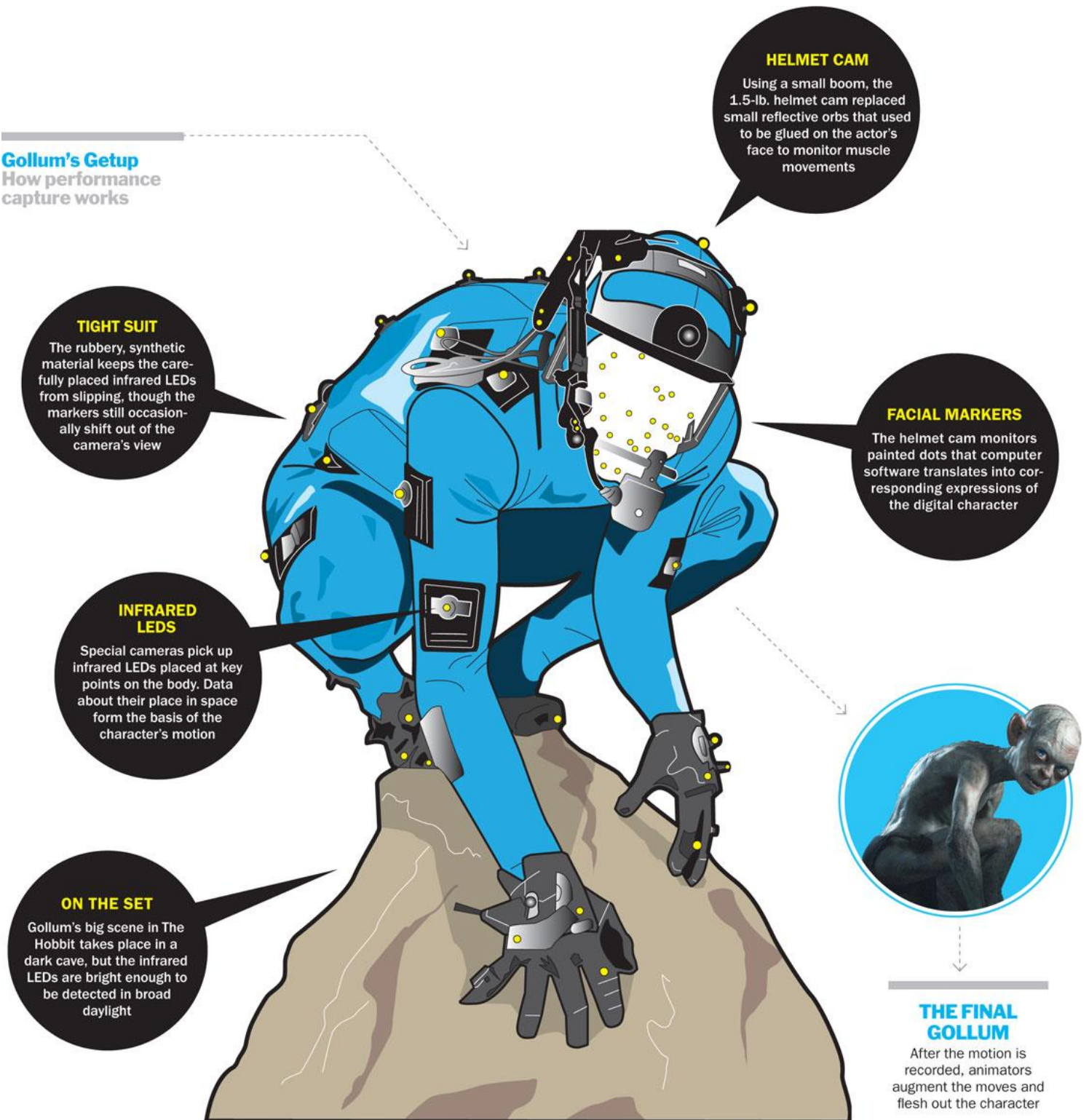


Figure D-1 State of the art motion capture rig as worn by actor Andy Serkis as he played the creature “Gollum” in the 2012 movie “The Hobbit: An Unexpected Journey”. Illustration by heather jones for time; Warner Brothers. (Time-Magazine, 2012)



Figure D-2 Actor Andy Serkis wearing the mocap suit (above) and as the digital Gollum (below), in *The Hobbit: An Unexpected Journey* (Image: Warner Bros. Pictures). Images sourced from (Gilsdorf, 2012)



Figure D-3 Mobcap face rig being used by the actress Zoe Saldana in the 2009 film Avatar. Image source: (Cawley, 2010)

Appendix E – interview questions

Phase 1 - Establishing questions

Personal Details

Participant ID:

Age: 0 – 12, 13 – 18, 19 – 30, 31 – 40, 41 – 50, 51 – 60, 61+

Sex: M / F

Home Town: _____

ISL and literacy background

1. How would you rate your SL skills 0 1 2 3 4 5
@School? | 3rd level? | Qualified interpreter? | Duration?
2. Do you have Deaf siblings? Y / N
3. Would you consider your level of English: good | middle | poor?

Technology background

1. When using software/web, is it a problem for you to read the text? Y / N
2. Would you prefer the content translated into SL video? Why? Y / N
3. Would you consider yourself a gadget/technical person? Y / N
4. Do you like animated movies such as “Shrek” or “the Incredibles”? Y / N
5. Have you ever used a sign language avatar before? Y / N
6. Do you enjoy using this technology? Y / N
7. Do you find them easy to understand? Y / N
8. Do you think that there are any difficulties with this technology? Y / N
9. Do you fear that this technology will replace ‘real’ interpreters? Or ‘human’ video?
Y / N
10. Do you think avatar technology has a place in your life? Y / N

Phase 2 - Comprehension Questions

Video 1

		1 st viewing	2 nd viewing
1.	Was that easy to understand?	0 1 2 3 4 5	0 1 2 3 4 5
2.	The Avatar mentioned a number. What was that number? Can you remember if she was talking about age, quantity or time?		
3.	Did the avatar mention a sibling? Brother or sister? Younger or older?		
4.	Did you see any emotion?	y/n	y/n

Video 2

1	Was that easy to understand?	0 1 2 3 4 5	0 1 2 3 4 5
2	Where did the avatar start her journey? What was her destination?		
3	How did she get there?		
4	Who/What did she see on her journey?		
5	What was the avatar asked to produce at the checkpoint?		
6.	Did you see any emotion?	y/n	y/n

Video 3

1	Was that easy to understand?	0 1 2 3 4 5	0 1 2 3 4 5
2	Where was the Avatar?		
3	How was the Avatar travelling?		
4	What was the avatar going to do?		
5	Who went shopping with the Avatar?		
6	What did they see while shopping?		
7	Was the area busy?		
8	How did the avatar feel?		
9	How old was the Avatar?		
10.	Did you see any emotion?	y/n	y/n

Video 4

1	Was that easy to understand?	0 1 2 3 4 5	0 1 2 3 4 5
2	Who was with the Avatar this time?		
3	Where they in a rush?		
4	They walked into a building. What was that building?		
5	What did the man say to them?		
6	How old was the Avatar?		
7	What did the father do?		
8	Do you remember what the sister's reply to the Father was?		
9	What did the father then tell the sisters to do?		
10	How did the Mother feel about that?		
11	After the argument, where did the avatar go?		
12	What did the avatar do at the table?		
13	What was the avatar holding?		
14	Where did the avatar look?		
15	What did the Avatar see?		
16.	Did you see any emotion?	y/n	y/n

Video 5

1	Was that easy to understand?	0 1 2 3 4 5	0 1 2 3 4 5
---	------------------------------	-------------	-------------

2	Who is the Avatar talking about in the video?		
3	Where did the sister put her hand?		
4	What was her sister doing?		
5	What did her sister think?		
6.	Did you see any emotion?	y/n	y/n

Phase 3 - Post questions

1. Did you like **Anna**? Y/N
2. What were **Anna's** best and worst points?
3. Did you like **Luna**? Y/N
4. What were **Luna's** best and worst points?
5. Where the avatars easy to understand? Which avatar did you find easier to understand: Luna
| Anna
6. Does the avatars movement look natural?
7. In Your opinion, where would this technology best suit:
 - a. To translate text - Novel size / paragraph size.
 - b. To translate websites.
 - c. In a classroom environment
 - d. Official meetings with Doctor/Bank/Solicitor etc...
 - e. TV signing
 - f. Other
8. Which is preferable the avatar or video or read English text? (rank 1-3)
9. Would you use this avatar if there was no ISL video available?