

Investigating the Intelligibility of Synthetic Sign Language Visualization Methods on Mobile Phones

Mehrdad Ghaziasgar, James Connan

Department of Computer Science

University of the Western Cape, Private Bag X17, Bellville, Cape Town, 7535

Tel: +27 21 9593010, Fax: +27 21 9593006

email: {2438188, jconnan}@uwc.ac.za

Abstract—More than 300,000 South African deaf people, whose first and only language is South African Sign Language (SASL), face difficulty in communicating with hearing people on a daily basis, despite the pressing need to do so. The SASL project at the University of the Western Cape is in the process of designing a Machine Translation system on mobile phones to automate the translation between SASL and English as a supplement to the use of highly skilled interpreters that are scarce and expensive. This research investigated the intelligibility of the sign language rendered by four methods of displaying SASL to be integrated into the SASL MT system on mobile phones. Three of these methods make use of 3D humanoid Avatars and the fourth method makes use of videos of SASL-speaking humans. We found that on average, a recognition rate of 65% was achieved across all methods, clearly demonstrating that synthetic sign language rendered on mobile phones can be intelligible. However, we also found that most of the recognition errors were caused by SASL dialect differences amongst test subjects. We showed that eliminating this factor increased the average accuracy to 85% across all methods.

Index Terms—SASL, Sign Language, Avatar, Visualization, Feasibility, Intelligibility.

I. INTRODUCTION

There are over one million deaf people in South Africa [1]. Approximately 300,000 of these are profoundly deaf in both ears and use South African Sign Language (SASL) as their first and only language [1]. Contrary to common belief, sign languages are not gestural representations of spoken languages but are fully-fledged languages of their own [2]. Additionally, different countries have sign languages of their own such as Greek Sign Language (GSL) in Greece, Japanese Sign Language (JSL) in Japan, Arabic Sign Language (ArSL) in Saudi Arabia, British Sign Language (BSL) in the United Kingdom and South African Sign Language (SASL) in South Africa [3]. Each of these languages also experience variations from region to region that is analogous to dialect and idiolect differences in spoken languages. The majority of deaf people in South Africa are illiterate in spoken languages [1].

These facts have contributed to the creation of a strict communication barrier between the deaf and hearing [4]. Nevertheless, there is a pressing need for the deaf to communicate with the hearing on a daily basis in numerous contexts. Research has shown that 90% of all deaf children

are born to hearing parents [5]. These children need to communicate with their hearing parents as a necessity of life. Other circumstances include seeking medical and dental treatment, purchasing groceries, purchasing train or bus tickets, to mention only a few. Another significant problem arising from this communication barrier is discrimination in employment opportunities [6]. Employers are reluctant to employ people that are illiterate and are difficult to communicate with. This, in turn, has caused a state of poverty amongst the deaf in South Africa [1].

The use of interpreters to remedy this situation has proven inefficient and inadequate. Interpreters are in short supply [1] [7]. Their services are also very costly and beyond the financial capability of most deaf people.

The SASL project [8] at the University of the Western Cape is in the process of designing a mobile phone-based Machine Translation system that can automate the translation between SASL and English. One of the requirements of such a system is to synthesize sign language. D. van Wyk produced a high quality 3D humanoid sign language visualization system (SLV) for desktop computers [7]. He refers to his Avatar as *Man*. *Man* can visualize both manual and non-manual sign language gestures. Manual gestures are mainly hand motion and shapes. Non-manual gestures include facial expressions, eye gaze and motions of the body, neck and head. It is thought that non-manual gestures affect sign language intelligibility, although the extent of this is unknown. Our research produced a reduced 3D humanoid SLV system based on *Man* that can be rendered on mobile phones as a 3D model. Our Avatar is called *Phlank*.

It was not known whether sign language produced by these systems (and SLVs in general) and rendered on mobile phones was intelligible to deaf viewers. We also wished to compare the intelligibility of the two SLVs as part of a feasibility comparison. *Man*, while complete, cannot be rendered on a mobile phone as a 3D model and therefore lacks flexibility arising from this. *Phlank*, on the other hand, can be rendered on a mobile phone but with less detail. This paper describes the experiment we devised to, primarily, determine whether sign language produced by SLVs is intelligible to deaf people when rendered on mobile phones, and secondly, to compare different methods of doing so.

II. RELATED WORK

Very limited work has been done to assess the intelligibility of synthetic sign language. Furthermore, to our knowledge, no work whatsoever has been done to assess the

intelligibility of synthetic sign language on mobile phones. In this respect, our work is a pioneer study.

To our knowledge, the only work that has been done to assess the intelligibility of synthetic sign language has been done by Cox et al. of the ViSiCAST project [9]. The ViSiCAST project produced a 3D humanoid sign language visualization system that was part of an English to British Sign Language Machine Translation system to be implemented in post offices in the UK. Their 3D Avatar was called *Tessa*.

They conducted a study to assess the intelligibility of British Sign Language (BSL) signs rendered using *Tessa*. *Tessa* was capable of performing 133 BSL phrases made up of 444 BSL words (with repetitions). They conducted two sets of experiments, one of which was an objective evaluation of the intelligibility of the sign language displayed, and the other, a subjective evaluation of the partiality of the deaf towards their Avatar. We limit our discussion to the objective evaluation they carried out, since this is applicable to our work.

Six profoundly deaf people whose first language was BSL took part in the experimentation. Each test subject was shown all 133 phrases in blocks of between 20 and 24 phrases per viewing [9]. Each time, the subject was instructed to write down what they had understood from the viewing or to mark it as unknown. The subject was allowed to replay phrases up to a maximum of 5 times after which the phrase was marked as unknown.

Thereafter, those phrases which the subject had identified incorrectly were re-presented to the subject along with the text of the phrase. They were asked to state whether their error was due to the sign language being inappropriate, such as in a different dialect, or just not clear.

The results were evaluated, first, by determining the percentage of signs recognized correctly in each phrase. This provided a recognition rate per phrase. The percentage of correctly identified signs across all phrases was also determined for each sign.

On average, 61% of complete phrases were identified correctly, ranging from 42% to 70%. On average, 81% of signs were identified correctly, ranging from 67% to 89%. They also found that subjects required, on average, 1.8 viewings before making an attempt at identification. Analyzing the causes of incorrectly identified signs revealed that 30% of these were as a result of inappropriateness of the sign. Cox et al. remark in [9] that variations in BSL dialect played a significant role in this figure. They also remark that variations in sign language dialect pose a great challenge to implementation of sign language visualization systems. The remaining 70% of errors were due to unclear signing.

A greater number of studies have focused on assessing the effects of various compression parameters and techniques on the intelligibility of sign language video. This research is not applicable to our research but we refer the reader to [10], [11] and [12] for information on this subject.

III. EXPERIMENTAL PROCEDURE

Our experiment was aimed at determining the intelligibility of synthetic sign language on mobile phones as

well as comparing four methods of sign language visualization (SLV) on mobile phones.

The first method was the use of our low-detail mobile Avatar *Phlank* depicted in figure 1a. *Phlank* only has a set of arms, hands and a head. It is not capable of performing any non-manual gestures. It was built as the optimum model that could be rendered as a 3D model on the mobile phone we used, the Sony Ericsson C905.

The second and third methods made use of the Avatar *Man*. We have mentioned before that *Man* is capable of full-body animation, including all non-manual gestures. We wished to investigate the importance of non-manual gestures in SLV. Therefore, the only difference between the second and third methods is that they, respectively, did and did not incorporate non-manual gestures. Figures 1b and 1c depict the word “Stomach pain” as rendered by the second and third methods. The final method was used as a comparative base and was the display of sign language videos of SASL speakers. Figure 1d is a depiction of this method.

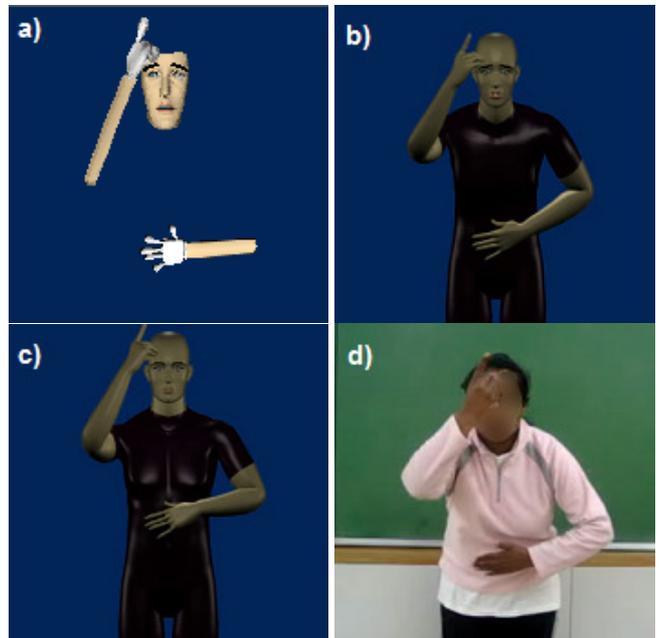


Figure 1: Sign language visualization using a) the low-detail avatar *Phlank* b) the high detail Avatar *Man* with non-manual gestures c) the Avatar *Man* without non-manual gestures and d) a sign language video.

A. Collection of Signs and Animation of Avatars

In animating our Avatars, we selected 16 short commonplace phrases from a phrase book, most of which consisted of single words. The SASL signs of these phrases were collected by enlisting the help of the Dominican School for the Deaf in Wynberg, Cape Town. Five different profoundly deaf students whose first language was SASL were instructed to sign each of the 16 signs under the supervision of a hearing teacher that had expert knowledge of SASL. Each of these signs was recorded and labeled as sign 1 through 16 as seen in table I.

The Avatars *Phlank* and *Man* were then animated to perform each of those 16 signs by means of keyframing.

TABLE I
LIST OF SASL WORDS RECORDED AND USED

No	Phrase	No	Phrase
1	Bus	9	Medicine
2	Doctor	10	Restaurant
3	Good Evening	11	Right
4	Hello	12	Sick
5	Help Me	13	Soccer
6	Help You	14	South Africa
7	How	15	Toilet
8	Left	16	Water

Keyframing, as applied to sign language animation, is a manual animation technique that involves identifying important frames in the video sequence at which the hands undergo significant changes in speed and direction of motion. These frames are known as keyframes.

The Avatars were animated in Blender for methods 1, 2 and 3 by identifying keyframes in each sign language video and creating the poses for each keyframe with attention to detail. Each keyframe was then placed in the correct timing order. This method of posing followed by timing was in accordance with the research findings of Terra et al. [13] in keyframing. Blender then interpolated between keyframes. The non-manual gestures for method 3 were also animated using keyframing in a similar manner.

The animations for method 1 were exported as M3G files which were subsequently imported into the Mobile 3D Graphics API in Java ME on the mobile phone. They were then rendered as a 3D model.

The animations of methods 2 and 3 and the videos of method 4 were exported to the MPEG-4 Part 14 (MP4) format and imported into the Mobile Media API of Java ME and played as a video file.

B. Viewing Sequence

It was decided to use 16 test subjects such that showing each test subject all 16 signs would yield an equal overall number of viewings per sign and a total of 256 viewings.

Each group of 4 signs were shown using one of the 4 methods for each test subject. Each test subject would therefore view 4 signs in each of the 4 methods in randomized order. Table II shows 4 viewing groups constructed such that each of the words in the ‘word’ array was assigned to each of the methods in the corresponding ‘method’ array. This determined the word-method combinations for each group and ensured that each word-method would be viewed exactly 4 times by 4 different subjects. The viewing order was then randomized for each subject to obtain 16 distinct viewing sequences.

C. Experimental Setup

As mentioned previously, Cox et al. [9] found that variations in sign language dialect posed a serious challenge to the implementation of a sign language visualization system. For this reason, we decided to conduct our experim-

TABLE II
ARRANGEMENT OF METHOD-WORD VIEWING GROUPS

Subjects	Arrangement
1,5,9,13	Word: {1,2,3,4,5,6,7,8,9,10,...,14,15,16} Method: {1,2,3,4,1,2,3,4,1, 2 ,..., 2 , 3 , 4 }
2,6,10,14	Word: {1,2,3,4,5,6,7,8,9,10,...,14,15,16} Method: {2,3,4,1,2,3,4,1,2, 3 ,..., 3 , 4 , 1 }
3,7,11,15	Word: {1,2,3,4,5,6,7,8,9,10,...,14,15,16} Method: {3,4,1,2,3,4,1,2,3, 4 ,..., 4 , 1 , 2 }
4,8,12,16	Word: {1,2,3,4,5,6,7,8,9,10,...,14,15,16} Method: {4,1,2,3,4,1,2,3,4, 1 ,..., 1 , 2 , 3 }

ent in the same school from which we collected our signs. This would eliminate the effects of variations in dialect as far as possible.

The 16 test subjects consisted of 9 students, 3 teachers and 4 staff members. All 9 students were profoundly deaf and used SASL as their first language. 5 of them were females and 4 males. 2 were Grade 9 students, 2, Grade 8, 1, Grade 7 and Grade 6 each and 3, Grade 5. The teachers were all hearing but had expert knowledge in SASL. 2 were female and 1 male. Of the staff members, all were female. 3 were profoundly deaf and used SASL as their first language and 1 was hearing but had extensive experience and knowledge in SASL.

In constructing our experiment, we adapted the experimental model of the ViSiCAST project [9]. One of the teachers was requested to explain the experimental procedure to the students. Each subject was shown the viewing sequence explained in a previous section. Each of the words in the sequence was shown without context. For each viewing, the subject was instructed to write down the phrase they had seen in each of 16 rows in an answer form provided, or to write “I don’t know” in the same row. The subject was allowed to replay the sign up to a maximum of 10 times, after which the phrase was marked as unknown.

Those signs that had been identified incorrectly were then re-presented to the subject and the subject was informed of the meaning that we had assumed. He/she was then asked whether this discrepancy was due to inappropriateness of the sign such as a varied dialect or whether it was due to unclearness of the sign on the mobile phone screen.

IV. RESULTS AND DISCUSSION

It was found that 168 viewings, about 65% of the total, of were identified correctly. On average, 1.8 viewings were required before an attempt at recognition was made. 181 attempts, about 70% of the total, were made after 1 viewing and 58 attempts, about 22% of the total, after 2 viewings. This seems to suggest that the majority of recognition attempts were made with conviction. Only 1 attempt was made after more than 5 viewings. It was made after 6 viewings.

Figure 2 shows the percentage of correctly identified phrases per person. The number of correctly identified signs ranged between 13, about 81% of the total, and 6, about 37% of the total.

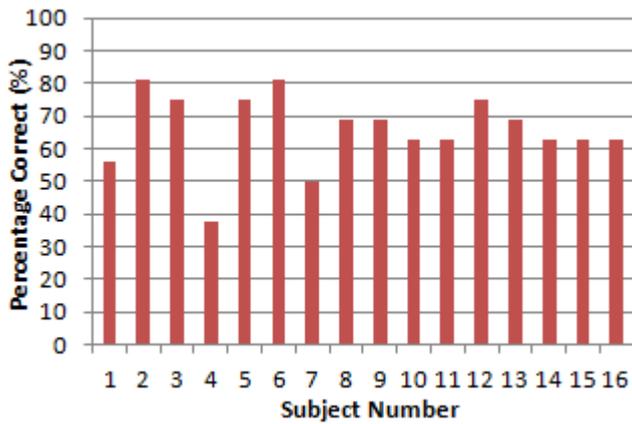


Figure 2: Percentage of correct answers per subject.

This suggests that a wide range of subject types were spanned. Henceforth, we refer to method 1 that used the *Phlank* Avatar as ‘LowRes’, methods 2 and 3 that used the *Man* Avatar with and without non-manual gestures as ‘Facial’ and ‘NoFacial’, respectively, and method 4 that used sign language videos as ‘SLVid’.

Table III summarizes the number and percentage of correctly identified signs per method and this data is depicted in figure 3. Surprisingly, the sign language videos (SLVid) were not 100% recognizable. The SLVid and Facial methods had comparable correctness percentages of about 81% and 73% respectively. The two remaining methods had much lower and comparable accuracies of around 50%.

At first glance, it appeared that non-manual gestures played a major role in sign language intelligibility. However, we found that the apparently huge difference of about 20% between Facial and the two methods without non-manual gestures LowRes and NoFacial was composed of a difference of about 13 viewings. It was found that the latter two methods fell short in a few specific words rather

TABLE III
CORRECTLY IDENTIFIED SIGNS PER METHOD

Method	Number Correct	Percentage Correct (%)
SLVid	52	81
Facial	47	73
NoFacial	34	53
LowRes	35	54

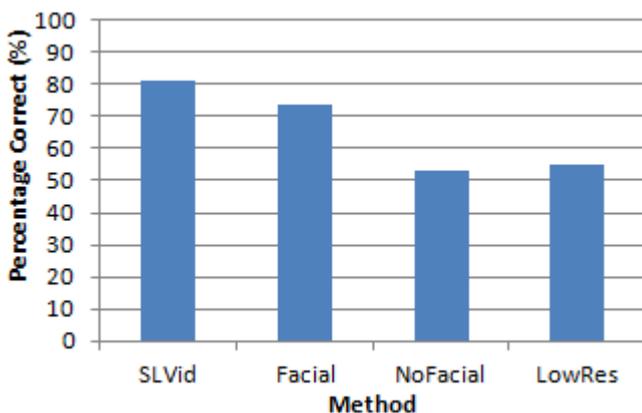


Figure 3: Percentage of correct answers per method.

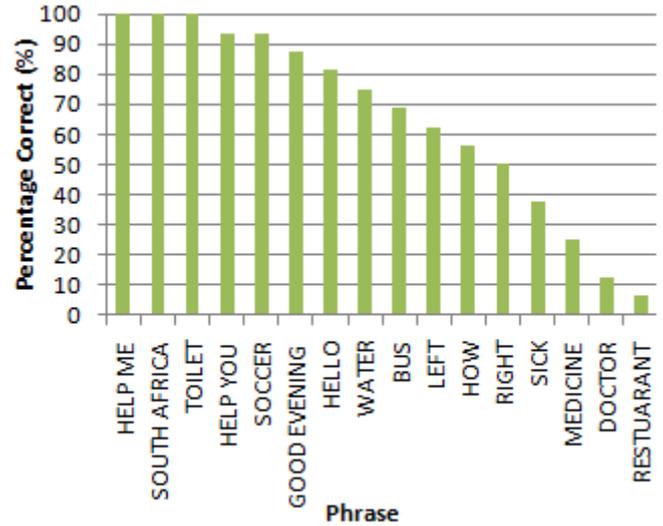


Figure 4: Percentage of correct identifications per phrase.

than in general, ‘How’ and ‘Right’ being the most prominent sources of such errors. This suggests that non-manual gestures may play a significant role only in specific phrases.

Even though the percentage of correct signs per method, per subject and overall are very encouraging and suggest that we can in fact display intelligible SASL on mobile phones, we analyze the reasons for incorrect identifications. Figure 4 depicts the number distribution of the two reasons given for incorrect identifications per word. Those words that achieved 100% correct recognition have been omitted.

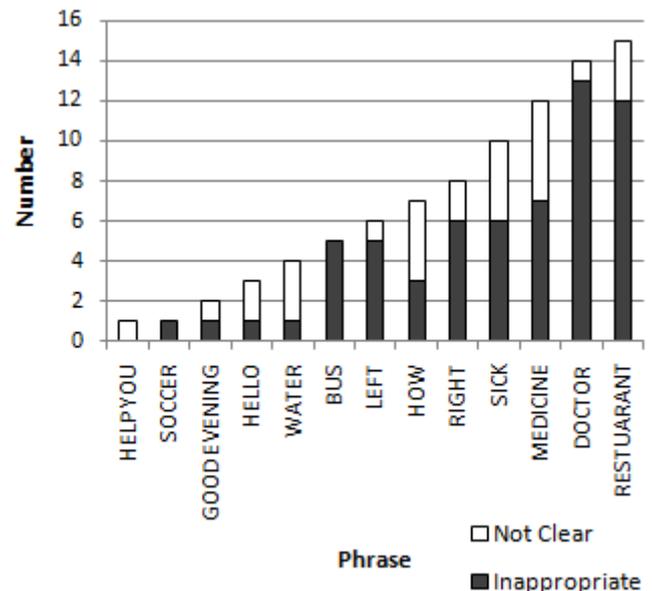


Figure 5: Number distribution of reasons given for incorrect identifications per word.

It is very evident that the biggest factor causing recognition errors was the inappropriateness of the sign – variations in dialect – and not unclearness arising from the mobile phone display. This was, in fact, what we found as we conducted the experiment. Comments such as “I don’t use that sign”, “I do that sign like this...” and “That sign is mostly used by this other group” were oft repeated. Subject 4, that had the lowest number of recognitions, revealed that she had only recently joined that school and the dialect she used in many ways differed from the one used in the school.

We also learned from teachers that students had their own sub-dialect that had non-standard variations of certain signs.

We have stated our recognition rates as is. For investigative purposes we now take out all samples that registered recognition errors solely arising from dialect issues. We find that the average recognition rate increases to 86%. Figure 6 depicts the updated recognition rates for each of the methods.

TABLE IV
RECOGNITION DISTRIBUTION OF IDENTIFIED SIGNS PER METHOD AFTER REMOVING DIALECT-AFFECTED SAMPLES

Method	Correct	Incorrect	Percentage Correct (%)
SLVid	52	1	98
Facial	47	5	90
NoFacial	34	8	81
LowRes	35	13	73

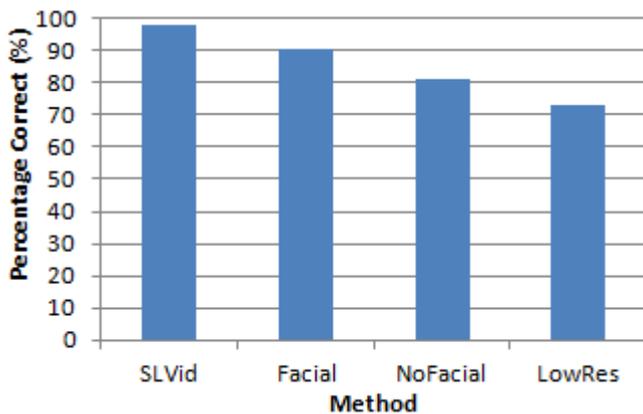


Figure 6: Percentage of correct identifications per method after removing dialect-affected samples.

The recognition rate of SLVid increases to very close to 100% which is what we expect if there is no unclearness in the signs resulting from the display of the mobile phone. All other methods increase in accuracy drastically as well. It is found that the recognition rate decreases approximately proportionally according to the detail of the Avatar, which is also what is intuitively expected.

V. CONCLUSION

In this paper we explained the experiment we conducted to, first, determine whether or not it is possible to display intelligible synthetic sign language on mobile phones and, second, to compare the intelligibility of four methods of sign language visualization on mobile phones.

We found that it was, in fact, possible to display intelligible synthetic sign language on mobile phones. We achieved a recognition rate of about 65%. We showed that the 35% of recognition errors were predominantly caused by variations in dialect between SASL speakers, despite efforts to rule this factor out, and not because of the unclearness of the display. Having removed the samples that were identified incorrectly solely due to dialect issues, we showed that the average recognition rate increased to 86%.

We also showed that, with dialect issues present, the use of the SASL project's *Man* Avatar with non-manual gestures achieved comparable results to sign language video, with average recognition rates of 73% and 81% respectively. In the same circumstances, the same Avatar without non-manual gestures achieved very comparable results to the low-detail Avatar *Phlank*, with average recognition rates of 53% and 54% respectively.

Having removed the samples identified incorrectly due to dialect issues, we showed that the recognition rates of all methods increased to 98%, 90%, 81% and 73% for sign language video, *Man* with and without non-manual gestures and *Phlank*, respectively.

It also appeared, before the removal of samples with dialect issues, that non-manual gestures have a very significant effect on sign language intelligibility. However, we showed that, with the removal of dialect issues, non-manual gestures do not have as significant an effect in the intelligibility of individual signs except in specific signs and contexts.

ACKNOWLEDGEMENT

We would like to thank the Dominican School for the Deaf in Wynberg, Cape Town for facilitating our experiments and their cooperation in general. This work was funded by Telkom/Cisco/THRIP.

REFERENCES

- [1] M. Glaser and W.D. Tucker. "Telecommunications bridging between Deaf and hearing users in South Africa". *Proc. CVHI 2004*.
- [2] W.C. Stokoe. "Sign language structure: an outline of the visual communication systems of the American deaf. 1960". *J Deaf Stud Deaf Educ*, 10(1):3-37, 2005.
- [3] R.G. Gordon and B.F. Grimes. "*Ethnologue: Languages of the world*". SIL International Dallas, TX, 2005.
- [4] S. Foster. "Communication as social engagement: implications for interactions between deaf and hearing persons". *Scandinavian Audiology*, 27(4):116-124, 1998.
- [5] E. Dolnick. "Deafness as culture". *Atlantic Monthly*, 272(3):37-53, 1993.
- [6] F.G. Bowe. "Deaf and hard of hearing Americans' instant messaging and e-mail use: A national survey". *American Annals of the Deaf*, 147(4):6-10, 2002.
- [7] D. van Wyk. "Virtual human modelling and animation for sign language visualisation". Master's thesis, University of the Western Cape, 2008.
- [8] SASL Project. "Integration of Signed and Verbal Communication: South African Sign Language Recognition and Animation". [Online] Available at <http://www.coe.uwc.ac.za/index.php/SASL.html>, [Accessed: April 2010].
- [9] S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, and S. Abbott. "TESSA, a system to aid communication with deaf people". In *Proceedings of the fifth international ACM conference on Assistive technologies*, page 212. ACM, 2002.
- [10] A. Cavender, R.E. Ladner, and E.A. Riskin. "MobileASL:: intelligibility of sign language video as

constrained by mobile phone technology”. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, page 78. ACM, 2006.

- [11] F. Ciaramello, A. Cavender, S. Hemami, E. Riskin, and R. Ladner. “Predicting intelligibility of compressed American Sign Language video with objective quality metrics”. In *2006 International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. Citeseer, 2006.
- [12] M. Pavel, G. Sperling, T. Riedl, and A. Vanderbeek. “Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language”. *J. Opt. Soc. Am. A*, 4(12):2355–2365, 1987.
- [13] S.C.L. Terra and R.A. Metoyer. “Performance timing for keyframe animation”. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 253–258. Eurographics Association, 2004.

Mehrdad Ghaziasgar received his BSc in Physics and Computer Science in 2007 from the University of the Western Cape Summa Cum Laude and his BSc (Honours) in Computer Science from the same institution Cum Laude in 2008. He is presently completing his MSc at the same institution. His research focuses on sign language synthesis and mobile-based communication applications for the South African deaf.

James Connan heads up the SASL project at the University of the Western Cape. He has a wide range of research interests that include computer vision, machine learning and machine translation between South African sign language and English.