Project proposal:

Open mobile video telephony for the Deaf in South Africa

By: Chris Laidler
LDLCHR004

1. Contents

2. Project Description

Introduction

Many members of the South Africa the Deaf community do not only suffer directly from their physical disability, but suffer from a number of related socio-economical complications. The biggest of these being: illiteracy and poverty, these stem from a legacy of discrimination and poor educational structures tailored for the Deaf in South Africa [10].

What may people do not realize, is that sign language (SL) is not a spatial form of another language such as English. It is a distinct language of its own, with its own structure and grammar. Not only is it distinct but there are many forms of sign language, South African Sign Language (SASL) being the primary language used by the Deaf in South Africa. SASL is the first language of most of the South African Deaf community. There have been a number of attempts to create a written form of sign language; these include HamNoSys, a phonetic transcription tool developed in 1984. The second is SignWriting, developed by Valerie Sutton in 1974, which is a graphical bi-dimensional representation using symbols [6]. The major disadvantage of theses is the many varying forms of SL, thus they have not been widely adopted. Thus any written communication between the Deaf is often not in their first language.

Information and Communications Technology (ICT) is a field that has revolutionized the way most of us communicate, with a large drive towards mobile hand held devices. Much of the ICT development is based on and was developed for audio communication. This naturally precludes the Deaf. The other alternative is text communication using ICT but as the previous section highlighted this is not suitable for the Deaf either as it forces them to communicate in a second or third language, not only that but it is often considered impersonal and slow [3]. This means many of the advances in ICT have not been available to the Deaf.

With the development of camera enabled smart phones, video communication is now a real alternative. This opens up a new communication channel for the Deaf, allowing them to benefit from new ICT developments. This highlights the need for a mobile visual form of communication tailored for the Deaf.

This need has led to the formation of a project, funded by a Dutch funding agency (SANPAD). The project aims to provide Deaf users with a practical way of communicating in their own language, SASL, and at the same time highlight policy impediments to the widespread adoption of such a solution.

The ultimate goal is produce real time video communication between Deaf users. Our project aims to lay down some of the ground work required for the realization of the goals of the greater project.

Problem statement

There has been a substantial body of work, devoted to the needs of sign language video. Recently much of this has been focused on the requirements and associated difficulties of using mobile devices to facility sign language video.

Firstly mobile devices are a naturally restrictive environment. The three major factors that affect the project are; the limited computational power, limited bandwidth available for data transfer and the power requirements and limited battery life of mobile devises.

Sign language video has a number of unique requirements, the major factors are that the video has to be of high enough quality that the sign language is intelligible, and that the video stream is continuous and uninterrupted. Video jitter has a substantial negative affect on the intelligibility and usefulness of sign language video.

These two aspects the available technology and the communication requirements are opposed. Communication requires high quality video and the technology limits achievable quality, this leads to the need for our work.

There has been much work done in the field to determine optimizations that can be made to video and the way it is encoded to maximize intelligibility. Limited computational power means that the encoding of the video needs to be efficient, with minimal computation. The limited available bandwidth means that video will have a limited bit rate. Much of the resulting work has focused on region of interest (ROI) encoding. Most of the segmentation for these ROI encoders is based on skin detection. Unfortunately may methods of skin detection have poor results when considering dark skin colors. Meaning these ROI encoders will have poor results in our South African context.

Research question

The first objective of this project is to **investigate optimizations that can be made to the encoding of sign linguae video, in a rate constrained environment. Particularly grey scale encoding and its effects on intelligibility and bit rate.**

The second objective is to **implement a region of interest (ROI) encoder, based on face detection, and investigate differing levels of ROI encoding and their effects on intelligibility and bit rate for this encoder.**

One of the goals of the greater project is that this system utilizes open software and system to achieve its goals.

3. Background

The Local Deaf community

This project is run as part of a greater project; this project has been working in conjunction with *Deaf Community of Cape Town* (DCCT) a nongovernmental organization (NGO) based at the Bastion of the Deaf in Newlands Cape Town

[10]. DCCT serves a disadvantaged Deaf community of approximately a thousand people [10]. The community members are united by their disability, there are a number of socio-economic factors that disadvantage this community. Their disability often leads to a lack education and unemployment and many of them come from previously disadvantaged race groups, which all lead to cycle of poverty in the community. One of the unifying features the community is that they communicate in SASL and that this is their first language [10]. Many members of the community have poor levels of spoken, written and reading literacy in any other language [10]. This is a common theme in most of the literature on the Deaf, their first language is a form of SL and understandably this is their preferably form of communication [10] [3]

Deaf communication

There have been a number technological developments to facilitate Deaf communication. Many of these are based on the concept of a relay, this is essentially an intermediary that translates a message to another form and then passes it along. This can be a computer or as is often the case with a deaf communication, a person. Most of these relays are centered on communication between the hearing and the Deaf.

There have been numerous forms text telephony (TTY) around the world; these include the *Teldem* in South Africa [10] and the more widely used short message service (SMS) for mobile devices. There have been a number of text to audio relays, allowing the Deaf to communicate to the hearing. Text to audio relays are ether automatic or manual, but they have mostly been limited to the developed world and even here they are not widely available. In recent times there have been a number of these implemented for mobile devices, these include Vodaphone's service for the Nocia 9210i and Mobile TextPhone [10]. Due to the difficulty in interpreting spoken language most of these are still manual, which makes them slow and expensive. These relays do not allow the Deaf users to communicate with other Deaf users.

There are a number of international video relay services (VRS) as well, these are currently exclusively manual, but there are a number of projects that are attempting to automate this system, but again there is still no effective way to automatically interpreting SL, thus these relays are relatively expensive and thus exclusive.

There are a number of systems based on avatars, these allow users to enter to text that is converted into a structure that is displayed as a character signing, these include, Signeuse Virtuelle, eSighn VSigns and WebSign [6], but again these require the sender to enter their message as text.

The aim of this project is to allow the Deaf to communicate in SL. Recently there has been a big move towards using internet based video conferencing. Here many of the standard video conferencing tools are not suitable for SL communication, the reasons for this will be outlined in the next section. There are a number of SL orientated videoconferencing tools as well. Most of these rely on various forms of video or web cameras, a computer and a connection to the internet, all of these rely on set infrastructure [3]. There is currently no mobile option specifically for Deaf users in South Africa. Our aim is to investigate the feasibility of using hand held mobile devises to facilitate this type video communication in South Africa.

Needs and characteristics of SL video

The natural question that is asked is: why don't the Deaf just use existing video conferencing tools? The short answer is that the video quality just simply isn't high enough. Existing tools may have video but they are still predominantly focused on audio communication, thus the video quality simply isn't high enough to make sign language intelligible.

There has been much research in the field of video communication for the Deaf, and much of the current focus is on mobile video communication. Here specifically the constraints are low bitrates and the lack of computational power and the small nature of handled devices. Cavender *et al* [3] conducted a series of focus groups regarding the requirements of mobile video SL communication. Here is a brief summary of their findings:

- The camera and screen should face the same direction to facilitate two way communications.
- The mobile device should have a stand.
- There is scope for holding the device in one hand for short conversations.
- There is the possibility of using a separate external camera.
- The users want the ability to supplement the video message with text input.
- The device should have existing text features such as: e-mail, Instant Messaging and SMS.
- The device should have a "ring" functionality to indicate incoming video call.
- The user should have the ability to accept or reject a call.
- The system should have the ability to leave a video message, they dubbed this SignMail.
- Video messaging must be accusable from other technology such as standard computers.
- Privacy is not as big a concern as was expected this was attributed to the inherent public nature of a SL conversation.

As mentioned previously, when considering the mobile environment, there is a constraint on the available bandwidth if we consider synchronous video communication this will put an upper limit on possible video bitrate. Thus we are working in a rate constrained environment, and the objectives are to maximize video quality or more importantly intelligibility, or frame rate. Thus if an optimization is described as "decreasing bitrate" this would be considering a constant frame rate, but in the specific mobile environment, which has a set bit rate this will be realized as an increase in achievable video quality or frame rate.

Intelligibility tests

Prof. Judy Harkins of Gallaudet University's Department of Communication Studies, was consulted with regard to testing the intelligibility of sign language video. She has worked in the area of intelligibility of sign language video. She highlighted a number of key features required for intelligibility testing of sign language video. Firstly there is no standardized test for SL intelligibility that Prof Harkins knows of. Here work in the field has highlighted a number of key features that are needed in test videos. These are broad hand gestures that encompass the full signing space, proper names that require finger spelling, and a natural rate of signing.

There are two subjective metrics that have been adopted to evaluate intelligibility of SL. The first is a direct intelligibility test in which comprehension is directly tested by a written questionnaire. The second is an opinion test in which users are asked to evaluate their perception of the intelligibility 5 point Mean opinion score (MOS) scale. The mean result is taken to be a numeric measure of perceived intelligibility. MOS are widely used in multimedia especially in audio and voice telephony, where there are standards specified by International Telecommunication Union (ITU) in Recommendation P.800, but to the best of our knowledge no standards exist for SL communication. Prof. Harkins, advises that MOS be used to rate intelligibility. This is supported by the work of Nakazono *et al* [9] who did some of the initial work on ROI encoding. They found that the results of the two methods; comprehension testing and MOS are highly correlated, and the second is prefer as it is less labor-intensive, and does not rely on the writing capability of the participants.

4. Related Work

Eye tracking

One of the major themes that much of the sign language specific research is based on is that much of the content and meaning of SL is conveyed by the face and fine facial gestures. Facial expressions have a big impact on the context of what is being said. This was popularized by the work of Muir *et al* and Agrafiotis *et al* [8] [1] [2]. They performed an eye tracking studies in which subjects took part in a number of experiments. In these experiments subject watched video narratives in BSL. They used the Eyelink eye tracking system to record the gaze points of the participants. Results of their experiments show that experienced signers focused on the face of the signer, especially the mouth, whereas inexperienced signers and no signers often looked at the hands [2][1]. The reason for this is that much of the contextual information obtained from the hands and arms is relatively easily picked up in peripheral vision. While much of the fine contextual information is conveyed through subtle facial expressions [3]. The fact that for sign language most if not all the attention of the viewer is focused on the face forms the bases for most of the work on codec optimization that will be discussed.

Foveated and Region of interest coding

With their eye tracking results Agrafiotis *et al* then went on to propose the use of foveated video coding to reduce overall bitrates [2]. They had established that the viewer focused on the face and more specifically the mouth of the signer. Foveated compression aims to exploit the falloff is spatial resolution of the human visual system away from the point of fixation. They developed an algorithm to separate the image into 8 areas around the fixation point of the viewer. Each area is defined at the macroblock level and each area has a similar maximum visually detectable spatial frequency. These are roughly concentric rings around focal point. They proposed implementing a variable quality coding process by specifying the QP of each MB depending on the foveaion area in which it fell within. Thus manually forcing an affective high bitrates around the face, and thus improving the video quality around the face. The idea of adding weight to areas of the image later became known as Region of interest (ROI) encoding, we use this term to describe it through the rest of this document. They proposed two methods for locating the Fixation point, firstly using the assumption that the signer would be roughly in the centre of the frame they proposed simply fixing the focal point in the centre. The second method they proposed was to use face tracking to find the face and thus the focal point. They performed a test in which they did face tracing and ROI encoding. They reported decreased a bitrate without substantial loss of intelligibility [2] although they gave no substantial details of their experimental method or results.

The next step in the world of encoding for SL was work done by Nakazono *et al* [9]. Their paper gives a good outline of the work up to that point. They noted a problem with simply changing the QP of a MB while encoding. If a target bit rate is set, and QP's are changed on the fly the remaining MB will simply consume the freed bits, thus these bits will not be available for later MB's with high QP's, the QPs need to be evaluated for the whole frame and then weighted accordingly, this ensures that the target bitrate will be maintained. They proposed an ordering in which the MBs be encoded to solve this problem. They specified that first an order be set for GOB's and then an ordering for MB in those GOB's. The ordering of the GOBs was centred in the centre of the ROI, thus if the allocated bits were exhausted the areas of highest significance would have been encoded. They also maid and important distinction that the background in signed environment may be busy, this would consume a large number of bits in the traditional video encoding system. They proposed defining a RIO around the signer and then marking blocks not in this region not to be 'not coded'. Essentially dropping the bits allocated to these regions to almost zero thus drastically reducing the required bits, even in a location with a busy background. They did note that setting MB close to the point of focus to be 'not coded', "gave a strange feeling" to the viewer so the suggested only using it on the extremities. They implemented all their optimizations in the H.263 encoding and performed fairly rigors experimentation, using subjective MOS to evaluate their results. They showed that at set low bitrates their optimized encoding far outperformed the base encoding.

MobileASL

From this point most of the dominant work in the field of encoding SL video was done by the MobileASL group, headed by Prof. R. Ladner of the University of Washington.

Cavender *et al* [3] did some investigations into the interplay between different ROI encodings FPS and bitrates. Through preliminary studies they found that at framer rate of 5fps finger spelling was difficult to interpret, while the difference between 15 and 30 fps was negligible . Thus they decided to use 10 and 15 as their experimental value, evaluating whether fewer better quality frames are better than more, worse quality frames. They decide to test bitrates of 15, 20 and 25kbps. They used a ROI encoding simply using a square around the face of the signer, and varying the QP in this area between three values which we will refer to here as high medium and low, where high refers to a higher quality difference between the ROI and the rest of the frame. They customized the x264 codec to encode their video and tested its intelligibility on a smart phone. They evaluated their results using a subjective MOS questionnaire. They had some interesting results, showing that unsurprisingly the users prefer higher bitrates across the board, but that they prefer lower frame rates preferring 10fps to 15fps, which allows fewer high detail frames, and a medium ROI, which gives sufficient detail to the face while at the same time not obscuring the hands [3].

The work of *Ciaramello et al* [4] aimed to developing an objective metric for measuring the intelligibility of encoded SL video. They developed a metric to assess the intelligibility of SL video, their metric is based on the Mean square error (MSE) in the hands and face. They found their results were highly correlated with the results obtained in the work of Cavender *et al* [3]. They concluded that their objective measure for the intelligibility was an effective one.

The work presented in the three papers [5] [12] [11] all deal with varying the encoding parameters of and adapted x264 codec to optimize it for encoding SL Video in the most intelligible way. They predominately gained results by optimizing the values of QF and other parameter for given MB's. The problem was that the process was prepossessed intensive and they wish to run it in real time on a mobile device. They solved this problem by realizing that the results they were getting were falling in a predictable manner. They thus prospered a solution by generating the results on a powerful online server and then storing them as a lookup table on the local device, thus cutting down on computation time and still gaining coding optimization.

The MobileASL project has made much some significant findings and developments in the field. They are a big well-funded long term project, with the goal of real time video communication at very low bitrates. They may achieve their goals but I believe that by the time they do, the constraints of the bitrates will be irrelevant as the mobile infrastructure will have developed well beyond what it was when the project was started, although this does not decrease the value of their findings.

Local work

There was a paper produced by Ma *et al* [7] which delta with the optimization of the x264 codec for asynchronous communication for the deaf. This paper was tested in the community for which this project is intended. The research question was a valid one that is very applicable to our project. A number of metric were used, these were a number of industry standards such as the MSE of PSNR, the SSIM index and the VQM. A number of custom metrics were used as well, these include compression rate (CR), Compression time (CT), transmission time (TT) and delay time (DT).

5. Procedures and Methods

Overview

The main objective is to maximize intelligibility of sign language video in a bitrate constrained environ.

One optimization we believe that has not been fully explored is the color space of the sign video. We believe that using gray scale video will not significantly decrease intelligibility. While the rate saving that can be made using a reduced color space can be used to increase the quality of the video. Our belief is that the increase in quality will outweigh any detail lost due to the decreased color space. We will test this experimentally.

Much of the current work on intelligibility optimized sign language video, focuses on region of interest (ROI) encoding; see the section on related work for a full description. The local team working on sign language video, has not implemented a ROI encoder, as yet. We plan to implement a ROI encoder, that can be used in the greater project, for this and future work. We will test the affectedness of the codec experimentally.

As outlined earlier most of the existing ROI encoding systems user skin detection to segment the video, and this will not be effective in our system. We will instead investigate the possibility of using face detection to segment the video for ROI encoding.

Ok Hung will ummmmm I have no idea!!!!! Investigate other optimizations to x264 codec and/or implement it on the HTC?

Perform objective tests on the encoding system.

We will conduct a user study to determine the effectiveness of our encoding system. This user study will take the form of a subjective intelligibility test using MOS questioners.

Encoding system

Codec

From our background research we have selected open source x264 video codec as the codec we will use in our encoding system. The latest version of the x264 codec and any required libraries and related tools will be obtained and compiled to a binary.

Gray scale color space

We will investigate any rate gains that can be obtained from the selection of the output color space. This will be achieved by testing various combinations of encoding parameters of the base x264 codec. Our initial belief is that the use of a gray scale color space will not degrade intelligibility but will decrease rate used to encode the video. This additional rate can be utilized to increase either frame rate or quality, thus increasing intelligibility.

ROI codec

We will modify the x264 codec to implement a ROI encoding scheme. We believe this can be done by my manually weighting the QP values of the macroblocks.. Thus using the existing block based structure of the H264 standard as the base unit to be used as the regions of the ROI codec. The weighting of the QP variable of these macroblocks will effectively assign more or less rate to the various macro blocks. In the initial stages, before the face based segmentation is implemented, we will simply use a static segmentation, assuming the face of the signer will be roughly in the center of the frame. Thus we will use a basic square or circular region for the ROI codec.

Segmentation detection

We will develop a segmentation module based on face detection. The input video will then be analyzed by a face detection module. We do not have the time resources to create our own face detection system. Thus we will utilize a standardized open source face detection library. Facelib and the openCV media library, are potential candidates. We will investigate whether we can implement them in the context of our project. The results of the face detection will then be used to segment and weight the video frames, using the base 16x16 pixel macroblocks as the segmentation unit. The segmentation will be using the facial location and an appropriate geometric interpretation of where the signer's torso and hands may be. This segmentation will then be used instead of the previous static segmentation, implemented in the modified x264 codec.

Hung's optimizations to the codec

Details of what Hung is hoping to do. Some form of optimization to the x264 codec and/or implementation on the phone?

Overall system

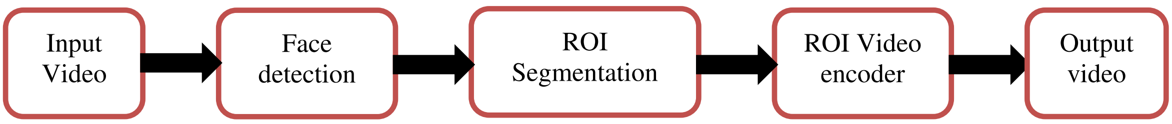

Figure 1

The sections outlined above will be combined to create the encoding system. This system is depicted by Figure 1. The input will be the base test videos crated for the intelligibility test. These will then be analyzed by a face detection module. Each frame will then be segmented, using the facial location. This segmentation will then be used by the ROI codec. The ROI codec will be used, with the various desired combinations if settings, to encode the video. The resultant video will be ROI encoded video that can be evaluated for objective and subjective intelligibility.

Experimentation

Test video segments

For the intelligibility experiment test video segments will be required. These video segments will be encoded with various settings, to create the experimental videos.

To reduce experimental error these test video segments will need to be of similar length and quality. Thus they will need to contain similar content.

For these videos, a two way conversation will be posed, with one person asking demographic questions and prompting the other person to answer. The answers will be in the form of a short story or description. These responses should make sense in a standalone context. The questions will be chosen to encourage answers that will incorporate the key features of SL; board signs and finger spelling where applicable. The "conversational" structure is aimed at regulating signing speed. The 'answering' member of the conversation will be filmed. A number of short extracts will be taken from their replies or stories. These will then be rated by fluent signers and an appropriate selection taken; these will be the base test videos.

The participants used to produce these test videos will be DCCT staff, or paid members of the DDC community, sourced with the help DCCT staff. A sign language interpreter will be needed when the video is recorded and segments rated for similar intelligibility. The participant shown in the video will be asked to sign a waiver.

Intelligibility experiments

For this project the results of the encoding system will be tested subjectively with a user intelligibility experiment. This intelligibility test will be performed by approximately six selected members of the community severed by the DCCT and with the help of the DCCT staff. The experiments will be performed at the Bastion of the.

Participants will be seated in a room with a sign language interpreter and an experiment facilitator who will assist them and answer and questions they may have. They will be given a brief description of the experiment and what is required of them, this will be as both a written document and an interpreted version read by the facilitator from a script. It will be emphases that the experiment is voluntary and that participants may stop at any point. They will then be asked to fill out a consent form.

They will then be asked to fill out, with the assistance of the interpreter if required, a brief demographic questionnaire. Answering this demographic questionnaire will be voluntary, if the participant ids uncomfortable answering any question they will be permitted to omit it.

They will then be shown two demonstration videos. These two test videos will serve to familiarize the participants with the format of the experiment, and will at the same time give them reference points. As the two test videos will be examples of the best and worst quality videos that will be shown to them during the rest of the experiment.

The participant will then be asked to view each test video, after each video the participant will be asked to rate the intelligibility of that video using a 5poinbt MOS questionnaire.

The videos will be sown on a HTC legend mobile phone. Participants will be asked to ether hold the phone or place it on a stand on a table, depending on their preference. It will be recorded on the questionnaire by the facilitator whether they held the phone or placed it on the table.

6. Ethical and Legal Issues

We will be conducting user experiments and thus we will need to get ethics clearance for these, we do not foresee any difficulties in this regard. The participants will be members of the community served by DCCT, and will be selected with the assistance of staff members from the DCCT.

This project will be released as an open project in the interest of the greater project which will utilize open systems and software.

7. Anticipated Outcomes

This project aims to lay groundwork to implement a local ROI encoder that can be utilized on a mobile platform.

Systems

This project will utilize open source software where ever possible.

The core sections will be written in C / c++. The final ROI encoder will where ever possible utilize existing open source components. These include facelib and OpenCV for face detection and the x264 codec to be modified as the ROI encoder.

Impact

Key success factors

The key factor of this project are an investigation into the effects of gray scale video, the effects of resolution and a face detection based ROI encoder.

This project will be considered successful if these three aspects are tested, and their effects on intelligibility and bit rate are subjectively and objectively established.

8. Project Plan

Risks

This project has a number associated risks factors, many of these are related to time, and what implementation can be achieved in the limited time.

We may be unable to implement a custom x264 encoder for the initial user tests. In this case we will use a pre-compiled version of the x264 codec, or any other codec or video editing tool that has the capability to encode the video with the desired test properties.

We may be unable to implement a standard face detection library. In the event that we are unable to effectively implement a programmatic means of face detection, we will manually locate facial positions in the required video frames.

We may be unable to customize the x264 encoder to act as a ROI encoder. In this case we will manually segment the video using postproduction video editing software. We then apply a blurring filter to various sections of the video, this will emulate ROI encoding.

Resources

- Mobile devices
  The department will be supplying us with phones to use in this project. These will be HTC legend running Android 2.1.

- Video recording
  We will have access to the department video camera, to be used in the recording of test video.

- Financing
  This project is funded by SANPAD and thus there will be money available for experimentation and other project related expenses.

- The deaf community
  We will have access to the Deaf community serviced by the NGO DCCT, and based at the Bastion for the Deaf, we will utilized these community members to do any user test or experiments we have to perform. Our liaison with this community will be Meryl Glazer. We will have access to a sign language interpreter at the community center.

- Development computers
  Chris has a PC and a laptop and access to the UCT computer science Honors lab.

- Development environment
  Development will be done in c / c++, there are a number of free IDE's for c++, and these include eclipse.

Deliverables and Milestones

Mon: 13/09/2010
Final proposal submitted.

Tue: 15/09/2010
Submit the questions to use in test video, to Meryl Glazer.

Wed: 15/09/2010
Record test video at the Bastion.

Fri: 17/09/2010
Have the test videos segmented into a number of short segments, suitable for encoding testing.
Application for ethics clearance submitted.
Base x264 codec compiled.

Mon: 20/09/2010
Color space encoding system finalized.

Fri: 22/09/2010
Rate the test videos segmented at the Bastion, to select suitable candidates to be used as the test videos.

Fri: 24/09/2010
First ROI encoder prototype complete.

Mon: 27/09/2010
ROI codec complete.
Draft of Experimental design chapter complete.

Thu: 30/10/2010
Face detection module complete

Fri: 01/10/2010
Encoding System complete

Sat: 02/10/2010
ROI test videos encoded.

Sun: 03/10/2010
Pilot intelligibility experiment.

Mon: 04/10/2010
Draft of Systems design and implementation chapter complete

Wed: 06/10/2010
Intelligibility experiments performed at the Bastion.

Mon: 11/10/2010
Report outline complete.

Mon: 18/10/2010
Report first draft complete.

Mon: 25/10/2010
Report final draft complete.

Thu: 28/10/2010
Report submitted for proof reading.

Mon: 01/11/2010
Final Report submitted.

Thu: 04/11/2010
Poster submitted.

Mon: 08/11/2010
Webpage complete.

Fri: 12/11/2010
Reflection paper submitted.

September

Mon	Tue	Wed	Thu	Fri	Sat	Sun
30 Design Chapter	31	1	2	3 Proposal	4 Proposal	5 Proposal
6 Proposal	7	8	9	10	11	12
13 Final Proposal	14	15 Record test video	16	17 x264 compiled	18	19
20 Encoder 1 DCCT UWC Students First Implementation Experiment	21 DCCT UWC Students	22 Evaluate video segments	23	24 Prototype ROI encoder	25	26
27 ROI codec	28	29 DCCT Workshop Final Prototype	30 Face detection	1 Encoding System	2	33 Pilot experiment

October

Mon	Tue	Wed	Thu	Fri	Sat	Sun
4 Implementation Testing Coding complete!	5	6 Intelligibility experiments	7	8	9	10
11 Report Outline Report: Outline	12	13	14	15	16	17
18 Report first Draft	19	20	21	22	23	24
25 Report final Draft Report: Final Draft	26	27	28 Report proof reading	29	30	31

November

Mon	Tue	Wed	Thu	Fri	Sat	Sun
1 Report: Final	2	3 Demo	4 Poster Demo	5 Demo	6	7
8 Web Page Open Day Demo	9	10	11	12 Reflection paper	13	14
15	16	17	18 Final Presentations	19 Final Presentations	20	21
22	23	24 External Examiner	25	26	27	28
29	30	1	2	3	4	5

9. References

[1] Agrafiotis, D. 2003. Perceptually optimised sign language video coding based on eye tracking analysis. *Electron. Lett.* 39, 24 (2003), 1703.

[2] Agrafiotis, D., Canagarajah, C. N., Bull, D. R., Kyle, J., Seersb, H. and Dyec, M. 2006. A perceptually optimised video coding system for sign language communication at low bit rates. *Signal Process Image Commun.* 21, (2006), 531-549.

[3] Cavender, A., Ladner, R. E. and Riskin, E. A. 2006. 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* (Portland, Oregon, USA, October 22–25). ACM, New York, NY, USA, 71-78. 10.1145/1168987.1169001.

[4] Ciaramello, F. and Hemami, S. 2007. 'Can you see me now?' An Objective Metric for Predicting Intelligibility of Compressed American Sign Language Video. In (28 January). San Jose, 6492.

[5] Ciaramello, F. M. and Hemami, S. S. 2007. Complexity constrained rate-distortion optimization of sign language video using an objective intelligibility metric. In *Presented at IEEE Western New York Image Processing Workshop (WNYIP)* (). Citeseer, .

[6] Jemni, M., Elghoul, O. and Makhlouf, S. 2007. A Web-Based Tool to Create Online Courses for Deaf Pupils. In *IMCL Conference, Amman* ().

[7] Ma, Z. and Tucker, W. 2008. Adapting x264 to asynchronous video telephony for the Deaf. In *Proc. South African Telecommunications Networks and Applications Conference,(SATNAC 2008)* ().

[8] Muir, L. J. 2005. Perception of sign language and its application to visual communications for deaf people. *Journal of Deaf Studies and Deaf Education.* 10, 4 (2005), 390.

[9] Nakazono, K., Nagashima, Y. and Ichikawa, A. 2006. Digital encoding applied to sign language video. *IEICE Trans. Inf. Syst.* 89, 6 (2006), 1893-1900.

[10] Tucker, W. D. *Softbridge: a socially aware framework for communication bridges over digital divides.* PhD Thesis, Department of Computer Science, University of Cape Town, , 2009.

[11] Vanam, R., Riskin, E. A. and Ladner, R. E. 2009. H. 264/MPEG-4 AVC encoder parameter selection algorithms for complexity distortion tradeoff. In *Proceedings of the 2009 Data Compression Conference* (Snowbird, UT, March 16 - 18). DCC. IEEE Computer Society, Washington, Seattle, WA, 372-381. 10.1109/DCC.2009.53.

[12] Vanam, R., Riskin, E. A., Ladner, R. E., Ciaramello, F. M. and Hemami, S. S. 2009. JOINT RATE-INTELLIGIBILITY-COMPLEXITY OPTIMIZATION OF AN H. 264 VIDEO ENCODER FOR AMERICAN SIGN LANGUAGE. (2009).