Improving fluency in sign language to text systems

Sam Black
524689

Supervisor
Prof Martin Russell

Initial Report

Table of Contents

Introduction

Communication has been the greatest triumph of human development; be it speech and language, stone tablets, hand written parchments, printed documents or electronic transmission. The ability to distribute information in a format or by a method that multiple people can understand and access has enabled the accelerated growth of ideas and culture around the world. Conversely, a lack of communication leads to misunderstanding, hostility and conflict. Good and consistent communication is important, especially in this connected world.

The misunderstandings occur mostly when two differing methods of communication are used; a French speaker conversing with a Russian speaker for example. Problems occur if they cannot communicate, requiring they either speak a common (second) language, or hire an interpreter.

Technology can help in these circumstances by being the interpreter. The advantage of using technology as the interpreter instead of a human is that the technology can be updated or adjusted to differing methods of communication.

Previous systems to convert sign language to text have concentrated on translating individual letters rather than using whole words; whilst this limits the corpus needed, it is uncomfortable and inconvenient to spell each word rather than just sign it (imagine phonetically spelling each word). Most systems currently cannot translate sign language to text in real time fluently, requiring the user to sign each letter or word discretely, reducing the usefulness of the system in a real world setting.

Thus, the aim of the system is to use HMM to implement a fluent British sign language (BSL) to text translator for a limited corpus, such as an information help point for the University campus (for example, “Where is the Guild?”).

Research

Automatic sign language recognition (ASLR) is a combination of gesture recognition and facial recognition (Edwards, 1997). Various methods exist to capture the data, most systems utilise an image capture system that extracts the vector data. The unobtrusive nature of this method is the major factor to its wider adoption and basis of research. Image capture is less accurate than using a motion tracking system, such as using a data glove or optical motion tracking system, as there is a lower amount of noise and segmentation of the data (Dreuw, Rybach, Deselaers, *et al*, 2007).

Accuracy in the system is approaching that of automatic speech recognition (ASR) systems of 85% - 90% (Holt, Hendriks, and Andringa, pp. 8), but lower word error rates (WER) can be achieved by using multiple layers of HMMs. Layered HMM are best for systems which require low response times with a large corpus (Zhang, Yao, Jiang, *et al*, 2005).

Whilst other systems are relatively successful with identifying finger spelt words (Travieso, Alonso, and Ferrer, 2003), discrete isolated words signed (Grobel and Assan, 1997) with no attempt at grammatically correct sentence structure (Akmeliawati, Ooi and Kuang, 2007), few are successful in fluent sentence recognition with a grammar. Computer processed visual data is still susceptible to noise and interference, such as the background being of similar colour to the users hands, adverse lighting conditions or other people in motion providing spurious inputs (Je, Kim and Kim, 2007). Computer vision based systems offer the most comprehensive coverage of a sign language user's movements, but are computationally expensive to calculate the vectors off the user's body and limbs (Holt, Hendriks, and Andringa, pp 11).

Outcomes

To construct a fluent ASLR system based on HMM, using a motion tracker to collect the sign data for PCA. Adding facial expressions as vectors to the data set would be advantageous.

Requirements

The requirements to achieve the outcomes are as follows;

1. Create the corpus to be signed
2. Create grammar, models and pronunciation dictionary for HTK
3. Record the signing using a motion tracker
4. Compile the vector data from the motion tracker
5. Conduct principle component analysis (PCA) on the vectors
6. Run training data through HTK
7. Improve the grammar, models and pronunciation dictionary
8. Re-run the training data to test improvements

Plans

Gantt Chart

The gantt chart for the project is in Appendix 1.

Priorities

The priority for the next few weeks will be to construct the corpus, the language model and the grammar for the HMM (as detailed in Appendix 2). As stated in the gantt chart, the vector data will be collected in the final week of the Autumn term, to be processed over the Christmas break.

The major workload area will be the PCA of the data set and converting this into a format which HTK can apply the grammar, pronunciation dictionary and language model to (in effect replacing the sound file sampling normally employed).

Health and Safety

As most of the work for this project is software based, the risks are minimal.

Using the motion tracking equipment in the Psychology department to record the signing data will be risk assessed before use, taking advice from the Psychology department in this regard.

Ethics implications

Recording the sign language data will require a fluent BSL user, the ethics of involving a human must be considered.

As the purpose of working with a BSL user is to acquire vector data associated with the movement of their arms and hands only, and requires no psychological testing or deception of the user, the ethical implications are minimal. The user will be briefed before hand as to the nature of the work to be done and informed of any risks involved (as identified by the risk assessment). Further guidance will be sought from the Human Research Ethics Committee before the data is recorded.

Costing

There shouldn't be any expenditure in pursuing the project as most of the work is software based. The hiring of a person who can sign the corpus may incur a cost.

References

Akmeliawati, R. Ooi, M. P. L. and Kuang, Y. C. (2007). **Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network.** Proceedings of Instrumentation and Measurement Technology Conference, Warsaw, Poland, 1-3 May 2007.

Dreuw, P. Rybach, D. Deselaers, *D. et al.* (2007). **Speech Recognition Techniques for a Sign Language Recognition System.** In Interspeech, pages 2513-2516, Antwerp, Belgium, August 2007.

Edwards, A. D. N. (1997). If the glove fits: Progress in sign language recognition. **Ability**, 22: pp. 12-13

Grobel, K. and Assan, M. (1997) **Isolated Sign Language Recognition using Hidden Markov Models.** Systems, Man, and Cybernetics, 'Computational Cybernetics and Simulation'. IEEE International Conference 12-15 Oct. 1997. Volume 1, pp. 162-167

t. Holt, G. Hendriks, P. and Andringa T. (unknown date). **Why don't you see what I mean?** [online]. http://ict.ewi.tudelft.nl/pub/gineke/ASR05.pdf [Accessed 07 October 2008]

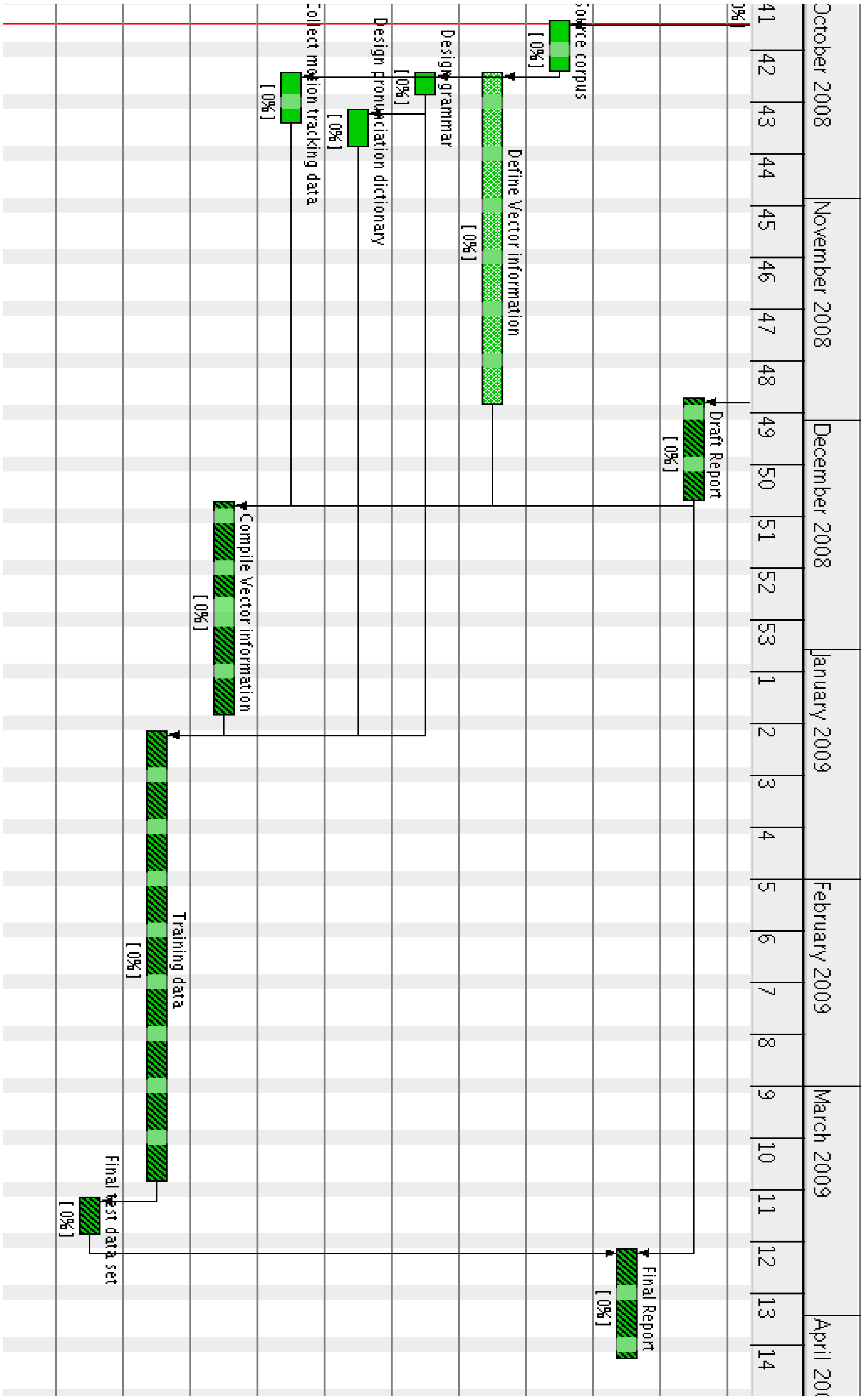
Je, H. Kim, J. and Kim, D. (2007). **Vision-Based Hand Gesture Recognition for Understanding Musical Time Pattern and Tempo.** In Industrial Electronics Society, 33rd Annual Conference of the IEEE, 5-8 Nov. 2007, pp. 2371-2376

Travieso, C.M. Alonso, J.B. and Ferrer, M.A. (2003). **Sign language to text by SVM.** Proceedings of the Signal Processing and Its Applications, Seventh International Symposium 1-4 July 2003, Volume 2, pp. 435-438

Zhang, C. Yao, H. Jiang, F. *et al* (2005). **Multilayer Method Based On Multi-Resolution Feature Extracting and MVC Dimension Reducing Method for Sign Language Recognition.** Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005

Appendices

Appendix 1: Gantt Chart



Appendix 2: HTK Model

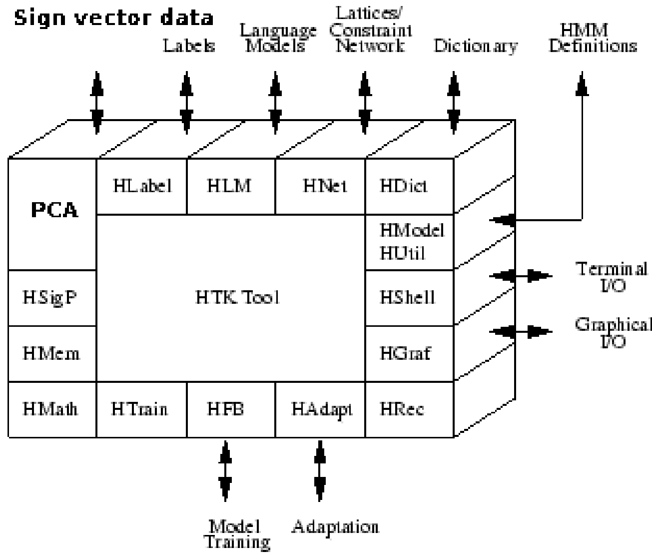


Figure 1: HTK tool software design, with speech analysis replaced with sign vector data and PCA. (Based on the HTK Tutorial, pp. 24).

Appendix 3: Simple grammar

A simple grammar for an information help point or tourist information centre would be as follows;

```
$location = GUILD [OF STUDENTS] |
            CLOCK TOWER |
            GREAT HALL |
            BUS STOP |
            TRAIN STATION;
```

(SENT-START (WHERE IS THE (NEAREST|) <$location>) SENT-END)

The grammar is very simple, but could cover a large amount of location questions by expanding the $location variable. It could be improved by moving the (NEAREST|) to locations that make sense, as the grammar above could generate

Where is the nearest Great Hall?

whilst grammatically correct, it would sound (and sign) wrong; a better method would be to alter $location to

```
$location = ...
              [ NEAREST ] BUS STOP |
              [ NEAREST ] TRAIN STATION;
```

and remove the (NEAREST|) option in the sentence.

Appendix 4: Access to source code

The HTK source code, research, reports, language models, grammars, dictionaries and scripts are accessible using source code management software package GIT (http://git.or.cz/) by;

```
git clone http://lapwing.homelinux.org/repos/personal.git
```

and can be update by

```
git pull
```

in the directory **personal/**.

Source code for this project is in the **personal/uni/ee4p/** directory.

Trac (http://trac.edgewall.org/) is used to keep track of issues, work to do and a simple wiki. This is available from http://samwwwblack.lapwing.org/ .