Speech Training and Recognition for Dysarthric Users of Assistive Technology

M. Parker[3], P. Green[2], P. Enderby[3], M. Hawley[1], S. Cunningham[1], R. Palmer[3], A. Hatzis[2], J. Carmichael[2], S. Brownsell[1], & P. O'Neill[1]

[1]Department of Medical Physics & Clinical Engineering, Barnsley District General Hospital NHS Trust, [2]Department of Computer Science, University of Sheffield, [3]Institute of General Practice, University of Sheffield.

Barnsley District General Hospital NHS Trust

Abstract

Severe dysarthria can be associated with concomitant physical disability necessitating the use of adapted input devices to operate Environmental Control Systems (ECS) and other Electronic Assistive Technology (EAT). Switching systems often control the ECS via a scanning pattern, taking the user through a hierarchy of menu options. This requires the user to have sight of the options menu or to memorise a series of audible tones to track the commands. The process can be a time consuming one. EAT users suggest that a speech operated control system would be an attractive alternative to traditional switch systems, and some are now commercially available.

Research suggests that the use of commercially available computerised automatic speech recognition (ASR) systems by people with severe dysarthria is of limited functional benefit. Research into the use of 'text inputting' programs using speaker independent recognisers illustrate that recognition rates decline rapidly as speech intelligibility deteriorates. There have also been examples of assistive technology being operated by speaker dependent models. This allows the recogniser to be trained with samples of the users own speech which is a better option for dysarthric speech where the output may bear little resemblance to a 'normal' production. However, ASR systems are intolerant of wide variations in speech production. It has been suggested that there is a decrease in computer recognition rates due to the variability of motoric output associated with severe dysarthria.

The STARDUST project brings together expertise from speech pathology, computer sciences and medical engineering to develop an ASR system that can be accessed by those with severe dysarthria and physical disability. The STARDUST team has developed an ASR using Continuous Density Hidden Markhov Models. The ASR is structured around a small vocabulary speaker dependent system trained with a limited corpus of the client's speech, selected to operate assistive technology. The recogniser uses isolated words that can be combined into command strings.

We are currently working alongside a small group of volunteers, all of whom have cerebral palsy or MS, to produce a functionally useful product.

Aims of project

➢To develop small vocabulary speaker dependent ASRs for use by people with severe dysarthria.

➢To link ASR with EAT in a small number of demonstration sites and evaluate the effectiveness of the technology in situ.

➢To develop a suite of recording and visual feedback displays of clinical use in speech training.

ASR with Severe Dysarthria

Problem

Speech recognition is difficult with variable speech production, frequently associated with severe dysarthria. Speech production may also change over time.

Training sets for the ASR in this project are comparatively small in size. Due to the physical problems of the project volunteers, the collection of speech samples is time consuming, laborious and repetitious.

ASR for dysarthric speech

STARDUST solutions

•Aim for *CONSISTENT* RATHER THAN *INTELLIGIBLE* speech output

•Small vocabulary sets targeted at specific EAT commands selected by the client. Target maximum command flexibility for the minimum number of utterances.

•Facility to predict which productions within a client's vocabulary set are likely to be confused with one another by the recogniser.

•Each vocabulary item must:

- •be comprised of enough phonetically distinguishable tokens to make it unique from the production of other vocabulary items
- •show limited variability of production (consistency) from the recognisers target model over time.

In the very simplest of terms:

OPTIMAL RECOGNITION = ⇩ CONFUSABILITY + ⇧ CONSISTENCY

The Confusability Matrix

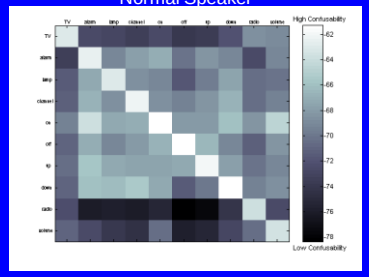
STARDUST programming allows the visualisation of a Confusability Matrix, illustrating the probability of specific productions being confused with other word items.

To reduce confusion requires either:

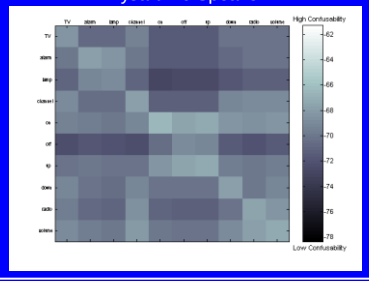
- **•the changing of a vocabulary item to one that will contain distinguishable phonetic tokens from those contained within the other vocabulary items, or…**
- **•training motor output to reduce variability in dysarthric speech output for specific word items**

Visualising Confusability

Normal Speaker



Dysarthric Speaker



Speech Training

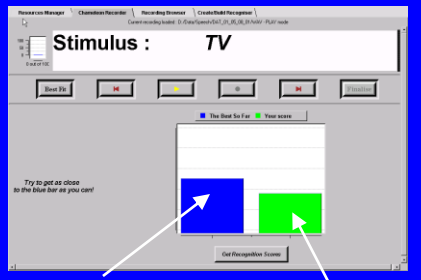
Speech training is seen as a way of attempting to reduce the variability (i.e increase the consistency) of single word output.

'Training' is conducted as a remote activity with the client utilising visual and auditory feedback to try and match their production with a target selected by the computer from their own corpora of data.

The 'target' is called the "Best Fit". This is the one utterance from the training set which the model would be most likely to produce. It is not necessarily the most *intelligible* example of the word, but the one that best approximates the person's most *likely* production.

All subsequent repetitions of the word should be as close to this model as possible to increase the likelihood of recognition by the computer.

Speech Training Trial – Client Display



Current Results

Increased recognition results for severely dysarthric speech compared with a 'commercial' recogniser.

Examples of 93% recognition for combined commands when linked to EAT in the home. Full trial for effectiveness within the home is ongoing.

Summary

Severely dysarthric output shows consistent, distinguishable phonetic features for any given speaker.

•Articulatory patterns have shown change as a result of auditory and visual feedback in some cases of cerebral palsy, where there has been no directive 'speech' intervention for many years. This has allowed the introduction of specific, stable and distinguishable phonetic tokens within single word utterances.

•Current results suggest that ASR can be a viable augmentative system for EAT for those people with severe dysarthria.

Acknowledgements

This research was sponsored by the UK Department of Health New and Emerging Application of Technology (NEAT) programme and received a proportion of its funding from the NHS Executive. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health or the NHS Executive.