Speech Training and Recognition for Dysarthric Users of Assistive Technology

M. Parker¹, P. Green², P. Enderby³, M. Hawley¹, S. Cunningham¹, R. Palmer³, A. Hatzis⁴, J. Carmichael², S. Brownsell¹, & P. O’Neill¹

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

Abstract

Several dysarthrias can be associated with concomitant physical disability, necessitating the use of selected vocal tract devices to operate Environmental Control Systems (ECS) and other Disability Aids. Some techniques require direct vocalisation of commands, and such systems are particularly laborious and repetitive for those with severe motoric dysarthrias. Conventional approaches to assistive communication, based on traditional voice-activated systems, are of limited functionality for those with severe speech variability.

Research suggests that the use of commercially available computerised automatic speech recognition (ASR) systems by people with severe dysarthrias is at limited functional benefit. It is thought by the use of an 'eating pattern'-driven computerised training system that will illustrate that recognition rates decline rapidly as speech variability deteriorates. It has two modes of operation, one based on vocalisation of commands, the other on the recitation of numerical digits, on a normal production basis. Both ASR systems are robust to variations in speech variability, a feature which has to be addressed in speech recognition research. The STARDUST project brings together expertise from speech technology, computer science and medical engineering to develop an ASR system that can be accelerated to those with severe dysarthrias and acquired speech.

The STARDUST matrix was developed as a means of estimating the confusability of items using a Continuous Density Hidden Markov Model. The ASR is structured around a small vocabulary speaker dependent system interfaced with a computerised training system for the STARDUST environment. The system was operated in a clinical setting to determine the feasibility of using an ASR system in a clinical setting. The recognition accuracy of the system was assessed using the classification of words that can be combined into the correct string.

The authors are currently working alongside a small group of volunteers, all of whom have cerebral palsy, to determine the functional and clinical outcomes.

Aims of project

➢ To develop small vocabulary speaker dependent ASRs for use by people with severe dysarthrias.
➢ 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 a single phonetically distinguishable token or, tokens, that make it unique from the production of other vocabulary items
• show limited variability of production (consistency) from the recogniser’s target model over time.

In the very simplest of terms:

OPTIMAL RECOGNITION =

CONFIGURABILITY + 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

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 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.

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.