Published paper

Palmer, R., Enderby, P. and Cunningham, S. P. (2004) *The effect of three practice conditions on the consistency of chronic dysarthric speech*, Journal of Medical Speech Language Pathology, 12 (4). pp. 183-188.

The Effect of Three Practice Conditions on the Consistency of Chronic Dysarthric Speech

Rebecca Palmer and Pam Enderby

Institute of General Practice and Primary Care

University of Sheffield, UK

Stuart Cunningham

Department of Medical Physics & Clinical Engineering,

Barnsley District General Hospital NHS Trust

Barnsley, UK

Abstract

This study investigated whether it is possible for people with chronic dysarthria to adjust their articulation in three practice conditions. A speaker dependent, speech recognition system was used to compare participants' practice attempts with a model of a word made from previous recordings to give a recognition score. This score was used to indicate changes in production of practice words with different conditions. The three conditions were: repetition of written target words, visual feedback, and an auditory model followed by visual feedback. For eight participants with dysarthria, the ability to alter speech production was shown, together with a differential effect of the three conditions. Copying an auditory target gave significantly better recognition scores than just repeating the word. Visual feedback was no more effective than repetition alone. For four control participants, visual feedback did produce significantly better recognition scores than just repetition of written words and the presence of an auditory model was significantly more effective than visual feedback. Possible reasons for differences between conditions are discussed.

Introduction

People with chronic dysarthria have often been discharged from speech and language therapy as further improvement in their speech is not expected. However, a small number of case studies suggest that people with chronic dysarthria do have potential for change and that treatment may in fact be beneficial sometime after the onset of dysarthria (Enderby & Crowe, 1990; Keatly & Wirz, 1994)

This study investigated the potential for people with chronic dysarthria to change their articulation of single words with treatment, despite a plateau in spontaneous recovery or little progress with traditional methods. The specific question asked was: In individuals with chronic dysarthria, which is the most efficient feedback condition in assisting their attempts to adjust speech production to a target (habitual) production?

a) Auditory feedback alone,
b) visual feedback of proximity to the target (habitual) production, and
c) auditory model of the target (habitual) production as well as feedback of proximity to the target production.

It was hypothesised that participants would be able to change their articulation and that this change would be greater with visual feedback than simple repetition, and that the additional information provided by an auditory example would further improve pronunciation performance.

Method

An alternating treatments design was used to compare the relative effects of the different practice conditions within individual participants. Counterbalancing the order of presentation of experimental conditions between participants is inherent to an alternating treatments design, accounting for any order effects (Hedge, 1994).

Participants

Eight participants were recruited to the study enabling observations of participants across a continuum of severe, moderate and mild dysarthria. Four non-dysarthric speakers were also recruited to observe effects of different practice conditions on the complete spectrum of motor speech abilities. Participants with dysarthria were excluded from the study if the dysarthria was rapidly deteriorating, if they had other speech, language, hearing or cognitive impairments, if they were under 18 years of age or if they were unable to read short sentences.

The Frenchay Dysarthria Assessment was completed by a speech therapist to identify the type and severity of the dysarthria (Enderby, 1983). Ratings were confirmed by an independent verificant listening to tape recordings of the speaker saying the words and sentences and conversation sections of the Frenchay Dysarthria Assessment. Table 1 shows four men and four women between 24 and 44 years of age participated in the study. A range of dysarthria types are presented and aetiologies are developmental or of sudden onset with the exception of participant 8 who had Unverricht Lundborg disease. She was recruited as her speech therapy notes stated that there had been no perceived deterioration in speech between May 2001 and May 2003.

Four non-dysarthric speakers, two men and two women were age matched to the group of speakers with dysarthria. The non-dysarthric speakers had no hearing loss, learning or cognitive disability and no speech or language impairment.

<u>Treatment content and equipment</u>

Five real words were selected for practice according to a range of criteria ensuring the words were phonetically balanced to account for individual difficulties with different sounds. The five words were; bed, food, money, garden and telephone.

The use of speech recognition technology is a central component of the study allowing small changes in articulation performance to be measured. Any changes in production are reflected by changes in this probability score produce by a speaker dependent speech recogniser.

Speaker-dependent recognisers were trained consisting of whole-word, 11 state, hidden Markov models (HMMs) using 12 Mel-frequency cepstral coefficients with first-order temporal derivatives and energy normalisation as input features. The HMM toolkit (Young et al., 2001) was used to operate the recognition models, in unison with a specially written software interface (Hatzis et al., 2003).

The software was installed onto a Toshiba SP6100 laptop computer. An Andrea DA-400 array microphone was used to record the participants from a distance of approximately 50 cm from the participant's mouth.

<u>Procedure</u>

Each participant was visited at home for 45-60 minutes, once a week for four weeks. In the first visit recordings were made and at the following three visits, the participants practised pronouncing the words used in the study under three different practice conditions.

The first visit was a pre-practice visit in which the five words were presented to each participant twenty times each in a random order. When the written word presented on the computer screen turned green, the participant was asked to say the word. These recordings were used to train a speaker dependent speech recogniser tailored to each individual participant. The recogniser compares the incoming sound wave to the model it made of a word and produces a probability score indicating the likelihood that the incoming signal was that word. It is the ability of the recogniser to compare pronunciation of a word with a model pronunciation that is being exploited in this study as a potential way of providing speakers with dysarthria with a target pronunciation and feedback about their practice pronunciations in relation to this target. The best fit example (target pronunciation) was one of the twenty recorded examples that had the highest recognition score and can therefore be assumed to represent the notional average pronunciation of a word for a given speaker (Hatzis et al. 2003). Hence, in this study, the targets are not related to normal articulatory goals but represent the most common habitual production of the word by the individual speaker.

In each experimental practice condition, the participant was presented with each word ten times in a row to practise before presentation of the next word.

<u>Condition 1: No computer generated feedback</u>

The participant practised each word receiving only auditory and kinaesthetic feedback on the pronunciation before practising it again.

<u>Condition 2: Immediate visual feedback</u>

The participant received immediate visual feedback about the pronunciation two or three seconds after the practice attempt. Two coloured bars were displayed on the computer screen. One bar, shown in blue, represented the probability score of the *best fit* example and, the other bar, shown in green represented the probability score for the practice pronunciation. The participant then had the opportunity to use this information to change the pronunciation in an attempt to match the height of the green bar to the blue target.

<u>Condition 3: An auditory target followed by visual feedback on the practice pronunciation</u>

Before the word was practised, the investigator played the target pronunciation (best fit example) three times. The participant was instructed to listen to the example and then copy or imitate it as closely as possible. Green and blue bars providing visual feedback on the practice pronunciation followed as in condition 2 before the target model was played three times again.

Each participant practised the words in each of the three conditions in each experimental session. To control for any effects of one condition on the next one, the presentation of conditions was counterbalanced.

In addition, to reduce the possibility of carryover effects from one condition to another, a speech task, conversation, unrelated to the practice conditions was carried out for ten minutes between the presentation of each condition to avoid internal rehearsal of the words based on the feedback that had been provided in the previous condition.

Results

The computer generated 150 speech recognition scores, determined from the aligned likelihood for the target word, for each condition for each participant. Higher scores indicate better recognition of a word than lower scores. Changes in the recognition scores represent changes in the production of the word. As log values of the scores were generated, they appear as negative numbers.

A univariate two way analysis of variance was used to test for significance between the mean scores in each condition, using a Tukey as a post hoc test. Comparisons between conditions were made for the eight participants with dysarthria and for the four participants with no motor speech disorder. For dysarthric participants there was a significant effect of condition [$F(2,7) = 20.831$, $p < 0.05$]. For the non-dysarthric participants the effect of condition was also significant [$F(2,3) = 148.585$, $p < 0.05$].

For the participants with dysarthria, the visual feedback with an auditory model (condition 3) was significantly more effective in helping speakers to change their production of words than just reading words (condition 1), or being given visual feedback (condition 2) ($p < 0.05$). For this group, visual feedback was not significantly more effective than reading alone. However, for the speakers without motor speech disorders (non-dysarthric speakers), visual feedback (condition 2) was significantly more effective in changing production of words than reading words alone (condition 1), and the combination of an auditory model and visual feedback (condition 3) was also significantly more effective than visual feedback (condition 2) ($p < 0.05$).

The box plots in figure 1 enable a closer inspection of results of different practice conditions for individual speakers. Group results showed that visual feedback was no more effective than repetition alone in achieving a target pronunciation for speakers with dysarthria.

Taking a closer look at participants 1 and 3, recognition scores are slightly lower (closer to the target) with visual feedback than when simply repeating words. An auditory model is indicated as being beneficial to speakers with dysarthria by group results. This is clearly shown for participants 2, 4, 5 and 8. The benefit of an auditory model appears to be only marginal for participants 1, 3 and 7 and no better than visual feedback for participant 6. There are no apparent differences in terms of type of dysarthria or severity between those who show a clear effect of an auditory model and those who only show a marginal effect.

For the normal speakers, participants 9, 11 and 12 all show clear differences between the median recognition scores in each practice condition as reflected by the group result. Figure 1 shows that participant 10 does not follow the same pattern as the other normal speakers as the addition of an auditory model did not appear to be of benefit above visual feedback alone.

Figure 1 shows that the range and inter-quartile range of speech recognition scores are similar for each participant irrespective of severity. Interestingly, participant 9 has no motor speech disorder but has the largest range and inter-quartile range of recognition scores representing a greater amount of variation in pronunciation of words than the other normal speaker and the eight participants with dysarthria.

For participants 7 and 8, the inter-quartile range of recognition scores is visibly smaller when practising under condition three (auditory model) than conditions one and two, suggesting a greater consistency in production of words with an auditory model to copy for these two participants. For participants 4, 7 and 11, the inter-quartile ranges are greatest for condition 2 (visual feedback) indicating that visual feedback increased the variability of word pronunciation for these speakers. For participants 2, 8 and 10, the inter-quartile range of scores in condition 2 were smaller than for condition 1 indicating that visual feedback may have helped reduce variability in pronunciation of words for these participants.

Discussion

The results show that the speakers with chronic dysarthria in this study could use auditory models and visual feedback in order to adjust their practice pronunciations to a target (habitual) pronunciation. Contrary to the hypothesis, the presence of visual feedback was no more effective than simple repetition for the speakers with dysarthria as a group, however, the presentation of an auditory model in addition to visual feedback had a significant effect on enabling participants to pronounce words closer to their target. For the non-dysarthric participants, visual feedback was more effective than repetition of written words alone, unlike for the participants with dysarthria, and the presence of an auditory model had a significantly greater effect than visual feedback alone.

A consideration of the information provided by the auditory targets and visual feedback, and the mental processing required to use them, can suggest explanations for the findings. The auditory model provided a target for the participants to imitate. Imitation is involved in early motor speech development and as such, may be a more innate or natural process by which to alter articulation than visual feedback (Ingram, 1989). It is also possible that, as auditory imitation is used by infants, the process does not require as much conscious cognitive involvement as the interpretation of the visual feedback. Differences in the information content of the visual feedback and auditory model could also explain the result. Blackwell and Newell suggested that the information of feedback serves to calibrate a task for an individual (Blackwell & Newell, 1995). The visual feedback in this experiment consisted of a bar chart indicating only how close a production was to the target without any indication of reasons for differences between the score and the target pronunciation. Suggesting a goal to a patient has been found to help motivation in practice and a specific goal has been shown to be more effective than a non specific goal such as ‘do your best’ (Carr & Shepherd, 1987; Gauggel & Fischer, 2001). The auditory model suggests the goal specifically, whereas the visual feedback in this experiment does not. The ranges in recognition scores were generally smaller when imitating the auditory model than when responding to visual feedback and for two participants, the range of scores with visual feedback was bigger than when simply repeating words read from the screen. This supports the idea that specific information helps participants to pronounce a word closer to the target, while non specific feedback alerts the participant to how much the pronunciation needs to change on the next trial, but does not indicate how to do this.

In this study, speakers with dysarthria changed their pronunciation towards one that was already achievable. It can be debated whether or not producing a pronunciation that is already achievable constitutes change or whether this would be more suitably termed *control* of pronunciation. Speech recognition technology has proved useful in identifying these small changes or ability to control pronunciation which leads to a further question of whether speakers with chronic dysarthria can make big enough changes in their pronunciation to be perceivable, and therefore clinically significant.

In conclusion, the clinical implications of this study are that speech recognition technology can be used to identify ability to control speech and hence a potential to improve pronunciation with therapy.

References

Blackwell, J. & Newell, K. (1995). The informational role of knowledge of results in motor learning. Acta Psychologica, 92, 119-129.

Carr, J. & Shepherd, R. (1987). A Motor Relearning Programme for Stroke. Heinemann: London.

Enderby, P., (1983). Frenchay Dysarthria Assessment. Pro-ed: Texas.

Enderby, P. & Crow, E. (1990). Long-term recovery patterns of severe dysarthria following head injury. British Journal of Communication Disorders, 25, 341-354.

Gauggel, S. & Fischer, S. (2001). The effect of goal setting on motor performance and motor learning in brain damaged patients. Neuropsychological Rehabilitation, 11(1), 33-44.

Hatzis, A., Green, P., Carmichael, J., Cunningham, S., Palmer, R., Parker, M. & O'Neill, P. (2003) An Integrated Toolkit Deploying Speech Technology for Computer Based Speech Training with Application to Dysarthric Speakers. Proceedings of the 8th European Conference on Speech Communication and Technology, Geneva, 2213-2216.

Hedge, M. (1994). Clinical Research in Communicative Disorders: Principles and Strategies, 2. Pro-ed: Texas.

Ingram, D. (1989). First Language Acquisition: Method, Description and Explanation Cambridge University Press: Cambridge.

Keatly, A. & Wirz, S. (1994). Is 20 years too long? Improving intelligibility in long-standing dysarthria – A single case treatment study. European Journal of Disorders of Communication, 29, 183-201.

Koenigsberg, R. (1989). Churchill's Medical Dictionary Churchill Livingstone Inc.: New York.

Young, S., Evermann, G., Kershaw, D., Moore, G., Odell, D., Valtchev, V. & Woodland, P. (2001). The HTK Book, Cambridge University Engineering Department: Cambridge.

Author Note

Rebecca Palmer, Institute of General Practice and Primary care, University of Sheffield, UK; Pam Enderby, Institute of General Practice and Primary care, University of Sheffield, UK; Stuart Cunningham, is now at Department of Human Communication Sciences, University of Sheffield, UK.

Development of the technology used in this study was supported by Mark Hawley, Department of Medical Physics & Clinical Engineering, Barnsley District General Hospital NHS Trust, Barnsley, UK, and Phil Green, Department of Computer Science, University of Sheffield, UK and funded by the Department of Health New and Emerging Applications of Technology programme. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health.

Correspondence concerning this article should be addressed to Rebecca Palmer, Institute of General Practice and Primary Care, Community Sciences Centre, Northern General Hospital, Herries Road, Sheffield, S5 7AU, UK. E-mail: r.l.palmer@sheffield.ac.uk

Table 1

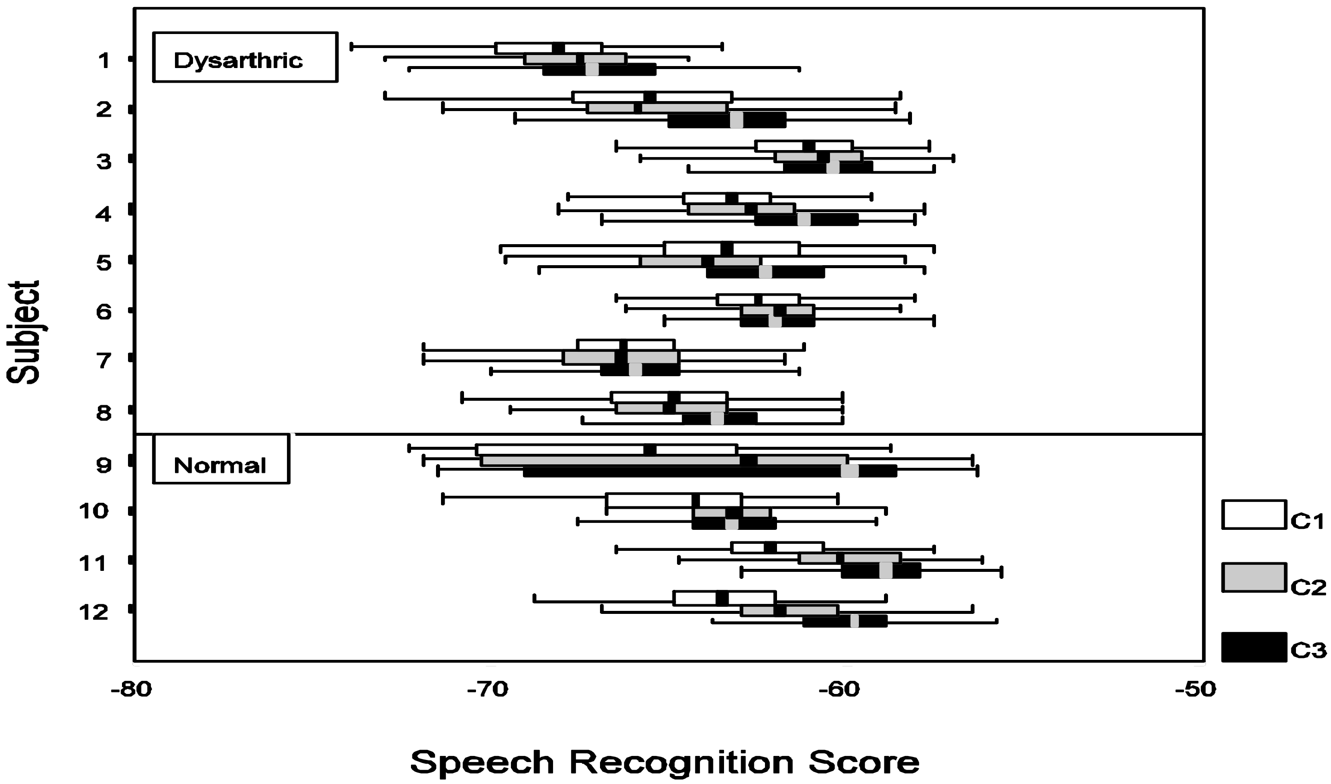
Information about the participants with dysarthria

Participant	Initials	Sex	Age (years)	Severity	Type	Aetiology	Time post onset (years)
1	MR	M	35	Severe	Spastic	CP	35
2	KD	F	40	Severe	Spastic	CP, CVI	40,4
3	AJ	M	39	Severe	Spastic-ataxic	TBI	19
4	PM	M	43	Severe-moderate	Spastic-ataxic	unknown	35
5	PB	M	39	Moderate-severe	Spastic	CP	39
6	RH	F	35	Moderate	Mixed	CP	35
7	KM	F	44	Mild-moderate	Ataxic	Brain tumour, CVI	2.5
8	NB	F	24	Mild	Spastic	Unverricht Lundorg disease[a]	14

Note. M = male F = female CP = cerebral palsy CVI = cerebrovascular incident TBI = traumatic brain injury.

[a]Unverricht Lundborg disease is a progressive myoclonus epilepsy which can also involve dysarthria, intention tremor and mild dementia (Koenigsberg, 1989).

Figure 1. Speech recognition scores for each dysarthric and normal participant in condition 1 (repetition), condition 2 (repetition with visual feedback of performance and condition 3 (listening to an auditory model target). The scores are determined from the aligned likelihood values of the target model for each target word (Hatzis et al. 2003). Scores towards the right of the axis indicate a closer fit to the target model.


Dysarthric
Normal
Subject
1
2
3
4
5
6
7
8
9
10
11
12
C1
C2
C3
-80
-70
-60
-50
Speech Recognition Score