City Research Online

City, University of London Institutional Repository

Citation: Whiteside, S. P., Inglis, A. L., Dyson, L., Roper, A. H., Harbottle, A., Ryder, J., Cowell, P. E. & Varley, R. A. (2012). Error reduction therapy in reducing struggle and grope behaviours in apraxia of speech. Neuropsychological Rehabilitation, 22(2), pp. 267-294. doi: 10.1080/09602011.2011.639614

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/13941/

Link to published version: https://doi.org/10.1080/09602011.2011.639614

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

ERROR REDUCTION THERAPY IN REDUCING STRUGGLE AND GROPE BEHAVIOURS IN APRAXIA OF SPEECH

Sandra P. Whiteside, A. L. Inglis, Lucy Dyson, Abigail Roper, Andrew Harbottle, Jennifer Ryder, Patricia E. Cowell & Rosemary A. Varley

Department of Human Communication Sciences, University of Sheffield, Sheffield, S10 2TA, UK

Corresponding author: Dr Sandra P Whiteside, Department of Human Communication Sciences, University of Sheffield, 31 Claremont Crescent, Sheffield, S10 2TA, UK. Telephone: +44 (0) 114 222 2447. Fax: +44 (0) 114 273 0547. Email: s.whiteside@sheffield.ac.uk

Abbreviated title: Error reduction therapy in Acquired Apraxia of Speech

Keywords: Acquired apraxia of speech; Error reduction principles; computer-based speech therapy; two-phase cross-over treatment design.

ABSTRACT

We report an intervention study focused on the speech production difficulties present in acquired apraxia of speech (AOS). The intervention was a self-administered computer therapy that targeted whole word production and incorporated error reduction strategies. The effectiveness of the therapy was contrasted to that of a visuo-spatial sham computer program and performance across treated words, and two sets of matched words was assessed. Two groups of participants completed the study which employed a two-phase cross-over treatment design. Participants were randomly assigned to a speech first or sham first condition. Treatments were administered for 6 weeks, with a 4 week rest between interventions. Participants were assessed five times in total; twice at baseline, once following each of the intervention phases, and once following a lapse of 8 weeks after the end of the second phase of intervention. The occurrence of accurate word production and speech characterised by struggle and groping behaviours was recorded on a repetition task. Participants showed significant gains in speech accuracy and fluency, and reductions in articulatory groping and struggle behaviours following the use of the speech program. These gains were largely maintained once the therapy was withdrawn.

INTRODUCTION

Apraxia of speech (AOS) is an impairment of speech production in which output contains numerous articulatory errors that reduce intelligibility. It is viewed as a disruption to the motor plans that control speech production and as a result, speech appears to lose its automaticity, instead appearing under conscious control (Lebrun, 1990; Lecours & Lhermitte, 1976). The condition is characterised by articulatory groping, which involves preparatory visible and sometimes audible speech gestures. Groping is not unique to AOS and a similar phenomenon can be observed in the 'conduite d'approache' behaviours of speakers with conduction aphasia (McNeil, Robin & Schmidt, 2009). However, it represents a core feature of speech apraxia and features in many diagnostic typologies for the impairment (e.g., Wertz, LaPointe, Rosenbek, 1984; Dabul, 1986). In apraxia, speech appears effortful and features can include an altered voice quality and changed durational characteristics such as longer inter-syllabic pauses and prolonged segment and syllable durations (Kent & Rosenbek, 1982, 1983; Whiteside & Varley, 1998a; Varley, Whiteside & Luff, 1999). There is also reduced cohesion or overlap in articulatory gestures, reflected in reduced levels of coarticulation (Ziegler & von Cramon, 1985, 1986; Whiteside & Varley, 1998b; Whiteside, Grobler, Windsor & Varley, 2010).

There is debate regarding the size of the motor plans for speech and the lack of agreement on this issue motivates different forms of therapy for AOS. One proposal is that all speech output proceeds from activation and concatenation of individual segmental plans (Shattuck-Hufnagel, 1979). The segmental view has motivated many treatment approaches for AOS, and sound production therapies address articulatory gestures with the aim of increasing the inventory of postures that can be achieved (Rosenbek, 1985; Peach 2004). Once a repertoire of differentiated gestures is established, the therapy moves to combining gestures in order to produce cohesive syllables. The approach involves assisting the patient to acquire conscious, declarative knowledge of the targets and dynamics of articulatory gestures and employs repeated practice of slow, effortful and often errorful speech production. Although such therapies may lead to patients acquiring the targeted gestures, there is limited evidence for the generalisation of this learning to the same gesture in different syllables and to other motorically similar articulatory postures (Wambaugh, Duffy, McNeil, Robin, & Rogers, 2006). An alternative proposal is that complete movement gestalts, or schemas, are formed for outputs that are produced with some frequency, thus by-passing the need for segment-by-segment assembly (Crompton, 1981; Keller, 1987; Levelt & Wheeldon, 1994; Whiteside & Varley, 1998a; Varley & Whiteside, 2001).

Alternative approaches to therapy for AOS are motivated by the schema view of speech production and focus on the production of whole words and utterances. Rather than developing explicit awareness of the subcomponents of speech, schema-approaches aim to maintain automatic and implicit movement control. In this way, ‘learned misuse’ of the usual mechanisms of speech control is avoided (Taub, Miller, Novack, Cook, Fleming, Nepomuceno, et al., 1993; Pulvermüller & Berthier, 2008). Errorless learning principles by which patients practice fluent, automatic and error-free behaviour, may be particularly useful in avoiding learned misuse. In the case of therapies for AOS, the use of automatic and implicit mechanisms that govern healthy speech control is maintained, and entraining compensatory mechanisms that might deliver only slow and effortful speech production is avoided.

Errorless learning techniques have been incorporated into a number of interventions aimed at ameliorating the impairments that follow acquired neurological injury (Fillingham, Hodgson, Sage, & Lambon Ralph, 2003). The effectiveness of the method in comparison to errorful learning has been explored across conditions such as aphasia, and the memory problems of amnesia and Alzheimer's Disease, and across a range of different cognitive-behavioural domains. With regard to interventions in amnesia, there is a series of reports of advantages of errorless over errorful learning techniques (e.g., Wilson, Baddeley, Evans & Shiel, 1994) and the amnesia literature contains ongoing debate regarding the source of the errorless learning advantage (Baddeley & Wilson, 1994). In amnesia, automatic implicit memory is less impaired than conscious declarative memory (Corkin, 2002). One proposal is that implicit learning in particular is disrupted under errorful conditions, with variable responses hampering the formation of stable neuronal assemblies that encode new learning. However, this proposal is controversial and other investigators claim that both implicit and explicit memory systems support errorless learning (e.g., Tailby & Haslam, 2003).

Although the amnesia literature provides some support for the effectiveness of errorless learning techniques, the evidence in support of their use in remediating aphasic language impairment is more equivocal. Fillingham, Sage, & Lambon Ralph (2006) and Conroy, Sage & Lambon Ralph (2009) explored the impact of errorless learning in contrast to trial-and-error methods on remediation of anomic lexical deficits. Their results showed that, despite a patient preference for the errorless technique, the outcomes of the two forms of intervention were equivalent. Raymer, Strobel, Prokup, Thomason & Reff (2010), in an exploration of the effectiveness of errorful and errorless intervention in acquired dysgraphia, showed better outcomes for the errorful technique, although again participants preferred the errorless training procedures. The different findings with regard to comparisons of learning techniques across amnesia and aphasia might be linked to the claim that errorless learning facilitates automatic, procedural learning in particular. While some linguistic knowledge is available for conscious report, such as explicit knowledge of the link between a word and its referent in naming, other elements remain implicit (Ullman, 2001; Ardila, 2010). Speakers do not have explicit or conscious knowledge of the processes involved in generating a grammatical sentence or of the adjustments of articulators necessary to produce fluent and accurate word forms. Speech production represents a motor skill that involves unconscious procedural memory. One possibility is that incorporating errorless learning principles into therapies for acquired speech impairments, such as apraxia of speech, may be particularly important in designing effective interventions. By contrast, interventions directed at explicit knowledge such as the link between an object and its name may show less advantage for errorless learning methods.

We report the outcomes of an intervention for the speech impairments of AOS that focuses on production of whole words rather than isolated articulatory gestures. The intervention was motivated by a number of principles, one of which was errorless learning. Another important principle was that therapy is delivered at a sufficient level and intensity to stimulate reorganisation of underpinning neural systems (Pulvermüller & Berthier, 2008; Varley, 2011). However, delivery of intensive therapy may represent a challenge to resource-limited healthcare systems and the strategy we have adopted in achieving high intensity in a cost-effective manner is to design software programs that allow the user to self-administer therapy without a clinician being present. The use of a self-administered therapy model inevitably has the consequence that the specifics of therapy implementation are lost. Therefore, although the speech program aimed to incorporate errorless learning strategies, it might be better described as designed on ‘error reduction’ principles (Fillingham et al., 2003).

The aim of the speech program was that, as much as possible, the user would practise automatic, fluent and errorless speech production and non-fluent speech attempts including struggle and groping towards articulatory targets would be reduced. In order to achieve this aim, the following components were incorporated into the therapy. The therapy involved a defined set of 35 words, and before any attempts at speech production, there was a significant sensory-perceptual training phase. Prior to the user making judgments, the program modelled errorless performance. The first task involved spoken word-picture matching (Input Level 1), and after the presentation of the auditory stimulus, the computer highlighted the correct picture. Subsequently the program modelled correct spoken-written word matching (Input Level 2). In the errorless models, there was a gradual increase in the delay between the stimulus and indication of the response in order to allow the users to make their own decision on the target even though no overt response was required. Once the errorless modelling tasks had been presented five times, the participant attempted the same tasks (Input Levels 3 and 4). If the participant met a 90% accuracy criterion, the program progressed to the next level. If the criterion was not met, the participant was returned to the errorless modelling tasks at Input Level 1 or 2. The final input tasks involved auditory perceptual judgments. Again the program modelled errorless decisions as to whether an auditory presentation of the target word was correct or incorrect (Input Level 5), and then the participant made the same judgments (Input Level 6). If the 90% accuracy criterion was not met, the participant was returned to the Input level 5 task. Although AOS is an impairment of speech production, the rationale for including the input/sensory-perceptual phase was, first, that apraxia is often accompanied by aphasia. The input training therefore stabilised pre-phonetic processing of the target words. Second, strong functional connectivity between sensory-perceptual and motor speech systems has been observed (Watkins, Strafella, & Paus, 2003; Rizzolatti, 2005). Hence listening to words results in 'mirror' activations in motor control systems. The input components aimed to prime the subsequent word production phases of the intervention. Thus when the participant was required to produce the target words, there would be greater likelihood of fluent, error free production. The perceptual phase of the program also enabled participants with severe apraxia to establish stable input models for the target vocabulary prior to production attempts, and so increase the likelihood of some error-free production.

Error reducing strategies were also incorporated into the design of the production phases of the intervention. The participant was initially required to observe video clips of a speaker saying the target words (Speaking Level 1). The program then moved to imagined production of words, with the request that the participant silently produces the word (Speaking Level 2). At the next stage, overt production was required and, in order to minimise errors, the user repeated the word after audio and video presentation (Speaking Level 3). Initially immediate repetition after the model was required, but then increasing delay was introduced between the model and the request for an overt response. The final stages of the program allowed the user to record and listen back to their repetitions (Speaking Level 4) and then to practise the target words in a variety of sentence frames (Speaking Levels 5 & 6), and finally in isolated naming (Speaking Level 7). However, errors were again minimised by availability of cues and the user could access written word, initial sound or whole word sound file prompts.

In summary, we explored the effectiveness of the speech intervention that incorporated error-reduction principles in reducing the signs of errorful, non-automatic and effortful speech production in participants with AOS. We hypothesised that the intervention would result in reductions in articulatory groping and struggle, and increased levels of accuracy and fluency in a word repetition task.

MATERIALS AND METHODS

Study and therapy design

Fifty participants with AOS were recruited to the study. The study was granted ethics approval under the NHS LREC procedures. All participants gave their informed consent. Participants were classed having AOS by two independent speech and language therapists using standard diagnostic criteria such as reduced speed of articulation, reduced speech fluency, struggle, groping and dysprosody (Wambaugh, Duffy, McNeil, Robin & Rogers, 2006). All participants had some degree of coexisting aphasia. Participants were not receiving any speech and language therapy directed at apraxic or aphasic impairment during the course of the study.

The study used a two-phase cross-over design (Cowell, Whiteside, Windsor & Varley, 2010) where participants were allocated to one of two treatment groups (speech first (SPF) or sham first (SHF)). Participants were blind to the experimental hypothesis, and had no knowledge of the status of the sham program. The design of the study included a series of speech and language assessments which were conducted at different points across the study. The initial evaluation included background speech and language assessments, and measures of two behaviours that were not predicted to change as a result of the intervention were taken at two points: i) at the beginning of the study at baseline (B1); and ii) at the maintenance assessment (M) at the end of the study. The 2 control assessments were: i) the CAT Comprehension of spoken sentences (possible maximum of 32) (Swinburn, Porter & Howard, 2004); and ii) PALPA 48 (written word to picture matching sub-test (possible maximum - 40)) (Kay, Lesser & Coltheart, 1992). Baselines on two repeated measures (a picture naming task and a word repetition task) were conducted at two points (B1, B2) with a 3 to 4 week interval between B1 and B2.

These repeated measures were also carried out at three additional points in the study. The baseline assessments for the repeated measures were designed to determine performance prior to intervention and the stability of speech behaviours. B2 was immediately followed by an initial treatment period which was either a speech therapy computer program or a visual sham computer program. The sham program had identical human-computer interfaces as the speech program, but contained minimal speech or language content. The tasks involved visuo-spatial problem solving, for example, delayed matching of complex designs. Participants were informed that the visual sham program was designed to improve attention and memory. Both programs used a simple interface, and were automatically booted up when the computer was switched on. A device was built into the speech program to ensure that it was booted up to the point at which participants had left off in any prior sessions. The number of sessions and active time spent on the programs were monitored automatically within both the speech and sham programs. Both programs were self-administered for a period of six weeks. Participants were contacted by a therapist after the initial induction session to discover if they were experiencing any difficulties in using the programs. Additional support was provided to assist use where necessary. On the whole, minimal additional support was required by participants. At the end of the initial therapy a third assessment session took place (Post Treatment 1 (Tx1)). For the SPF group the evaluation was labelled the Post Speech (PTx-Speech) assessment, and for the SHF group, the Post Sham (PTx-Sham) assessment. The participant then entered a four week rest period, which was included to maintain levels of motivation due to the length and cross-over design of the intervention study. After the rest period, there was a second treatment period of six weeks. During this second treatment period, the SPF participants were given the visual sham therapy, and the SHF participants received the speech-based therapy program. The end of the second period of therapy was marked by a fourth assessment session (Post Treatment 2 (PTx2)), which was the Post Sham

(PTxSham) assessment for the SPF group, and the Post Speech (PTxSpeech) assessment for the SHF group. Participants then had a rest period of eight weeks with no intervention or contact, followed by a final test session to assess therapy maintenance effects (Maintenance).

Participants

Data for forty four participants with AOS who completed the entire treatment study and all five assessment sessions (B1, B2, PTxSpeech, PTxSham, Maintenance) are reported here. The participants included 20 females and 24 males with a mean age of 65 years (SD = 15 years; age range = 28 to 86 years). The participants were at least 5 months post-onset of a cerebrovascular accident (CVA), with a mean time post-onset time of 22 months (SD = 21, range=5 to 105 months). All participants except for five were either right or predominantly right-lateralised. Of the five non-right handers, three displayed mixed laterality, and two were predominantly left-handed.

Participants were block randomised to either SPF or SHF conditions using a blind envelope system and by an investigator who was blind to case. A total of 21 participants (8 females and 13 males) with a mean age of 60 years (SD = 17 years; range = 28 to 84 years), and mean post-onset time of 19.0 months (SD=14.9 months) were in the SPF group. A total of 23 participants (12 females and 11 males) with a mean age of 68 years (SD = 13 years; range = 36 to 86 years), and a mean post-onset time of 25.1 months (SD=25.4 months) were in the SHF group. There were no significant age differences (t (42) = 1.757, p=.086), or post-onset time differences (t(42)= -.965, p=.340) between the participants within the two groups.

Speech and Language Assessments

The extent of aphasic difficulties and severity of apraxia were assessed at the beginning of the study. The assessments which are most relevant for the current study are reported here with the maximum possible score (where relevant) and the mean scores indicated for the SPF and SHF groups, respectively. Independent t-tests showed that there were no significant differences (p>.05) between the scores for the two groups, suggesting that levels of severity were matched across both groups. The results of these t-tests are also provided:

i) Spoken Picture Naming (a subset of 10 low and 10 medium frequency words from PALPA 54 (Picture Naming x Frequency) (Kay, et al., 1992) (max score 20: SPF mean 14.24 (SD 7.44); SHF mean 12.52 (SD 8.10); t(42)=.729, p=.470);
ii) Spoken Reversible Sentence Comprehension (in house, unpublished test) (max score 20: SPF mean 14.00 (SD 4.38); SHF mean 14.52 (SD 3.73); t(42)=-.427, p=.672);
iii) Auditory Lexical Decision (subset of stimulus items from Franklin, Turner & Ellis, 1992) (max score 20: SPF mean 15.00 (SD 2.71); SHF mean 14.4 (SD 3.31); t(41)** =.653, p=.518, ** 1 participant in the SPF group was unable to complete the assessment);
iv) Auditory Minimal Pairs (subset of stimulus items from Franklin et al., 1992) (max score 24: SPF mean 20.65 (SD 4.36); SHF mean 19.26 (SD 4.91)); t(41)** =.975, p=.335, ** 1 participant in the SPF group was unable to complete the assessment);
v) Non word repetition accuracy (max score 20: SPF mean 4.80 (SD 3.21); SHF mean 3.10 (SD 2.87); t(42) =1.827, p=.08);

vi) Repetition of words of increasing syllable length (max score 30: SPF mean 20.05 (SD 8.60); SHF mean 19.39 (SD 9.64); t(42) =.237, p=.814);

vii) Non-speech oromotor tasks (e.g., voluntary cough; lateral tongue movement; alternating lip rounding and spreading) - max score 15: SPF mean 8.07 (SD 5.57); SHF mean 8.57 (SD 4.81); t(42) =-.348, p=.729);

viii) Mean phonation time in seconds across 3 trials (SPF mean 9.24 (SD 6.34); SHF mean 8.4 (SD 5.27); t(41)** =-.348, p=.729, ** 1 participant in the SHF group was unable to complete the assessment);

ix) Diadochokinetic (DDK) rate for multiple repetitions of /t/ (‘tuh’) - (percentage correct: SPF mean 39.16% (SD 40.16%); SHF mean 31.73% (SD 29.43%); t(42) =-.704, p=.485); syllables per second: SPF mean 3.12 (SD 1.55); SHF mean 2.56 (SD 1.62); t(42) =1.163, p=.251);

x) DDK for multiple repetitions of the sequence /p t k/ (‘puh tuh kuh’) - (percentage correct: SPF mean 6.58% (SD 13.68%); SHF mean 3.93% (SD 8.29%); t(42) =.786, p=.437); syllables per second: SPF mean 2.30 (SD 1.03); SHF mean 2.04 (SD 1.01); t(42) =.823, p=.415).

Speech assessment materials: speech repetition task

Outcomes of therapy were evaluated on a number of tasks and one of these – a word repetition task – is the focus of the current report. Participants repeated a set of 105 words after a single presentation by the experimenter. The set included 35 items that were treated at all levels in the speech program (T), 35 items which were matched as closely as possible on phonetic complexity (Phonetically Matched - PM), and 35 items matched on frequency (Frequency Matched - FM) using spoken lexical frequency data from the British National Corpus (BNC) (Leech, Rayson & Wilson, 2001). The treated items were chosen for their functionality in everyday communicative contexts. The three word sets are given in Appendix I.

Assessment of speech accuracy and performance

All productions elicited via the speech repetition task were scored for accuracy by a researcher who had not been involved in the testing of participants, and who was blind to the order of treatment (i.e., SPF vs. SHF). The accuracy scoring protocol assessed the presence of fluent and error free production, groping and struggle prior to and during production, and phonetic accuracy. The details of the accuracy scoring protocol are provided in Appendix II.

Inter-rater reliability measures

Accuracy scoring was repeated for a subset of the speech repetition data by a second rater who was blind to the first rater's scores. This was done for 558 speech samples drawn from 16 participants with AOS. They were sampled from the five different assessment sessions and included 8 participants from the SPF group, and 8 participants from the SHF group. In addition, the samples represented individuals with different levels of AOS severity (mild: n=3, moderate: n=3; and severe: n=10). Spearman's Rank Correlation indicated a high level of inter-rater reliability (n=558, *rho*=.895, p<.0001).

Accuracy scores and determining therapy outcome measures

The accuracy scores were used to characterise output that was fluent and error free, and that which was errorful and contained struggle and articulatory groping behaviours.

These two sets of behaviours were selected to represent speech production ability along the continuum within the speech production accuracy assessment protocol (see Appendix II). Frequency totals (the number of occurrences) were calculated for each of the 13 categories of scores (Appendix II) for all participants for the T, PM, FM items for all five test sessions. Responses that contained articulatory groping and/or struggle, or failure to produce the target (scores 0, 1, 2, 2c, 3b, 4b, 5) were categorised as Struggle. The number of responses that reflected prompt, fluent and error free production (scores of 7) were coded within the Fluent category. Responses that were phonetically aberrant either in segmental accuracy or duration, but contained no struggle or grope behaviours were not analysed in this study.

Program usage levels

The program usage levels are given in Table 1 for both programs across all participants, and by group (SPF and SHF) to the nearest minute. Program usage levels per week are also indicated. There was a tendency to use the first program more than the second. However, there were no significant group differences in the usage levels of either the speech ($t(42) = 1.294$, $p=.203$), or the sham program ($t(42)= -.863$, $p=.393$). A total of 12 participants completed the entire speech program (i.e., all levels of both the Input and Speaking tasks). A further 21 participants completed the speech program up to Speaking Level 4 at which stage the word repetition tasks are completed. Eleven participants failed to reach this therapy stage. We report results for all 44 participants regardless of therapy achievement.

<PLACE TABLE 1 ABOUT HERE>

Statistical Analysis

The Struggle and Fluent data sets were analysed separately. Figures 1a to 3a and Figures 4a to 6a illustrate the means (+/- 1.0 Standard Errors) for the Struggle and Fluent scores, respectively, by test session (Baseline 1, Baseline 2, PTxSpeech, PTxSham, Maintenance), item set (T, PM, FM), and order of treatment (speech first (SPF) and visual sham first (SHF)).

Figures 1b to 3b and Figures 4b to 6b provide the means (+/- 1.0 Standard Errors) by test session (Baseline 1, Baseline 2, PTxSpeech, PTxSham, Maintenance), and item set (T, PM, FM) for both SPF and SHF groups combined.

The categories used in the frequency counts for the Struggle and Fluent scores were based on a rating scale that was both ordinal and nominal in nature. Non-parametric analyses were therefore used to assess differences in the scores across the five test sessions. However, due to the two-phase, cross-over design of the study, it was first necessary to examine whether the Struggle and Fluent scores for the PTxSpeech and PTxSham sessions were affected by order of treatment to determine whether data from the two phases could be combined. A mixed model ANOVA was conducted to test for order of treatment effects. A parametric approach was adopted to investigate key interaction effects that are not easily assessed with standard non-parametric statistical methods. Key comparisons between outcomes of the parametric and non-parametric analyses are noted in the results below.

RESULTS

Investigating order of treatment effects

A mixed model ANOVA was conducted for the Struggle scores with test session (B1, B2, PTxSpeech, PTxSham, Maintenance) and items (T, PM, FM) as repeated measures and order of treatment (SPF and SHF) as the between subjects factor. Results of the ANOVA revealed significant main effects for test session (F(4, 168)=8.496, p<.0001) and items (F(2, 84)=8.802, p<.0001), but no significant order of treatment effects (F(1, 42)=.824, p=.369). There were significant interaction effects for test session x items (F(8,336) =3.270, p=.001), but no significant interaction effects for test session x order of treatment (F(4, 168)=.093, p=.985), items x order of treatment (2, 84)=.561, p=.573) or test session x items x order of treatment (F(8, 336)=1.317, p=.234). Due to the lack of significant interaction effects between test session and order of treatment, the data for the SPF and SHF groups were combined for further statistical analysis.

A mixed model ANOVA was also conducted for the Fluent scores with test session (B1, B2, PTxSpeech, PTxSham, Maintenance) and items (T, PM, FM) as repeated measures, and order of treatment (SPF and SHF) as the between subjects factor. Results of the ANOVA revealed significant main effects for test session (F(4, 168)=15.234, p<.0001) and items (F(2, 84)=40.544, p<.0001), but no significant order of treatment effects (F(1, 42)=.981, p=.328). There was a marginal significant interaction effect for test session x items (F(8,336) =1.970, p=.05), and test session x items x order of treatment (F(8, 336)=1.937, p=.054) due to the SPF group displaying higher levels of treatment maintenance effects for the T, PM and FM items. These results were replicated in the non-parametric statistics which are described in more detail below. There were no significant interaction effects for test session x order of treatment (F(4, 168)=.1.978, p=.100), or items x order of treatment (F(2, 84)=.433, p=.650). As was the case for the ‘Struggle’ scores, due to the lack of significant interaction effects between test session and order of treatment, the Fluent data were combined for the SPF and SHF groups for further statistical analysis.

Struggle scores

Two-tailed Wilcoxon Signed Ranks tests were conducted to test for differences across the five test sessions for the T, PM, and FM data sets. Some additional tests were conducted to examine differences between the SPF and SHF groups. The results of these tests revealed effects which were parallel to those of the mixed model ANOVA test (see *Investigating order of treatment effects*). Due to the number of multiple comparisons conducted (n=15) using the Wilcoxon Signed Ranks tests, the alpha level (p=.05) was adjusted to avoid type-1 errors. The adjusted alpha level *(p = .05/15 = .0033*) was used. There were no significant differences between the 'Struggle' scores for B1 and B2 for the T (Z=-.690, p=.490 – see Fig 1b), PM (Z=-1.849, p=.064– see Fig 2b) and FM (Z=-1.222, p=.222– see Fig 3b) items. This indicates that performance on the speech repetition task was stable across the two baselines before the first treatment period.

Comparisons between B2 and PTxSpeech revealed significant reductions in the Struggle scores for the T items (Z=-3.764, p<.0001 – see Fig 1b), but not the PM items (Z=-2.049, p=.040 – see Fig 2b), or the FM items (Z=-1.351, p=.177 – see Fig 3b). This indicates that the speech intervention resulted in participants displaying a significant reduction in struggle in repetition for those items which were directly treated within the speech program. Although there was a trend indicating that the reduction of struggle generalised to phonetically similar untreated items (see Fig 2b), this did not reach significance using the adjusted alpha level.

<PLACE FIGURES 1a, 1b, 2a, 2b, 3a, 3b ABOUT HERE>

Effects of the sham program were examined by comparing Struggle scores for B2 and PTxSham. Analyses of the combined group data revealed no significant reductions in the scores for the T items (Z=-2.084, p=.037 – see Fig 1b), the PM items (Z=-1.955, p=.051 – see Fig 2b), or the FM items (Z=-1.303, p=.193 – see Fig 3b). However, on examination of the T items by individual groups, the SPF group displayed evidence of a reduction in their scores (Z=-3.265, p=.001 – see Fig 1a), which was not the case for the SHF group (Z=-.406, p=.684 – see Fig 1a). Rather than indicating a reduction in struggle following the sham program in the SPF group, this finding provides evidence for enduring speech treatment effects on the T items. This was confirmed by a further comparison between the PTxSpeech and PTxSham scores for the T items for the SPF group which revealed no significant changes in the 'Struggle' scores (Z=-1.736, p=.083 – see Fig 1a).

The maintenance of treatment effects was examined by comparisons between the PTxSpeech and Maintenance Struggle scores for both groups combined. Results indicated that there were no significant differences for the T items (Z=-.914, p=.361), PM items (Z=-.279, p=.780) or the FM items (Z=-1.265, p=.206). This indicates that reductions in Struggle scores following treatment were maintained. Given the two-phase design element of the study, this effectively amounted to a significant maintenance effect after 18 weeks for the SPF group, who received the speech treatment in the earlier period of the study, and 8 weeks for the SHF group who received the speech treatment in the later period.

Fluent scores

A further series of two-tailed Wilcoxon Signed Ranks tests was conducted to test for differences in the Fluent scores across the five test sessions for the T, PM, and FM data sets. Some additional tests were conducted to examine differences between the SPF and SHF groups. Similar to the Struggle scores, the results of these tests paralleled those of the mixed model ANOVA summarised above (see *Investigating order of treatment effects*). The alpha level (p=.05) was again adjusted to control for type-1 errors due to the multiple comparisons conducted (n=26). The adjusted alpha level *(p=.05/26 = .0019*) was used. There were no significant differences between the Fluent scores for B1 and B2 for the T items (Z=-2.209, p=.027 – see Fig 4b), PM items (Z=-467, p=.641 – see Fig 5b), or the FM items (Z=-1.619, p=.106 – see Fig 6b). The trend for an improvement in the Fluent scores for the T items between B1 and B2 (see Fig 4b) was examined further by group. This trend for a practice effect appeared to be carried by the SPF group (see Fig 4a). However, this was not significant using the adjusted alpha level (Z=-2.346, p=.019). There was no significant difference between B1 and B2 for the SHF group (Z=-.829, p=.407) (see Fig 4a). This indicates that performance on the speech repetition task was largely stable across the two baselines before the first treatment period.

<PLACE FIGURES 4a, 4b, 5a, 5b, 6a, 6b ABOUT HERE>

Comparisons between B2 and PTxSpeech revealed significant increases in the Fluent scores for the T items (Z=-3.825, p<.0001 – see Fig 4b), the PM items (Z=-3.611, p<.0001 – see Fig 5b), but not the FM items (Z=-2.884, p=.004 – see Fig 6b). This indicates that the speech program resulted in a significant increase in speech accuracy and fluency in repetition for the treated items, with generalisation to the phonetically matched items.

Effects of the sham program were examined by comparing the Fluent scores for B2 and PTxSham. There were trends but no significant increases in the scores for the T items (Z=-2.970, p=.004 – see Fig 4b) or the PM items (Z=-2.089, p=.037 – see Fig 5b). The FM items displayed no significant increases in the scores (Z=-1.577, p=.115 – see Fig 6b). The trends for increases in the Fluent scores were examined further by group for both the T and PM items. For the T items the marginally significant effect was carried by the SPF group (Z=-3.476, p=.001 – see Fig 4a), but not the SHF group (Z=-.900, p=.368 – see Fig 4a). This was confirmed by a further comparison between the PTxSpeech and PTxSham scores for the T items for the SPF group which revealed no significant changes in the Fluent scores (Z=-.629, p=.489 – see Fig 4a). For the PM items the increases in Fluent scores across both groups were not significant (Z=-2.089, p=.037 – see Fig 5b). This result was reflected in the data for both the SPF (Z=-1.181, p=.238) and the SHF groups (Z=-1.633, p=.103) (see Fig 5a).

The speech program treatment maintenance effects were examined by comparisons between the PTxSpeech and Maintenance 'Fluent' scores for both groups combined. Results indicated that there were trends for decreases in the scores for the T items (see Fig 4b), but these did not reach significance (Z=-2.402, p=.016) using the adjusted alpha levels. Further examination of the Fluent scores by group, showed that neither the SHF group (Z=-2.705, p=.007 – see Fig 4a) nor the SPF group (Z=-.732, p=.464 – see Fig 4a) displayed a significant decrease in the Fluent scores between PTxSpeech and Maintenance using the adjusted alpha levels. This suggests that treatment effects were maintained for the T items in the SPF group, but there was some evidence for only temporary gains in the SHF group. Results for the PM items indicated no significant difference between the PTxSpeech and Maintenance Fluent scores for both groups combined (Z=-1.515, p=.250 – see Fig 5b), indicating that the treatment effects were maintained. Again however, there was a difference between the two participant groups with stronger maintenance displayed by the SPF group (Z=-1.515, p=.250 – see Fig 5a) than the SHF group (Z=-2.447, p=.014 – see Fig 5a). Results for the FM items indicated no significant differences between the PTxSpeech and Maintenance Fluent scores for both groups combined (Z=-.875, p=.382 – see Fig 6b). This stable performance was, once again, carried by the SPF group (Z=-1.444, p=.149 – see Fig 6a), while the SHF group showed more fragile behavioural gains (Z=-2.361, p=.018 – see Fig 6a). In summary, the results indicate that evidence of maintenance was apparent for the SPF group, and, given the two-phase cross-over design of the study, the results for the SPF group amounted to significant maintenance of behavioural gains for up to 18 weeks after the intervention.

Control assessments

There were no significant differences between the CAT scores at the start of the study (T1), and the end of the study (T2) (CAT 32 - T1: Mean=20.75, SE=1.07; T2: Mean=20.82; SE=1.05; t(43)= -0.113, p=.911). Similarly, no significant differences were found between T1 and T2 for the PALPA 48 sub-test scores (T1: mean=34.00 (SE=1.33); T2: mean=34.48 (SE=1.33); t(42)=-1.017, p=.315). This indicates that neither the speech nor the sham programs had any effect on untreated language behaviours.

Correlations between 'Fluent' and Struggle' Scores

A series of Spearman's Rank Correlation tests was run on the 'Struggle' and 'Fluent' scores within assessment points (i.e., B1, B2, PTxSpeech, PTxSham, Maintenance) and word item sets (i.e., T, PM, and FM). These results showed a pattern of significant negative correlations between the 'Struggle' and 'Fluent' scores which ranged from *rho*= -.70 (n=44, p<.0001) for the PTxShamTreated items, to *rho*=-.82 (n=44, p<.0001) for the

B2Treated items. The negative values in the correlation coefficients are indicative of lower levels of struggle being significantly correlated with higher levels of fluency. However, the range in the *rho* values indicates that while the two scales are related, the two measures tap distinctively different behaviours and there was variation in the 'struggle' and 'fluency' scores for participants across phases of the study.

DISCUSSION

We report the outcomes of an intervention study for the speech production impairments of AOS using a word repetition task. The study employed a two-phase cross-over design (Cowell et al., 2010), in which the effects of a self-administered speech program based on errorless and error reducing principles were compared to those of a visuo-spatial sham intervention. The speech program targeted whole word production, in contrast to traditional therapies for AOS which train the deliberate and conscious production of isolated articulatory gestures (Varley & Whiteside, 2001). Articulatory therapies have shown limited generalisation, and may only serve to reinforce the high levels of conscious control and effort in speaking, and the loss of automaticity observed in AOS (Lebrun, 1990; Whiteside & Varley, 1998). To some degree, articulatory therapy might result in learned misuse of the automatic, procedural routines of healthy speech control. The aim of the speech intervention in this study was that participants would practice fluent, automatic, cohesive and error free speech, and that struggle and effort would be minimised.

The effects of the intervention on articulatory groping and struggle on one hand, and accuracy and fluency on the other were evaluated for word repetition. The results indicated significant treatment gains following the speech intervention. This was reflected in significant reductions in the Struggle scores (see Fig 1b). In addition, significant increases were found in the Fluent scores (see Fig 4b) for the treated words, and this treatment effect was present in both groups of participants, irrespective of the order in which programs were administered. The results also supported generalised treatment effects beyond the treated items, with the phonetically matched words displaying significant increases in Fluent scores. The reduced levels of struggle and the increased levels of fluency lend support to the effectiveness of the speech intervention program and suggest that the principles underpinning the program contributed to reducing error and struggle, and in enhancing fluent and error free speech production in apraxic speakers.

The treatment effects were not due to more general placebo-type effects, for example, resulting from intensive stimulation or the positive effects of participating in a research study of a computer therapy. Although some behavioural change was evident in the sham phase, these were evident only in the participant group who were administered the speech program first. The change in scores during this phase could be attributed to the continuing consolidation of the effects of the speech treatment. The participant group who received the sham program first displayed no change in speech following this intervention.

In addition to gains in speech control after the speech intervention, longer term maintenance effects after 18 and 8 weeks were evident in the speech first and sham first groups, respectively. Maintenance effects for the reduction in Struggle scores appeared to be robust for both groups across the all word sets. This suggests that the speech program was effective in reducing struggle. In the case of the Fluent scores, the maintenance effects were robust for the speech first group across all word sets. This result might be due to effects of the order of the interventions interacting with individual differences in the severity profiles of participants within each group, or with the individual differences in program usage levels. Future research will explore the profiles of participants who differ in their maintenance performance. Furthermore, therapy outcome in relation to stages achieved in the intervention programme will be determined.

The results of the current study suggest that designing interventions for motor speech impairments incorporating the principles of errorless learning and error reduction represents a useful therapeutic strategy (Fillingham et al., 2003; 2006). Such programs appear to result in improvements in word accuracy and fluency, together with reduced levels of articulatory groping and struggle. These improvements show a degree of generalisation beyond treated word forms, and also maintenance in some cases after the intervention is withdrawn. However, the intervention reported here involved a number of components, not only errorless learning. It may be that the outcomes were due to any number of these principles and their interactions. For example, the outcomes might be due to working at the level of whole words, or the intensity of treatment that is possible under a computer, self-administered therapy model (Varley, 2011). However, previous research into intervention for AOS has not always revealed encouraging results in terms of treatment gains and their generalisation and maintenance. We therefore suggest that therapies that emphasise fluent, normalised speech production, and how they compare to traditional articulatory therapy for AOS, deserve further investigation.

References

Ardila, A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes. *Aphasiology, 24*, 363-394.

Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: amnesia and the problem of error elimination. *Neuropsychologia*, *32*, 53-68.

Conroy, P., Sage, K. & Lambon Ralph, M. A. (2009). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology*, 23 (11), 1311-1337.

Corkin, S. 2002. What's new with the amnesic patient HM? *Nature Reviews Neuroscience, 3,* 153-160.

Cowell, P.E., Whiteside, S.P., Windsor, F., & Varley, R.A. (2010). Plasticity, permanence and patient performance: study design and data analysis in the cognitive rehabilitation of acquired communication impairments. *Frontiers in Human Neuroscience*, Special Issue on Cognitive Neurorehabilitation, Volume 4, Article 213, 1-12.

Crompton, A. (1981). Syllables and segments in speech production, in A. Cutler (ed.) *Slips of the Tongue and Language Production*, (pp.109-162), Berlin: Mouton.

Dabul, B. (1986). Apraxia Battery for Adults. Tigard, OR: CC Publications.

Fillingham, J.K., Hodgson, C., Sage, K. & Lambon Ralph, M.A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation, 13*, 337-363.

Fillingham, J.K., Sage, K. & Lambon Ralph, M.A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation, 16*, 129-154.

Franklin, S., Turner, J. E., & Ellis, A. W. (1992). *The ADA Comprehension Battery*. London: Action for Dysphasic Adults.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic Assessments of Language Processing in Aphasia.* Hove, UK: Psychology Press.

Keller, E. (1987). The cortical representations of motor processes of speech, in E.Keller and M.Gopnik (eds.) *Motor and Sensory Processes of Language*, Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Kent, R. D., & Rosenbek, J. C. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language*, *15*, 259-291.

Kent, R. D., & Rosenbek, J. C. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research, 26*, 231-249.

Lebrun, Y. (1990). Apraxia of speech: a Critical Review. *Journal of Neurolinguistics*, *5 (4),* 379-406.

Lecours, A. R. & Lhermitte, F. (1976). The "pure form" of the phonetic disintegration syndrome (pure anarthria); Anatomo-clinical report of a historical case. *Brain and Language*, *3*, 88-113.

Leech, G., Rayson, P. & Wilson, A. (2001). Word frequencies in written and spoken English: based on the British National Corpus. London: Longman, Pearson Education.

Levelt, W.J.M. & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239-269.

McNeil, M.R., Robin, D.A., & Schmidt, R.A. (2009). Apraxia of speech: definition and differential diagnosis. In M.R. McNeil (Ed.) Clinical Management of Sensorimotor Speech Disorders (pp. 249-67). 2nd Edition. New York: Thieme Medical Publishers.

Peach, R. K. (2004). Acquired apraxia of speech: features, accounts and treatment. *Topics in Stroke Rehabilitation*, 11(1), 49-58.

Pulvermüller, F. & Berthier, M. L. (2008). Aphasia therapy on a neuroscience basis. *Aphasiology, 22*, 563-599.

Raymer, A., Strobel, J., Prokup, T., Thomason, B., & Reff, K. (2010). Errorless versus errorful training of spelling in individuals with acquired dysgraphia. Neuropsychological Rehabilitation, 20, 1-15.

Rosenbek, J. C. (1985). Treating apraxia of speech, in D. F. Johns (Ed.), *Clinical Management of Neurogenic Communicative Disorders* (pp. 267-312). Boston, USA: Little Brown.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology, 210*, 419-421.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production, in W.E Cooper and E.C.T Walker (eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*, (pp. 295-342), Hillsdale, N.J.: Lawrence Erlbaum.

Swinburn, K., Porter, G. & Howard, D. (2004). *Comprehensive Aphasia Test*. New York, Psychology Press.

Tailby, R., & Haslam, C. (2003). An investigation of errorless learning in memory-impaired patients: improving the technique and clarifying the theory. *Neuropsychologia, 41,* 1230-1240.

Taub, E., Miller, N. E.,Novack, T. A., Cook, E. W., Fleming, W. C., Nepomuceno, C. S. et al., (1993). Technique to improve chronic motor deficit after stroke. *Archives of Physical Medicine and Rehabilitation, 74*, 347-354.

Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews Neuroscience, 2*, 717-726.

Varley, R. (2011). Rethinking aphasia therapy: a neuroscience perspective. *International Journal of Speech-Language Pathology*, *13,* 1, 11-20.

Varley, R. A. & Whiteside, S. P. (2001). What is the underlying impairment in acquired apraxia of speech? *Aphasiology, 15*, 39-49.

Varley, R. A., Whiteside, S. P., & Luff, H. (1999). Dual-route speech encoding in normal and apraxic speakers: some durational evidence. *Journal of Medical Speech - Language Pathology*. *7*, 127-132.

Wambaugh, J. L., Duffy, J. R., McNeil, M. R., Robin, D. A., & Rogers, M. A. (2006). Treatment guidelines for acquired apraxia of speech: a synthesis and evaluation of the evidence. *Journal of Medical Speech-Language Pathology*, 14, xv-xxxiii.

Watkins, K.E., Strafella, A. P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41,* 989-994.

Wertz, R.T., LaPointe, L.L., & Rosenbek, J.C. (1984). Apraxia of Speech in Adults: The Disorder and Its Management. Orlando: Grune & Stratton.

Whiteside, S. P., Grobler, S., Windsor, F. & Varley, R. A. (2010). An acoustic study of vowels and coarticulation as a function of utterance type: a case of acquired apraxia of speech. *Journal of Neurolinguistics*, *23,* 145-161.

Whiteside, S. P., & Varley, R. A. (1998a). A reconceptualisation of apraxia of speech: a synthesis of evidence. *Cortex, 34*, 221-231.

Whiteside, S. P., & Varley, R. A. (1998b). Coarticulation in apraxia of speech: an acoustic study of non-words. *Logopedics Phoniatrics Vocology*, *23*, 155-163.

Wilson, B.A., Baddeley, A., Evans, J., & Shiel, A. (1994). Errorless learning in the rehabilitation of memory impaired people. *Neuropsychological Rehabilitation, 4,* 307-326.

Ziegler, W., & von Cramon, D. (1985). Anticipatory coarticulation in a patient with apraxia of speech. *Brain & Language*, *26*, 117-130.

Ziegler, W., & von Cramon, D. (1986).Disturbed coarticulation in apraxia of speech: acoustic evidence. *Brain & Language*, *29*, 34-47.

ACKNOWLEDGMENTS
The study reported here was funded by the BUPA Foundation Specialist Grant. We wish to thank the participants and their families for assisting with the study. We also wish to thank the NHS therapists who assisted the project.

DISCLOSURE OF CONFLICTS OF INTEREST
Drs. R. A. Varley and S. P. Whiteside are co-authors of a commercially available computer speech program used to treat post-stroke speech production impairments. They receive royalties from the sale of the software. This software was used in the treatment study reported in this paper. The remaining authors have no commercial interest in the software.

Table 1. Usage levels for both the speech and the visual sham programs. The number of sessions is indicated together with statistics of the active time (all time values are to the nearest minute, except for Hours per week) spent on both computer programs. Data are provided across all participants and by group (SPF and SHF).

	Speech Program active time	Number of Speech Program sessions	SHAM active Time	Number of SHAM sessions
ALL PARTICIPANTS (n=44)				
Minimum	212	3	71	4
Maximum	3029	85	3129	80
Total	47341	1904	42682	1679
Mean	1076	43	970	38
S.E. Mean	82.76	2.96	105.92	3.14
Hours per week	3		2.7	
SPF GROUP (n=21)				
Minimum	254	3	71	4
Maximum	3029	85	2322	80
Total	24930	948	18357	733
Mean	1187	45	874	35
S.E. Mean	135.20	5.56	144.93	4.85
Hours per week	3.3		2.4	
SHFGROUP (n=23)				
Minimum	212	19	137	10
Maximum	1996	64	3129	76
Total	22411	956	24325	946
Mean	974	42	1058	41
S.E. Mean	97.23	2.62	154.22	4.05
Hours per week	2.7		2.9	

Figure 1a. ‘Struggle’ scores by test session and order of treatment – Treated Items

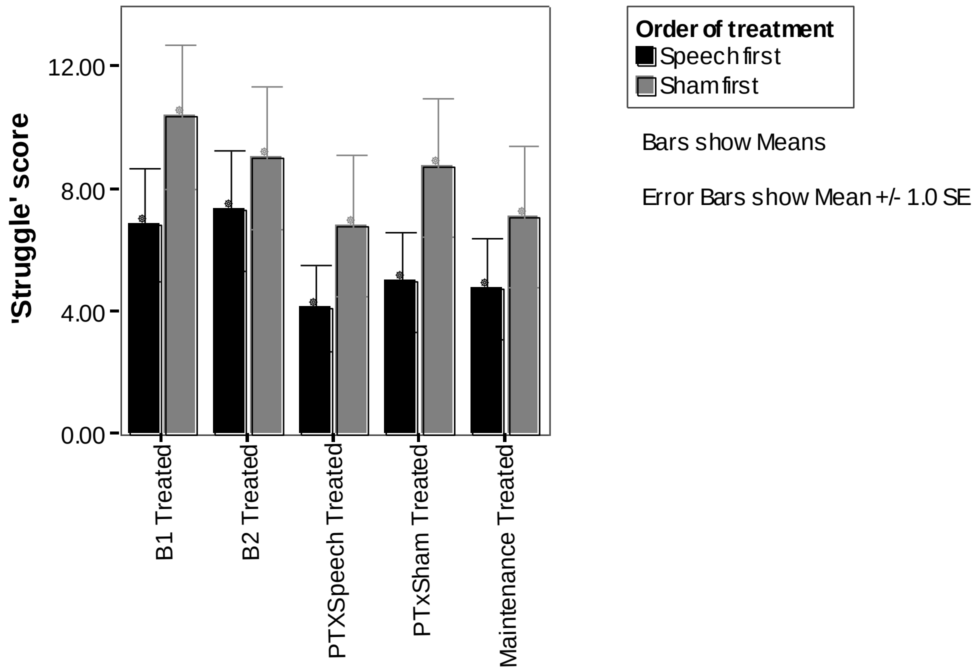


Figure 1b. ‘Struggle’ scores by test session for SPF and SHF groups combined – Treated Items (bars show means and error bars show +/- 1.0 SE)

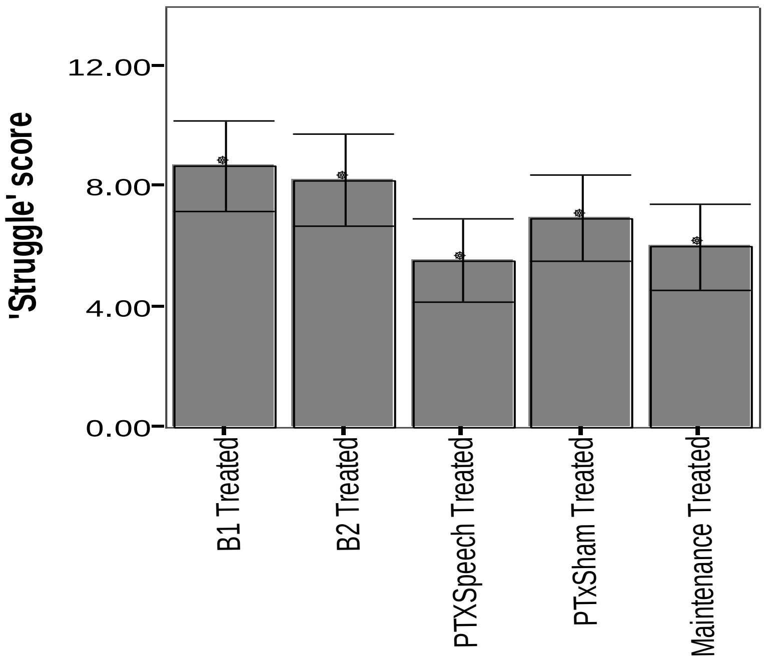


Figure 2a. ‘Struggle’ scores by test session and order of treatment – Phonetically Matched Items

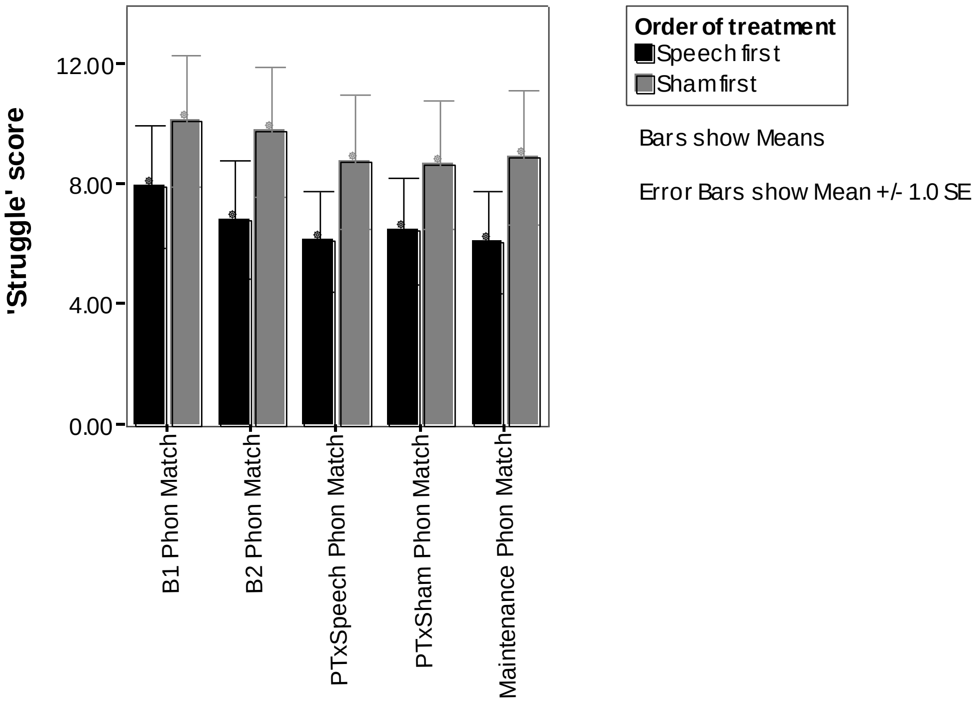


Figure 2b. ‘Struggle’ scores by test session for SPF and SHF groups combined – Phonetically Matched Items (bars show means and error bars show +/- 1.0 SE)

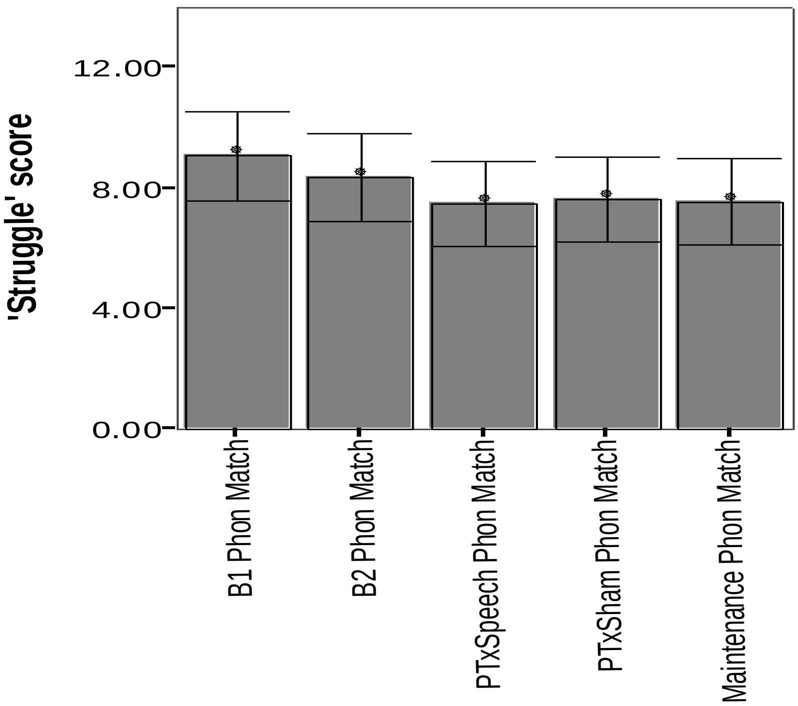


Figure 3a. ‘Struggle’ scores by test session and order of treatment – Frequency Matched Items

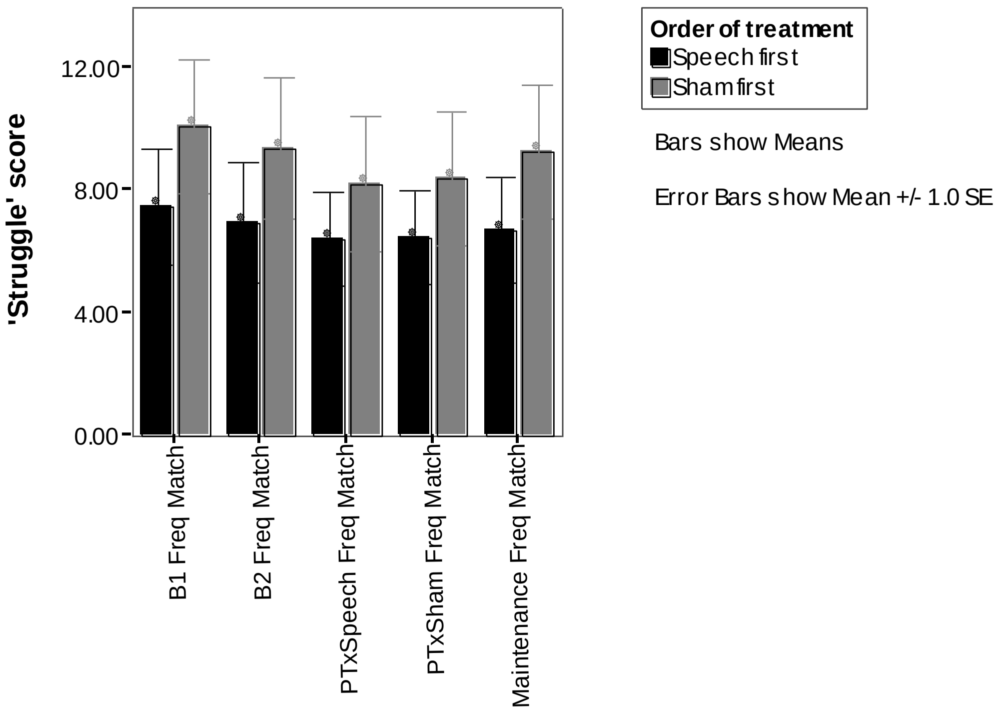


Figure 3b. ‘Struggle’ scores by test session for SPF and SHF groups combined – Frequency Matched Items (bars show means and error bars show +/- 1.0 SE)

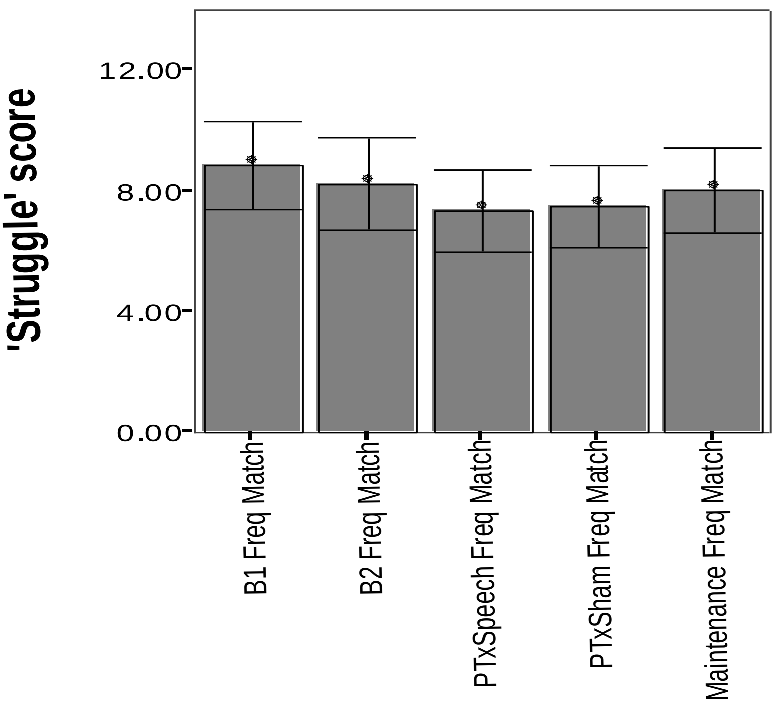


Figure 4a. 'Fluent' scores by test session and order of treatment – Treated Items

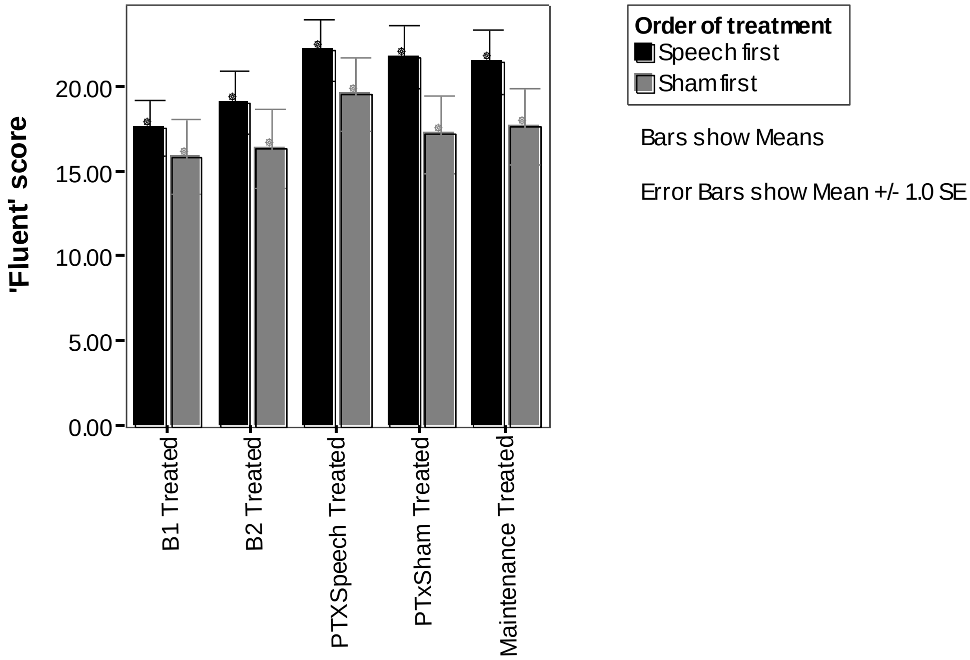


Figure 4b. 'Fluent' scores by test session for SPF and SHF groups combined – Treated Items (bars show means and error bars show +/- 1.0 SE)

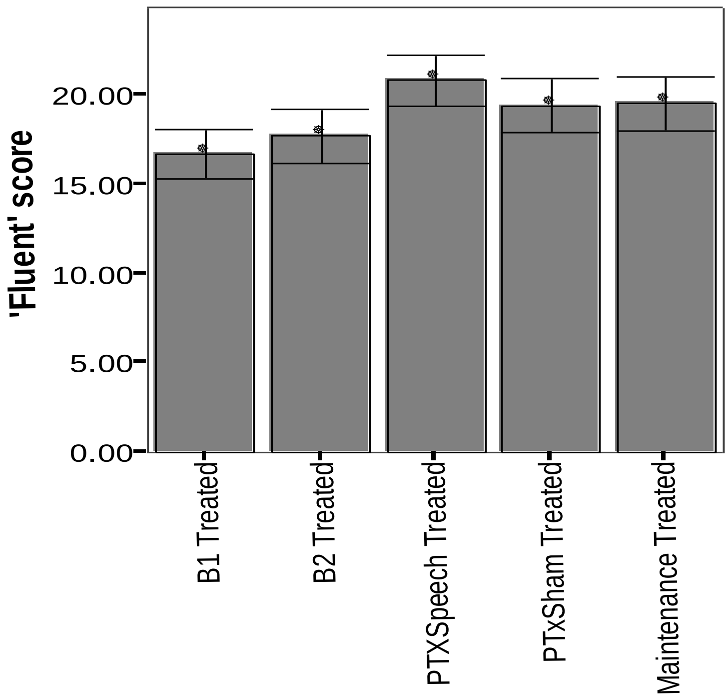


Figure 5a. ‘Fluent’ scores by test session and order of treatment – Phonetically Matched Items

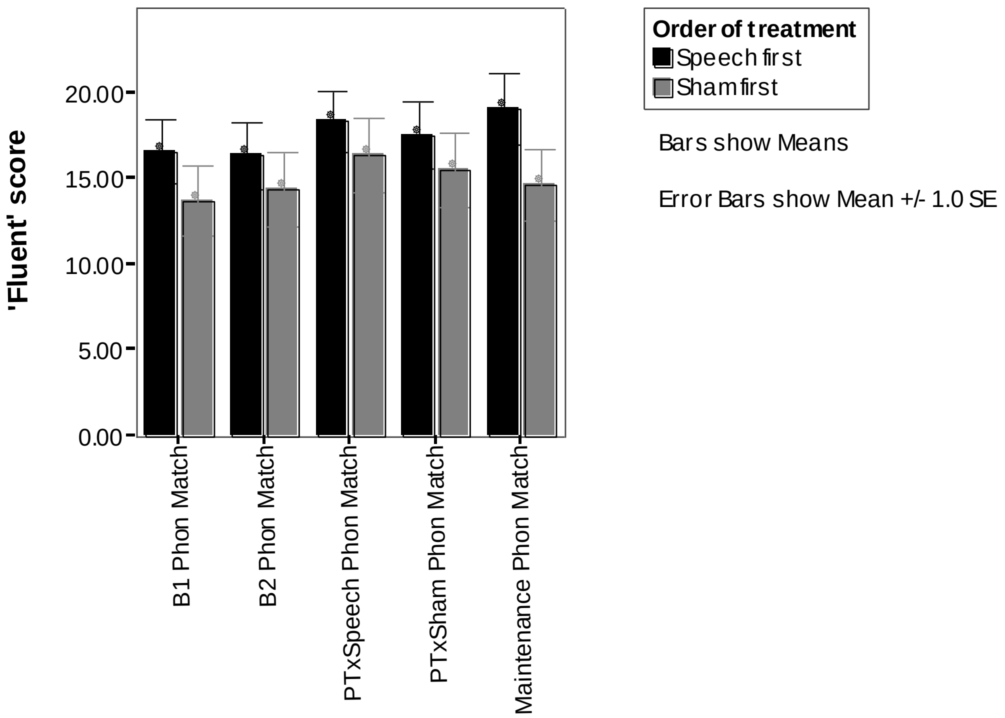


Figure 5b. ‘Fluent’ scores by test session for SPF and SHFgroups combined – Phonetically Matched Items (bars show means and error bars show +/- 1.0 SE)

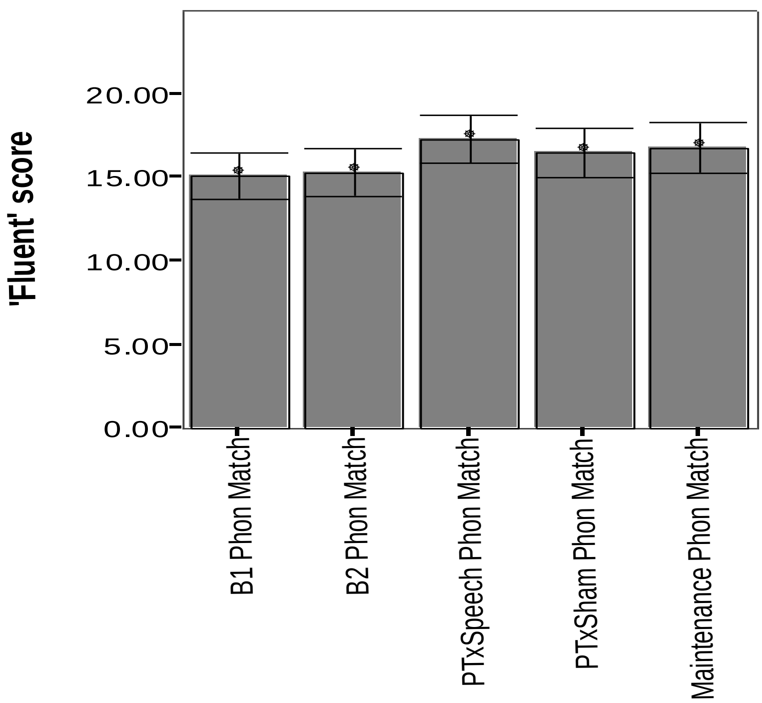


Figure 6a. ‘Fluent’ scores by test session and order of treatment – Frequency Matched Items

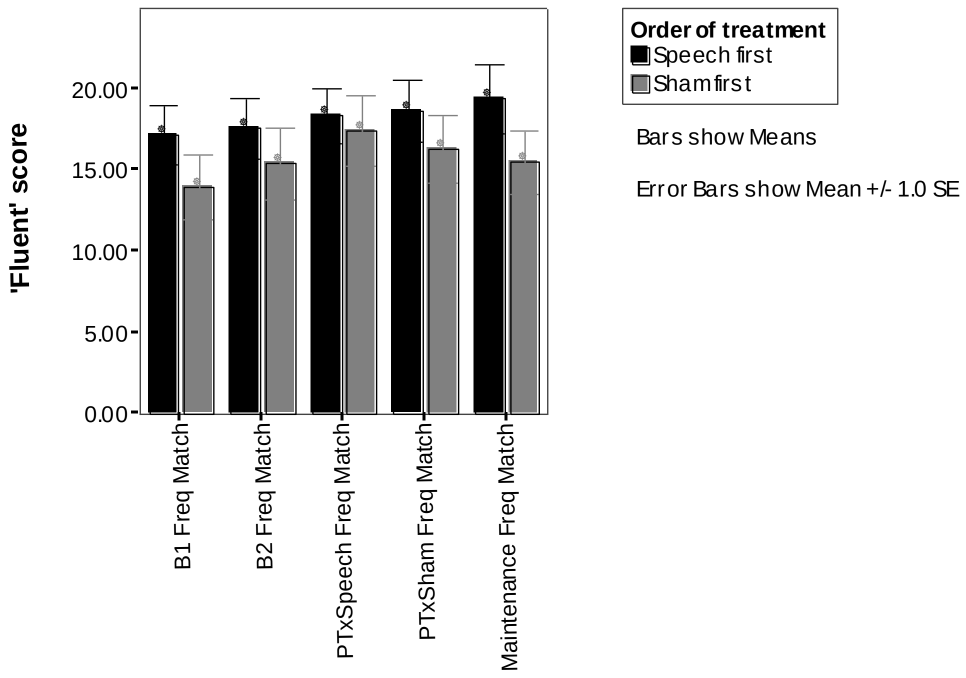
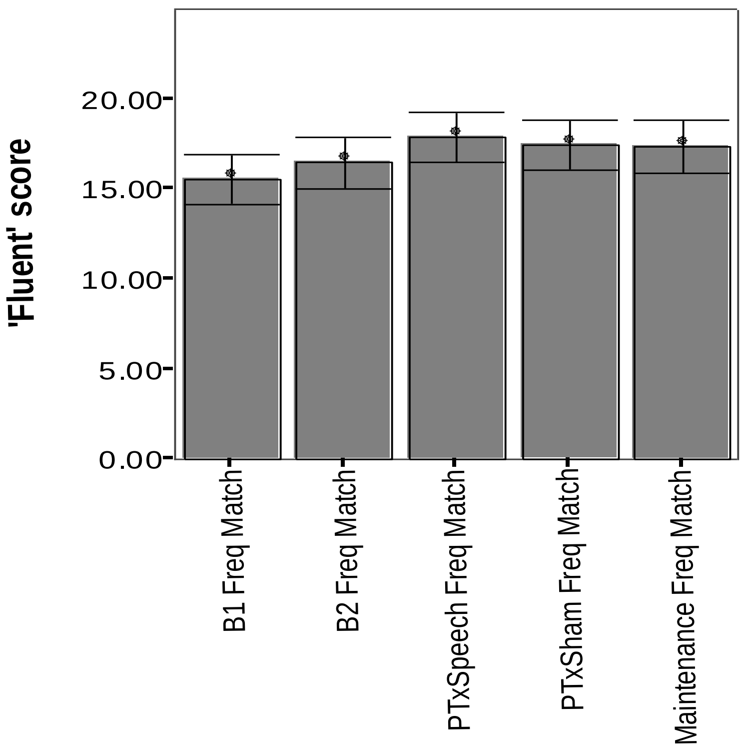


Figure 6b. 'Fluent' scores by test session for SPF and SHFgroups combined – Frequency Matched Items (bars show means and error bars show +/- 1.0 SE)



Appendix I. Treated, Phonetically Matched and Frequency Matched lexical items included in the repetition task. Lexical frequencies (spoken – Leech, Rayson & Wilson, 2001) are provided in brackets where available.

TREATED ITEMS	**PHONETICALLY MATCHED ITEMS**	**FREQUENCY MATCHED ITEMS**
MEAT (41)	SHEET (33)	FISH (69)
HOLIDAY (97)	SATURDAY (147)	ANIMAL (26)
PARTY (227)	FORTY (388)	WATER (278)
HOSPITAL (97)	FESTIVAL (13)	COMPUTER (85)
TEA (172)	BEE	BED (179)
BRAIN (26)	TRAIN (56)	TEETH (33)
DOG (128)	DOOR (225)	GAME (112)
BATH (44)	PATH (22)	WASH (13)
WORK (601)	WORM	MAN (405)
HEAD (164)	SHED (16)	SHOP (188)
PHONE (144)	BONE (16)	TOWN (159)
FRIEND (102)	FRENCH (40)	DRINK (70)
LIGHT (90)	KITE	FOOT (84)
GIRL (156)	PEARL	BOY (159)
NIGHT (465)	WHITE (131)	HOUSE (460)
PRICE (121)	ICE (31)	LETTER (132)
SLEEP (35)	SHEEP (35)	SPORT (28)
CAR (379)	JAR	DAY (746)
BODY (109)	BABY (93)	TEACHER (67)
HELP (79)	HARP	BANK (85)
TABLET (22)	TOILET (40)	COOK (24)
HEALTH (154)	WEALTH	HAND (221)
GARDEN (88)	WARDEN	MUSIC (76)
BOOK (220)	HOOK (11)	BACK (270)
NURSE (21)	PURSE	PAIN (39)
COFFEE (90)	TOFFEE	PUDDING (20)
WALK (33)	CHALK	WINE (39)
HAIR (111)	CHAIR (88)	FOOD (117)
EYE (53)	TIE (14)	EAR (21)
KEY (38)	SEA (45)	BEER (20)
DOCTOR (106)	FACTOR (26)	RADIO (78)
LEG (50)	EGG (33)	HEART (56)
MORNING (459)	WARNING (13)	PAPER (232)
MEETING (178)	EATING (41)	WOMAN (138)
TALKING (428)	TOURING	CHILDREN (367)
Mean Frequency: 152.23 (SD:143.58; Range:21-601). Phonetic Complexity 1 syllable – n=24 2 syllables – n=9 3 syllables – n=2 Consonant Clusters - n=10	Mean Frequency: 61.1 (SD: 82.3; Range: 10-388) Phonetic Complexity 1 syllable – n=24 2 syllables – n=9 3 syllables – n=2 Consonant Clusters - n=6	Mean Frequency:145.54 (SD:154.96; Range: 13-746) Phonetic Complexity 1 syllable – n=24 2 syllables – n=8 3 syllables – n=3 Consonant Clusters - n=8

Appendix II. Accuracy Scoring Protocol for the Speech Repetition Task

Score	Description
7	*Accurate response – quick*: (a) response latency < 2 seconds. No verbalisation, groping and/or struggle behaviour is observed while response is formulated. (b) word duration is perceived to be similar to that of the experimenter.
6	*Accurate response – slow*: (a) response latency > 2 seconds. No verbalisation, groping and/or struggle behaviour is observed while response is formulated; **or** (b) word duration perceived to be rather longer than that of experimenter, and patient as own control, either due to slow release of consonants **or** lengthened vowels (i.e., can be one segment).
5	*Accurate response – false start*: verbalisation, groping and/or struggle behaviour is observed while response is formulated.
4b	*Accurate response – false start – slow*: word duration perceived to be rather longer than that of experimenter, and patient as own control, either due to slow release of consonants **or** lengthened vowels (i.e., can be on one segment). Verbalisation, groping and/or struggle behaviour is also observed.
4	*Inaccurate response – 1 segment*: (a) same place, distortion of consonant (b) vowel distortion (c) same place, same manner, voicing error on consonant (d) insertion of additional segment (e) deletion of segment (f) movement of segment (g) consonant substitution (h) lengthened closure phase of plosive
3c	*Inaccurate response – slow:* word duration perceived to be rather longer than that of experimenter, and patient as own control, either due to slow release of consonants **or** lengthened vowels (i.e., can be one segment). No verbalisation, groping and/or struggle behaviour is observed. In addition, response falls into category of 'inaccurate response – 1 segment', above.
3b	*Inaccurate response – false start*: verbalisation, groping and/or

	struggle behaviour is observed while response is formulated. In addition, response falls into category of 'inaccurate response – 1 segment', above.
3	*Inaccurate response – 2 segments*: (a) same place, distortion of consonant(s) (b) vowel distortion (c) same place, same manner, voicing error on consonant(s) (d) insertion of additional segments (e) deletion of segments (f) movement of segments (g) consonant substitutions (h) lengthened closure phase of plosive
2c	*Inaccurate response – false start – slow:* word duration perceived to be rather longer than that of experimenter, and patient as own control, either due to slow release of consonants **or** lengthened vowels (i.e., can be one segment). Verbalisation, groping and/or struggle behaviour is observed while response is formulated. In addition, response falls into category of 'inaccurate response – 1 segment', above.
2b	*Inaccurate response – slow:* word duration perceived to be rather longer than that of experimenter, and patient as own control, either due to slow release of consonants **or** lengthened vowels (i.e., can be one segmenst). No verbalisation, groping and/or struggle behaviour is observed. In addition, response falls into category of 'inaccurate response – 2 segments', above.
2	*Inaccurate response – false start*: verbalisation, groping and/or struggle behaviour is observed while response is formulated. In addition, response falls into category of 'inaccurate response – 2 segments', above.
1	*Inaccurate response – false start – slow:* word duration perceived to be rather longer than that of experimenter, and patient as own control, either due to slow release of consonants **or** lengthened vowels (i.e., can be one segment). Verbalisation, groping and/or struggle behaviour is observed while response is formulated. In addition, response falls into category of 'inaccurate response – 2 segments', above.

0	*No response < 10 secs.* *Three segmental errors.* *Completely off-target.* *Produced in a phrase.* *More than one presentation of a word is required (except for experimenter or technical error).*