Self-Administered Computer Therapy for Apraxia of Speech
Two-Period Randomized Control Trial With Crossover

Rosemary Varley, PhD; Patricia E. Cowell, PhD; Lucy Dyson, BMedSci; Lesley Inglis, PhD; Abigail Roper, MMedSci; Sandra P. Whiteside, PhD

Background and Purpose—There is currently little evidence on effective interventions for poststroke apraxia of speech. We report outcomes of a trial of self-administered computer therapy for apraxia of speech.

Methods—Effects of speech intervention on naming and repetition of treated and untreated words were compared with those of a visuospatial sham program. The study used a parallel-group, 2-period, crossover design, with participants receiving 2 interventions. Fifty participants with chronic and stable apraxia of speech were randomly allocated to 1 of 2 order conditions: speech-first condition versus sham-first condition. Period 1 design was equivalent to a randomized controlled trial. We report results for this period and profile the effect of the period 2 crossover.

Results—Period 1 results revealed significant improvement in naming and repetition only in the speech-first group. The sham-first group displayed improvement in speech production after speech intervention in period 2. Significant improvement of treated words was found in both naming and repetition, with little generalization to structurally similar and dissimilar untreated words. Speech gains were largely maintained after withdrawal of intervention. There was a significant relationship between treatment dose and response. However, average self-administered dose was modest for both groups. Future software design would benefit from incorporation of social and gaming components to boost motivation.

Conclusions—Single-word production can be improved in chronic apraxia of speech with behavioral intervention. Self-administered computerized therapy is a promising method for delivering high-intensity speech/language rehabilitation.

Clinical Trial Registration—URL: http://orcid.org/0000-0002-1278-0601. Unique identifier: ISRCTN88245643.

(Stroke. 2016;47:822-828. DOI: 10.1161/STROKEAHA.115.011939.)

Key Words: aphasia ■ apraxias ■ computers ■ cross-over studies ■ speech ■ therapeutics

Speech/language impairments after stroke are subcategorized into aphasia, dysarthria, and apraxia of speech (AOS). AOS is a disorder at the interface of language and speech production, involving breakdown in mapping from abstract linguistic representations to motor plans. Typical behaviors include speech errors, loss of automaticity and fluency, and altered timing parameters. In severe cases, patients may be nonverbal. Lesions causing AOS usually occur within the left cortical motor or somatosensory areas. Because of the proximity of speech control regions to left perisylvian cortex, AOS often co-occurs with aphasia.

Behavioral interventions for AOS involve 2 broad classes of therapies: bottom-up articulatory-kinematic therapies focus on individual speech sounds and top-down interventions aim to reestablish fluent production of larger linguistic units. Comparisons of outcomes for the 2 approaches are not conclusive. Intervention research has largely used quasi-experimental designs with nonrandom assignment. A meta-analysis and systematic review conclude that there is no randomized controlled trial evidence in support of intervention for AOS.

We report outcomes of an intervention for AOS combining these 2 therapeutic traditions. The intervention aimed to improve word production, with target forms ultimately placed in sentence frames. This approach acknowledges the common comorbidity of AOS with aphasia, allowing both linguistic and phonetic processes to be targeted. Trials of aphasia therapies indicate that lower intensity interventions have limited outcomes. Attempts to increase face-to-face therapy dose can result in high attrition rates because attending multiple appointments can challenge participants. The use of software programs, allowing participants to self-administer intervention, may circumvent this difficulty. A feasibility study reported that computer therapy is cost-effective and acceptable to patients with poststroke anomia.

We used a software therapy for AOS. It involved a perceptive stage (spoken word-picture matching, auditory-written
word matching, and auditory lexical decision), followed by a production stage. The perceptual component aimed to consolidate form-meaning representations of target vocabulary and facilitate feedforward input to motor representations. The production stage consisted of hierarchical speech activities. First, participants observed videos of word production, followed by blocks of trials requiring imagined production. The program then moved to overt word repetition with increasing delays between stimulus and response. Responses were audiorecorded by the software. The final stages involved more autonomous word production. Participants used trained words in sentence frames, followed by independent word retrieval/picture naming (program detail).

We explored the effectiveness of this intervention in a randomized controlled trial with a subsequent crossover period. Participants were randomly assigned to 1 of 2 order conditions: speech-first condition or sham-first condition. The sham intervention was another self-administered software program with identical interfaces but minimal speech/language content, involving visuospatial activities, for example, pattern matching and timed jigsaw completion. We report outcomes for the first intervention period and descriptively profile effects of the crossover. Power calculations based on an initial pilot study indicated that, for medium-to-small effect sizes (eg, 0.5–0.33) and \( \alpha = 0.05 \) (2-tailed), a sample size of 50 pairs of cases was required to ensure sufficient power for repeated measures comparisons (>95% for medium effects and 80% for small effects).

Our objective was to determine effectiveness of the speech intervention in improving communicative/functional adequacy of word production in comparison with a sham intervention for chronic AOS. Two baseline measures of speech before intervention were recorded to evaluate behavioral stability. Participants were profiled on a range of measures to establish AOS severity and the presence of comorbidities. The primary outcome measure was communicative adequacy of spoken naming. The secondary outcome measure was phonetic accuracy of words in repetition. Three word sets were developed, each containing 35 items. One set appeared in the intervention (treated). Two untreated sets consisted of matched items (phonetically similar to treated words) or control items (phonetically dissimilar). They allowed identification of generalization of treatment to similar or remote forms. Other outcome measures were collected but not reported here (repetition word duration, health economic analysis, and connected speech). The primary hypothesis was that speech intervention would result in significantly greater improvement in naming adequacy than sham. Secondary research predictions were as follows: (1) speech intervention would result in improved repetition accuracy; (2) effects of speech intervention would generalize to phonetically related untreated forms but not unrelated control words; (3) speech improvements would be maintained through a no-intervention period to final assessment (18 weeks post intervention for the speech-first group and 8 weeks for the sham-first group).

Methods

This study was granted ethical approval by an National Health Service panel. Volunteers gave consent to participation. Some deception was involved because participants were blinded to the sham nature of the visuospatial program. Participants were told that the program aimed to improve attention and memory. Participants were offered debriefing on completion of the study. It was a single-center, community-based trial (Sheffield, United Kingdom). Participants self-administered interventions, supported by speech and language therapists (SLTs), in their homes.

Participants

Participants were recruited from community SLT services across the South Yorkshire region over a 25-month period. The inclusion/exclusion criteria were adults with chronic AOS (at least 5 months post onset of apraxic stroke); unilateral left hemisphere lesion(s); the absence of neurodegenerative condition; premorbid competence in English; sufficient auditory/visual acuity to interact with a laptop; and not receiving impairment SLT. AOS diagnosis was independently confirmed by 2 SLTs using standard diagnostic criteria: disrupted speech intelligibility (distortions/substitutions) with intact gross oral movements and reduced speed/fluency and effortful speech (hesitations, groping, and prosodic disruption). In cases of uncertainty, a third assessor evaluated behavior. All assessors were registered SLTs. Fifty participants were recruited (29 men and 21 women). Figure 1 displays progression through the study. After baseline evaluation, participants were randomly allocated to speech-first/sham-first conditions by a researcher blind to case via block randomization (block sizes, 20-20-10). Assessors were aware of block sizes. An unpredictable allocation sequence was generated via computer randomizer. The sequence was transferred to opaque numbered envelopes and consecutive referrals allocated to condition via these envelopes. A subsequent allocation check revealed that 1 participant, allocated to sham-first group, did not receive interventions in planned order. An intention-to-treat criterion was used, and data from this participant were analyzed as per initial randomization. No stratification/minimization was used. Subsequent comparison of baseline AOS severity, aphasia severity, age, years of education, time post onset, and laterality using independent samples, \( t \) tests (2-tailed; \( \alpha = 0.05 \)), revealed no significant differences across the 2 order conditions (Table 1). There was a sex imbalance in the speech-first condition, with more men than women (17 versus 8).

Procedure

Before randomization, there were 2 baseline evaluation sessions (B1 and B2) to assess stability of naming and repetition behavior. The gap between baselines was 7 to 34 days (\( M = 18 \)). There were 3 word sets, each containing 35 items (Table I in the online-only Data Supplement). One word set (treated) appeared in intervention. Treated words, and nonmatched controls, represented vocabulary of high functional value and were roughly matched on word frequency and imageability. The 2 control sets did not appear in treatment and were either phonetically matched or phonetically dissimilar to treated forms (eg, treated: night; matched: white; control: house). All sets were roughly matched on word length and syllable structure. In repetition, participants repeated items after live presentation by an experimenter. Words were presented in a fixed pseudorandom order, with no phonetically similar items appearing in sequence. Only first responses were scored. The repetition task included all 35 items from each set. Naming performance was scored on 23 word triplets (only triplets with good name agreement by healthy speakers were included to avoid treatment effects being inflated by disambiguation of images during therapy). No cues were given in either task other than orientation cues to key elements of photographs in naming. Speech data were audiorecorded for subsequent analysis by an assessor who had no participant contact and was blind to allocation and period. Naming responses were scored as correct/incorrect (1/0). Correct responses were target words or appropriate synonyms (eg, children-kids). Problematic responses were scored in a consensus fashion by a group of 5 to 6 raters, the majority of whom were blind to allocation and period. Phonetic errors were not penalized if a listener could unambiguously identify the intended target. Repetition responses were coded on a 0- to 7-scale (eg, 0=no/entirely off-target response;
Responses scored at 6/7 were recorded as correct. An inter-rater reliability check on a subset of 558 samples was performed by a further member of the research team who was blind to period, allocation, and assessor 1’s ratings. The reliability sample was drawn from 16 participants with different levels of AOS severity, randomly selected across assessment points, and with equal numbers from both order conditions. Spearman’s rank correlation indicated a high level of inter-rater reliability (n=558; ρ=0.895; P<0.0001).

Immediately after the second baseline, participants were loaned a laptop for ≈6 weeks (speech-first range: 36–64 days, M=45 [SD, 5.1] and sham-first range: 42–50 days, M=44 [SD, 1.97]). Participants could access only their allocated program. An SLT researcher assisted participants with program use for initial sessions, followed by phone contact to check progress. Further support visits were arranged as needed (face-to-face visits in period 1: speech-first range, 1–6; M=4 [SD, 1.45] and sham-first range, 1–7; M=3 [SD, 1.17]). The regular use of software was encouraged (once or twice a day for at least 20 minutes). The actual intensity of treatment was
determined by the participant. The program recorded interactions and compliance with recommendations could be tracked. After six weeks, the laptop was withdrawn, and speech was reevaluated (outcome 1 [O1]).

After a four-week rest phase, the crossover period began. The speech-first group received sham intervention, and the sham-first group received the speech program. Programs were again available for six weeks. Laptops were then withdrawn, and further reassessment was completed (outcome 2 [O2]). Final reassessment (Maintenance [M]) took place after an eight-week no-treatment period.

### Results

There was no significant difference in program usage across the 2 groups in period 1: speech-first range, 355 to 1888 minutes and \( M = 1142 \) (SD, 439.54); sham-first range, 137 to 3129 minutes; \( M = 1026 \) (SD, 726.17); \( t(46) = 0.66; P = 0.512 \). The use of the first program tended to be higher than the second (period 2: speech-first range, 0–2322 minutes and \( M = 832 \) [SD, 677.55]; sham-first range, 103–2106 minutes and \( M = 996 \) [SD, 529.06]).

Statistical analyses are reported for period 1, with naming accuracy (Table 2) and repetition accuracy (Table 3) as dependent measures. Period 2 results are profiled for treated item type (treated, matched, and control) as the repeated measures, and treatment (sham-first; speech-first) as the between-group factor. Main effects of item type were significant (\( F = 3.35; df = 2, 92; P < 0.05 \)) with more treated items correctly named than control items (\( t = 3.02; df = 47; P < 0.01 \); Bonferroni corrected, \( c = 0.0167 \) for this and subsequent post hoc analyses of item effects). There were no other significant effects. Naming accuracy was stable at baseline and comparable across treatment groups.

### Naming

Means and SEs for correctly named items are presented in Table 2. Baseline stability was investigated by ANOVA with assessment point (B1 and B2), item type (treated, matched, and control) as repeated measures and treatment (sham-first; speech-first) as the between-group factor. Main effects of item type were significant (\( F = 3.35; df = 2, 92; P < 0.05 \)) with more treated items correctly named than control items (\( t = 3.02; df = 47; P < 0.01 \); Bonferroni corrected, \( c = 0.0167 \) for this and subsequent post hoc analyses of item effects). There were no other significant effects. Naming accuracy was stable at baseline and comparable across treatment groups.

Period 1 treatment effects were investigated using ANOVA with assessment point (B2 and O1) and item type (treated, matched, and control) as repeated measures and treatment (sham-first; speech-first group) as the between-group factor. Results revealed a main effect for assessment point (\( F = 18.82; df = 1, 46; P < 0.001 \)) and a significant interaction between assessment point and treatment group (\( F = 5.66; df = 1, 46; P < 0.05 \)). The assessment point effect was because of better naming at O1 collapsed across item type and treatment group (estimated marginal mean±SE: B2=12.71±1.20 and O1=14.07±1.18). Assessment point interacted with treatment group with greater improvement in naming for the speech-first (estimated marginal mean±SE: B2=13.54±1.73 and O1=15.64±1.70; \( t = 3.68; df = 22; P < 0.01 \)) than the sham-first group (estimated marginal mean±SE: B2=11.88±1.66 and O1=12.49±1.63; \( t = 2.10; df = 24; P < 0.05 \)). The main effect of item type (\( F = 7.68; df = 2, 92; P < 0.01 \)) and the interaction between item type and treatment group (\( F = 3.12; df = 2, 92; P < 0.05 \)) were also significant. Overall, treated items were named more accurately than matched (\( t = 2.59; df = 47; P < 0.014 \)) and control items (\( t = 3.99; df = 47; P < 0.001 \)). However, only the difference between treated and control items was significant for both the sham-first (\( t = 2.60; df = 24; P = 0.0167 \)) and the speech-first groups (\( t = 3.01; df = 22; P < 0.01 \)). Post hoc analysis of the significant interaction between assessment point×item type×treatment group (\( F = 6.82; df = 2, 92; P < 0.01 \)) showed little change between B2 and O1 for the treated, matched, or control items for the sham-first group (differences...
of estimated marginal means between B2 and O1 for treated: \(d=0.12\); matched: \(d=0.76\); control: \(d=0.96\); Bonferroni corrected \(\alpha=0.008\) for 6 post hoc comparisons). By contrast, increases in accuracy between B2 and O1 for the speech-first group were larger for treated and control items (differences of estimated marginal means between B2 and O1 for treated: \(d=3.30\); \(t=3.71\), \(df=22\), \(P<0.005\); matched: \(d=1.04\); control: \(d=1.96\); \(t=3.35\), \(df=22\), \(P<0.005\)). Figure 2 shows the effect of crossover, with increased naming accuracy for the sham-first group after exposure to the speech program (statistical analysis is given in Table III in the online-only Data Supplement).

### Repetition Accuracy

Table 3 presents means and SEs for the number of treated, matched, and control items with accuracy ratings of 6 or 7. Baseline stability was investigated by ANOVA with assessment point (B1 and B2) and item type (treated, matched, and control) as the repeated measures and treatment (sham-first and speech-first groups) as the between-group factor. Main effect of item type was significant \((F=17.96\); \(df=2\), 92; \(P<0.001\)) with greater accuracy for treated compared with matched and control items (treated versus matched: \(t=5.73\), \(df=47\), \(P<0.001\); treated versus control: \(t=3.82\), \(df=47\), \(P<0.001\)). There were no other significant effects, indicating that repetition accuracy was stable across baselines and comparable across treatment groups.

Period 1 treatment effects were investigated by ANOVA with assessment point (B2 and O1) and item type (treated, matched, and control) as repeated measures, and treatment (sham-first and speech-first group) as the between-group factor. Main effects for assessment point \((F=15.18\); \(df=1\), 46; \(P<0.001\)) and item type \((F=25.32\); \(df=2\), 92; \(P<0.001\)) were significant. Assessment point effects were because of higher accuracy after intervention, collapsed across item type and treatment group (estimated marginal mean±SE: B2=16.85±1.50 and O1=18.51±1.45). Item type effects resulted from significant differences among all 3 word sets, with the highest accuracy for treated, followed by control, and then by matched items (treated versus matched: \(t=6.35\), \(df=47\), \(P<0.001\); treated versus control: \(t=4.05\), \(df=47\), \(P<0.001\); matched versus control: \(t=3.17\), \(df=47\), \(P<0.001\)). Post hoc analysis of the significant interaction between assessment point/item$type treatment group \((F=3.98\); \(df=2\), 92; \(P<0.05\)) showed relatively little change between B2 and O1 for treated, matched, and control items in the sham-first group (differences of estimated marginal means between B2 and O1 for treated: \(d=0.36\); matched: \(d=1.60\); control: \(d=1.32\)). For the speech-first group, there were significant increases in accuracy for treated and matched items (differences of estimated marginal means between B2 and O1 for treated: \(d=3.13\); \(t=3.22\), \(df=22\), \(P<0.005\); matched: \(d=1.96\), \(t=2.96\), \(df=22\), \(P<0.008\); control: \(d=1.57\) (Figure 1 in the online-only Data Supplement displays the crossover effects showing an increase in repetition accuracy for the sham-first group after exposure to the speech program; Table III in the online-only Data Supplement, statistical analysis).

### Maintenance Effects

Maintenance of gains in naming and repetition of treated items was examined with paired \(t\) tests, comparing immediate post-speech intervention performance with the maintenance assessment (speech-first group: O1 versus M; sham-first group: O2 versus M). For naming, there were no significant changes in the speech-first group \((t=1.61\); \(df=19\); nonsignificant) or the sham-first group \((t=1.49\); \(df=23\); nonsignificant), indicating maintenance of treatment gains. For repetition, there were no changes in the speech-first group \((t=0.75\); \(df=19\); nonsignificant) but a significant decrease in performance in the sham-first group \((t=3.06\); \(df=23\); \(P<0.01\)).

Dose–response correlations were computed (Figure 3). Response was measured as the difference in naming of treated items between B2 and maintenance. Dose was measured in terms of minutes of speech program use. The correlation for

---

Table 3. Mean (±SE) Repetition Accuracy for T, M, and C Words Across Assessment Points by Group; Medians in Brackets

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Item Type</th>
<th>Baseline 1</th>
<th>Baseline 2</th>
<th>Postintervention 1 (O1)</th>
<th>Postintervention 2 (O2)</th>
<th>Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sham-first group</td>
<td>n 25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>T</td>
<td>16.92±2.19 (19)</td>
<td>17.60±2.31 (21)</td>
<td>17.96±2.23 (20)</td>
<td>20.64±2.09 (22)</td>
<td>19.33±2.19 (21)</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>14.64±2.01 (16)</td>
<td>14.92±2.20 (16)</td>
<td>16.52±2.18 (19)</td>
<td>17.84±2.17 (23)</td>
<td>16.29±2.05 (18)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>15.04±2.00 (17)</td>
<td>16.12±2.15 (17)</td>
<td>17.44±2.09 (21)</td>
<td>18.60±2.16 (19)</td>
<td>17.38±2.00 (18)</td>
<td></td>
</tr>
<tr>
<td>Speech-first group</td>
<td>n 23</td>
<td>23</td>
<td>23</td>
<td>22</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>T</td>
<td>17.48±1.87 (20)</td>
<td>18.87±2.11 (22)</td>
<td>22.00±2.06 (25)</td>
<td>21.55±2.13 (24.50)</td>
<td>22.75±1.94 (24)</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>16.35±1.96 (18)</td>
<td>16.26±2.03 (17)</td>
<td>18.22±1.89 (20)</td>
<td>17.36±2.02 (17.50)</td>
<td>19.95±2.27 (22.50)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>16.96±1.99 (19)</td>
<td>17.35±2.02 (19)</td>
<td>18.91±1.93 (18)</td>
<td>18.91±2.00 (20)</td>
<td>19.75±2.21 (22.50)</td>
<td></td>
</tr>
</tbody>
</table>

C indicates control; M, matched; O, outcome; and T, treated.
both groups was positive, indicating an increase in correctly named items as a function of increased time using the speech intervention (speech-first group: $r=0.45$, $P<0.05$; sham-first group: $r=0.42$, $P<0.05$).

### Discussion

In this 2-period, crossover study, we observed improvements in both naming and repetition in speakers with chronic and stable AOS impairments. Treatment effects were generally specific to trained vocabulary, with only limited transfer to phonetically similar words in repetition accuracy. The effects of intervention were largely maintained when interventions were withdrawn, and in the speech-first group, this retention period was 18 weeks. There was some loss of gains in repetition accuracy in the sham-first group, which might be because of lower baseline performance or the lower use of the period 2 intervention affecting the speech program. Treatment effects were specific to the speech program. The period 1 results, equivalent to a randomized controlled trial design, revealed no significant speech change in response to the sham program. The period 2 profiles reflect the manipulation of the crossover, with increased scores on treated items in the sham-first group for naming and repetition. Furthermore, the significant relationship between speech treatment dose and response is an indicator that behavioral change might be linked to the speech intervention.

The item-specific improvement in naming is similar to that found in successful therapies for anomic aphasia. One possibility is that the effects we observed resulted from lexical facilitation rather than enhancement specifically at the phonetic level. It is evident from the aphasia severity scores (Table 1) that most, but not all, participants had significant accompanying aphasic impairment. Given the strong interconnectivity between lexical and phonetic levels, top-down activation from the lexical level may enable access to motor plans. Importantly, in the face of item-specific effects, the use of functionally relevant vocabulary in therapy is essential.

The results provide evidence that computer therapy and development of programs enabling patients to self-administer interventions are important directions in rehabilitation of post-stroke speech and language disorders. This model of intervention may allow administration of high-intensity therapies in a cost-effective manner. Some participants had little or no previous experience in using computers; however, design of programs with simple interfaces enabled computer novices to access interventions with SLT support. Family members were largely positive about the intervention, some reporting reduced burden of care in that they felt able to pursue their own activities, knowing that the participant was engaged in purposeful activity. Participants were also generally positive about the software, although many commented on the repetitive nature of stimulation. The dose levels administered by participants were varied and sometimes modest. An important future direction for software design is to incorporate game and social elements to maximize motivation and achieve higher usage levels. This refinement would benefit engagement with the later stages of the program in particular, which focus on the use of trained words in sentence frames. Practice at this level is likely to be crucial in achieving transfer to spontaneous speech.

### Acknowledgments

We thank Andrew Hardbottle and Jennifer Ryder for contributions to data collection and Rotherham NHS Trust for assistance in participant recruitment.
Sources of Funding
This research was funded by the Bupa UK Foundation specialist grant programme. The funders had no role in design or conduct of the study.

Disclosures
Drs Varley and Whiteside are coauthors of a commercially available program used to treat speech impairments. They and the University of Sheffield (employers of Cowell, Dyson, and Whiteside) receive royalties from sales of software. The software used in this study was a pilot version of this program. The other authors report no conflicts.

References
Self-Administered Computer Therapy for Apraxia of Speech: Two-Period Randomized Control Trial With Crossover
Rosemary Varley, Patricia E. Cowell, Lucy Dyson, Lesley Inglis, Abigail Roper and Sandra P. Whiteside

Stroke. 2016;47:822-828; originally published online January 21, 2016;
doi: 10.1161/STROKEAHA.115.011939

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/47/3/822

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2016/01/21/STROKEAHA.115.011939.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/
SUPPLEMENTAL MATERIAL

Self-administered computer therapy for apraxia of speech: a two-period randomized control trial with crossover

Rosemary Varley, PhD\textsuperscript{1}; Patricia E. Cowell, PhD\textsuperscript{2}; Lucy Dyson, BMedSci\textsuperscript{2}; Lesley Inglis, PhD\textsuperscript{2}; Abigail Roper, MMedSci\textsuperscript{3}; Sandra P. Whiteside, PhD\textsuperscript{2}

Supplemental Tables

Table I*. Treated, Phonetically Matched and Control items. Lexical frequencies\textsuperscript{1} (spoken) are provided in brackets where available.

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

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
### Table II*. Accuracy Scoring Protocol for the Speech Repetition Task

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
</table>
| 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;  
(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 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 |
<table>
<thead>
<tr>
<th></th>
<th>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.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2c</td>
<td>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 – 2 segments’, above.</td>
</tr>
<tr>
<td>2b</td>
<td>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.</td>
</tr>
<tr>
<td>2</td>
<td>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.</td>
</tr>
<tr>
<td>1</td>
<td>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.</td>
</tr>
</tbody>
</table>
| 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). |

Table III. Statistical analysis of period 2 outcomes

**Naming Accuracy:** The effects of treatment for the first and the second period of the study were investigated by ANOVA where the second baseline measurement, the first and second post-intervention measurements for naming accuracy (Periods 1&2: B2, O1, O2) and Item Type (Treated, Matched, Control) were the repeated measures, and Treatment (Sham-First; Speech-First) was the between-group factor. Main effects for Periods 1&2 (F=13.28, d.f.=2, 90, p<0.001) and the interaction between Period 1&2 and Treatment group (F=4.10, d.f.=2, 90, p<0.05) were significant. Period effects were due to better naming after intervention at O1 and O2 compared to B2 collapsed across Item Type and Treatment group (estimated marginal means ± S.E.s: B2 = 12.55±1.22; O1 = 13.91± 1.19; O2 = 13.85±1.21; Bonferroni corrected pairwise comparisons, p-values <0.001). Period effects interacted with Treatment group due the greatest relative improvement in naming occurring for the Speech-First group following speech intervention in Period 1 (estimated marginal means ± S.E.s: B2 = 13.23±1.77; O1 = 15.32±1.74; O2=14.53±1.76) in contrast to the Sham-First group where the greatest relative improvement occurred following the speech intervention in Period 2 (estimated marginal means ± S.E.s: B2 = 11.88±1.66; O1 = 12.49± 1.63; O2=13.17±1.65). Thus, both treatment groups showed a larger improvement at the measurement point following speech intervention compared to the sham intervention. Item Type (F=14.98, d.f.=2,90, p<0.001), Item Type x Period (F=3.33, d.f.=4,180, p<0.05) and Item Type x Period x Treatment effects(F=4.83, d.f.=4,180, p<0.01) were significant. Effects were due to better naming for Treated compared to Matched and Control items (estimated marginal means± S.E.s for Treated: 14.23±1.20; Matched: 13.09±1.19; Control: 12.99±1.22) and the more marked nature of these differences after Period 1 for Speech First and after Period 2 for Sham First groups (Table 2).

**Repetition Accuracy:** The effects of treatment for the first and the second period of the study were investigated by ANOVA where the second baseline measurement, the first and second post-intervention measurements for repetition accuracy (Periods 1&2: B2, O1, O2) and Item Type (Treated, Matched, Control) were the repeated measures, and Treatment (Sham-First; Speech-First) was the between-group factor. For the independent variable Test Item (Treated, Matched, Control), Mauchly’s test of sphericity identified a violation of the assumption for equal variances of differences (W=0.86, chi-square=6.54, d.f.=2, p<0.05). Conservative interpretation of all ANOVA terms involving Item Type was applied by way of reporting Greenhouse-Geisser adjusted values. Main effects of Periods 1&2 (F=20.17, d.f.=2, 90, p<0.001) and Item Type (F=31.69, Greenhouse-Geisser adjusted d.f.=1.76, 79.08, p<0.001) were significant. Period effects were due to improved rates of repetition accuracy after intervention at O1 and O2 compared to B2 collapsed across Item Type and Treatment group (estimated marginal means ± S.E.s: B2 = 16.66±1.52; O1 = 18.27± 1.46; O2 = 19.15±1.46). Item Type effects were due to significant differences among all three types (estimated marginal means± S.E.s for Treated: 19.63±1.51; Matched: 16.71±1.47; Control: 17.74±1.45).
Supplemental Figure I: Means (±S.E.s) for repetition accuracy of treated items in the Speech-First and Sham-First conditions across assessment points.

Supplemental References