Published in final edited form as:

Aphasiology. 2014 January 1; 28(5): 515–532. doi:10.1080/02687038.2014.886323.

Treating apraxia of speech with an implicit protocol that activates speech motor areas via inner speech

Dana Farias¹, Christine Herrick Davis¹, and Stephen M Wilson²

¹Davis Medical Center, Physical Medicine and Rehabilitation, University of California, Sacramento, CA, USA

²Department of Speech, Language, and Hearing Sciences and Department of Neurology, University of Arizona, Tucson, AZ, USA

Abstract

Background—Treatments of apraxia of speech (AOS) have traditionally relied on overt practice. One alternative to this method is implicit phoneme manipulation which was derived from early models on inner speech. Implicit phoneme manipulation requires the participant to covertly move and combine phonemes to form a new word. This process engages a system of self-monitoring which is referred to as fully conscious inner speech.

Aims—The present study aims to advance the understanding and validity of a new treatment for AOS, implicit phoneme manipulation. Tasks were designed to answer the following questions. 1. Would the practice of implicit phoneme manipulation improve the overt production of complex consonant blends in words? 2. Would this improvement generalize to untrained complex and simpler consonant blends in words? 3. Would these treatment tasks activate regions known to support motor planning and programming as verified by fMRI?

Method & Procedures—The participant was asked to covertly manipulate phonemes to create a new word and to associate this newly formed word to a target picture among 4 phonologically-related choices. To avoid overt practice, probes were collected only after each block of training was completed. Probe sessions assessed the effects of implicit practice on the overt production of simple and complex consonant blends in words. An imaging protocol compared semantic baseline tasks to treatment tasks to verify that implicit phoneme manipulation activated brain regions of interest.

Outcomes & Results—Behavioral: Response to implicit training of complex consonant blends resulted in improvements which were maintained 6 weeks after treatment. Further, this treatment generalized to simpler consonant blends in words. Imaging: Functional imaging during implicit phoneme manipulation showed significant activation in brain regions responsible for phonological processing when compared to the baseline semantic task.

Conclusions—Implicit phoneme manipulation offers an alternative to traditional methods that require overt production for treatment of AOS. Additionally, this implicit treatment method was

shown to activate neural areas known to be involved in phonological processing, motor planning and programming.

The Apraxia of Speech Treatment Guidelines Committee of the Academy of Neurological Communications Disorders and Sciences (ANCDS) developed guidelines for the treatment of apraxia of speech (AOS). They performed an exhaustive review of 59 published studies for this purpose. Articulatory-Kinematic Treatments were found to be the most widely accepted behavioral approach for the treatment of AOS. These treatments incorporate overt practice such as modeling, repetition, and articulatory cueing as a means of improving accuracy of the spatial and temporal aspects of speech production (Wambaugh, Duffy, McNeil, Robin, & Rogers, 2006). The use of overt practice has shown generalization within targeted sound groups from complex to simpler consonant clusters (Maas, Barlow, Robin & Shapiro, 2002; Schneider & Frens, 2005). This investigation asks if an implicit-based approach would replicate the generalization response of these prior researchers.

Implicit-based treatments that do not require an overt verbal response have been applied successfully in individuals with fluent aphasia (Davis, Harrington & Baynes, 2006). Extending this work to individuals with AOS, Davis, Farias, and Baynes (2009) proposed a novel treatment which focused on implicit phoneme manipulation as a means of improving motor planning and programming of speech. This implicit-based treatment offered an alternative to an individual who was reluctant to engage in traditional therapy requiring overt repetition. The level of interest and duration of therapy sessions increased as a result of allowing the patient to select the correct response without the need for overt speech. An added benefit was that verbal errors were not reinforced through overt production. This approach, which required the retrieval, manipulation, and internal monitoring of sounds in various phonetic contexts, was based on Van Der Merwe's Four Level Framework (1997) of speech sensorimotor control. The repeated implicit practice of phoneme manipulation through tasks such as rhyming, deletion and alliteration was hypothesized to improve the efficiency of transition between phonological plans and the motor plans for speech sounds. Van Der Merwe (1997) describes this process as a gradual transition from abstract phonological knowledge to spatial and temporal coordinated patterns of muscle or motor commands for speech movements.

This implicit approach for the treatment of AOS depends on the inner monitoring of speech rather than correcting errors after they are produced. Baars, Motley, and MacKay (1975) and Levelt (1989, 1983) describe this internal monitor as a mechanism which detects and intercepts speech errors before they are articulated. They suggest that inner speech is heard through an inner loop that transmits the speech plan at the phonetic and/or phonological level back into the speech comprehension system for monitoring. This monitor operates on a pre-articulatory representation of the utterance and has access to phonological representations being constructed in production such as the phonetic plan, phonemes and metrical information (Wheeldon & Levelt, 1995). According to Baars et al. (1975), when a speech error is detected during pre-articulatory editing, it is suppressed prior to overt speech and repaired. Since Baars' and Levelt's early work on inner speech there have been varying accounts of how this mechanism monitors, detects and repairs speech errors (Dell & Repka, 1992; Hartisuker, 2001; Postma, 2000; Postma & Noordanus, 1996).

Recently Oppenheim and Dell (2010) presented arguments for and challenges to this understanding of inner speech. They propose that inner speech is on a continuum from abstract inner speech, which lacks articulatory detail, to a more fully realized overt speech that includes both sound and motor movement. The manner in which inner speech includes motor aspects of speech is reflected in the stages of lexical processes that are engaged. Oppenheim and Dell (2010) investigated the different levels of lexical processes activated by mouthed and unmouthed inner speech in 80 normal participants. Two error effects were observed in their investigation; lexical bias effects (the preponderance of word versus non words in speech errors) and phonemic similarity effects (speech errors that are produced by an exchange of phonemes with shared features). Both mouthed and unmouthed inner speech activated lexical-phonemic representation to produce lexical bias effects. However, unmouthed silent speech failed to engage detailed articulatory representations necessary for phonemic similarity effects. In contrast, mouthed inner speech showed phonemic similarity effects reflecting the engagement of articulatory phonetic information. Mouthed inner speech is therefore more similar to overt speech in these two speech-error effects.

Geva, Jones, Crinion, Price, Baron and Warburton (2011) investigated the inner and overt speech of individuals with chronic aphasia. They defined inner speech as “the ability to create an internal representation of the auditory word form and to apply computations or manipulations to this representation” (p. 3072). In related work, Geva, Bennett, Warburton and Patterson (2011) suggest that their tasks, rhyming and homophone judgments, demand more attention and focus on inner speech which they describe as “fully conscious inner speech.” Tasks requiring more ‘conscious’ inner speech, such as rhyme and homophone judgments, have shown greater activation of areas associated with phonological processing such as the LIFG (Burton, LoCasto, Krebs-Noble, & Gullapalli, 2005; Geva, et al., 2011; McDermott, Petersen, Watson & Ojemann 2003; Price, Devlin, Moore, Morton & Laird, 2005). Oppenheim and Dell suggest that as “speakers engaged in more detailed articulatory planning, their inner speech reflects that information” (p. 1158). Therefore, on a continuum, fully conscious inner speech may be more closely associated with Oppenheim and Dell’s (2010) “mouthed inner speech” than abstract unmouthed inner speech.

Findings in both normal participants and those with aphasia have relevance for the design of this implicit-based treatment approach to AOS. The focus of this treatment is on the use of fully conscious inner speech for treating AOS rather than the traditional focus on overt articulated speech. If we are in fact using conscious inner speech then we would expect that our tasks would produce similar activation patterns to those findings from Geva et al., (2011).

Although the application of implicit treatments to improve speech production in AOS has shown some success, it is acknowledged that further support for this approach is warranted. A limitation of the prior investigation on the use of implicit treatments for treating apraxia of speech (Davis, et al., 2009) was the lack of imaging to support that neural networks involved in speech motor planning are activated during covert speech without overt production. As such, this current design includes a functional imaging component. If functional magnetic resonance imaging (fMRI) shows that implicit phoneme manipulation tasks activate areas

known to be necessary for phonological encoding and motor planning this treatment technique may find wider acceptance.

First we hypothesized that treatment, which focuses on fully conscious inner speech, will improve overt speech production in individuals with AOS. Second, we hypothesized that implicit phoneme manipulation will activate regions known to support phonological processing, motor planning and programming such as the ventral premotor cortex, the posterior inferior frontal gyrus, and the anterior insula. We will answer the following questions:

1. Would phoneme manipulation that relies on fully conscious inner speech, improve production of targeted complex consonant blends as measured by accuracy of probe word productions and as measured by the *d* statistic size effects?
2. Would this treatment applied to complex consonant blends generalize to non-treated simpler consonant blends as measured by production accuracy of probe words and measured by the *d* statistic size effects?
3. Would f MRI results provide support that the phoneme manipulation task activates neural areas known to support phonological processing, motor planning and programming?

METHOD

Participant

The participant (SB) is a 56 year-old right-handed, college-educated, African-American, English-speaking male who suffered a hypertensive stroke. He was administered intra-arterial Tissue Plasminogen Activator in the emergency room. Magnetic imaging at the time of his admission revealed a left middle cerebral artery stroke with a lesion involving the left postcentral gyrus, insula, supramarginal gyrus, and underlying white matter tracts (see Figure 1). This participant received speech therapy twice a week for 3 months prior to the start of the intervention. The focus of therapy was on articulatory cueing as a means of improving accuracy of the spatial and temporal aspects of speech. Six months after his stroke he was recruited to participate in this research project. His primary diagnosis was AOS based on analysis of errors on reading, repetition and naming and subtests of Apraxia Battery for Adults (ABA -2) (Dabul, 2000). SB's speech, although relatively fluent with mild anomia, was characterized by articulatory groping, difficulty initiating utterances, frequent hesitations, speech sound distortions and prosodic disturbances characterized by significantly decreased rate and extended inter-word and vowel durations. No dysarthria as described by Duffy (2005) was observed. SB had normal hearing and vision with no hemiplegia, hemianopsia, articulatory weakness, non-verbal oral or limb apraxia as reported by the Neurologist in outpatient clinic. IRB approval and consent was obtained prior to initiation of treatment.

Procedure

Pre- and post-intervention assessment—Standardized measures of aphasia and apraxia were administered. The pre-tests included the Western Aphasia Battery-Revised

(WAB-R) (Kertesz, 2006), The Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 2001), the Pyramids and Palm Trees Test (PPT) (Howard & Patterson, 1992), the Auditory Word Discrimination Subtest of the Test of Auditory-Perceptual Skills-Revised (TAPS) (Gardner, 1996) and the ABA -2. These tests were administered to provide a detailed description of the deficits associated with apraxia and aphasia. Pre- and post-testing was administered by a Speech Language Pathologist (SLP) who was blind to the intervention.

The BNT was administered and SB's score of 34/60 was 4 standard deviations below the mean for his age. On error analysis SB produced 9 errors due to distortions, deletions, substitutions or omissions (e.g., "pentil" for pencil, "hander" for hammer). The PPT was administered to measure conceptual semantic knowledge from word and picture stimuli. Two subtests were administered, the all picture and all word versions, which yielded scores of 50/52 and 52/52 respectively. There was no significant difference between the scores on these two subtests applying the McNemar's Test. These scores indicated that SB's conceptual knowledge was intact. The ABA -2 was administered to aid in the diagnosis of apraxia. On the subtest measuring performance on the repetition of words of increasing length and complexity SB had mild-moderate difficulty with the phoneme sequencing of words of increasing length, for example, "zipping" for zippering, "pleaser" for pleasingly. The examiner noted prominent vowel prolongations/distortions during this task. Deterioration in performance was 1.0, one syllable average minus three syllable average on the ABA-2. Further examples of error distortions on oral reading of the Grandfather's passage include "wist" for wished, "accent" for ancient, "seferal" for several, "beer" for beard, "sifilly" for skillfully. No limb or oral apraxia was exhibited on subtests from the ABA-2. The Auditory Word Discrimination of the TAPS measured the participant's ability to discriminate paired words with phonemically similar consonants, cognates or vowel differences. SB's score on Auditory Word Discrimination score was 33 which gave a T score of 46 and indicated that SB's auditory discrimination was within 1 standard deviation from the norm. The WAB-R Aphasia Quotient was 68, with subtests scores as follows: Comprehension scores of 8.6, Repetition score of 4.8 (e.g., "pite" for pipe, "pasty cook" for pastry cook, "rinting" for ringing), Naming score of 7.6 and a fluency score of 5. The combination of errors on repetition, naming, oral reading and in spontaneous speech confirmed that our participant had problems planning or programming directions for speech movements for accurately selected and sequenced phonological representation of word forms.

Development of test and treatment stimuli—The initial screening words were obtained from 40,000 Selected Words, (Blockcolsky, Frazer & Frazer, 1987). From this text a corpus of 118 words with initial consonant blends comprised of two and three consonants were selected for baseline trials. These words were presented for repetition over 3 separate sessions to establish a baseline. Performance on these words determined which consonant clusters were most difficult for the participant to articulate. Each session was digitally recorded with an Olympia 100. Two SLPS, blind to the order of treatment, listened to the recorded sessions and phonetically transcribed the words to determine error type. A word was scored as incorrect if any portion of the word was judged to contain distortions,

omissions, additions, substitutions, prolongations and/or intrusive schwa. Probes were selected from those words in which the consonant cluster was judged to be in error on one or more productions of the 3 baselines.

Selection of probes and control task—Eleven complex consonant blends and 11 simpler consonant blends were selected for training based on results from the set of 118 screening items. The complex consonant blends consisted of 3 initial consonants (e.g. squish, splash) and the simpler consonant blends consisted of two initial consonants (e.g. swish, slash). An additional 22 words that met criteria for inclusion comprised the untrained probes and were selected with initial consonant blends that matched the training stimuli. Additionally, all trained and untrained words were matched based upon CV construction (e.g. initial complex blend words squash/squish were in CCCVC construction; initial simpler blend words snip/snap were in CCVC construction). Untrained stimuli were used for a measure of generalization of training (see trained and untrained stimuli in Table 1).

The Object fluency subtest on the WAB-R operated as the control task for the experiment. Object fluency required the participant to name as many animals as he could in one minute. This task, known as animal or semantic fluency in the literature, is a measure of executive function (Burgess, Alderman, Evans, Emslie & Wilson, 1998). As a control it was assumed that performance would be unaltered by the phoneme manipulation treatment. Generalization of treatment effects to this task would be unexpected unless there was a loss of experimental control.

Development of phoneme manipulation templates—An intervention was designed to encourage the participant, SB, to covertly manipulate phonemes to create a new word, and to associate this newly formed word to a target picture among 4 choices. This intervention consisted of therapist-designed templates developed with web-based Google TM images arranged on Microsoft PowerPoint® templates. Each template consisted of 4 pictures, one of the target and three phonologically-related foils (see Figure 2). The intervention required SB to make a choice of one of the four pictures based on instructions given to him to perform a phoneme manipulation. The manipulation required the subject to hold a word ending in auditory memory while adding various consonant or consonant blends to form a new word. The word ending and consonants were given auditorially and the subject was required to perform the manipulation without overt verbal output. When the pictures were shown to SB the instructions were “I want you to point to the correct picture, think about it, don’t say it out loud.” For example when training complex consonant blends, SB was shown 4 pictures and asked to point to the target when given this auditory instruction: “‘Eak,’ what would it be if you added ‘str’ to the beginning?” All pictures were phonologically related to the target such as the foils “seeks,” “peaks,” and “sneaks” for the target “streaks.” When training simpler consonant blends, SB was shown 4 pictures with the auditory instruction similar to the complex training. For example, “ ‘Ip’ what would it be if you added ‘sn’ to the beginning?” Pictures of “sip,” “slip,” and “skip” were presented as foils for the target picture “snip” (see Figure 2). To ensure fidelity of treatment, adherence to this predetermined script was followed. Additionally, the administration of treatments tasks was provided by the same therapist. Untrained probes were never used in the templates as foils during treatment. Two

individuals with Master degrees in Speech Pathology, blind to the development of the templates, independently performed the phoneme manipulation tasks for all the treatment templates. The judges were in 100% agreement on the target response for the phoneme manipulation templates.

Treatment—Treatment was given three times a week for 1–1.5 hours by an SLP. In this multiple-baseline design three baselines were obtained for all 44 trained and untrained words selected for probes prior to training. Training was given in two blocks: a two-week block training on initial complex consonant blends (consisting of 3 initial phonemes e.g. “str”) and a second two-week block of training on simpler consonant blends (consisting of 2 initial phonemes e.g. “st”).

This design employed a probe schedule that reduced the influence of overt speech on training. Probes were collected only after each block of training was completed. Therefore three consecutive probe sessions were completed after the first two weeks of training initial complex consonant blends and another three sessions were completed after the second two week block of training words with initial simpler consonant blends. This reduced probe schedule decreased the unintended repetition of errors during training which according to Wambaugh, (2006) may reinforce speech errors. Six weeks after training ceased probe data for three maintenance sessions was obtained.

Judgment of probes—Each probe session was digitally recorded. Two SLPS, blind to the order of treatment probes, listened to the recorded sessions and phonetically transcribed the words to determine error type. A word was scored as incorrect if any portion of the word was judged to contain distortions, omissions, additions, substitutions, prolongations and/or intrusive schwa. This scoring criterion was used for baselines, as well, as probes and following treatment. Errors such as distortions, prolongations, or intrusive schwas are typically associated with AOS. To further analyze error type in our participant we obtained a distribution of errors from the final baseline. Prolongations comprised 37%, substitutions 30%, omissions 18% and distortions 15% of the total errors. Although we cannot definitively rule out the contribution of phonological errors, there is a possibility that distortion errors are mis-categorized as substitutions due to faulty auditory perception of the error by the judge (Kent, 1996). Given the high prevalence of prolongations, consistent location of errors, and both consonant and vowel distortions, we are confident that this participant has AOS.

The *d* statistic was obtained from the probe data and was calculated according to Beeson and Robey (2006). The *d* statistic is derived from the mean score of probes post-treatment minus the mean of baseline probes over the standard deviation of the baseline probes. When applying the *d* statistic, all data preceding the intervention for the specific sound are considered baseline, and all data following the intervention are considered maintenance. Effect size benchmarks for the *d* statistic on lexical retrieval treatments were used for this interpretation (small= 4.0, medium = 7.0, and large = 10.1).

Imaging—The fMRI was completed a month after the completion of treatment and was used to compare brain activation patterns under two conditions: while performing implicit

phoneme manipulation (experimental task) and while performing semantic-based tasks not requiring phoneme manipulation (control task). The semantic-based tasks were developed to control for activation associated with naming and language unrelated the process of phoneme manipulation.

Development of the imaging templates—Two sets of twelve templates were developed for the imaging component of this study: twelve templates for the experimental tasks and twelve for the control tasks. The experimental tasks for imaging were similar to the treatment phoneme manipulation tasks. These consisted of twelve blending tasks, each comprised of two pictures that were taken from stimuli used during implicit treatment. Another twelve semantic templates were developed as a control for the phonological manipulation tasks.

For the experimental task the subject was required to perform an implicit phoneme manipulation as in the treatment phoneme manipulation tasks. For example, the subject was shown pictures representing “slice” and “splice” and given an auditory prompt “Splice, slice ... ‘ice’, what would it be if you added ‘spl’ to the beginning” (see Figure 3).

The control stimuli (semantic templates) required the subject to make a decision comparing a semantic aspect of two pictures presented. The presentation of the picture stimuli for the control (semantic) task matched the experimental task in that two pictures were displayed with an auditory prompt. There were no words on the semantic templates to avoid any activation of regions associated with phonological processing. The semantic questions were related to a sensory, functional or heuristic (encyclopedic) comparison of the two pictures shown. The subject was shown a template with 2 pictures, for example, the Liberty Bell and the Washington Monument, and given the auditory prompt “the Liberty Bell, the Washington Monument, which one would be heavier?” (See Figure 4). This decision was based on a sensory and or encyclopedic comparison of the two pictures. The semantic tasks were developed in a similar manner to semantic tasks developed by Heath, McMahon, Lyndsey, Angwin, MacDonald, van Hees, Johnson and Copland (2012). However, instead of asking a single question about a semantic feature of a picture presented, this design asked the participant to make a comparison. The twelve two-picture semantic templates were judged independently by 2 SLP’s who arrived at a 97% agreement on the control tasks.

Imaging Procedures—Prior to entering the scanner the participant was trained on all procedures and practiced using a button press to select a picture.

The picture templates for both the experimental and control tasks were presented to the subject while in the scanner. An auditory prompt was given through headphones. The duration of the auditory prompt was 9–10 seconds and the subject was asked to respond as soon as possible after the end of the auditory prompt. He had 5 seconds to respond. The response consisted of a button push which corresponded to the number (either “1” or “2”) below the picture. No feedback was given to the participant on the accuracy of the responses. The inter-trial interval (from the end of one trial to the beginning of the next) was variable, with a mean of 10 seconds.

The participant was scanned on a Siemens Trio 3T scanner at UCSF. We acquired 300 T2*-weighted echo-planar volumes (plus an additional 2 initial discarded volumes) with the following parameters: 32 AC/PC-aligned axial slices in interleaved order; slice thickness = 3.6 mm with 0.9 mm gap; field of view = 230 x 230 mm; matrix = 96 x 96; TR = 2000 ms; TE = 28 ms; flip angle = 90°. A T1-weighted MPRAGE sequence was also acquired for anatomical reference.

The fMRI data were preprocessed with standard methods in SPM5. Data were corrected for slice timing differences, realigned to account for within-scan head movement, smoothed with a Gaussian kernel of 8 mm FWHM, and high-pass filtered (cutoff = 128 s) to remove slow signal drift.

The blending and semantic task blocks were each modeled as boxcar functions convolved with a standard hemodynamic response function. The boxcar onsets were aligned with the stimuli rather than the responses, because the phonological or semantic nature of each trial became apparent early in the auditory prompt, so we assumed that phonological or semantic processing respectively would begin at that time. The data were fit with a general linear model including these two explanatory variables, along with 6 covariates of no interest based on head motion during the scan. The contrast of interest was blending blocks versus semantic blocks. This contrast was thresholded at voxelwise $p < 0.001$, then corrected for multiple comparisons at $p < 0.05$ based on cluster extent according to Gaussian random field theory (Worsley et al., 1996) as implemented in SPM5. An inclusive mask was used comprising voxels that were active for the blending task relative to baseline at voxelwise $p < 0.05$.

RESULTS

Treatment results

Response to training was verified by d statistic analysis. Training complex consonant blends (List A) resulted in a d statistic of 5.7 on trained probes and 4.2 on untrained probes (see Figure 5). When applying the d statistic, all data preceding the intervention for the specific sound are considered baseline, and all data following the intervention are considered maintenance. For example, when applying the d statistic in the complex phoneme cluster training there are 3 baselines used for the calculation of the d statistic. However in the simple phoneme cluster training all data preceding that intervention is are considered for calculating that d statistic which is 6 baseline data points. Therefore baseline 4, 5 and 6 were obtained during complex phoneme cluster training. Training generalized to all the probes of simpler consonant blends (List B) giving a d statistic of 4.8. However due to the response of simpler consonant blends to complex training the baseline used to calculate the d statistic for the second phase of training simpler consonants was high and therefore the results of yielded a smaller d on trained (1.37) and untrained probes (1.83). Improvements were maintained on both complex and simpler consonant blends of trained and untrained probes 6 weeks after training.

Post-intervention results

After the implicit phoneme intervention SB scored 50/50 on the Auditory Word Discrimination Subtest of the TAPS, compared to 33/50 on pretest. His BNT score was 37/60 compared to 34/60 on pre-test. His deterioration score on the ABA-2 decreased to .6 compared to 1.0 on pre-test. He achieved a WAB-R AQ score of 89.2 compared to an AQ of 68 on pre-test. WAB-R subtests scores improved as follows: Comprehension score of 10.0 compared to 8.6 on pre-test, Repetition score of 7.5 compared to 4.8 on pretest, Naming score of 9.1 compared to 7.6 on pre-test. The Object fluency subtest of the WAB-R operated as the control task. SB scores on this subtest were unchanged by treatment. He scored 11 on this subtest on pre and post intervention testing.

Imaging results

Structural MRI revealed that the lesion was centered in the left postcentral gyrus, extending ventrally to the precentral gyrus of the insula and encroaching on the anterior supramarginal gyrus and on the superior temporal gyrus just posterior to Heschl's gyrus and including underlying white matter, likely the superior longitudinal fasciculus. Functional imaging showed that the blending task recruited predominantly left-lateralized brain regions, when compared to the baseline semantic task (Fig. 6, Table 2). Activated left hemisphere regions comprised the precentral gyrus, inferior frontal gyrus (pars triangularis), anterior insula, posterior superior temporal gyrus, supramarginal gyrus and angular gyrus. The supplementary motor area was activated bilaterally, and in the right hemisphere there was activity in the insula and inferior frontal gyrus (pars opercularis). These activated regions included brain areas immediately anterior and immediately posterior to the lesion, suggesting that surrounding tissue had retained its functionality.

DISCUSSION

The purpose of this study was to investigate the effects of implicit phoneme manipulation on the production of words with consonant blends. This implicit approach targeted the interface of phonological codes to speech sounds which is thought to be impaired in AOS (Van de Merwe, 1997). This training employed a type of fully conscious inner speech which focuses attention on the transition between phonological plans and the initial phases of motor planning for speech sounds.

Results support the first hypothesis that implicit phoneme manipulation improves production of targeted complex consonant blends as measured by accuracy in probe word productions. These findings expand on prior research (Davis et al., 2009) which showed improvement in the overt production of consonant blends in an individual with a primary diagnosis of apraxia of speech following implicit-based therapy. A limitation of this prior study was the assessment of progress through overt probes, which provided the opportunity for overt practice thereby compromising the implicit focus. In the current study overt probes were restricted to pre- and post-training probes in an effort to limit overt practice during implicit training.

Additionally we asked, would treatment generalize to simpler consonant blends as measured by production accuracy of non-treated probe words. This participant showed generalization of treatment effects to simpler consonant blends. This supports prior research that found overt training of complex consonant clusters generalized to singletons (Maas et al., 2002).

Size effects as measured by the *d* statistic verified the benefits of this implicit treatment. Treatment probes for whole word production of complex consonant clusters gave a *d* statistic of 5.7 which equated to a large size effect. Untrained probes of complex consonant clusters yielded a *d* statistic of 4.2 which is associated with a medium size effect. These results suggest that treatment aimed at the targeted complex consonant cluster generalized to untreated words with complex clusters. Additionally, simpler consonant blends responded to complex training (*d* = 4.8). Smaller size effects were obtained from the second phase of training of simpler consonant blends. This may have been the result of a higher baseline obtained from responses to the previous complex consonant blend training.

Unlike prior studies (Maas et al., 2002; Schneider & Frens, 2005) which examined only the accuracy of the targeted clusters, here the accuracy of probes was based upon accurate repetition of the whole word. Whole word analysis was used because apraxic errors are typically influenced by complexity, such as co-articulation, consonant vowel construction and length. Errors, although most common in initial complex blends, may occur throughout the word. Therefore, the complexity of the final consonant blend may adversely affect the production of the initial consonant cluster, for example in the trained stimuli, “squashed” is more complex than “squash.” To examine this further, five matched pairs of stimulus words (from both trained and untrained lists) were selected. These pairs were identical except for the final consonant cluster (e.g., “scratch” and “scratched”). Analysis of these five pairs showed no significant difference on accuracy of word production following intervention. These results suggest that either the added complexity of the final consonant cluster was not sufficient to adversely affect production or that the majority of errors were confined to the initial target cluster which improved equally across these pairs.

The second hypothesis, that implicit phoneme manipulation activates neural areas known to be involved in phonological processing, motor planning and programming, was also supported. According to Moser, Fridriksson, Bonilha, Healy, Baylis, Baker and Rorder (2009) the supplemental motor area, insula, premotor ventro-lateral frontal areas, and primary sensorimotor cortex are involved in motor speech planning. These areas were found to be activated even in the absence of overt speech (Aleman, Formisano, Koppenhagen, Hagoort, de Haan, & Kahn, 2005; Fiez, Tallal, Raichle, Miezin, Katz & Petersen, 1995). Consistent with these findings the treatment tasks, although devoid of overt speech, activated these same regions. Although imaging studies have well established that inner speech cannot simply be described as overt speech without a motor component (Barch et al., 1999; Basho et al., 2007; Geva et al., 2011; Indefrey & Levelt, 2004; Palmer, Rosen, Ojemann, Buckner, Kelley, & Petersen, (2001); Shuster & Lemieux, 2005) we suggest that the phoneme manipulation tasks utilize fully conscious inner speech as described by Geva et al. (2011). They state that fully conscious inner speech “requires one to monitor, or listen to, one’s own inner speech in order to successfully perform the task.” Further, conscious inner speech is associated with activation in the left inferior frontal gyrus, a region essential for

phonological processing (Paulesu, Frith & Frackowiak, 1993) and which was significantly active during our experimental task.

A finding in addition to the hypothesized results was an improvement in the participant's auditory word discrimination. The Auditory Word Discrimination subtest of the TAPS measured the participant's ability to discriminate paired words with phonemically similar consonants, cognates or vowel differences (e.g. miss-mess). The participant improved from one SD below the norm at pre-test to within normal range (50/50 correct) at post-test presumably due to the course of treatment. We propose this may be the result of a focus and attention to minimal speech sound differences which is inherent within this implicit treatment approach. Alternatively, it may reflect improved working memory such that the subject, through repeated practice of phoneme manipulation, was able to use subvocal rehearsal to improve the phonological store of the referent for later judgment comparison.

In addition to an improvement in auditory discrimination, marked improvements were shown in all subtests of the WAB-R on post-test. The total AQ score increased from 68 to 89.2 following this intervention. This suggests that the effects of a focus on phoneme manipulation extend beyond our initial targeted words to improvements in naming, comprehension, repetition and expression. These improvements were unexpected. Naming improvements which consisted of a decrease in errors may be the direct result of our intervention. For example previous distortions such as "hander" for "hammer" and "pentil" for "pencil" were produced correctly on post-test.

A limitation to the design of this study was the lack of a counterbalanced condition. An attempt was made to control for order effects between training complex and simpler consonant blends. After the treatment of complex consonant clusters the subject's production of simple consonant clusters improved such that our criteria of 1–2 errors over 3 baselines could not be established. Consequently, it could be that the treatment itself rather than the focus on complex clusters resulted in the generalization to simpler clusters. Therefore, we cannot with confidence state that treating complex consonants generalized to simpler consonant clusters.

An additional limitation was that our participant was aware that the treatment tasks were designed to improve his speech. Therefore it is possible that our participant overtly practiced these target words outside of therapy as no specific instruction not to practice was provided.

CONCLUSIONS

The results of this study advance the understanding and validity of a new treatment for AOS, implicit phoneme manipulation. This implicit treatment method was shown to activate neural areas known to be involved in phonological processing, motor planning and programming.

Implicit phoneme manipulation offers an alternative to traditional methods for the treatment of AOS. Treatment for AOS is well suited to implicit interventions as it requires the individual to focus on conscious inner speech, a necessary precursor to self-correction and self-monitoring for accurate overt speech. This approach used colorful, web-based images

presented on a computer which were arguably more engaging than traditional repetition drills. Aside from being engaging, this approach has the potential to be adapted as a home, computer-based activity, thereby increasing intensity and allowing practice without the need for face-face sessions with a SLP. Additional improvements in auditory discrimination, repetition, and comprehension may be a byproduct of this approach.

Acknowledgments

We are indebted to SB, a motivated participant who was eager to attend therapy sessions and contribute to advancing our understanding of treating AOS.

We acknowledge the support in part by the National Institutes of Health (NIDCD R03 DC010878 to SMW).

References

- Aleman AE, Formisano H, Koppenhagen P, Hagoort EHF, de Haan RS, Kahn. The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*. 2005; 15:221–228. [PubMed: 15269107]
- Baars B, Motley M, MacKay D. Output editing for the lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:382–391.
- Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC, Cohen JD. Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. *Neuroimage*. 1999; 10(6):642–657. [PubMed: 10600410]
- Basho S, Palmer ED, Rubio MA, Wulfeck B, Muller RA. Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech. *Neuropsychologia*. 2007; 45(8):1697–1706. [PubMed: 17292926]
- Beeson PM, Robey RR. Evaluating single-subject treatment research: lessons learned from the aphasia literature. *Neuropsychol Rev*. 2006; 16(4):161–169. [PubMed: 17151940]
- Blockcolsky, VD.; Frazer, JM.; Frazer, DH. 40,000 selected words. Tuscon, AZ: Communication Skill Builders, Inc; 1987.
- Burgess PW, Alderman N, Evans J, Emslie H, Wilson BA. The ecological validity of tests of executive function. *J Int Neuropsychol Soc*. 1998; 4(6):547–558. [PubMed: 10050359]
- Burton MW, Locasto PC, Krebs-Noble D, Gullapalli RP. A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage*. 2005; 26(3):647–661. [PubMed: 15955475]
- Dabul, BL. Apraxia Battery for Adults. 2. Austin, TX: PRO-ED; 2000. (ABA-2)
- Davis C, Farias D, Baynes K. Implicit Phoneme Manipulation for the Treatment of Apraxia of Speech with co-occurring Aphasia. *Aphasiology*. 2009; 23(4):503–528.
- Davis CH, Harrington G, Baynes K. Intensive semantic intervention in fluent aphasia: A pilot study with fMRI. *Aphasiology*. 2006; 20:59–83.
- Dell, GS.; Repka, RJ. Errors in inner speech. In: BBJ, editor. *Experimental slips and human error: Exploring the architecture of volition*. New York: Plenum; 1992. p. 237–262.
- Duffy, JR. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. 2. St. Louis: Elsevier Mosby; 2005.
- Fiez JA, Tallal P, Raichle ME, Miezin FM, Katz WF, Petersen SE. PET studies of auditory and phonological processing: effects of stimulus characteristics and task demands. *Journal of Cognition Neuroscience*. 1995; 7:357–375.
- Gardner, MF. *Test of Auditory-Perceptual Skills-Revised*. Psychological & Educational Publications, Incorporated; 1996.
- Geva S, Bennett S, Warburton EA, Patterson K. Discrepancy between inner and overt speech: Implications for poststroke aphasia and normal language processing. *Aphasiology*. 2011; 25(3): 323–343.

- Geva S, Jones PS, Crinion JT, Price CJ, Baron JC, Warburton EA. The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain*. 2011; 134(Pt 10):3071–3082. [PubMed: 21975590]
- Hartsuiker RJ, Kolk HHJ. Error Monitoring in Speech Production: A Computational Test of the Perceptual Loop Theory. *Cognitive Psychology*. 2001; 42(2):113–157. [PubMed: 11259106]
- Heath S, McMahon K, Nickels L, Angwin A, MacDonald A, van Hees S, et al. Priming picture naming with a semantic task: an fMRI investigation. *PLoS One*. 2012; 7(3):e32809. [PubMed: 22412928]
- Howard, DaPK. The Pyramids and Palm Trees Test. Suffolk, UK: Thames Valley Test Company; 1992.
- Indefrey P, Levelt WJ. The spatial and temporal signatures of word production components. *Cognition*. 2004; 92(1–2):101–144. [PubMed: 15037128]
- Kaplan, E.; Goodlass, H.; Weintraub, S. Boston Naming Test. 2. Philadelphia, PA: Lippincott, Williams & Wilkins; 2001.
- Kent RD. Hearing and Believing Some Limits to the Auditory-Perceptual Assessment of Speech and Voice Disorders. *American Journal of Speech Language Pathology*. 1996; 5:7–23.
- Kertesz, A. Western Aphasia Battery-Revised (WAB-R). San Antonio, TX: Pearson; 2006.
- Levelt WJ. Monitoring and self-repair in speech. *Cognition*. 1983; 14(1):41–104. [PubMed: 6685011]
- Levelt, WJ. Speaking: From intention to articulation. Cambridge, MA: MIT Press; 1989.
- Maas E, Barlow J, Robin D, Shapiro L. Treatment of sound errors in aphasia and apraxia of speech: Effects of phonological complexity. *Aphasiology*. 2002; 16(4–6):609–622. [PubMed: 22787286]
- McDermott KB, Petersen SE, Watson JM, Ojemann JG. A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia*. 2003; 41(3):293–303. [PubMed: 12457755]
- Moser D, Fridriksson J, Bonilha L, Healy EW, Baylis G, Baker JM, et al. Neural recruitment for the production of native and novel speech sounds. *Neuroimage*. 2009; 46(2):549–557. [PubMed: 19385020]
- Oppenheim GM, Dell GS. Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*. 2008; 106(1):528–537. [PubMed: 17407776]
- Oppenheim GM, Dell GS. Motor movement matters: the flexible abstractness of inner speech. *Mem Cognit*. 2010; 38(8):1147–1160.
- Palmer ED, Rosen HJ, Ojemann JG, Buckner RL, Kelley WM, Petersen SE. An event-related fMRI study of overt and covert word stem completion. *Neuroimage*. 2001; 14(1 Pt 1):182–193. [PubMed: 11525327]
- Paulesu E, Frith CD, Frackowiak RS. The neural correlates of the verbal component of working memory. *Nature*. 1993; 362(6418):342–345. [PubMed: 8455719]
- Postma A. Detection of errors during speech production: A review of speech monitoring models. *Cognition*. 2000; 77(2):97–132. [PubMed: 10986364]
- Postma ANC. The production and detection of speech errors in silent, mouthed, noisemasked, and normal auditory feedback speech. *Language and Speech*. 1996; 39:375–392.
- Price CJ, Devlin JT, Moore CJ, Morton C, Laird AR. Meta-analyses of object naming: Effect of baseline. *Human Brain Mapping*. 2005; 25:70–82. [PubMed: 15846820]
- Schneider SL, Frens RA. Training four-syllable CV patterns in individuals with acquired apraxia of speech: Theoretical implications. *Aphasiology*. 2005; 19(3/4/5):451–471.
- Shuster LI, Lemieux SK. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain Lang*. 2005; 93(1):20–31. [PubMed: 15766765]
- Thompson, CKaS; Lewis, P. Linguistic-Specific Approach to Treatment of Sentence Production Deficits in Aphasia. *Clinical Aphasiology*. 1994; 22:307–323.
- Van der Merwe, A. A theoretical framework for the characterization of pathological speech sensorimotor control. In: McNeil, MR.; Robin, DA.; Schmidt, RA., editors. *Clinical management of sensorimotor speech disorders*. New York: Thieme; 1997. p. 1-25.
- Wambaugh JL, Duffy JR, McNeil MR, Robin DA, Rogers M. Treatment guidelines for acquired apraxia of speech: Treatment descriptions and recommendations. *Journal of Medical Speech Language Pathology*. 2006; 14(2):xxxv–ixvii.

- Wambaugh JL. Treatment guidelines for apraxia of speech: Lessons for future research. *Journal of Medical Speech-Language Pathology*. 2006; 14(4):317–321.
- Wheeldon LR, Levelt WJM. Monitoring the time course of phonological encoding. *Journal of Memory and Language*. 1995; 34(3):311–344.
- Worsley KJ, Marrett S, Neelin P, Evans AC. Searching scale space for activation in PET images. *Hum Brain Mapp*. 1996; 4(1):74–90. [PubMed: 20408187]

Figure 1.
Acute MRI showing left MCA stroke

1

2

3

4

Figure 2.
Foils “sip,” “slip,” and “skip” for the target “snip.”

Figure 3.
“Splice, slice ... ‘ice’, what would it be if you added ‘spl’ to the beginning?”

1

2

Figure 4.

Pictures representing the Liberty Bell and the Washington Monument were given with the auditory prompt “the Liberty Bell, the Washington Monument, which one would be heavier?”

Figure 5. Results of implicit phoneme manipulation on trained and untrained probes. B = baseline, M = Maintenance, LT Long-term Maintenance. List A = complex phoneme clusters, List B = simple phoneme clusters

Figure 6.
Functional imaging of blending tasks when compared to baseline semantic tasks.

Table 1

Trained and untrained probes

Complex consonant blends		Simpler consonant blends	
Trained	Untrained	Trained	Untrained
scribble	scrabble	sniffle	spackle
squealed	squalled	flick	fleece
splits	splats	slide	slush
squashed	squished	sleepless	sleeveless
squash	squish	sleds	truffle
squeal	squall	snip	snap
scratch	screech	snips	clamp
streamline	strychnine	grasp	flossed
scratched	strapped	crossed	slashed
streaks	strives	cross	clam
strife	strive	slingshot	slipknot

Table 2

Blending task compared to baseline semantic task by ROI

Brain region	MNI coordinates			Volume mm ³	Max t	p
	x	y	z			
Left precentral gyrus and bilateral supplementary motor area	-24.9	5	45.1	21544	5.78	<0.001
Precentral gyrus (ventral peak)	-56	0	42		5.78	
Precentral gyrus (dorsal peak)	-38	0	62		5.15	
Supplementary motor area	-6	8	52		5.72	
Left intraparietal sulcus	-21.8	-65.9	46.4	6408	4.74	<0.001
Right insula and inferior frontal gyrus (pars opercularis)	42.6	10.8	8.8	4640	5.04	<0.001
Left inferior frontal gyrus (pars triangularis) and anterior insula	-40.6	30.6	7.3	4584	5.22	<0.001
Inferior frontal gyrus	48	30	6		5.22	
Anterior insula	32	26	6		4.85	
Left supramarginal gyrus	-41.1	-40.5	33.9	3064	4.67	0.001
Left posterior superior temporal gyrus	-60.2	-46.7	18.5	2080	4.29	0.007