

Study Title: Treatments of Acquired Apraxia of Speech

NCT #: NCT01483807

Document: Study Protocol

Date of last protocol approval by IRB: November 22, 2015

## Protocol:

### Participants

#### Inclusion Criteria:

- At least 6 months post onset of focal brain injury
- Chronic acquired apraxia of speech and aphasia
- Native English speaker
- Age: 21 – 85 years
- Pure tone hearing adequate to pass screening at 35 dB at 500, 1000, and 2000 Hz least one ear, or if aided, or verification from an audiological examination that hearing was adequate for conversational level speech.
- Negative histories for alcohol and/or substance abuse and neurological conditions other than stroke; self-reports verified by medical record review
- Performance within normal limits on the *Test of Nonverbal Intelligence-4* (Brown, Sherbenou, & Johnsen, 2010).

#### AOS diagnostic criteria (McNeil, Robin, & Schmidt, 1997; 2009)

- Slow rate of speech
- Sound errors which included a predominance of distortions
- Syllable segregation
- Relatively consistent trial-to-trial articulation errors
- Prosodic abnormalities.

A battery of tasks was administered to elicit speech samples from which to judge presence/absence of AOS characteristics: 1) *Apraxia Battery for Adults-2* (Dabul, 2000), 2) word and sentence repetition tasks (Wambaugh, Kalinyak-Fliszar et al., 1998; Wambaugh, West,

& Doyle, 1998; and Mauszycki & Wambaugh, 2008, and 3) discourse production (Nicholas & Brookshire, 1993). AOS symptoms were evident in at least some, but not necessarily all, of the preceding speech elicitation tasks. The diagnosis of AOS was initially made by each participant's primary clinician (one of the authors) and then was verified by the first author.

Language assessments:

- *Western Aphasia Battery-Revised (WAB-R; Kertesz, 2007)*
- *Porch Index of Communicative Ability (PICA; Porch, 2001)*
- Nicholas & Brookshire Discourse Elicitation Battery (Nicholas & Brookshire, 1993)

Demographic data collected:

- Age
- Months post-onset
- Etiology
- Area of brain damage
- Years of Education
- Premorbid handedness
- Race/ethnicity
- Hemiparesis
- Marital status

## **Experimental Design**

A combined group and SCED was used. The group design component was a two-phase, cross-over design and the SCED component was a multiple baseline designs (MBD) across behaviors and participants. This report is focused on the results of the group design. However,

probe data from the SCEDs will be provided in the supplemental materials for the 10 participants whose data have not been reported previously.

Accuracy of articulation of two sets of experimental words (see “Experimental Stimuli”) was measured repeatedly prior to treatment (see “Probes” for additional detail). For each participant, treatment (SPT-Random [SPT-R] or SPT-Blocked [SPT-B]) was then applied to one set of experimental words while probing continued with both experimental sets. A two week, no-treatment interval followed the first treatment phase and a short-term maintenance probe was completed at the end of the interval. Repeated probing of the set of items to be treated in the second treatment phase during the no treatment interval (as part of the SCED) ensured that performance was stable prior to the application of treatment with that set. Then, the alternate treatment was applied to the remaining set of items. Follow-up probes were conducted at 2, 6, and 10 weeks following the second treatment phase. Please see Wambaugh et al. (2014, 2016) for description of the SCED component.

**Treatment Order and Experimental Set Assignment.** Quasi-randomized assignment was combined with participant matching to assign participants to one of two treatment orderings (i.e., SPT-R → SPT-B or SPT-B → SPT-R). We aimed to have a balance of aphasia type, aphasia severity and AOS severity reflected in the treatment orders. The first participant enrolled was randomly assigned to a treatment condition. Each subsequently enrolled participant was “matched” to a previously enrolled participant and assigned to the opposite group if he/she had the same aphasia type, had an AQ within 10 points of the match, and had the same AOS severity rating. When a match did not exist, then the participant was randomly assigned to a treatment order. Treatment order assignment is shown in Table 2. Participant matching for treatment order is shown in Supplemental Appendix B: aphasia severity was within 10 AQ points for 8/10

matches, aphasia type was the same for 9/10 matches, and AOS severity was the same for 8/10 matches. Please note that the purpose of the investigation was not to compare performance across groups as all participants received both treatments. The matching and random assignment was conducted to minimize possible confounds associated with treatment order.

As indicated previously, each participant received both SPT-R and SPT-B. Two sets of experimental words were devised for each participant (described below). Those sets were randomly assigned to treatment condition to assist in controlling for influence of stimuli on treatment outcome.

### **Experimental Stimuli and Probe Procedures**

**Probe and Treatment Stimuli.** Two sets of words were individually devised for each participant. For each set, there were two subsets, each of which represented a different sound target. Pretreatment assessment of word production was used to identify specific sounds that were difficult for the participant to produce (i.e., tasks described in “Participants” with results shown in Table 2). For 18 of the 20 participants, there were a total of 30 words per set, with 20 items designated for treatment and 10 items designated for generalization measurement. Half of the 30 words contained one sound target and the remaining half contained a different sound target (10 treatment items and 5 untreated items per sound target). For the remaining four participants (two from each treatment ordering group), the number of stimuli were reduced (see Wambaugh et al., 2016). This reduction was made to allow for application of a modified form of blocking (see “Treatment”). Across the participants, stimuli included monosyllabic words, multisyllabic words, and words in phrases (See Supplemental Appendix C for lists of stimuli for each participant). The type of stimuli selected was determined by the pretreatment testing results that provided an indication of level of difficulty for that participant (i.e., Table 2).

In the development of the stimuli, it was not expected that participants would be able to produce all other sounds in the experimental word, other than the target sound(s), correctly. Errors could, and did, occur on non target sounds. All target sounds for a given participant were considered in stimuli selection. For words representing a given target, we tried to avoid inclusion of the other targets. For example, if the targets were /b/, /sw/, /m/, or /n/, in the selection of /b/ words, we avoided words that contained /sw/, /m/, or /n/. It was not always possible to exclude other target sounds, particularly with multisyllabic words. Within participant, lists were balanced as closely as possible for syllable length and position of the target sound within the word (e.g., initial, medial, or final). The experimental sets were also balanced as much as individual variation allowed for singletons versus clusters, manner, and place of production. For example, if one experimental set had a subset representing a particular cluster, then the other experimental set also had a subset with a cluster. Possible generalization effects were also considered in the selection of stimuli with choices made to avoid generalization across lists. For example, if /pl/ was selected for one list, then none of the other lists would be comprised of /l/ blends or /p/ or /b/ words. Words were not balanced for frequency or familiarity due to other constraints. It is recognized that such balancing would have been preferable and was initially attempted, but could not be achieved. The nature of the experimental tasks (i.e., repetition) and the experimental design assuaged our concerns over familiarity and frequency somewhat; these potential confounds would have been present across all study phases (e.g. increased familiarity through repeated exposure should have been revealed in the baseline phase). A variety of databases and dictionaries were used to develop stimuli lists.

**Probes.** Production of the experimental words was elicited in probes. During probes the examiner presented each word verbally, one at a time in random order, and asked the participant

to repeat the word as accurately as possible. No feedback concerning accuracy of production was provided; only general encouragers (e.g., you're trying hard) were used. The treatment items were included in probes to measure the *acquisition effects* of treatment and the untreated exemplars of trained items were used to measure the *response generalization* effects of treatment (see Dependent Variable). Note that the term "acquisition" is used to represent probe performance with treated items and not performance during treatment and as such is analogous to the term "retention" as used in the motor learning literature. All probe sessions were audio recorded.

**Blinded examiner probes.** An investigator who was blinded to the assignment of experimental words to treatment condition conducted probes identical to the probes described above. One blinded probe was completed during the baseline phase and at the end of each treatment phase.

### **Dependent Variable**

The dependent variable was accuracy of articulation of a target sound or sounds produced in words, with a single score of correct or incorrect being used for each experimental item. Scoring was completed using on-line, modified narrow transcriptions supplemented and verified with audio recordings (e.g., Haley, Bays, & Ohde, 2001). The target sound(s) for each experimental word was/were identified and determined to be accurately or inaccurately produced on the basis of the transcription. Then, the word containing the target sound(s) was given a score of "correct/incorrect". For each subset of items for each participant, the percentage of words in which the target sounds(s) was produced correctly was tallied from the verified, online productions; the subsets were further separated into trained and untrained subsets. These percentages served as the dependent variables for all analyses.

Target sounds were required to be produced without distortion in the correct syllable of the word. Multisyllabic words were required to have the correct number of syllables so that location accuracy could be judged. For words embedded in a phrase, the phrase word had to be attempted in order for the target sound (in the other word member of the phrase) to be scored as accurate. However, accurate production of the non-targeted phrase word was not required.

The *first complete* production was scored. If the entire target item was produced and there was an error on the target sound(s) and then the participant self-corrected the error, that item was scored as incorrect. However, if there was a false-start (the entire target item was not produced initially) followed by a correct production, then the complete production was scored as correct.

For the majority of the participants, accuracy of production of the one target sound/cluster per word was scored. However, for a subset of four participants with multisyllabic words targets (two participants from each of the treatment orderings), accuracy of production of all consonants in the word was required to achieve an overall score of correct production (Wambaugh et al., 2016). These four participants had mild to mild-moderate AOS and sometimes had greater than 50% accuracy of production of specific target sounds in baseline. In order to have sufficient numbers of errors to demonstrate change with treatment, all consonants were scored/targeted.

The experimenter who conducted the probe completed the online scoring of that probe. During the treatment phases, these experimenters also provided treatment. These same experimenters also conducted the baseline and follow-up phase probes when there was no treatment. That is, the same experimenters were used to conduct all probes throughout all phases of the study for a given participant. Ideally, experimenters who were blinded to the assignment



of items to treatment would conduct all probes, but this was logistically and economically infeasible. Consequently, a limited number of blinded probes were conducted (as described previously). With the blinded probes, the examiner was not involved in treatment.

## **Treatment**

SPT utilizes the therapeutic techniques of modeling/repetition, contrastive practice, orthographic cueing, integral stimulation, and articulatory cuing. These techniques are applied using a response-contingent hierarchy as follows:

1. The speech/language pathologist (SLP) produces a verbal model of the word or phrase and requests a repetition. If monosyllabic words are the treatment targets, substeps are used for the purposes of contrastive practice upon an incorrect production (e.g., Wambaugh & Mauszycki, 2010; Wambaugh & Nessler, 2004). If multisyllabic words are the target, then contrastive practice is not used and the next step is attempted.
2. The SLP uses printed letters/word to indicate the sound in error, directs the participant to attend to the target sound, provides another verbal model, and requests a repetition.
3. The SLP uses integral stimulation - “watch me, listen to me, say it with me” - and attempts simultaneous production until a correct production is achieved, with a maximum of three attempts. In cases where a phrase is used, the entire phrase is attempted.
4. The SLP provides articulatory cuing appropriate for the sound production error and then repeats the procedures used in the previous step. Only the target word is practiced (not the entire phrase).

5. The SLP presents the next item.

Verbal feedback concerning the correctness of production of the sound target is provided immediately after each response. Since the hierarchy is response-contingent, ensuing steps of the hierarchy are completed *only* after an erroneous production. Following a correct response and provision of feedback, the next item is presented. The steps of the hierarchy are not reversed following a correct response. In previous SPT investigations (Wambaugh & Mauszycki, 2010; Wambaugh & Nessler, 2004), multiple productions of the target word were elicited upon a correct response at any step of the hierarchy. In the current study, the multiple repetitions following a correct response were not elicited because these could be considered a form of blocked practice.

As indicated in Step 1, when SPT is applied to monosyllabic words, the first step includes a sub step in which a minimal contrast word that reflects the sound error is practiced (see Wambaugh & Mauszycki, 2010 or Wambaugh & Nessler, 2004). Minimal contrast is not used when the targets are multisyllabic words because real word minimal pairs are rarely possible. Participants with monosyllabic treatment stimuli completed this sub step of Step 1.

The SPT hierarchy was applied to each treatment item repeatedly in a treatment session. For SPT-R, SPT was applied to all 20 treatment items (10 words for each target) by alternating between the subsets randomly. For SPT-B, SPT was applied to each of the 10 words in a subset before treatment was applied with the other subset. Within subset, the words were presented in random order for SPT-blocked. For four of the participants (two from each of the treatment ordering groups), a modification was made to the blocking procedure; blocking was applied to each word so that each word was practiced for five minutes (please see Wambaugh et al., 2016 for rationale and description of procedure).

The treatment sessions were 50 to 60 minutes long excluding probes. One treatment trial consisted of application of the SPT hierarchy to each of the treatment items. As many treatment trials as possible were completed in the 50-60 minute time period. The number of treatment trials per session varied within and across participants and corresponded to the number of errors that occurred because more errors required use of more steps of the hierarchy. Therefore, participants completed more trials per session toward the end of a treatment phase in comparison to the initial treatment sessions with a phase. Although number of treatment trials varied, the length of each treatment session was carefully controlled and was consistent within and across participants.

Number of treatment sessions was equivalent across treatment phases for each participant. Eighteen of the participants received 20 treatment sessions per phase and two participants received 10 treatment sessions per phase (P5 & P15; Wambaugh et al., 2016). The shorter length of treatment was due to a rapid treatment response. For clinical purposes, it would likely be desirable to administer the necessary number of treatment sessions to achieve a desired level of performance (rather than have a pre specified number of sessions). However, in the current investigation, lack of equivalency of number of sessions per condition for each participant could have confounded our findings.

ASHA certified research SLPs conducted treatment sessions three times per week. For one participant (P2), a master's student provided treatment for a portion of the sessions; this student was supervised 100% of the time by the third author who also provided treatment. Accommodations to treatment schedule were made as needed (e.g., illness, vacations, holidays). Each participant selected his/her treatment location which remained constant throughout the study (i.e., participant's residence, research laboratory, or university clinic).

## **Reliability**

**Dependent Variable.** Twenty percent of all probes (including all SCED probes) were quasi-randomly selected for rescoring by a research SLP who had not provided treatment; random selection occurred for all phases of the study. The audio recordings of probe sessions were used for rescoring. The reliability SLP was blinded to the assignment of lists to treatment phase and to assignment of items as treated or untreated. Point to point agreement was calculated for scoring of each item and percent agreement was calculated for each probe. Agreement for scoring each word/target as correct/incorrect ranged from 85% to 96%, with the average being 91%.

**Independent Variable.** A research SLP who had not participated in a given participant's treatment scored audio recorded treatment sessions for accuracy of implementation of treatment. Ten percent of all treatment sessions were randomly selected and were scored for administration of specific treatment components: 1) correct application of steps of the treatment hierarchy – 99.6 %; 2) presentation of target words (all words presented in each trial) – 99.8%; 3) session length of 50-60 minutes within +/- 1 minute – 100%; and 4) accurate application of blocking and randomization – 100%.