Research Note

The Effects of Behavioral Speech Therapy on Speech Sound Production With Adults Who Have Cochlear Implants

Frances M. Pomaville* and Chris N. Kladopoulos**

Purpose: In this study, the authors examined the treatment efficacy of a behavioral speech therapy protocol for adult cochlear implant recipients.

Method: The authors used a multiple-baseline, across-behaviors and -participants design to examine the effectiveness of a therapy program based on behavioral principles and methods to improve the production of target speech sounds in 3 adults with cochlear implants. The authors included probe items in a baseline protocol to assess generalization of target speech sounds to untrained exemplars. Pretest and posttest scores from the Arizona Articulation Proficiency Scale, Third Revision (Arizona-3; Fudala, 2000) and measurement of speech errors during spontaneous speech were compared, providing additional measures of target behavior generalization.

Results: The results of this study provided preliminary evidence supporting the overall effectiveness and efficiency of a behavioral speech therapy program in increasing percent correct speech sound production in adult cochlear implant recipients. The generalization of newly trained speech skills to untrained words and to spontaneous speech was demonstrated.

Conclusion: These preliminary findings support the application of behavioral speech therapy techniques for training speech sound production in adults with cochlear implants. Implications for future research and the development of aural rehabilitation programs for adult cochlear implant recipients are discussed.

Key Words: adults, aural rehabilitation, speech production, speech-language pathology, cochlear implants

As a result of technological advances over the past 4 decades, cochlear implants (CIs) are now a viable option for many adults who experience bilateral, moderate-to-profound sensorineural hearing loss that impairs speech recognition and functional communication abilities and who demonstrate limited improvement with traditional amplification (Gifford, 2011). Adult CI candidates may experience a range of communication deficits, depending on the age of onset and on the severity of their hearing loss. Adults with prelingual deafness often demonstrate significantly impaired speech production (Busby, Roberts, Tong, & Clark, 1991; McGarr & Osberger, 1978), while those with postlingual deafness often experience a gradual deterioration of speech sound production (Abberton et al., 1985; Leder & Spitzer, 1990; Ramsden, 1981; Zimmerman & Rettaliata, 1981). This speech deterioration may result in marginally normal speech, which can carry significant social and vocational penalties and can negatively affect an individual’s self-image (Cowie, Douglas-Cowie, Phil, & Kerr, 1982; Edgerton, 1985).

The focus of most research that examines the speech production outcomes of adult CI users has been with postlingually deafened adults who have experience producing intelligible speech (e.g., Edgerton, 1985; Kishon-Rabin, Taitelbaum, Tobin, & Hildesheimer, 1999; Langereis, Bosman, van Olphen, & Smoorenburg, 1997; Svirsky, Lane, Perkell, & Webster, 1992; Vick et al., 2001; Waters, 1986). Little research has been conducted to examine the speech production characteristics of prelingually deafened adult CI users (e.g., Busby et al., 1991; Dawson et al., 1995; Vick et al., 2001; Wong, 2007). There is substantial individual variability in the rate of improvement and the level of speech performance attained by adult CI users (National Institutes of Health, 1995).

Although technological advances have resulted in improved patient outcomes for many CI users, technology is only one contributor to a user’s success. Aural rehabilitation (AR) refers to the treatment of hearing loss and often includes components such as auditory training, speechreading,
speech therapy, education, and counseling (Boothroyd, 2007; Sweetow & Palmer, 2005). AR programs vary greatly in terms of their content and implementation, and empirical evidence regarding the efficacy of individual components is lacking. Recent advances in CI technology have led to a renewed interest in the application of AR programs to CI recipients (Bloom, 2004; Rossi-Katz & Arehart, 2011). The present study focused on one aspect of AR: speech therapy. To date, the effects of various therapy approaches that seek to improve speech production in adult CI users have not been well documented (Busby et al., 1991; Pantelemidou, Herman, & Thomas, 2003; Waters, 1986; Wong, 2007). Several researchers have emphasized the need to consider speech therapy as part of the AR process (Dawson et al., 1995; Edgerton, 1985; Waters, 1986; Wayner, 2000; Wong, 2007). Yet, studies conducted by Rossi-Katz and Arehart (2011) and by Tucci, Lambert, and Ruth (1990) found that many CI recipients received a limited amount of AR because they lived far from their CI center or lacked motivation, family support, or funding.

The purpose of this study was to evaluate the efficacy of behavioral speech therapy for improving speech sound production in adult CI recipients with prelingual or early deafness. Behavioral speech therapy procedures that have been effective in treating a variety of communicative disorders were extended to adult CI recipients. This study addressed two research questions. First, does the percent correct production of target speech sounds improve with participation in a behavioral speech therapy program when compared with baseline measures? Second, does the percent correct production of target speech sounds during spontaneous speech improve after participation in a behavioral speech therapy program when compared with pretest measures?

Data Collection and Treatment Procedures

This research was conducted with the approval of the Institutional Review Board of the governing institutions and conformed to all standards of human subject research. All sessions were conducted in an individual treatment room at the Speech, Language, and Hearing Clinic at CSUF. All sessions were audiotaped.

Prescreening and posttest procedures. Prescreening (pretest) and posttraining (posttest) procedures consisted of administering the Arizona Articulation Proficiency Scale, Third Revision (Arizona-3; Fudala, 2000) and measuring speech errors during spontaneous speech. Administration of the Arizona-3 was consistent with the instructions provided in the test manual, resulting in a list of speech sound errors that occurred during the test. The researcher then engaged the examinee in conversation to collect at least 50 utterances of spontaneous verbal communication. Sample lengths ranged from 50 to 66 utterances. The researcher used the audiotape to create a written transcript of the examinee's speech sample. Misarticulations were transcribed using the International Phonetic Alphabet. This resulted in a list of speech sound errors that occurred during spontaneous speech. Next, the number of correct productions for each of the error speech sounds was divided by the total number of opportunities for production of that sound to calculate the percent accuracy for the production of each error speech sound. The results of the prescreening procedures showed that all three participants demonstrated multiple speech sound errors and reduced intelligibility. Table 1 presents a list of speech sound errors identified during the prescreening. For
each participant, three speech sounds from their list of speech sound errors were selected and targeted for therapy. The selection of target speech sounds was based on choosing sounds that were least likely to influence one another and most likely to affect intelligibility. Clients’ preferences regarding the sounds that bothered them the most were also considered. In addition, pretest and posttest results from the Arizona-3 and the spontaneous speech sample provided a means to compare speech sound production, at the word and conversational levels, before and after treatment.

**Baseline procedures.** A baseline protocol was created for each participant that consisted of 90 stimulus items: 20 words that were used for training and 10 words that were not used for training (i.e., generalization probe items) for each of the three target speech sounds. Generalization probe words were included to measure stimulus–response generalization to untrained exemplars. On the basis of the findings of Skelton (2004), training and probe exemplars were selected that contained the target speech sound in all positions of the words, as well as in blends where appropriate. Words also varied from one to five syllables in length. A list of the treatment and probe exemplars for each participant is provided in the online supplemental materials. Untrained exemplars were intermixed with the trained exemplars, and all exemplars were presented in a randomized order during the baseline phase and after each treatment session. All exemplars required evoked, word-level productions of the target speech sound. No corrective feedback or reinforcement was provided during the baseline or treatment probe procedures. Using this baseline measure, a stable baseline level of percent correct production of each of the three target speech sounds was established for all participants prior to initiating treatment. Each participant was asked to read the words on the list, and each production was scored as correct (+) or incorrect (−). The obtained scores were used to calculate the percent of correctly produced target speech sounds. A stable baseline performance across sessions was defined as no more than a 10% difference in percent correct productions for three consecutive sessions. Generalization probe items were not considered when determining the stability of baseline measures.

**Treatment phase procedures.** Participants were scheduled for two 1-hour sessions per week for the duration of the study (4–5 weeks). Throughout treatment, the audibility of all verbal models and cues was confirmed through client report. Once stable baselines were established for all participants, treatment was initiated for the first target speech sound with the first participant. Correct production of the first target sound was modeled, and visual, verbal, and tactile prompts were provided to facilitate correct articulatory placement to produce that sound. Once a participant demonstrated correct articulatory placement, therapy proceeded using the following discrete trial procedures.

The researcher modeled the target speech sound in isolation or in a consonant-vowel (CV) syllable, and the participant provided an imitative response. Each imitated response was followed by immediate positive feedback for a correct production (i.e., the target response) or corrective feedback for an incorrect production. Positive feedback included verbal praise or nonverbal confirmation such as a head nod. Corrective feedback included saying, “No, that is not correct,” describing what was incorrect about the production, and possibly providing verbal instructions or a visual demonstration to facilitate a correct production. Each modeled response was scored as correct (+) or incorrect (−). Once a participant produced five consecutive correct modeled responses, the response mode shifted to evoked responses. The participant was then provided with a written symbol representing the target speech sound or CV syllable. The participant produced the target speech sound or syllable in response to the written symbol. Each evoked response was followed by immediate feedback, as previously described. If, at any time, the participant produced two consecutive incorrect evoked responses, modeling was reinstated until five consecutive correct modeled productions were obtained. Once the participant achieved 80% accuracy for evoked productions of the target speech sound in isolation or CV syllables, treatment shifted to the word level.

Word-level treatment progressed from modeled to evoked trials, as previously described. Initially, the participant was provided with a card containing the written target word while the researcher modeled the first target word. Once the participant produced five consecutive correct modeled responses, the response mode shifted to evoked responses in which the participant was provided with a card containing the written target word but no model. Once the participant demonstrated 10 consecutive correct evoked productions of the first target word, training began on the next word for that target speech sound. Treatment continued in this manner with additional exemplars until the treatment criterion was obtained for the first target speech sound. The treatment criterion was reached when the percent accuracy for production of the target speech sound in trained words increased by a minimum of 50% as compared with the session with the highest percent correct speech production during the baseline phase. For each participant, treatment ended when the treatment criteria for all three target speech sounds were met.

Once the treatment criterion was obtained for the first target speech sound, treatment began with the second target speech sound. The treatment protocol for the second and
third target speech sounds was identical to the procedure used for the first target speech sound. The initiation of treatment for the first target speech sound was then staggered across sessions from one participant to another. This procedure resulted in a staggered initiation of treatment across target responses for each participant and a staggered initiation of treatment across participants. After the study was concluded, data summary sheets for all three target speech sounds that indicated the percent correct speech production for all baseline, treatment, and probe measurements taken during the study were produced and are available in the online supplemental materials.

**Interrater Reliability**

Interrater reliability of correct and incorrect speech productions for each target response was established by having a second researcher independently score responses to baseline and probe items using an audiotape of the session. This occurred for one randomly selected baseline session and two randomly selected treatment sessions for each participant and for all administrations of the Arizona–3. For each of the sessions, word-by-word recordings of correct and incorrect speech sounds were compared among raters, and a percentage agreement score was calculated. Interrater agreement for the pretest and posttest Arizona–3 measures, and for baseline and treatment phase measures, ranged from 87% to 95.6%, from 93.3% to 96.7%, and from 90% to 93.3%, respectively, thus demonstrating a high level of interrater reliability.

**Results**

**Multiple Baselines Across Behaviors**

A comparison of percent correct speech production between baseline and treatment phases provided evidence to support the overall effectiveness of the therapy program. Figures 1, 2, and 3 present percent correct speech production of treatment (solid lines) and probe (dashed lines) words as a function of sessions for Participants A, B, and C, respectively. Baseline phase data for each target behavior appear to the left of the vertical line, while treatment phase data appear to the right of this line. These figures show that for each participant, baseline performance for all three target speech sounds was stable prior to the initiation of treatment and that each target behavior met the treatment criterion prior to the initiation of treatment for the next target behavior. The data presented for all three participants show a desired treatment effect in that percent correct production of the target speech sounds remained stable during the baseline phase and increased when the intervention was applied.

Target speech sounds selected for Participant A were /s/, /l/, and /r/, introduced in that order. Figure 1 shows that treatment was initiated for the first target speech sound (/s/) during Session 4, for the second target speech sound (/l/) during Session 5, and for the third target speech sound (/r/) during Session 6. The percent correct speech production for /l/ increased to criterion during Session 10, at which time treatment for Participant A was discontinued.

Target speech sounds selected for Participant B were /l/, /l/, and /l/, introduced in that order. Figure 2 shows that treatment was initiated for the first target speech sound (/l/) during Session 5, for the second target speech sound (/l/) during Session 6, and for the third target speech sound (/l/) during Session 9. The percent correct speech production for /l/ increased to criterion during Session 11, at which time treatment for Participant B was discontinued.

Target speech sounds selected for Participant C were /l/, /l/, and /l/, introduced in that order. Figure 3 shows that treatment was initiated for the first target speech sound (/l/) during Session 6, for the second target speech sound (/l/) during Session 7, and for the third target speech sound (/l/) during Session 9. The percent correct speech production for /l/ increased to criterion during Session 11, at which time treatment for Participant C was discontinued.

In addition to the visual inspection of graphical data, the calculation of effect size can quantify the outcomes for individual participants when a single-subject research design is used (Beeson & Robey, 2006). Effect size was calculated to quantify the magnitude of change in level of performance for each investigated target behavior with the exception of target behavior 2 for Participant 3 because of a lack of baseline variance in this case (see Table 2). Effect size was calculated using a variation of Cohen’s d statistic:

\[ d_i = \frac{\bar{X}_A - \bar{X}_A}{S_A}, \]

where \(A_2\) and \(A_1\) designate posttreatment and pretreatment periods, respectively, \(\bar{X}\) is the mean of the data collected in a period, and \(S_A\) is the corresponding SD. This version of Busk and Serlin’s (1992) \(d\) statistic (\(d_i\)) is a reliable estimator of effect size when pretreatment variance is a nonzero value. In addition, the effect sizes for each series of data points can be averaged to represent the treatment effect for each single participant by using a weighted mean, as described by Beeson and Robey. Using this calculation, weighted \(d\) statistics of 20.97, 24.32, and 27.95 were determined for Participants A, B, and C, respectively. It is not possible to code these treatment effects as small, medium, or large because standards for interpretation (i.e., benchmarks) for speech production research have not yet been established. Nevertheless, these results combined with visual inspection of the data presented in Figures 1, 2, and 3 reveal a positive trend in percent correct speech production as a result of treatment. Furthermore, the magnitude of change between prescreening and posttreatment measures (effect size) appears to be meaningful and may contribute to the establishment of benchmarks in this area.

**Multiple Baselines Across Participants**

Figure 4 shows percent correct speech production as a function of session for each target speech sound and each participant. The top, middle, and bottom panels illustrate Participants A, B, and C, respectively. The vertical lines...
represent the initiation of treatment for each target behavior. For each participant, stable baselines were established for all target behaviors, and the initiation of treatment was then staggered across participants. Visual analysis of the data presented for all three participants shows a desired treatment effect in that the percent correct target speech sound productions increased to criterion when the intervention was applied, while target speech sound production of participants that were not treated (i.e., baseline measures) remained stable.

**Generalization of Correct Speech Sound Production to Untrained Words**

Overall, visual inspection of the data represented in Figures 1, 2, and 3 supports the generalization of correct production of target speech sounds from trained words to untrained (probe) words as indicated by a corresponding increase in the percent correct speech production for both word types at the time treatment was introduced. Notably, for all the participants, there were sessions when the percent
correct speech production for the untrained words was greater than for the trained words. This may be the result of the number of probe items included on the baseline probe protocol (10 probe items vs. 20 treatment exemplars). Despite these exceptions, the percent correct speech production and overall pattern of performance for the trained and untrained words was similar during most treatment sessions.

**Generalization to Spontaneous Speech**

Pretest and posttest measures of the percent correct productions of target speech sounds taken during spontaneous speech were compared to determine whether production of target speech sounds during spontaneous speech improved after participation in the behavioral speech therapy program. Table 3 shows the number of opportunities each
participant had to produce each target speech sound and the percent correct production of target speech sounds in each speech sample. The results presented in the table show that all participants improved their productions of the target speech sounds at the conversational speech level even though treatment focused on single-word productions. The overall amount of improvement ranged from 12.7 percentage points (/n/ for Participant C) to 83.3 percentage points (/z/ for Participant B).

Discussion
In this preliminary study, the efficacy of using behavioral speech therapy to improve speech sound production in three adults with CIs was investigated. According to the American Speech-Language-Hearing Association’s levels of evidence for studies of treatment efficacy, this multiple-baseline design study would be rated at Level IIb: a well-designed quasi-experimental study (ASHA, 2004).
Visual inspection of the baseline and treatment data presented for the three participants shows an increase in percent correct speech productions after treatment onset for each target speech sound using a multiple baseline across target speech sounds for each participant (see Figures 1, 2, and 3) and across participants (see Figure 4), thus demonstrating a desired treatment effect. Effect size calculations further support the magnitude of change. These findings are consistent with several other studies that suggested speech therapy, in conjunction with the CI, may provide many adult CI recipients an opportunity to improve their speech production and communicative abilities (Busby et al., 1991; Dawson et al., 1995; Panteleimou et al., 2003; Waters, 1986; Wong, 2007; Zwolan, Kileny, & Telian, 1996). This is relevant because deaf speakers, including those with marginally normal speech, may experience lower self-image or significant social and vocational penalties (Cowie et al., 1982; Edgerton, 1985). Thus, improved speech production may potentially improve quality of life in these areas.

The replication of this finding using multiple baselines across behaviors and participants provided evidence that the intervention rather than some extraneous event was responsible for the change in percent correct speech production. Prior to the study, all three participants demonstrated multiple speech sound errors even after being implanted for at least 3 years. None of them reported recent changes in their speech production, and all three demonstrated stable baselines. Therefore, the findings were consistent with those of Dawson et al. (1995), Waters (1986), Wayner (2000), and Wong (2007), who concluded that the improved audition provided by the CI alone is often not adequate to ensure optimal communication.

In regard to treatment efficiency, each participant met the treatment criterion for his or her respective individual target speech sounds within four sessions and met the treatment criterion for all three target sounds within seven sessions. Treatment efficiency was also supported by generalization of target speech sounds to untrained stimuli and to conversational speech, even though treatment only occurred at the word level. Generalization of improved speech production to untrained words was confirmed through use of an intermixed probe (Hegde, 2003), as seen in Figures 1, 2, and 3. The generalization of correct speech sound production to spontaneous speech was established through comparing the speech errors identified during the analysis of spontaneous speech samples collected during prescreening and posttest procedures (see Table 3). The amount of improvement varied across target speech sounds and was likely influenced by the participants’ motivation and ability to monitor their spontaneous speech, the particular speech sound, and variability in the amount of training time devoted to each speech sound during treatment. Overall, however, these findings support the generalization of improved speech sound production in trained words to untrained words and to spontaneous conversational speech.

The concern over treatment efficiency has led to the development of concurrent treatment in which training occurs on the full range of presumed easy-to-hard exemplars within each session, intermixed and in random order (Skelton, 2004). Skelton found that concurrent treatment, as opposed to hierarchical task sequencing, promoted generalization of the target behavior and treatment efficiency. Training and probe exemplars for this study varied from one to five syllables in length and contained the target speech sound in all positions of words and in blends where appropriate. Although treatment in the present study focused on word-level productions, the findings further support the notion that phonemes do not need to be trained separately for various positions within the word, for blends, or for words of various lengths. The speed at which the participants met the treatment criterion for each target behavior was rapid despite the variability of exemplar types, as was generalization to untrained stimuli and spontaneous speech. Therefore, the present findings provide preliminary evidence to support the premise that behavioral speech therapy using mixed exemplars is both effective and efficient in improving the production of target speech sounds in adults with CIs.

Rossi-Katz and Arehart (2011) found that the majority of audiologists who responded to their survey did not consider speech-language evaluation results when directing
Figure 4. Percent correct target speech sound production as a function of session (baseline and intervention phases) for each target behavior and participant in a multiple-baseline, across-behaviors and -participants design. The staggered lines represent the initiation of treatment for each target behavior. TB = target behavior.

Participant A

Participant B

Participant C

rehabilitation for their older CI patients. Tucci et al. (1990) found that aural rehabilitation programs for adult CI recipients were sparse and that those offered by the CI centers often did not include speech production training as a primary component. In addition, many CI recipients did not receive additional AR services because of the long travel distances to their CI centers. The experiences of the participants in this study were consistent with results reported by Rossi-Katz and Arehart and by Tucci et al. in that they were treated at three different CI centers, yet none of them participated in comprehensive AR programs that included speech therapy. In addition, all of them had to travel at least 2.5 hr to their respective centers. These findings support the need for more community-based AR programs separate from CI centers.

Although these findings are preliminary, the identification of an efficacious approach to improving speech sound production in adult CI recipients contributes to the
establishment of an evidence-based speech therapy protocol for this group of individuals. This should encourage more speech-language pathologists to develop similar therapy programs for this population and should encourage CI centers to include behavioral speech therapy as a primary component in existing AR programs.

Limitations and Future Directions

The primary limitation of the present study is the inability to infer from the sample to a larger population owing to the small sample size. The multiple baselines across speech sounds and across participants do serve as a form of replication; however, inferring the present findings to the population of adult CI recipients is limited by using three participants. Therefore, additional replication of the presented methods will be required to demonstrate external validity of the findings. In addition, the study would have been strengthened by extending the duration of treatment so that the participants would be required to maintain the treatment criterion for the third target behavior across additional sessions. This would have permitted an analysis of the stability of percent correct speech productions over time for all target responses. Moreover, the study would have been strengthened by collecting follow-up measures of percent correct speech production for training and probe words and for spontaneous speech some time after the posttreatment assessment was completed to examine maintenance and generalization of trained speech productions over time.

The results of this study suggest several possibilities for future research. Replication of the present and similar research findings can further substantiate the efficacy of behavioral speech therapy for CI recipients. Specifically, discrete trial speech production training with different participants (e.g., ages, abilities), settings, and speech sounds can contribute to the scope of application of the present treatment procedure.

Table 3. Number of opportunities to produce target speech sounds and the percent correct production of target speech sounds in the spontaneous conversational speech samples during the prescreening (pretest) and posttest.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Number of opportunities to produce target speech sounds</th>
<th>Percent correct production of target speech sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prescreening</td>
<td>Posttest</td>
</tr>
<tr>
<td>Participant A</td>
<td>/s/</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>/r/</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>/l/</td>
<td>18</td>
</tr>
<tr>
<td>Participant B</td>
<td>/r/</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>/l/</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>/s/</td>
<td>30</td>
</tr>
<tr>
<td>Participant C</td>
<td>/r/</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>/l/</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>/s/</td>
<td>43</td>
</tr>
</tbody>
</table>

Acknowledgments

We thank Katie Nemeth, the graduate student who served as our second researcher during the collection of interrater reliability data, and Don Freed, department chair at California State University, Fresno.

References


