Clinical Focus

The Use of Visual Feedback to Elicit Correct /r/

Misarticulation of /r/ is one of the most commonly occurring speech disorders in schoolchildren (St. Louis, Ruscello, & Lundeen, 1992). It is also considered to be one of the most difficult problems to treat (Bernthal & Bankson, 1993; Clark, Schwarz, & Blakely, 1993; Goda, 1970; Pendergast, 1971; Secord, 1981; Shriberg, 1975, 1980). Although there are a number of useful programs for eliciting correct /r/ (e.g., Shriberg, 1975, 1980), there is a small group of clients who are unable to attain correct production despite intensive effort on the part of both client and clinician (Ruscello, in process). This may appear to be a minor problem; however, these individuals are identified by their peers as a result of their speech difficulty (Crowe Hall, 1991).

Many of the techniques for eliciting correct /r/ involve elicting /s/ first and then using it to elicit /r/. The techniques for eliciting /s/ rely heavily on the client’s ability to use auditory feedback to determine correct articulatory placement (e.g., Mowrer, 1975; Shriberg, 1975, 1980). For example, Shriberg (1980) developed a remediation technique for children with persistent /r/ errors. He wanted to address the problem of the exaggerated jaw, lip, and tongue movements he had noted in these children. The first step in the program involves the use of a bite stick to provide jaw stabilization and eliminate the exaggerated movement. Next, the clinician provides correct models of the target, but gives no directions regarding articulator placement. The child imitates the clinician’s utterances until his or her production is perceptually judged to be correct.

This reliance on auditory feedback appears necessary because there is little definitive tactile or kinesthetic feedback during the production of /r/ or /s/. Although the lateral borders of the tongue are touching the inner surfaces of the upper teeth, this type of contact occurs during the articulation of other sounds, and thus is not a sufficient cue for correct production. In fact, placement cues may be detrimental. Shriberg (1980) suggested that it may be clinicians’ placement cues that cause the exaggerated articulator movements he noted in children with persistent /r/ errors.

Biofeedback is a means of supplying an individual with information that is not normally available at a conscious level regarding the consequences of a behavior. It has been demonstrated that various forms of biofeedback are efficacious in the remediation of a number of different speech problems, including velopharyngeal closure problems (Hoch, Golding-Kushner, Siegel-Sadewitz, & Shprintzen, 1986; Lotz & Netsell, 1987; Ruscello, Shuster, & Sandwisch, 1991), stuttering (Guitar, 1975; Hanna, Wilfling, & McNeill, 1975; Lanyon, Barrington, & Newman, 1976; Moore, Dunster, & Lang, 1975), and articulation/phonological disorders (Guilford & Hnath-Chisholm, 1991; Hardcastle, Gibbon, & Jones, 1991; Netsell & Cleeland, 1973; Reisberg, 1968; Ruscello, Cartwright, Haines, & Shuster, 1993; Shuster, Ruscello, & Smith, 1992; Wolfe & Irwin, 1975).

Shuster, Ruscello, and Smith (1992) used biofeedback in the form of a real-time speech spectrograph to elicit correct production of /r/ from a 19-year-old college freshman who had been unable to attain correct production using other types of techniques (e.g., Mowrer, 1975; Shriberg, 1975). After 17 sessions, during which the client’s accuracy rates for production of /s/ and /r/ were...
0% (using the techniques cited above), the spectrograph was introduced. The client attained correct production of /r/ during the second session using the visual feedback from the spectrograph. We were able to elicit correct production of /r/ from the /s/. It appeared that, at least for this client, the visual feedback from the real-time speech spectrograph was effective in attaining correct production of /r/.

The purpose of the present investigation was to expand on the findings of Shuster, Ruscello, and Smith (1992). Although the results of that study indicated that spectrographic feedback was successful when more traditional methods were not, we had not controlled for the number of trials, since we did not know if the technique would be clinically useful. In addition, it was unclear whether this type of abstract visual feedback could be used by other clients, particularly children and adolescents.

Method

Subjects

There were two subjects. The first, Jerry (not his real name), was a 10-year-old boy, and the second, Tina (also not her real name), was a 14-year-old girl. Jerry was in a regular fourth-grade classroom, but was bused to another school twice a week for activities that were part of the gifted program. Tina was seen during the summer before she entered the eighth grade. She attended a middle school, where she was a straight-A student. Neither child had any history of language problems. An oral peripheral examination revealed no evidence of weakness or incoordination of the speech mechanism in either child. Both children demonstrated pure tone (Carheart & Jerger, 1959) and speech recognition thresholds (Martin & Dowdy, 1986) that were within normal limits (bilaterally), as well as speech discrimination (Martin, 1994) scores of 100% (bilaterally).

In addition to an inability to produce /t/, /s/ and post-vocalic /t/ [tot] (for), /at/ [tar], /et/ (ear), and /et/ (air)] correctly, Jerry produced /s/ with a lateral lisp. Tina produced no sounds in error other than /r/ and /s/. Jerry had been enrolled in articulation treatment in the public school for the remediation of /r/ for 2 years. Tina had been enrolled in articulation treatment for the remediation of /r/ for 4 years. Neither child was able to produce /r/ correctly in any context.

Procedure

Baseline assessment and prespectrographic intervention. Both children were given both the word and sentence versions of the McDonald Deep Test of Articulation for /t/ (McDonald, 1964). In addition, each was asked to produce a list of words and sentences that contained /t/ and /s/. Each child was asked to produce five trials of each word or sentence. Five tokens were elicited following Weismer’s (1988) suggestion that multiple examples provide a more stable statistical estimate of disordered speech when conducting acoustical analysis.

The treatment procedures implemented for the two clients differed slightly because of their geographic location to the clinic. Since Jerry resided in the area, he was seen more intensively. Consequently, it was decided to provide traditional treatment initially, and then introduce biofeedback if the former was not successful. The traditional treatment (which relied mainly on the use of auditory feedback to attain correct production) was provided for the first half of a clinic semester. He was seen twice a week for 50 minutes each session. During each session, he produced 30 trials of each of the following: /otr/, /tr/, /ar/, and /et/. In each trial, Jerry was instructed to try to glide from the initial vowel into an /r/. In addition, the technique in which /r/ is shaped from /l/ (Shriberg, 1975) was also used. Verbal feedback was provided by the speech-language pathologist after each trial. Jerry produced 30 trials using each of these target productions each session, which resulted in a total of 150 trials per session.

At the end of 12 sessions (1,800 trials) of traditional treatment, a tape was made of Jerry’s spontaneous production of the various target /r/’s in isolation and in words. The tape was played for two clinical supervisors, who independently made judgments regarding the accuracy of articulation. They were unaware of the purposes of the study. The judges held the Certificate of Clinical Competence (CCC) in speech-language pathology and had been practicing for 22 and 19 years, respectively. They were asked to judge each production as correct or incorrect.

Tina did not receive traditional treatment. Unlike Jerry, she lived an hour away from the clinic and could only attend treatment during the summer, so our time with her was limited. In addition, she had been working unsuccessfully (as indicated by the fact that she was still unable to produce /r/ correctly in any context) on /r/ for 4 years in public school treatment, which we thought was sufficient to indicate that she was unlikely to attain correct production through any traditional means. Before spectrographic intervention, two judges, both of whom held master’s degrees in speech-language pathology and the CCC, rated Tina’s audiotaped productions in the same manner as the judges who listened to Jerry’s tape.

Spectrographic intervention. The number of weekly treatment sessions during which spectrographic feedback was used differed for the two clients. Jerry continued to receive two 50-minute sessions each week as he had during the time he received traditional treatment. Because Tina lived so far from the clinic, she received treatment once a week for an hour. The procedure for using the spectrograph was identical with both subjects.

A Kay Elemetrics Model 5500 real-time spectrograph was used. This instrument displays the spectrogram on a monitor, and as long as the spectrograph is recording, the spectrogram moves continually across the screen from right to left. The time axis (which affects how quickly the image moves across the screen) was set at 2 seconds, because it had been determined from previous experience that at 1 second the image moved too quickly, and at time axes longer than 2 seconds, the delay between production and the appearance of the image on the screen was too long. During the initial session, the investigator produced a variety of vowel sounds, while the subject watched the
effect of the sound changes on the visual display of the spectrograph’s screen. The investigator periodically stopped the display and pointed out the formants, which were referred to as “lines.” When the subject was able to point out the formants of the vowels on five out of five different vowel productions, the /s/ was addressed. Each subject required no more than 10 examples in order to reach this criterion.

The investigator first produced an /s/, then stopped the display, then started the display moving again as the subject produced his or her /s/. In this way, both productions could be displayed simultaneously (as long as the display was not “enhanced”). The investigator pointed out that the important difference between the two productions was that her second and third lines were so close that they appeared to be one, while the subject’s production clearly had two different lines. The simultaneous display and discussion of the differences were repeated for post-vocalic /l/ in the context of the vowels /i/ and /a/ (/ir/ and /al/).

When the subject could consistently point out the differences between the position of the second and third lines of the investigator’s production and those of his or her production, the actual treatment began. The criterion for determining that the client understood the display was the same as for the vowel examples. Again, both clients required fewer than 10 trials to reach criterion.

The same number of training trials that had been used during traditional treatment were also completed during Jerry’s work with the spectrograph. Jerry produced 30 trials each of /ar/, /or/, /ir/, and /er/, as well as 30 trials using /l/ as a starting position to shape /s/, again resulting in 150 total trials per session. Since Tina was coming only once a week for 10 minutes longer, she produced 45 rather than 30 trials of each item, for a total of 175 practice trials per session. A correct production was defined as one which was both perceptually judged to be correct by the investigators and which yielded a visual pattern on the spectrograph which has been correlated with the perception of a correctly produced /l/.

### Results

**Baseline assessment and prespectrographic intervention.** At the end of 12 sessions of traditional treatment, the judges’ ratings indicated that Jerry was still unable to produce either /l/ or /s/ correctly. Agreement between the judges’ ratings was 93%. These perceptual judgments were supplemented with acoustic measurements. Measurements were made on the onset formant frequencies as well as on the formant frequencies at the midpoint of each /l/ periodic portion. The formant values and patterns of the vowels that Jerry produced before treatment did not differ from those of the vowels that he produced after 6 weeks of treatment. Figure 1 shows spectrograms of Jerry’s production of ‘heart’ before (on the top) and after traditional treatment. Neither production shows the pattern that is associated with the perception of correct /l/, which is a low third formant (sometimes below 2 kHz) and second and third formants that are very close in frequency, often so close as to appear one (Dalston, 1975).

The ratings of the judges who listened to Tina’s tape indicated that she was unable to produce /l/ or /s/ correctly at the time treatment was initiated. Agreement between the judges was 96%. These judgments were also confirmed acoustically. Figure 2 shows spectrograms of the target words “wing” and “ring” produced by Tina. The initial portion of the word ‘wing’ shows the pattern that has been associated with perception of /w/. The first and second formants are very low initially, and there is delayed onset of the third formant. In the target “ring” utterance, the pattern is characteristic of neither a production which yields the perception of /w/ nor of one which yields perception of correct /l/. Although the first formant is very low (as in the target /w/ production), the second formant is higher than in “wing”, and the third formant does not show a delayed onset. The pattern also is not characteristic of that which is normally associated with perception of correct /l/. The second formant is lower than one would expect; there is a wide separation in frequency between the
second and third formants, and the third formant is higher (above 3 kHz). This type of formant pattern for a so-called “distorted” /r/ has been described by Ohde, McCarver, and Sharf (1989).

**Spectrographic intervention.** Jerry attained consistently correct production of /ar/ (26/30 or 87%) during the sixth session using the spectrograph (session 18). Table 1 shows the accuracy rate (in percentages) for each token for each session using the spectrograph (sessions 13–22), and for the two sessions after regular use of the spectrograph was stopped (sessions 23 and 24). Although he was able to produce some of the other vowels correctly during session 18, the number of correct productions was fewer than for /ar/. During the next three sessions, accuracy for /ar/ fluctuated; however, /ar/ and (by the ninth session) /or/ were the contexts that were most facilitative of correct production. The use of the spectrograph was discontinued after the tenth session because Jerry was able to produce the target sounds correctly in words without the visual feedback. He was then transferred to the regular clinic for transfer and maintenance. At that time, we were able to elicit /r/ from /-r/. He progressed from correct production in words, to sentences, and then conversation during the next semester and was subsequently dismissed from treatment. Figure 3 shows spectrograms of Jerry’s productions of the vowel /ar/ produced before treatment was initiated with the spectrograph and after he attained correct production. It can be seen in the first spectrogram that the second and third formants are not close. In the second spectrogram, the second and third formants are so close as to appear one.

Tina attained consistently correct production of /er/ (33/45 or 73%) by the third session using spectrographic feedback. Unlike Jerry, who seemed to prefer the back vowels, the context of a front vowel seemed to be most facilitative for Tina. Although she was able to produce the back vowels /ar/ and /or/ correctly some of the time, the number of correct productions was much lower (4% and 22%) than for /er/ and /ir/ (73% and 47%). This pattern continued until the sixth session, when she began to produce all of the sounds with 70% accuracy or better (Table 2).

Figure 4 shows a spectrogram of Tina’s correct production of target “ring.” Unlike the spectrograms shown in

<table>
<thead>
<tr>
<th>Session Number</th>
<th>ar</th>
<th>or</th>
<th>er</th>
<th>lr</th>
<th>l - r</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>87</td>
<td>27</td>
<td>3</td>
<td>27</td>
<td>40</td>
</tr>
<tr>
<td>19</td>
<td>60</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>47</td>
</tr>
<tr>
<td>20</td>
<td>71</td>
<td>69</td>
<td>52</td>
<td>21</td>
<td>68</td>
</tr>
<tr>
<td>21</td>
<td>73</td>
<td>81</td>
<td>50</td>
<td>50</td>
<td>67</td>
</tr>
<tr>
<td>22</td>
<td>73</td>
<td>84</td>
<td>61</td>
<td>73</td>
<td>86</td>
</tr>
<tr>
<td>23</td>
<td>97</td>
<td>90</td>
<td>87</td>
<td>80</td>
<td>87</td>
</tr>
<tr>
<td>24</td>
<td>100</td>
<td>90</td>
<td>86</td>
<td>83</td>
<td>90</td>
</tr>
</tbody>
</table>
Figure 2, this production shows a pattern that is typically associated with the perception of a correctly produced /r/. During the /r/ periodic portion (which she sustained so that the perception was /ɔːr/) the second and third formants are very close in frequency, and the third formant frequency of about 2 kHz is much lower than in her error production of “ring”, shown in Figure 2.

The week that intervened between Tina’s sessions appeared to affect her performance when compared to that of Jerry, who experienced a maximum of 4 days between sessions. Once Jerry had attained consistently correct production of /ar/, he required only one or two trials at the beginning of a new session in order to “find” the correct articulation. During sessions four through seven, Tina’s first five or six trials at the beginning of each session involved some “searching” with her articulators in order to remember what she needed to do to attain a consistently correct production. The use of the spectrograph was discontinued after the eighth session. By the tenth (and final) treatment session, Tina was able to produce the target sounds correctly in words without the visual feedback of the spectrograph. School started and she began receiving treatment at her school. Her school clinician reported that transfer was achieved at the sentence level, and she was continuing to work on conversation.

Once these children were able to produce /r/ correctly without visual feedback, both clients required a period of time during which they needed to prolong the target portion of the stimulus words and insert a vowel before /ɔːr/ in order to attain correct production. When they uttered the stimulus words at a normal rate of speech, they were not correctly produced. This period of time during which slowed production was necessary was longer for some contexts. Spectrograms of Jerry’s production of the target word “heard” are shown in Figure 5. In the first spectrogram, the second and third formants following the frication are initially widely spaced and then come together. This production yielded the perception: /haʊd/. In the second spectrogram, the second and third formants are close immediately following the frication, and this production yielded the perception: /hɔːd/.

Discussion

The results of this investigation indicated that at least some children and adolescents may be able to use the visual feedback from a real-time spectrograph to attain correct articulation of /r/. In addition, the data provide further support for the hypothesis that this type of feedback can be beneficial for individuals who have been unable to attain correct production of /r/ using traditional, auditory-based treatment techniques (Shuster, Ruscello, & Smith, 1992).

An alternative explanation for Jerry’s progress could be that his success was due to the cumulative effect of the treatment rather than the use of the visual feedback. We did not think, however, that this was a likely explanation. It was found that once the use of visual feedback was discontinued, he needed periodic “tune-ups” with the spectrograph to attain correct production of /r/ in the
context of front vowels. Without the visual feedback, he was unable to correctly produce /ir/ or /Er/, despite repeated trials. We then employed the spectrograph. Jerry produced a word containing /ir/ or /Er/, which was displayed on the spectrograph screen. It was unnecessary for the clinician to remind Jerry of the pattern of a correct production, because he had no difficulty remembering what the formant pattern should look like. Within two or three trials he produced the word correctly. He was then given approximately 15 minutes of practice on words containing /ir/ and /Er/. If his success was merely due to the amount of treatment, he should not have needed this type of “refresher” with the visual feedback. The particular form of treatment should not have mattered.

Tina’s progress indicated that even a client who may appear to have a poor prognosis for acquisition of correct production (given her lack of progress after 4 years of treatment and her age) can benefit from the use of visual feedback. In addition, she demonstrated that it is possible for a client who establishes a target /r/ production from a real-time spectrograph to finish treatment in a setting where there is no access to such a device (Ruscello & Shuster, 1993). The cost of a real-time spectrograph is decreasing; however, until more facilities possess this device, the approach used in this report is a viable alternative to such technology on site.

An interesting question is why this type of biofeedback worked when other approaches had failed. One hypothesis is that by providing visual feedback, it becomes unnecessary for the client to use auditory feedback to attain correct production. In fact, the client is specifically instructed to ignore the way the articulation sounds with this approach. Although we have used the approach with a small number of clients, all three stated, when they first achieved correct production of /r/, that it did not sound “right” to them. As stated earlier, nearly all of the traditional approaches rely on the use of auditory feedback for attaining correct production. Shriberg (1980) noted that the client must “have an auditory percept that can be reliably associated with good /r/ production” (p. 104) in order to achieve correct production. Given the comments of our clients, it may be that they do not have a “good” auditory percept (or underlying representation) for /r/.

The underlying representation of a morpheme is the abstract representation that an individual holds for the morpheme. The abstract representation includes the meaning of the morpheme as well as “all idiosyncratic learned phonological properties of the morpheme” (Dinnsen, 1984, p. 5). Many investigators believe that children have two underlying representations for a morpheme. One is a receptive level of representation that consists of the auditory characteristics of words the child can understand (Dinnsen, 1984; Fey, 1989; Locke, 1988; Schwartz & Leonard, 1982; Straight, 1980). The second is an expressive level of representation that consists of the articulatory characteristics of the words the child can produce (Dinnsen, 1984; Fey, 1989; Locke, 1988; Schwartz & Leonard, 1982; Straight, 1980).

Straight (1980) suggested that these two representations are independent of one another, although they can influence one another. The characteristics of the receptive level of representation are derived from the speech the child hears, whereas the characteristics of the expressive level are derived from his or her own articulatory movements and include the anticipated auditory sensations produced by an articulation (i.e., what the individual expects to hear as a result of a particular movement). Straight hypothesized that changes in the child’s production during development reflect a gradual merger in the match between the receptive level and the expressive level of representation.

If Straight’s theory is correct, the clients who do not make progress in treatment for /r/ may be those who fail to develop a match between their receptive and expressive levels of representation. This would explain why they do not make progress in treatment that relies on auditory feedback, why they can discriminate the correct and
incorrect productions of others\textsuperscript{1}, and why, when they finally do achieve correct production, they state that it does not sound “right.” The /r/ they expect to hear in the speech of others differs from that which they expect to hear in their own speech. When asked to produce /r/ (or /ɾ/), they produce what sounds “right” to them, which is their own version of the sound. By providing visual feedback, the clinician can bypass the client’s auditory system. The possibility that such a basis for a phonological disorder may exist has been suggested by Vihman (1993).

Locke (1980a, 1980b) obtained data that support the notion that some children judge their error productions to be correct. He asked kindergarteners to name pictures. In producing these labels, a number of the children substituted or omitted phonemes. The children were then presented with these same labels and the labels were either correctly articulated, incorrectly articulated as the children had produced them, or incorrectly articulated in a “neutral” way (with an error that was not like those produced by the children). The children were asked to judge each label in terms of the correctness of the articulation. Most of the children accepted the correct productions and rejected the errors. Some of the children, however, accepted both the correct form and their own incorrect form, while rejecting the neutral incorrect form.

Regardless of why it works, however, visual feedback from the real-time spectrograph appears to be efficacious for the remediation of /r/ in individuals who have not made progress through traditional approaches. Although recently other types of biofeedback and oral modifications have been used to elicit correct production of /r/ (e.g., Gibbon, Hardcastle, & Suzuki, 1991; Clark, Schwarz, & Blakely, 1993), an advantage of the spectrograph is that it is not necessary to introduce a foreign object, such as an artificial palate, into the oral space, and thus articulation is more natural during the learning process. Also, once the equipment is purchased, no additional expense is involved, unlike approaches where it is necessary to make an artificial palate. This type of feedback can be used with a variety of clients, including those who have a hearing loss. Finally, the spectrograph allows the freedom to achieve correct production of /r/ in any manner the client chooses. There is a tremendous amount of variation in the articulation of /r/ both between and within individuals (Delattre & Freeman, 1968; Kent & Read, 1992; Ohde & Sharf, 1992; Shriberg & Kent, 1982; Zwadzki & Kuehn, 1980). In some productions the tongue is bunched, whereas in others it is retroflexed. Some speakers demonstrate pharyngeal, palatal, and/or labial narrowing, whereas others do not. Since the visual feedback from the spectrograph is not dependent on the movement of one specific articulator (e.g., narrowing in the palatal, pharyngeal, or labial regions can lower F\textsubscript{3}), the client may use any or all of these types of articulation to achieve the formant pattern that leads to the perception of a correct production.

\textsuperscript{1}All three clients with whom we have used spectrographic feedback have had no difficulty in discriminating the incorrect and correct productions of /r/ of other speakers.

References


Key Words: speech production, phonological disorders, articulation disorders, speech perception