The Effectiveness of a Multimodal Vowel-Targeted Intervention in Accent Modification

Kathryn W. Brady
Natalie Duewer
Amie M. King
Southern Illinois University Edwardsville

Accent modification is rapidly becoming a well-established area of practice for speech-language pathologists (SLPs). In 2000, 77% of accredited speech-language pathology graduate programs responding to a survey reported that they provide foreign accent modification services in their clinics (Schmidt & Sullivan, 2003). In 2007, the Los Angeles Times reported a “surge in accent reduction classes” (Gorman, 2007, para. 6); the next year, U.S. News and World Report identified accent reduction specialist as one of 11 best-kept-secret careers in the United States (Nemko, 2008); and in 2012, Feinstein-Whittaker, Wilner, and Sikorski indicated that accent modification was “a growing niche in corporate America” (2012, p. 28).

Despite the increasing prevalence of accent modification as an area of practice in speech-language pathology, little information is available concerning the effectiveness of these services. The American Speech-Language-Hearing Association (ASHA) provides limited information on empirically validated interventions targeting accent modification. ASHA’s evidence maps, for example, which serve as a resource for SLPs who are attempting to implement training, and vowel accuracy continued to improve as the training sessions continued. The accuracy of vowel production was maintained 2 weeks post training. In contrast, the accuracy of the untrained control vowel remained unchanged. Results suggest that visual feedback combined with traditional articulation training strategies is effective in targeting vowels in a nonnative speaker of English.

Conclusion: Visual feedback programs such as spectrograms may be beneficial as part of a larger instructional program to target vowel production in accent modification clients.

KEY WORDS: speech intelligibility, accent modification, vowels, pronunciation
evidence-based practice, do not include information on accent modification services (ASHA, 2015a). Nor is accent modification included on the ASHA practice portal, which provides SLPs with clinical and professional practice guidance (ASHA, 2015b).

Although research has been conducted on the efforts of nonnative speakers to change or improve their English pronunciation, most of it has come from researchers who are investigating the effectiveness of instruction strategies in English-as-a-second-language (ESL) classrooms (e.g., Derwing, Munro, & Wiebe, 1998; Derwing & Rossiter, 2003). As Behrmann (2014) pointed out, there are potential difficulties generalizing these findings to accent modification training provided by SLPs, where clients are typically more proficient in the English language than in an ESL setting, and training is more individualized than is possible in the context of an ESL classroom. Fortunately, research is beginning to emerge on the effectiveness of accent modification services provided by SLPs. Fritz and Sikorski (2013), for example, presented evidence that Korean participants in a university accent modification program made significant gains both in terms of overall intelligibility and confidence in their ability to communicate effectively. Similarly, in a single-subject, alternating-treatment design, Behrmann (2014) found that accent modification training resulted in gains in intelligibility and ease of understanding for four Hindi-speaking participants. They also found that directly targeting segmentals (in this case, consonants) had a direct effect on the accuracy of their participants’ consonant production, and directly targeting prosody had a direct effect on the accuracy of their participants’ production of American English prosody, suggesting the potential importance of targeting each of these broad areas.

Thus, research to date indicates that accent modification training can be effective. However, we have little evidence to indicate what specific aspects of training or what training strategies are particularly effective or ineffective. Data are sorely needed that can be used to guide SLPs’ clinical service provision in the area of accent modification. ASHA clearly identifies this service as part of the scope of practice in speech-language pathology (American Speech-Language-Hearing Association, 2016); however, specific therapeutic guidelines currently do not exist. This study will begin to fill this gap by examining the potential effectiveness of one specific approach to vowel training with a non-native speaker of American English.

The Nature and Impact of Foreign Accent
ASHA (2013) defines an accent as “the unique way that speech is pronounced by a group of people speaking the same language” (para. 2). Despite the apparent simplicity of the definition, accent is an extremely complex phenomenon. Speech production can vary in a number of ways, including the production of phonemes, sound patterns, and prosody (i.e., stress patterns, intonation, rhythm, etc.). Among native speakers of a language, accent can vary as a result of the region the speakers are from and/or the social group they belong to. For second-language (L2) learners, accent is even more complex. The age at which they are initially exposed to the L2 (Abu-Rabia & Iliyan, 2011; Flege, Frieda, & Nozawa, 1997; Flege, Munro, & MacKay, 1995; Piske, MacKay, & Flege, 2001; Tahta, Wood, & Loewenthal, 1981), and the frequency with which they continue to speak their first language (L1) as they learn the L2 (Flege, Frieda et al., 1997; Piske et al., 2001), appear to play a role, as do cognitive factors such as skill at encoding phonetic information (Carroll, 1981), perceptual abilities (e.g., the ability/inability to perceive phonemic contrasts that are not part of one’s L1), and psychological factors such as motivation (Stevick, 1976) and ego (Guiora, Beilt-Hallami, Brannon, Dull, & Scovel, 1972).

Particularly relevant to any discussion of the effectiveness of an accent modification program is the idea that phonology (i.e., pronunciation, accent) is among the most intransigent aspects of an L2 to master. Scovel (1969) pointed out that, although it is difficult for adult learners to master the syntax of their L2, it appears to be more achievable than mastering phonology. He offered the example of Joseph Conrad, the Polish American author whose mastery of English allowed him to write great works of literature, but whose spoken English was unintelligible enough to prevent him from lecturing publicly.

The relative difficulty of acquiring native-like pronunciation in an L2 has been supported by research. Flege, Yeni-Komshian, and Liu (1999), for example, found that, although other factors such as level of education and amount of L1 and L2 use played a role in Korean participants’ level of skill in morphosyntax, the age of acquisition of English alone predicted the level of accentedness of the participants’ spoken English. Thus, the hypothesis that there is a critical period (or in its weaker version, a sensitive period) during which L2 learning must occur if it is to be successful (Lenneberg, 1967) seems to be potentially even more relevant to phonology than to other aspects of language. Perhaps as a result of this, pronunciation teaching has been relatively neglected in ESL teaching (Celce-Murcia, Brinton, Goodwin, & Griner, 2010), and a common perception has arisen that accent (i.e., pronunciation, L2 phonology) is all but impossible to master when one comes late to L2 learning. When we add to that notion the cultural,
psychological, and political issues surrounding accent, we find that accent modification is indeed a very complex phenomenon.

Unfortunately, listener perceptions of a “foreign accent” can potentially carry with them some negative ramifications for the speaker. Hosoda, Stone-Romero, and Walter (2007) presented experimentally controlled recordings of speakers speaking standard American English or Vietnamese-accented English to small groups of diverse American students. The students were asked to fill out a rating scale measuring a variety of speaker attributes. The students rated the speakers with Vietnamese-accented English as having a stronger accent, exhibiting less proficiency at communication, employing less power, and exerting a more negative effect on listeners than the speakers who spoke American-accented English (Hosoda et al., 2007). In addition, Lindemann (2003) conducted a study in which native English speakers listened to English recordings from both native Korean and English speakers and were asked to rate the speakers’ voices on six status-related traits (e.g., intelligence, incompetence) and six solidarity-like traits (e.g., friendliness, insincerity). Results from the study revealed that all of the native Korean speakers were judged more negatively with respect to status-like traits when compared with the native English speakers (Lindemann, 2003).

Judgments that are made about a person based on his or her accent can carry over to the workplace. For example, the type of foreign accent a person has, combined with the communication requirements of a job, can have an effect on the likelihood of a person being hired for a position (Hosoda & Stone-Romero, 2010). Huang, Frideger, and Pearce (2013) asked 179 undergraduate and graduate students of different races to make hiring decisions about individuals with and without foreign accents. Participants were given a picture of the interviewee paired with an audio recording of an interview. Picture/audio pairs included a White male with a native accent, the same White male with a nonnative accent, an Asian male with a native accent, and the same Asian male with a nonnative accent. In relation to hiring for managerial positions, the participants rated the speakers with a nonnative accent, regardless of race, as having less political skill, and they indicated that they would be significantly less likely to hire these speakers than those with a more native-like accent. Huang et al. found that collaborative skill and communication skill did not have a significant effect on employment recommendations; accent alone had an impact.

For a variety of reasons, then, nonnative speakers of English—even speakers who are highly intelligible—might seek out the services of an SLP in an effort to modify their accents. ASHA clearly establishes that “no dialectal variety of English is a disorder or a pathological form of speech or language” (ASHA, 2003, para. 1), but includes accent modification under the scope of practice of SLPs as “elective communication modification services” (ASHA, 2004, para. 5). The underlying assumption of an elective speech service is that SLPs understand and respect the right of individuals to determine the extent to which they choose (or do not choose) to change their speech patterns. Some nonnative speakers may desire to improve their pronunciation of American English in order to increase the effectiveness of their communication or to sound more native like. SLPs’ extensive knowledge of speech-sound production, combined with their knowledge of clinical techniques and strategies, puts them in a unique and qualified position to provide accent modification services to clients who seek out these elective services.

Vowels as Viable Accent Modification Targets

Both segmental and suprasegmental (i.e., prosodic) aspects of English are typically targeted in accent modification training (e.g., Avery & Ehrlich, 1992; Celce-Murcia et al., 2010). Individual targets are generally selected on the basis of client error, and these errors, in turn, typically involve sounds or sound patterns in English that are absent from the client’s L1. Speakers of German as an L1, for example, who do not have the interdental fricative as part of their L1 phonetic inventory, might substitute /z/ for /ð/ when speaking English. Speakers of languages with predominantly simple syllable structures (e.g., Kiswa-hili, Japanese) might reduce clusters in their English speech. Thus, targets in accent modification are chosen on the basis of patterns associated with L1 influence and with individual speech patterns.

Vowels are important and viable targets in accent modification. The American English vowel system is more extensive than the vowel system in many other languages, creating potential difficulties for nonnative speakers of English. Swahili, for example, has only five phonemic vowels, whereas American English has as many as 15, depending on the particular dialect of the speaker (Ladefoged, 2005). Thus, nonnative speakers of English might fail to make the vowel distinctions that signal meaning differences in English because they lack the English vowels in their phonetic repertoire. Vowels also play a large role in speech intelligibility. Several studies have established that vowels play a greater role than consonants in spoken word recognition at the sentence level (Fogerty & Humes, 2010; Fogerty, Kewley-Port, & Humes, 2012; Kewley-Port, Burkle, & Lee, 2007).
Despite the important role they play in speech intelligibility, vowels have frequently been neglected as targets in accent modification. Vowels are more difficult to describe in terms of articulation than consonants because the articulators do not make contact with each other; rather, the distinctions between vowels are created by changing the shape and size of the oral cavity (Celce-Murcia et al., 2010). In addition, potential limitations in nonnative speakers’ ability to perceive differences between vowel pairs when the distinction does not exist in their own phonology make traditional articulation strategies problematic. Thus, vowels are difficult not only for the client, but also for the SLP.

As a result, a multimodal approach to vowel training incorporating visual as well as auditory feedback would seem to be a logical approach to targeting vowels in the speech of nonnative speakers. No studies have been conducted to examine the effectiveness of incorporating visual feedback in targeting vowels in accent modification services; however, interventions using visual feedback have been implemented with native speakers of English to improve vowel production and thus increase speech intelligibility (Byun & Hitchcock, 2012; Ertmer, Stark, & Karlan, 1996). Byun and Hitchcock (2012) suggested that biofeedback, a strategy in which clients match sounds and formant patterns in a visual display, can improve clients’ production of speech sounds. In their study, they used CSL Sona-Match software, which uses real-time linear predictive coding spectra, to target vocalic /ɛ/ in 11 native English–speaking children ages 6;0 (years;months) to 11;9. Ten of the 11 participants had been receiving traditional articulation treatment (for /r/ and vocalic /r/ errors) from 1 to 4 years before the visual biofeedback program was implemented.

In the Byun and Hitchcock (2012) study, a traditional articulation intervention was implemented for 4 to 6 weeks. At that point, a biofeedback component was added to the traditional treatment. After allowing the children to familiarize themselves with the biofeedback program, the clinician explained the formant properties of /ɛ/. Participants were asked to match the target sound /ɛ/ to the /ɛ/ template that was provided using the Sonamatch software. A total of 30 productions of /ɛ/ were obtained in a series of five trials. Productions of /ɛ/ were significantly more likely to be rated as perceptually accurate following the biofeedback intervention. In addition, posttest F3 frequencies were lower than midtest F3 frequencies, demonstrating that the /ɛ/ sounds that were produced after the biofeedback intervention were more acoustically similar to the adult target than those produced before the intervention. Byun and Hitchcock concluded that biofeedback, in conjunction with traditional articulation treatment, can facilitate the correct production of /ɛ/ in monolingual children with /ɛ/ errors.

Ertmer et al. (1996) also investigated the effectiveness of visual feedback in vowel production. In their study, two children with profound hearing loss received treatment aimed at facilitating correct vowel production. A clinician provided placement instructions and other visual cues to elicit vowel production. In addition, spectrograms were provided to give the children visual feedback about the accuracy of their productions. Three vowels were targeted, and one additional vowel, which served as a control, was monitored but was not targeted. Both of the study participants improved significantly in the accuracy of the trained vowels; no improvement was made in the untrained vowel.

These studies suggest that vowel training programs, specifically, programs incorporating visual feedback, have been shown to improve the vowel accuracy of native speakers. However, no researchers to date have examined a multimodal approach to vowel intervention with nonnative speakers. The purpose of this single-subject study was to evaluate the effectiveness of a vowel-targeted intervention incorporating traditional articulation strategies and visual feedback in increasing the accuracy of English vowel production in a nonnative speaker of English.

**METHOD AND DESIGN**

**Participant**

The participant was a 24-year-old Iranian male who was a native speaker of Farsi. He was a graduate student at a mid-sized university in the Midwest and had expressed interest in improving his pronunciation of American English, specifically for future employment purposes. When the study began, he had been in the United States for approximately 11 months. The participant reported that he had studied English at a private language institute in Iran for 4 to 5 years before coming to the United States. His English classes focused on oral communication, with little emphasis on pronunciation. He had not studied English or participated in accent modification services in the United States before or during the course of the study. He had no history of speech or language disorders in his native language.

The participant’s overall English language abilities were not directly assessed, but they were adequate for him to participate successfully in a graduate program, where a minimum score of 79 was required on the TOEFL iBT for admission. His spoken English was generally intelligible but was perceptually
accented. Two speech-language pathology student clinicians who had no prior accent modification training rated the participant’s speech on average as 92% intelligible and his degree of accentedness as 4.5 on a 6-point scale, where a rating of 1 indicates a strong degree of accentedness and 6 indicates little to no accent (Morley, 1987). A rating of 4 on this scale suggested that the participant’s “accent causes interference primarily via distraction; the listener’s attention is often diverted away from the content to focus instead on the novelty of the speech pattern” (p. 485).

The vowel system of the participant’s native language, Farsi, is considerably more limited than that of English. Specifically, Farsi has only six vowels, three front (/i/, /ε/, /æ/) and three back (/u/, /o/, /ɑ/), with no tense-lax contrasts, no central vowels, and no diphthongs (Keshavarz & Ingram, 2002). This differs strikingly from American English, which has approximately 15 vowels (depending on a speaker’s dialect) that are phonemic ( Ladefoged, 2005), including three diphthongs. Vowel errors are relatively common in the English of Farsi speakers (Avery & Ehrlich, 1992), with the lax and central vowels being particularly problematic. Thus, we anticipated that the participant would have a more limited English vowel inventory than that of a native English speaker, resulting in vowel substitutions in his spoken English. Specific information about his vowel productions is discussed in the Procedure section.

Design

We used a single-subject multiple-probe design across behaviors to examine the effectiveness of the intervention. The single-subject research design is commonly employed in treatment studies because it allows conclusions about treatment effectiveness to be drawn that could not otherwise be drawn from a single case (Kazdin, 2011). The multiple-probe design involves targeting a number of behaviors in a controlled, sequential fashion. If change occurs in a specific behavior during the period of time in which it is targeted—and only during the period in which it is targeted—and if little or no change occurs in the behavior that is not targeted (i.e., the control), we can conclude that the observed changes in behavior resulted from the training and, thus, that the training was effective. In this design, then, the participant acts as his or her own experimental control (Kazdin, 2011).

We targeted three vowels in our study, and an additional untrained vowel served as a control. The initiation of training of each vowel was staggered, with training of Vowel 1 beginning after baseline was established. When a minimum criterion of 80% accuracy was reached in the participant’s production of Vowel 1 at the word level, training of Vowel 2 began. When Vowel 2 was produced with a minimum of 80% accuracy at the word level, training of Vowel 3 began. Vowel 4 served as the control and thus was not trained.

Materials

Assessment. In order to assess the participant’s production of 15 American English vowels and diphthongs (i, i, e, e, æ, o, A, Ə, ɔ, u, o, a, 造林, ɔɪ, ɪ), we created an assessment list of 45 words (the vowel assessment list). Each vowel (with the exception of / ə/ and / ɔ/) was presented in three words, two of which were monosyllabic and one that was multisyllabic. Because / ə/ appears only in unstressed syllables and /ɔ/ in stressed syllables, the target words were selected differently: / ə/ was presented in the final syllable of three multisyllabic words, and /ɔ/ was presented only in monosyllabic words. The words on the assessment list were pseudorandomized such that no two consecutive words contained the same vowel.

Training. Session stimuli consisted of the target vowel in isolation and in one-syllable words, two-syllable words, phrases, and sentences. A variety of training words was used during the practice portion of the sessions, and the training words potentially varied from session to session. To elicit correct production of each vowel, the clinician used a vowel chart and illustrations of tongue placement for vowels. The vowel chart used during the training sessions was based on the traditional vowel quadrilateral, which contains International Phonetic Association symbols of each of the vowels. The position of the vowel symbol on the quadrilateral represents the general features of the vowel (e.g., high/low, front/back, tense/lax). The vowel chart was modified during each phase of the intervention so that it contained only the vowels that were part of the participant’s phonetic repertoire, the current target vowel (e.g., /i/), and any previously targeted vowels. In other words, vowels missing from the participant’s phonetic repertoire that had not yet been targeted (including the control vowel) were excluded from the vowel chart so the participant was not inadvertently made aware of the “location” of the untargeted vowels and of their relation to other vowels.

Baseline and generalization. We developed a probe list of 10 one-syllable words and 10 two-syllable words for each target vowel. This list was administered as a baseline and generalization measure.

1The vowel / ə/ was not included in the assessment because the / ə/ vs. / ɔ/ distinction was not present in the trainer’s phonemic repertoire.
Equipment

We used the Real-Time Spectrogram (RTSPG) module of the SONA-Speech II software, Model 3650, to generate spectrograms of the vowel productions. The RTSPG ran on a Dell Optiplex 780 laptop computer, running Windows 7. All sessions were recorded using a Canon Vixia HF530 digital video camera and a wireless lavaliere microphone that was attached to the participant’s shirt. Additional audio recordings were made of each session using a Marantz PMD661 digital recorder.

Procedure

**Target selection.** Before training began, the participant read the vowel assessment list three times. The accuracy of production of each vowel was judged by the first and second authors, who reviewed the audio file of the participant’s productions and determined by consensus whether each vowel was accurate or inaccurate. Four vowels that the participant produced inaccurately in 50% or more of instances (namely, /i/, /ɪ/ , /ʌ/, and /ə/) were chosen as targets.

**Baseline.** To establish baseline performance, a probe of vowel production was administered on three different days before training of the first vowel began. The probe list for each target vowel was used for baseline data collection. On each day, the participant read 10 words from each of the vowel probe lists (and the same 10 words within a carrier phrase) that had been randomly selected.

**Training.** The participant took part in individual, 25- to 30-min sessions, 2 to 3 days a week, for a total of 11 sessions. The second author, a graduate student clinician in speech-language pathology, conducted the sessions in a quiet room in a university speech and language clinic.

**Vowel 1 training sessions.** The initial training session differed from the other training sessions in that it included a general introduction to vowel production, the vowel quadrilateral chart, and the spectrogram. The clinician explained that speakers produce different vowels by changing the size and shape of the oral cavity and that these changes involve tongue/jaw height (high, mid, low), tongue advancement (front, central, back), the degree of tension of the articulators (tense vs. lax), and the position of the lips (rounded vs. retracted).

The participant was provided with the American English vowel quadrilateral chart, modified as indicated earlier to contain only vowels in his repertoire and /i/, the first target vowel. The clinician used the quadrilateral chart to explain the characteristics of the target vowel and the vowel that the participant substituted for the target (i.e., the error vowel), in this case /ɪ/, and to contrast the target vowel with the error vowel. Other illustrations (e.g., mouth pictures, the vowel quadrilateral superimposed on a drawing of the oral cavity) were also used to show articulatory posture.

In addition, during the initial session, the clinician provided a brief description of spectrograms, explaining that they are a visual representation of speech and that the dark bands in the display (i.e., formants) represent vocal tract resonance. When the size and shape of the oral cavity change to produce different vowels, those changes are reflected in the relation between the formants in the spectrogram. The clinician created spectrograms of the first target vowel and the error vowel (i.e., /i/ and /ɪ/) and showed the participant the difference in the formant patterns between the two displays.

Following the introduction, the participant practiced producing the first vowel contrast (i.e., target vowel vs. error vowel) using the spectrogram. The clinician created a spectrogram of the phonemes /i/ and /ɪ/ in one window of the RTSPG and instructed the participant to produce the same phoneme contrast in the bottom window and attempt to match the formant pattern.

Across sessions, the vowel was elicited in a hierarchy of difficulty: isolation, one-syllable words, two-syllable words, phrases, and sentences. The target vowel was also contrasted with the error vowel in meaningful minimal pairs (e.g., *sit* vs. *seat*). The clinician provided verbal models, feedback about the accuracy of the participant’s vowel productions, and cues about articulatory placement to help him modify his incorrect productions. In addition, the participant evaluated his own accuracy by listening to his recorded vowel productions and analyzing the spectrograms. The participant produced a minimum of 50 correct productions of the target vowel during each 25-min session and a minimum of 25 spectrograms of target vowel productions.

After the training segment, the participant read 10 words that had been selected randomly from each of the vowel probe lists (for a total of 40 words), and the same 10 words in the sentence, “He read XX out loud.” These probe words were not used during production practice in the training segment of the sessions and were thus untrained. No clinician cues or visual feedback were provided during administration of the probe. Training of Vowel 1 continued until the participant achieved 80% accuracy of production of Vowel 1 on the probe list.

**Vowel 2 and 3 training sessions.** Each session began with 5 to 8 min of production practice on the previously trained vowel(s). Clinician feedback and
visual feedback from spectrographic displays were included in the review practice. When a new target vowel was introduced, the clinician showed the participant its placement on the vowel quadrilateral chart and provided a brief orientation to its characteristics (in terms of articulatory placement) and its unique appearance on the spectrogram. Following this introduction, the sessions proceeded exactly as described earlier for Vowel 1.

**Maintenance.** Maintenance was evaluated 2 weeks post training on two separate occasions. As in the baseline and training sessions, the participant read 10 randomly selected words (and the same 10 words within a carrier phrase) from each of the vowel probe lists. Neither visual displays nor verbal cues and feedback were used during these probes.

**Fidelity.** The clinician adhered to the training protocol outlined earlier. The first author, a licensed and certified SLP, monitored the training sessions. Both the supervising SLP and the clinician (based on video playback) assessed the fidelity of each session using a fidelity checklist and reached 100% consensus.

**Analysis**

In each phase of the study (i.e., assessment, baseline, training, and maintenance), the accuracy of vowel production (i.e., correct or incorrect) was assessed as described earlier. Graphs were constructed to show the percentage correct scores for each vowel target in each phase of the study (baseline, training, maintenance). Separate graphs were created for vowels used in isolated words and vowels used in words in the context of a sentence. The data were evaluated using visual inspection criteria outlined by Kazdin (2011).

In order to determine whether training was effective, we examined the magnitude of change across the phases of the study and differences in the rate of change during each phase. First, to examine magnitude of change, we looked for two indicators: (a) a change in the average percentage of correct productions of the vowel from baseline to training, and (b) a shift or discontinuity in performance from the end of baseline to the beginning of treatment. Next, in order to determine whether there was a change in the rate of improvement after the initiation of training, we looked for two additional indicators: (a) a change in trend or slope from baseline to training, and (b) limited latency between the onset of training and the change in performance. We used the same criteria to assess maintenance. In general, if the training was effective, we would expect to see significant differences between performance during baseline and performance during maintenance; in contrast, if the training was not effective, we would expect to see limited differences between performance during training and performance during maintenance.

Finally, we examined the overall pattern of the results by calculating the percentage of nonoverlapping data across phases. The extent to which data points within each phase overlapped helps us determine the overall size of the training effect (Kazdin, 2011).

**RESULTS**

Accuracy data for each vowel in baseline and training conditions were based on the participant’s probe word production. Ten probe words for each target vowel were read during the baseline sessions, at the end of each training session, and during the maintenance sessions.

**Accuracy of Vowel Production**

Figure 1 presents percentage accuracy scores for each of the target vowels at the isolated word level across the three phases of the study: baseline, training, and maintenance. We first examined the magnitude of change. For Vowel 1, /i/, the participant’s performance changed from an average of 7% accuracy during baseline to 75% accuracy during training, to 80% during maintenance. Similar patterns were observed across the remaining two target vowels. Accuracy of production of Vowel 2, /ɔ/, changed from an average of 4% during baseline to 81% during training, to 90% during maintenance, and Vowel 3, /u/, was produced with an average of 15% accuracy during baseline, 60% during training, and 50% during maintenance. In contrast, the control vowel, /ʌ/, was produced with an average of 4% accuracy during baseline and only 10% accuracy during maintenance.

Next, we looked for discontinuity in performance from the end of baseline to the beginning of treatment. A dramatic improvement in performance occurred upon initiation of training of each vowel. The accuracy of production of Vowel 1 increased from 10% at the last baseline session to 50% at the first training session, production of Vowel 2 improved from 10% accuracy to 60%, and production of Vowel 3 from 30% to 50%. The control vowel remained relatively stable over the course of the study, never exceeding 20% accuracy.

Next, we examined potential changes in the participant’s rate of improvement in vowel production. As can be seen in Figure 1, the trend line characterizing the participant’s performance during baseline is relatively flat for all of the vowels. During the training phase, there was an upward trend in performance.
for all of the treated vowels. In addition, there was little to no latency between the onset of training and the increase in performance, suggesting that training had a fairly immediate effect. All of these observations taken together suggest that the training of each individual vowel was effective.

Finally, as a method of evaluating the overall effectiveness of the training, we calculated the percentage of nonoverlapping data across the three phases. Between baseline and training, 0% of the data points overlapped, and between training and maintenance, 100% of the data points overlapped. Single-subject experiments in which fewer than 10% of the data points in baseline and training overlap are considered to demonstrate training that is highly effective (Scruggs, Mastropieri, & Casto, 1987).

In summary, the study results suggest that vowel training incorporating visual feedback was effective in improving vowel production at the word level in a nonnative speaker of English. The changes that occurred were both sizeable and rapid.

Figure 2 presents percentage accuracy scores for each of the target vowels produced in probe words at the sentence level across the three phases of the study. Results were consistent with those at the word level. The accuracy of production of each target vowel for all of the treated vowels.
vowel improved quickly and significantly upon initiation of treatment, whereas the accuracy of the control vowel remained relatively constant.

Social Validity

To supplement the objective outcome data presented in this study, we also gathered data pertaining to the social validity of the outcomes. Foster and Mash (1999) described three components of a treatment study that require social validation, including the goals of the treatment, the treatment procedures, and the treatment outcomes. In this study, social validity data were gathered during session 10 and also the final session. When asked if the participant thought that participating in the treatment helped him, he responded that he definitely thought it helped him. The participant also provided an example of a recent trip to a supermarket where he was able to self-correct a word that was misunderstood due to a mispronunciation of a targeted vowel sound. When requested to repeat the word, the participant was pleased to report that he was able to produce the word with the correct production of the vowel. He indicated that he found the use of the spectrographic analysis software particularly helpful.

The final piece of information that suggests that the treatment was socially valid occurred after the treatment had ended. The participant was interested in learning the characteristics of the control vowel. He asked to see it on a spectrogram and to view its phonetic transcription, and he asked about the difference between /a/ and /ʌ/. Taken together, these data suggest that the treatment resulted in important and meaningful change to the participant and support the objective outcomes presented in this study.

Discussion

Vowel Production Accuracy

The question addressed in this study was whether a multimodal vowel-targeted intervention combining traditional articulation treatment techniques and visual feedback would be effective in increasing the accuracy of a nonnative speaker’s production of English vowels. The findings of this study suggest that the intervention was highly effective and socially valid. There was a sizeable improvement in the participant’s accuracy of production of each vowel immediately following the onset of training, vowel accuracy continued to improve as training sessions continued, and vowel accuracy was maintained over the short term. The first two vowels targeted (/i/ and /ɔ/) were produced with a higher level of accuracy during the maintenance phase than the last vowel targeted (u), potentially suggesting that more practice using the visual feedback and traditional articulation techniques with the target vowel resulted in greater accuracy in production. Because the onset of training of each vowel was staggered, and because training of each vowel continued through the end of the study, Vowels 1 and 2 were targeted significantly longer than Vowel 3 was. Vowel 1 was targeted over the course of 11 sessions, and Vowel 2 over seven sessions, but Vowel 3 was targeted over only three sessions.

This study, together with previous research, suggests that although the age at which an L2 is acquired is a strong indicator of the degree of foreign accent that an L2 speaker will have (Abu-Rabia & Iliyan, 2011; Flege, Frieda, & Nozawa, 1997; Flege et al., 1995; Piske et al., 2001; Tahta et al., 1981), changing one’s pronunciation as an adult is not impossible. In fact, an L2 speaker of American English can learn the American English vowel system fairly quickly.

Limitations and Future Directions

This single-subject study provides useful information about the effectiveness of multimodal vowel-targeted treatment with a 24-year-old male speaker of Farsi. The information provided by the study will potentially guide larger scale studies. There are limitations, however, to the single-subject design that was used in this study. Previous studies (e.g., Flege, Bohn, & Jang, 1997) have suggested that the vowel production of nonnative speakers of English is influenced not only by the age of acquisition of the L2, but also by the nature of the L1. Therefore, it is not possible to generalize the results of this study to other clients or to native speakers of languages other than Farsi. Future research should use a larger sample size with a control group and should investigate vowel training in speakers of a variety of L1s.

In addition, although we were able to conclude that this training had an effect on the participant’s vowel productions, we did not attempt to analyze how each specific aspect of the training (i.e., traditional articulation training techniques and visual display feedback) might have contributed uniquely to the increase in the accuracy of the participant’s vowel productions. It would be beneficial in future studies to determine whether adding visual feedback to traditional articulation training has benefits beyond those of traditional articulation training alone.

Another limitation of the study involved the lack of assessment of long-term maintenance and generalization to spontaneous connected speech. The duration of the study did not allow for data regarding
long-term maintenance of accurate vowel production. Similarly, connected spontaneous speech data would provide useful information about the overall impact of the training on a client’s ability to communicate effectively in a naturalistic social environment. Future research would benefit from the inclusion of post-training follow-ups over a longer period of time.

Clinical Implications

The results of this study, when considered together with previous research (e.g., Ertmer et al., 1996), suggest that it may be beneficial to incorporate visual feedback mechanisms such as spectrograms into a larger instructional program when targeting vowel production in a non-native speaker. The spectrograms provide objective visual feedback that allowed the participant to self-evaluate his vowel productions in a way that he could not otherwise do with traditional articulation treatment alone. People learn using a variety of modalities; some modalities, such as visual feedback, may be more effective and/or may increase the efficiency of intervention and thus, may be a beneficial component in training programs. It should also be noted that the intervention procedures in this study can be easily implemented by SLPs. A variety of spectrographic analysis software (e.g., Praat [Boersma & Weenink, 2015], WaveSurfer [Beskow & Sjolander, 2013], Audacity [2015]) is widely available free of charge for download on standard computers.

With a large number of universities, corporations, and private practitioners offering accent modification programs, it is important for speech-language pathology, as a field, to begin to discover the types of training protocols that might be most beneficial to clients. Effective and efficient training methods to change accent in relatively short amounts of time must be used. Considering the lack of research available to guide SLPs in the area of accent modification services, this study significantly contributes to the knowledge base in this area. The results of the study indicate that when training vowels, visual feedback along with traditional articulation methods can be both effective and efficient. Larger scale studies are needed to further investigate this methodology.

ACKNOWLEDGMENTS

This research represents the second author’s master’s thesis in the Department of Special Education and Communication Disorders at Southern Illinois University Edwardsville.

REFERENCES


Contact author: Kathryn Brady, Speech Pathology and Audiology Program, SIUE, Campus Box 1147, Edwardsville, IL 62026-1147. E-mail: kabrady@siue.edu