Speech, Language and Hearing

ISSN: 2050-571X (Print) 2050-5728 (Online) Journal homepage: http://www.tandfonline.com/loi/yslh20

Can intensive phonomotor therapy modify accent? A phase I study

Megan Oelke, Lauril Sachet, Kathy Nagle, Lauren Bislick, C. Elizabeth Brookshire & Diane L. Kendall

To cite this article: Megan Oelke, Lauril Sachet, Kathy Nagle, Lauren Bislick, C. Elizabeth Brookshire & Diane L. Kendall (2015) Can intensive phonomotor therapy modify accent? A phase I study, Speech, Language and Hearing, 18:4, 229-242, DOI: 10.1179/2050572815Y.0000000009

To link to this article: http://dx.doi.org/10.1179/2050572815Y.0000000009

Published online: 22 Jun 2015.

Submit your article to this journal

Article views: 64

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=yslh20

Download by: [University of Washington Libraries] **Date:** 12 October 2016, At: 13:57

Original Research Paper

Can intensive phonomotor therapy modify accent? A phase I study

Megan Oelke[1], Lauril Sachet[2,3], Kathy Nagle[3,4], Lauren Bislick[3], C. Elizabeth Brookshire[3], Diane L. Kendall[1]

[1]Puget Sound Department of Veterans Affairs Medical Center, Research and Development, University of Washington, Speech and Hearing Sciences, Seattle, WA, USA, [2]New Mexico Department of Veterans Affairs Medical Center, Albuquerque, NM, USA, [3]Speech and Hearing Sciences, University of Washington, Seattle, WA, USA, [4]Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA

A variety of accent modification approaches exist for non-native English speakers. The training program developed for this study, *phonomotor treatment*, is based on a parallel distributed processing model of phonology and was hypothesized to modify accent via improved sound production in a non-native English speaker. The participant was a 20-year-old native Wuhan and Mandarin Chinese speaker. In the context of a single-subject repeated probe design, he received probes prior to, during, and immediately following training. Training was intensive and delivered for 20 hours over 2 weeks. Post-training results revealed statistically significant improvements in trained phoneme production in isolation and reduced listener effort at the sentence level. Generalization results were limited, suggesting that future participants may benefit from an increased training period and additional program development. Data gathered from this study may also help to inform the development of subsequent trials of the same program.

Keywords: Accent modification, Phonomotor treatment, Adults, Phonology, Proof of concept

Introduction

A foreign accent is the result of a linguistic difference between a non-native speaker's native language (NL) and that of the second or target language (TL) (Chakraborty *et al.*, 2011). These discrepancies are often categorized as either segmental (i.e. relating to production of consonants and vowels) or suprasegmental (i.e. prosodic; Flege, 1995). Accents often reflect a speaker's cultural and linguistic background and can be a source of pride and identity. Importantly, an accent is considered a linguistic difference and not a disorder (ASHA, 2012). Foreign accents, however, can contribute to diminished comprehension of a speaker's message, increased effort on the part of the listener, and associations with negative stereotyping (Gluszek and Dovidio, 2010). As a result, many non-native speakers seek out accent modification training to improve communication for both personal and professional reasons.

Much of the accent modification literature has focused on differences between first and subsequent language acquisition (Eckman, 2008; Flege, 1995; Lado, 1957; Liberman *et al.*, 1967; Strange, 1995). Flege's Speech Learning Model asserts that we perceive the TL through the 'grid' of our NL, such that unfamiliar TL phonemes will be assimilated to fit into familiar NL phonetic categories. In this view, accent modification training consists of redefining pre-existing phonetic categories and/or establishing new ones. Alternatively, the Markedness Differential Hypothesis (Eckman, 2008) assumes the degree of difficulty a non-native speaker will have in acquiring a new sound depends on its level of 'markedness'. The Markedness Differential Hypothesis is a weighted system that takes into account universal frequency (i.e. commonly shared sounds across different languages), NL phonological effects (e.g. word final stop devoicing by German speakers learning English), and articulatory complexity, thus allowing for the varying degrees of difficulty observed in TL phoneme acquisition. In the development of the Speech Learning Model and the Markedness Differential Hypothesis, much has been learned in terms of predicting TL error types; however, there has been little discussion regarding the application of such analyses to programs designed to modify accent.

In order to apply the knowledge gained from investigation of TL error types toward accent modification, the scope of accent research has grown beyond

---

Correspondence to: Megan Oelke, University of Washington, 1417 N.E. 42nd Street, Box 354875, Seattle, WA 98105-6246, USA. Email: moelke@uw.edu

© W. S. Maney & Son Ltd 2015
DOI 10.1179/2050572815Y.0000000009

linguistics to also include applied phonetics, speech–language pathology (SLP), and teaching English to speakers of other languages. Accent training has simultaneously undergone vast pedagogical shifts from instruction focusing on native-like production of English (Burgess and Spencer, 2000) to a more functional approach, with goals like improving intelligibility (i.e. the quantity of message understood by the listener) and comprehensibility (i.e. the amount of effort required by the listener) (Sikorski, 2005). The shift toward a more functional approach has focused on selecting specific sounds that may most influence intelligibility and comprehensibility (Brown, 1988; Munro and Derwing, 2006), and has dominated a small but growing body of works aimed at approaching accent modification from a functional–instructional perspective.

Functional–instructional accent modification approaches have focused on the relationship between segmental and suprasegmental instruction (Champagne-Muzar *et al.*, 1993; Derwing *et al.*, 1998), direct explanation of sound features, especially those that most impact intelligibility in the TL (Saito and Lyster, 2012), and the role of visual, auditory, and corrective feedback (de Bot, 1983; Saito and Lyster, 2012) on accent modification training outcomes. The results of these investigations demonstrate that pronunciation of the TL can indeed be modified with multimodal input, explicit instruction in sound production or intonation, or corrective feedback.

The training program used in the present study incorporates many of the approaches listed above. Known as phonomotor treatment, the training program adapts elements from the Lindamood Phoneme Sequencing program (Lindamood and Lindamood, 1998) and was originally developed for adults with acquired phonological impairments due to aphasia (i.e. a language impairment resulting from left hemisphere stroke) (Kendall *et al.*, 2008). Phonomotor treatment uses a multimodal instructional approach, emphasizing intensity, repetition, and use of reactive feedback for correct and incorrect responses. The program has been described extensively in the aphasia literature; for more detailed information see Brookshire *et al.* (2014), Kendall *et al.* (2013), and Kendall *et al.* (2008).

The goal of phonomotor training is to focus on the underlying phonological system to train all English sounds and then combine those sounds in sequences to improve phonological processing and awareness. The idea is that TL phonemes and phoneme sequences may either need to be established (in the case where phonemes are not shared between TL and NL) or modified (in the case of shared TL and NL phonemes). A framework that may be useful in guiding such an approach is a parallel distributed processing (PDP) model of phonology (Nadeau, 2001). This model, though computationally untested, is neurally plausible and based on the Wernicke–Lichtheim (Lichtheim, 1885) information-processing model of language. Rather than serial processing of isolated units, various linguistic domains (i.e. acoustic, orthographic, articulatory-motor, and conceptual/semantic) activate simultaneously during the processing of phonological representations. A discrete phoneme representation is multisensory, contributing to both perception and production. As such, individual phonemes do not exist as static, localized units, but are rather a dynamic product of simultaneous, multimodal activation.

Three distinct aspects of the PDP model could be useful in both supporting and enhancing our understanding of accent modification. Firstly, instantiation of representations (e.g. semantics, lexical/word, or phonology/sound) depends on the strength of neural connections. The probability that a representation will receive sufficient activation depends on the frequency of access and amount/diversity of neural input received (i.e. activation from multiple cognitive domains). Kendall *et al.* (2008) stated that during language learning, a gradual adjustment is made to the strengths of the connections between the units, such that a pattern of activity involving the units in one domain elicits the correct pattern of activity in the units of another domain. For example, input into the acoustic domain (e.g. auditory input /b/) should simultaneously engage the orthographic representation (e.g. letter ‘b’). Within this framework a ‘foreign’ accent occurs when pre-existing NL phonological representations are activated in the context of the TL, resulting in perceived sound distortions, substitutions, omissions, and additions in the TL.

Secondly, the PDP model bolsters the Speech Learning Model and the Markedness Differential Hypothesis (Eckman, 2008; Flege, 1995). Flege’s Speech Learning Model is accounted for in that sounds shared between the NL and TL will have the strongest neural connections. On the other hand, TL phonemes with shared (but not exact) NL input will be difficult to modify, because the likelihood of activating the NL representation is greater. In addition, novel sounds will present a challenge, as they require new multisensory input to develop the discrete patterns of activation. The PDP model also accounts for Eckman’s Markedness Differential Hypothesis in that a sound may be deemed more ‘marked’ because the pattern of activation is stronger and its representation more widely distributed in the NL rather than in the TL, either due to frequency of access or wealth of input.

Previous theories (Eckman, 2008; Flege, 1995; Lado, 1957; Liberman *et al.*, 1967) viewed phoneme perception as an auditory phenomenon responsible for driving production; only a few studies in accent modification research have shown that perception is not exclusively responsible for driving production (Couper, 2006; Sheldon and Strange, 1982). This has important implications for accent modification instruction, such that mastery of perception of TL sounds does not preclude improvements in TL production. However, the multimodal framework of PDP is a crucial addition to these explanations of *how* to elicit change. By broadening the definition of perception beyond the auditory modality to also include visual, orthographic, and tactile–kinesthetic information, PDP accounts for changes in production in the absence of changes in auditory perception, and identifies additional avenues of instruction.

In this proof-of-concept study we used a single-subject repeated probe design to ask whether a multimodal training program would alter phonological representations in a non-native English speaker. Our research questions are concerned with the acquisition, generalization, and maintenance of phonological representations resulting from therapy:

RQ #1: Is there a significant difference in repetition accuracy of trained sounds following training?

RQ #2: Do effects of treatment generalize to the accurate production of trained sounds in untrained contexts (e.g. *real words*, *nonwords*, *sentences*)?

RQ #3: Do effects of treatment generalize to improved intelligibility and perceived listener effort during sentence and discourse production?

Materials and methods

Participant selection

Eight non-native speakers of English were screened to determine eligibility for the study and one participant met inclusion criteria and was enrolled. Inclusionary criteria included non-native English accent and baseline intelligibility in spontaneous conversation less than 80% as judged via consensus agreement by a certified SLP and SLP master's student. No standardized assessments were used to determine intelligibility for screening purposes. Exclusionary criteria included impairments in cognition and inability to adhere to treatment schedule.

Participant

The participant was a 20-year-old, right-hand dominant male from Mainland China. He spoke both Mandarin and a local dialect of Wuhan. He began learning English in grade school at the age of 12, from a native Chinese speaker, and most of his instruction was in reading and writing (i.e. pronunciation was not emphasized). He came to the United States to study at the university level in 2011 and reported spending approximately 50% of his day speaking English. At the outset of this study, he expressed a desire to improve his pronunciation for general communication purposes and specifically wanted to gain more awareness regarding how to produce sounds accurately.

The participant's accent was characterized by frequent errors at the sound and suprasegmental levels, which compromised his intelligibility and increased the effort required by listeners to understand him, especially in conversation when the context was not known. The participant's connected speech contained substitutions (/w/ for /v/, word final voiceless for voiced consonants), omissions (no plural marker /s/ on certain nouns), additions (schwa insertion between consonant clusters, i.e. 'glad' produced as /gə-lad/), and distortions (distorted lax vowels and diphthongs, syllable/word final /l/ produced as /oʊ/, 'the' produced as 'ler'). Primary suprasegmental differences included slow rate and inaccurate stress placement, mostly at the word level.

Characterization of participant

In order to ascertain information pertaining to cognition, underlying phonological processing, hearing, and self-perception of accent, the following measures were administered prior to the start of treatment: Raven's Progressive Matrices (Raven, 1978), The Standardized Assessment of Phonology in Aphasia (SAPA; Kendall *et al.*, 2010), a hearing screening, and a modified version of The Communication Participation Item Bank short form (CPIB-short form; Baylor *et al.*, 2013). Results from these pre-training measures can be found in Table 1.

The participant's hearing and cognitive functioning were within normal limits, based on results of the hearing screening, and the Raven's and behavioral observation, respectively. Results from the SAPA showed that the participant's pre-training phonological processing abilities of the TL differed from that of a native English speaker across all modalities assessed (speaking, reading, and listening), with particular difficulty with real word and nonword

Table 1 Participant characteristics and pre-training standardized assessments

Age	Education	Native Language	Raven's*	SAPA[†]	CPIB-short form[‡]	Hearing screen
20	14	Mandarin Chinese, Wuhan	36/36	91/151	20/30	Pass

*Coloured Progressive Matrices (Raven, 1978).
[†]Standardized Assessment of Phonology in Aphasia (Kendall *et al.*, 2010).
[‡]Communication Participation Item Bank (Baylor *et al.*, 2013).

reading and auditory rhyme judgments, as well as nonword repetition, parsing, and blending. It should be noted that the SAPA was administered according to standard procedures, which count distortions as correct. The participant reported via the CIPB-short form that he had an easier time communicating with familiar people, and that his accent interfered especially when trying to get his turn in fast-moving conversations. In his screening interview, intelligibility in conversation was judged by two separate raters (certified SLP and SLP master's student) to be approximately 60% when the context was known, and about 40% when the context was unknown.

Study design

A single-subject repeated probe design was used and the participant received multiple probes prior to, during, and immediately following completion of the training program. The SAPA, CPIB-short form, and measures of both intelligibility and perceived listener effort (sentence and discourse levels) were administered pre- and post-training over a span of 3 days each. Repeated probes included repetition of trained sounds in isolation, trained sounds in real words, trained sounds in nonwords, and a control measure. These were administered five times at baseline over the 3-day testing period. Training was delivered for 20 hours over a 2-week span (two 1-hour sessions per day, 5 days per week). During training, repeated probes were given following every 2 hours of instruction (i.e. nine times), always at the beginning of a training session. Post-training repeated probes as listed above were administered four times throughout the 3-day post-training period. See Fig. 1 for an outline of probe and treatment administration. All computer-based measures (repeated probes and SAPA) were administered via a Dell Optiplex 9010 computer with Alesis M1 Active 520 speakers. The participant was seated approximately 36 inches from a 22 inch Dell monitor. All participant responses were digitally recorded using a Marantz Professional Solid State Recorder PMD671 and AKG C535 EB Austria microphone, which was placed approximately 8 inches from the participant's mouth.

Outcome measures

Methodology regarding the creation and scoring of all primary and secondary outcome measures are described in detail below.

Primary outcome measure

Acquisition of trained sounds (RQ 1)

The primary outcome measure of this study was accuracy of repetition of trained sounds in isolation. Stimuli included 37 phonemes that were elicited by a pre-recorded video of a female speaker on a computer (Dell Optiplex 9010) using VLC software (VLC is a portable, free and open-source, cross-platform media player that is downloadable free on the Internet; http://www.videolan.org/vlc/index.html). Prior to data collection, video recordings from the model had been judged for accuracy by three other members of the research team; these were re-recorded and re-judged until 100% accuracy was determined by consensus agreement. All 37 phonemes were presented in a random order at each probe session (pre/post and repeated). Scoring of participant verbal responses was completed during the probe session by the clinician administering the probes. Distortions, omissions, substitutions, and additions (i.e. schwa insertion → /gə/ for /g/) were considered inaccurate productions. Inter- and intra-rater reliability was performed on 20% of the entire corpus and computed using single-measures intraclass correlation coefficients (ICCs; IBM, SPSS software).

Secondary outcome measures

Secondary outcome measures focused on generalization to trained sounds in real words, nonwords, and sentences, as well as intelligibility and perceived listener effort. All measures were administered and scored by a certified SLP and are described in detail below.

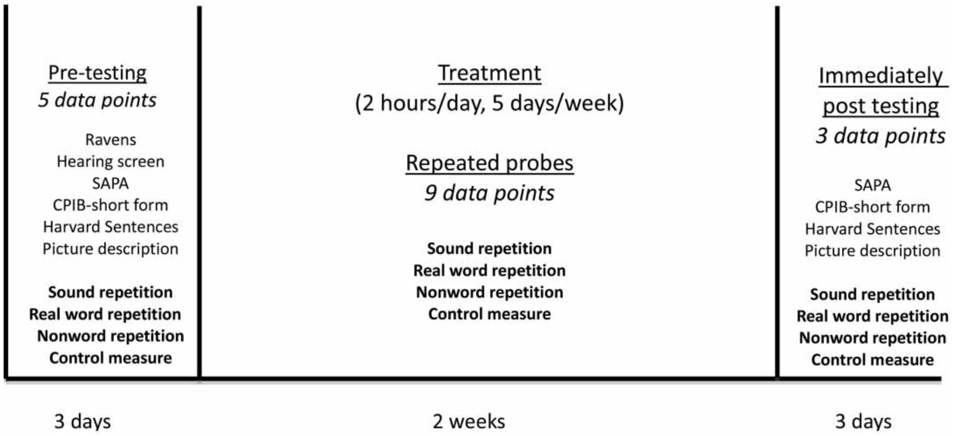

Figure 1 Treatment design.

Generalization of trained sounds to untrained real words (RQ 2.1)

The Assessment of Intelligibility of Dysarthric Speech (Yorkston and Beukelman, 1981) was used in order to determine accuracy of trained sounds embedded in real words. Stimuli consisted of novel lists of 50 one- and two-syllable words presented once during each probe session. The participant was asked to repeat the words aloud following a model from the testing SLP. Stimuli were also presented in written form to minimize errors due to unfamiliarity or memory. Participant responses were digitally recorded (Marantz Professional Solid State Recorder PMD671) and scored offline by a member of the research team. The number of possible correct phonemes was determined by counting phonemes per word and tabulating a grand total for each set of 50 words. Rhotic vowels (e.g. /ɚ/) were considered a single phoneme. Each word was scored individually and a point was subtracted for any phoneme in error (sound substitutions, omissions, additions, and distortions). A percent phoneme accuracy score was then derived for each data point.

Generalization of trained sounds to untrained nonwords (RQ 2.2)

In order to determine if effects of treatment generalized to a linguistic context not exposed in treatment, repetition of trained sounds in nonwords was assessed. The nonword stimuli list included 145 nonwords administered in two parts (i.e. a half-list consisting of 72 or 73 nonwords). Each half-list was administered every other probe session (i.e. List 1, List 2, List 1, etc.). Nonwords were phonotactically legal in English and were comprised of phonological sequences of low phonotactic probability (PP) and high neighborhood density (ND). Research by Storkel *et al.* (2006) suggests that phonological and lexical processing influence different aspects of word learning, and low PP is thought to assist with new learning while high ND is thought to assist in assimilating new lexical representations with existing representations. The combination of positional segment frequency (i.e. how often a phoneme occurs in a word position) and sum biphone frequency (i.e. probability of phoneme segments occurring together in a word) determine PP, whilst the numbers of words in a dictionary that differ from the target word by a single phoneme addition, deletion, or substitution determine ND (Brookshire *et al.*, 2014). Nonword stimuli were calculated via The Irvine Phonotactic Online Dictionary calculator Version 2.0 (IPHOD; Vaden *et al.*, 2009) by changing a single phoneme in a real word with low PP and high ND, and IPHOD was used to re-calculate PP and ND values for the nonword. These values were classified as high or low based on a median split (Storkel *et al.*, 2006).

The nonword stimuli were elicited by pre-recorded video of a male speaker presented on a computer (Dell Optiplex 9010) using VLC software. During probe sessions, the participant was instructed to watch the pre-recorded model and repeat each nonword. Participant verbal responses were scored during the probe session by the SLP administering the probes and were also digitally recorded (Marantz Professional Solid State Recorder PMD671) for reliability purposes. Accuracy was determined at a whole word level (i.e. accurate production of all phonemes was necessary for a word to count as correct). Sound substitutions, omissions, additions, and distortions were counted as errors. Because each half-list was phonetically balanced across the entire set (145 nonwords), data from each set of two half-lists were collapsed to create one data point. A final score of percent whole words correct out of total words was then calculated for each data point.

Generalization of trained sounds to untrained sentences (RQ 2.3)

To see if trained sounds generalized to the sentence level, untrained sentence stimuli were elicited from the Harvard Sentences (HS; Rothauser *et al.*, 1969), which includes 72 lists of 10 sentences each. Each list is phonetically balanced to match frequency of occurrence in English. Two different lists were presented at each pre- and post-training period for a total of four lists ($n = 40$ sentences). Since these sentences were to be scored for intelligibility, the objective was to sound as natural as possible. Thus, the participant was instructed to repeat the sentences following a model from the SLP administering the probes; stimuli were also presented in written form to minimize errors due to unfamiliarity or memory. All responses were digitally recorded (Marantz Professional Solid State Recorder PMD671) for subsequent analysis.

Scoring was performed offline by a member of the research team who was blinded to testing period (pre- versus post-training). Scorers listened to the participant recordings via headphones (Quiet Comfort 15 Acoustic Noise Cancelling Headphones) on a computer (Dell Optiplex 9010). Because content of the HS is considered decontextualized, the listener (scorer) must rely on word comprehension rather than logical deduction for understanding. Each sentence was transcribed broadly using the International Phonetic Alphabet (IPA), and then number of phonemes possible was tallied for each sentence. For each list of 10 sentences, a percentage of number of phonemes correct out of total number of phonemes possible was calculated. One point was subtracted for each inaccurate phoneme per sentence; again, inaccuracies included sound substitutions, omissions, additions, and distortions. The two list scores for each testing time point were then combined (because the original list had been split) to create one data point. Inter-rater reliability was performed on 20% of the stimuli (i.e. eight sentences) and calculated using ICCs (IBM, SPSS software).

Generalization of training to sentence and discourse level: speech intelligibility (RQ 3.1)

To assess whether any effects of treatment generalized to improvements in intelligibility, unfamiliar listeners were asked to transcribe and score sentence (HS) and discourse (picture description) level data. The HS, as described above, were used to assess sentence-level intelligibility. A picture description task (i.e. the picnic scene from the Western Aphasia Battery) (Kertesz, 1982) was used to elicit a three-minute discourse sample. Discourse-level intelligibility was assessed once at each testing period (pre- and post-).

For data analysis, digital recordings (Marantz Professional Solid State Recorder PMD671) from the participant speaking the HS and discourse were presented auditorily via headphones (Samson Stereo Headphones, RH600) to three unfamiliar listeners (i.e. undergraduate students from the University of Washington's Department of Speech and Hearing Sciences), who were blind to testing period. The listeners were asked to transcribe the samples using English orthography, replaying the samples as many times as needed. These listeners' transcriptions were then compared to those of a familiar listener, who followed the same transcription protocol as described above. For sentence-level scoring, lexical accuracy was measured in percent keywords correctly understood by the unfamiliar listeners. Keywords were defined as adjectives, nouns, adverbs, verbs, and negation words (i.e. *not*). Pre- and post-training averages were then derived by calculating percent keywords correctly transcribed. For discourse-level scoring, percent keywords correctly understood was calculated based on total possible keywords identified by the familiar listener.

Generalization of training to sentence and discourse level: perceived listener effort (RQ 3.2)

The HS and picture description samples described above were also used for determining perceived listener effort. Immediately after transcribing each sentence or passage, listeners rated the amount of effort required to understand the sample using a nine-point, equally appearing interval scale (1 = no effort, 9 = extreme effort) presented on a Dell Optiplex 9010 desktop computer (Munro and Derwing, 2006).

Rater reliability for sentence and discourse stimuli (intelligibility and perceived listener effort)

Inter-rater reliability for sentence stimuli was calculated by comparing the intelligibility score for each sentence from each listener to that of each other listener using ICCs (IBM, SPSS software), and by comparing the score from each listener to the group mean for each sentence using average-measures ICCs (Shrout and Fleiss, 1979). The same method was used for ratings of perceived listener effort. The single-measures ICC for sentence-level intelligibility scores was 0.665, indicating moderate agreement among the listeners; the average-measures ICC was 0.856, indicating very strong agreement with the mean (Portney and Watkins, 2000). The single-measures ICC for perceived listener effort was 0.438, indicating fair agreement among listeners; the average-measures ICC was 0.700, indicating strong agreement with the mean.

For the discourse samples, because of the small number of picture description stimuli, a measure of inter-rater agreement with the mean was more appropriate than ICCs. Inter-rater agreement with mean intelligibility scores was calculated using a criterion of ±0.025, or within 5% of the average for each sample. This value was chosen because the standard deviation of intelligibility scores was approximately 5% for each listener on this task. Mean inter-rater agreement was 50% (SD = 0.25).

Perceived listener effort ratings for each sample were considered to agree with the mean 'exactly' if they fell within one point (or 0.111 of the scale length) of the mean. The value of 0.111 was chosen because it is equivalent to one scale rating or one-ninth of the nine-point scale; a rating within one point would be ±0.55 of the mean. The probability of chance agreement within one point value was 0.1604 (Kreiman *et al.*, 1993). Mean agreement within one point for the three listeners was 37.5% (SD = 0.33); agreement for two listeners exceeded chance.

A second level of agreement, within two points of the mean, was also calculated for perceived listener effort. Two points on the scale represent 22.22% of the scale length, or ±1.11 of the mean. The probability of chance agreement within two scale points was 0.3086. Mean agreement within two points for the three listeners was 75% (SD = 0.25); agreement for all three listeners exceeded chance.

Control measure

The control probe was an auditory grammatical judgment subtest from the Test of Adult and Adolescent Language (Hammill *et al.*, 1994). This control measure was chosen because although syntactic knowledge is a linguistic skill, it was not expected to change as a result of phonological training. The subtest includes 35 items, though only 10 randomly selected items were administered at each testing session. Lists of the 10 items were presented at all pre- and post-training time points, as well as after every 2 hours of training. Items were read aloud by the administering SLP. The participant was instructed to listen to three separate statements per item and select which two of the three sentences meant the same thing (e.g. *(a) It was proven that Shirley was the winner*, *(b) Shirley was the proven winner*, and *(c) It was proven by Shirley who the winner was*). Scoring occurred during the probe session and a percentage of number correct was derived for each time point.

Training procedures (phonomotor program protocol)

As mentioned previously and in detail in Appendix A and the aphasia literature (Brookshire *et al.*, 2014; Kendall *et al.*, 2008, 2013), phonomotor treatment uses a multimodal instructional approach that emphasizes intensity, repetition, and use of reactive feedback for correct and incorrect responses. Only a brief overview will be presented here, including modifications made for this particular study (the first time it has been used for accent modification training).

The goal of phonomotor training is to train all English sounds (consonants (C) and vowels (V)) and then combine those sounds in sequences to improve phonological processing and awareness. Consonants and vowels are first trained in isolation, and are then trained in one- and two-syllable nonword combinations. In the full program (60 total hours of treatment) additional real and nonword stimuli consisting of low PP and high ND are also used. A list of the trained sounds used in this study can be found in Appendix B. All phoneme representations (e.g. tactile–kinesthetic, auditory, verbal, and orthographic) are engaged simultaneously throughout training to promote learning.

Treatment stimuli

In the context of this phase I study, training stimuli included phonemes in isolation and nonword combinations of increasing difficulty (e.g. VC, CV, CVC, CCVC, etc.). Nonword combinations were created during the session by the SLP administering training (e.g. feep, fop, fip, foop, etc.), based on participant performance during the session. Criterion for moving from Stage 1 (sounds in isolation) to Stage 2 (sounds in sequences) was reaching 85% accuracy for both perception and production tasks for each phoneme target across two training sessions.

Results

See Fig. 2 (graphs A–D) for repeated probe accuracy data pre-, during, and post-training. Specifically, Fig. 2 displays (A) sounds in isolation, (B) sounds in real words, (C) sounds in nonwords, and (D) auditory grammaticality judgment, respectively.

Repeated probe data were analyzed using effect sizes (ES) employing Cohen's *d* for interpretation: 2.60–3.89 (small), 3.90–5.79 (medium), and >5.80 (large) (Beeson and Robey, 2006). Calculation of *d* was achieved by taking the mean of the post-training value minus the mean of the pre-training value divided by the standard deviation of the pre-training value [$\mathrm{ES}(\mathrm{Mean}_{\text{post-training}} - \mathrm{Mean}_{\text{baseline}})/\mathrm{SD}_{\text{baseline}}$]. Pre- and post-training measures were calculated using paired Student's *t*-tests (significance $P \leq 0.05$).

RQ 1: Average repetition accuracy of trained sounds prior to training was 88% (SD = 1) and 97% post-training (*ES* = 6.21) (large). Intra-rater reliability was 83% and inter-rater reliability was 60%.

RQ 2.1: Average accuracy of trained phonemes embedded in real words was 95% (SD = 2) pre- and 96% post-training (*ES* = 0.77) (no effect).

RQ 2.2: Average accuracy of trained phonemes embedded in nonwords was 78% (SD = 6) pre- and 89% post-training (*ES* = 2.53) (no effect).

RQ 2.3: Average accuracy of phonemes embedded in sentences was 76% (SD = 0.35) pre- and 82% (SD = 1.45) post-training (*P* = 0.092). Inter-rater reliability was 87.5%.

RQ 3.1 – Intelligibility: Mean intelligibility for pre-training sentence-level probes was 82% (SD = 3), compared to mean post-training intelligibility of 87% (SD = 6). This difference was not statistically different (*P* = 0.123). Mean intelligibility for the pre- versus post-training discourse recordings also did not differ significantly (*P* = 0.678). Mean pre-training intelligibility was 95% (SD = 3) and mean post-training intelligibility was 95% (SD = 1).

RQ 3.2 – Perceived listener effort: There was a statistically significant difference in mean pre- and post-training ratings of perceived listener effort for sentence-level recordings (*P* = 0.002). The mean pre-training perceived listener effort rating was 3.87 (SD = 1.62) on a scale of 1–9; the post-training mean rating was 2.33 (SD = 1.71). For the eight picture descriptions, mean pre- and post-training ratings of perceived listener effort did not differ significantly (*P* = 0.76). Perceived listener effort for discourse showed a pre-training mean rating of 4.17 (SD = 0.64) and post-training 4.00 (SD = 0.61).

Pre- and post-training standardized tests: The SAPA results showed a pre-training raw score of 91/151 (60%) and post-training of 105/151 (70%). The subtests with the most improvement were subtest two (auditory phonologic processing) and subtest three (repetition, parsing, and blending).

Control measure: Pre-training scores on the sentence grammaticality judgment test revealed 46% (SD = 9) and post-training accuracy at 48% (*ES* = 0.17) (no effect).

Progression through training program: Training progressed through all stages presented in the training protocol, though only nonword, monosyllabic CVC

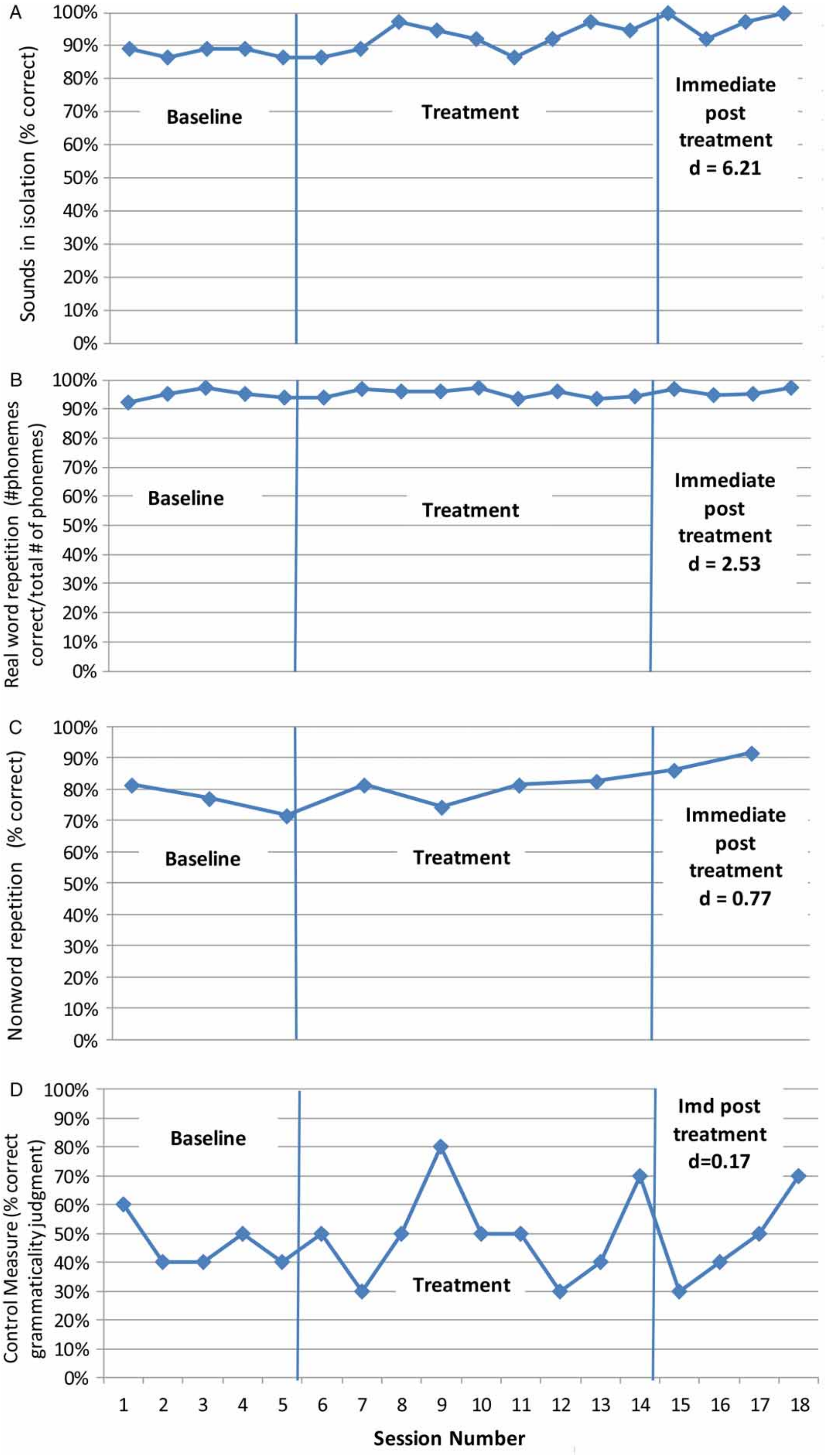

Figure 2 Repeated probe data. Effect sizes (ES): 2.60–3.89 (small), 3.90–5.79 (medium), and >5.80 (large) (Beeson and Robey, 2006).

syllable structures were mastered (greater than 85% accuracy across two sessions) in both perception and production tasks. In the last 4 hours of training, the participant reached 60% accuracy with production of nonword CVCC and CCVC structures, and in the last 2 hours of training he reached 75% accuracy with perception of CCVCC nonwords. It should be noted that the most typical errors made were in production of syllable final /l/ and perception and production of the vowel /ɛ/ in any context.

Discussion

This phase I, proof-of-concept study examined whether a phonologically based training program (i.e. phonomotor) motivated by a neurally distributed model of language function (i.e. PDP) would modify accent in a non-native English speaking individual. Our results, specifically the improvement of trained sounds, suggest that it does. Following training, individual sounds may have received sufficient activation as a result of increased frequency to the TL sounds and the multimodal nature of the neural input received. Additionally, training appears to have generalized more broadly, as mean ratings of perceived listener effort decreased significantly for one of the two measured tasks. In other words, while comprehension of the speaker's message did not change, the amount of effort expended to comprehend that message was deemed less by perceived listeners unfamiliar with the speaker. Mastery of the final stages of the treatment hierarchy and generalization to untrained contexts and behaviors were limited for this speaker within the training period. Further progress may have been restrained by a combination of the influence of NL phonologic knowledge on TL phonologic learning and an insufficient number of training hours. As well, speaker influences (e.g. amount of time speaking English, age of learning) may have prevented more robust gains. These points are discussed below.

Acquisition

Based on the combined theories stemming from accent modification research (e.g. Flege, 1995; Strange, 1995) and the PDP phonology model (Nadeau, 2001), we hypothesized that our multimodal phonomotor approach to training phonemes and phoneme sequence knowledge would be sufficient to improve TL phoneme repetition in a non-native English speaker following 20 hours of intensive instruction. Positive results revealing a large effect size support this hypothesis, indicating our participant improved TL phoneme-level production (repetition) with phonomotor training. These results are consistent with prior accent modification studies (Chakraborty *et al.*, 2011; Couper, 2006; de Bot, 1983; Derwing *et al.*, 1998; Flege, 1995; Saito and Lyster, 2012) and principles of neuroplasticity (Kleim, 2008; Nadeau, 2014). Given sufficient stimulation, feedback, and targeted perception/production opportunities, change occurred in our participant.

Generalization

We believe a few overall factors prevented more robust generalization gains. First and foremost, the participant likely would have benefited from additional training hours. Because phonomotor treatment has only previously been administered to adults with neurological damage due to stroke (Brookshire *et al.*, 2014; Kendall *et al.*, 2008, 2013), we assumed that a neurologically healthy brain would require fewer training hours to instantiate change. We likely underestimated the amount of time needed to produce changes in this participant's TL phonological system.

In addition to limited progression in the treatment hierarchy, we think that the participant's prior and current use of English may have prevented additional generalization gains. Having learned English from a non-native speaker, it may be that he experienced years of establishing and activating TL phoneme representations in the context of NL phonemes. It is possible he received limited accurate input for sounds not found in the NL (such as /ɪ/ and syllable final /l/). Furthermore, given that most of his English education reportedly focused on reading and writing, the diversity of initial input was also limited. The multimodal nature of the phonomotor treatment program, which provides opportunity for input and output in all language modalities, provided our participant with the diversity needed for sufficient phoneme activation, but perhaps not enough time in the program for wider activation of more complex linguistic networks. Or, perhaps gains would have been greater if therapy had only targeted those sounds that were not shared by the NL and TL. Change may also have been limited by the participant's reportedly speaking his native dialect half of the time throughout the duration of the study.

With regard to trained sounds embedded in real words, it may be that no change was detected in part because the participant's baseline performance was too high to allow for a statistically significant level of change. Performance may have been boosted by lexical/semantic knowledge. On the other hand, looking more closely at the errors produced, our participant was least successful during pre- and post-training probes with syllable final /l/ and /ϵ/, the same sounds that did not reach mastery in training.

Regarding trained phonemes in nonwords, which carry no lexical/semantic information and require purely phonetic–motoric skills, the probe items used in this study may have limited the results. Nonword probe items were both mono- and disyllabic; however, the participant mastered production and perception of only CVC nonwords in training. In other words, not practicing multisyllabic nonwords may have reduced his ability to accurately repeat two-syllable nonwords in probes.

As mentioned previously, phoneme accuracy in sentences was lower than accuracy within real single words. This finding is consistent with the increase in errors typically seen with a comparative increase in linguistic level and complexity. Due to the pervasive effects of NL phonotactic constraints on TL pronunciation, it may be more difficult for the speaker to access TL phonological representations in more complex and lengthy utterances.

Although sentence-level intelligibility increased marginally, it was insufficient to denote any significant change. However, perceived listener effort decreased significantly; although unfamiliar listeners' comprehension of the speaker's message did not change, listening to the message was deemed less effortful. This may signify a general shift in phoneme accuracy towards improved approximations of the targets. It is possible that rather than an entire corpus shift in phoneme accuracy, it is the same pervasive sounds identified in isolation that negatively affected sentence-level intelligibility. Further error analysis of the unfamiliar listeners' broad transcription would be required to determine whether this is the case. It may also be that wider changes affecting suprasegmental features of connected speech, unaccounted for in the current study, resulted from targeting isolated phonemes. Investigation into the comparative saliency of these single sounds on listeners' comprehension would be beneficial.

Finally, at the discourse level there were no significant changes seen in either intelligibility or perceived listener effort. The most likely conclusion is that training of sounds in isolation did not generalize to spontaneous speech with the given amount of training. It is also possible that suprasegmental factors such as prosody are more relevant in discourse and have a greater effect on intelligibility and perceived listener effort. A better understanding of the impacts of suprasegmental versus segmental errors on intelligibility and perceived listener effort is needed to explain generalization at this level.

Limitations

The intention of this phase I study was to determine if there were any effects of phonomotor training on the accent of a non-native English speaker. While our study was successful in this regard, there are limitations that warrant discussion.

Firstly, the content of probe stimuli may have been flawed. The real word probe stimuli may not have not been complex enough to capture phonological changes. Lexical–semantic involvement in real word production likely had top-down positive effects on production of one- and two-syllable real words at all testing periods. More complex real words (e.g. greater than three syllables, consonant clusters) may have been needed to challenge phonological processing abilities. Conversely, the nonword probe stimuli may not have captured more subtle changes in phonological processing, given that the participant did not master multisyllabic combinations in therapy. In other words, the nonword probe stimuli may have been too complex for this participant given his limited progression in the treatment program.

Secondly, perceived listener effort scores on the sentence task improved following treatment, but perceptual results should be interpreted with caution. All speech stimuli were produced by a single speaker and the number of raters was quite small ($n = 3$). Because listeners need to focus on their own internal processes to rate listener effort, it is subject to greater inter-rater variability than perceptual dimensions related specifically to the stimulus, such as speech acceptability or naturalness. Raters in this study showed strong adherence to the mean for both sentence- and discourse-level stimuli, with average-measures (ICC = 0.700) and mean agreement considerably above chance; it may not be appropriate to generalize these findings beyond our single case.

Finally, we believe collecting 3-month maintenance data would have been useful for determining training effects long-term. Unfortunately, the student raters who analyzed the sentence and discourse samples for pre- and immediate post-testing were no longer available at 3 months following therapy. The feasibility of re-training a new set of student raters was beyond the scope of this phase I study.

Implications and future directions

As mentioned above, we believe a greater number of treatment hours may be needed to increase patterns of activation at linguistic levels beyond single phoneme representations. Participants may even benefit from a program that is criterion – rather than time – based. It is not uncommon for accent modification training to be time-limited; on the other hand, speech–language therapy has specific criteria for dismissal, for good reason. Like speech and language disorders, speech differences such as foreign accent may not be suited to a one-size-fits-all approach to remediation.

It is possible that extrinsic variables such as the participant's NL and level of English proficiency influence the effectiveness of this program. Further replication is needed (manipulating variables of age of learning, NL, and level of English proficiency) across individuals with differing linguistic backgrounds and degrees of accentedness in order to identify optimal dosage and participant characteristics.

In thinking about future iterations of phonomotor treatment on accent modification training, the current order of phoneme training (see Appendix A) may not be of greatest benefit. There may be a number of phonemes shared by the NL and TL for which training can be reduced or eliminated. It may be more beneficial to begin training with difficult TL sounds, which can be identified prior to training by doing a phonological and phonetic inventory of the NL versus TL, and/or by assessing stimulability in various linguistic contexts (e.g. isolation, real words, nonwords, etc.). Since those sounds trained earlier in the program get more training time, it may be more effective to target those sounds that most affect the speaker's intelligibility.

To capture the effects of training on everyday communication, a more dynamic exit interview with the participant may be appropriate to qualitatively capture how s/he felt the training changed her/his behaviors. Finally, collection of maintenance data (i.e. 3 months post-training termination) would help to determine the long-term effects of training.

Acknowledgments

The authors would like to acknowledge the participant for his time and efforts, and the University of Washington Institutional Review Board for approving this study. They would also like to thank Dr. Amber Franklin for her contributions on this project. The first, second, and last authors are employed by the US Department of Veterans Affairs.

Disclaimer statements

Contributors All authors and contributor Dr. Amber Franklin participated in conceiving and designing the study. The second and last authors participated in obtaining funding and/or ethics approval. All authors participated in collecting, analyzing, and interpreting the data. The first and second authors did the majority of writing the article, and all authors participated in revising the article.

Funding None

Conflicts of interest The authors report no conflicts of interest.

Ethics approval University of Washington Institutional Review Board (IRB) approval was granted prior to initiation of this study (IRB# 43263).

Appendix A: Phonomotor treatment protocol

Treatment materials	Small mirror Line drawings of mouth postures, icons for voiced/voiceless consonants Letter tiles Wipe-off board with markers Small colored blocks	
	Stage 1: Sounds in Isolation	Stage 2: Sounds in syllables
Overview	The purpose of Stage 1 is to train sounds in *isolation* through multimodal instruction using tasks designed to engage distributed articulatory-motor, acoustic, tactile–kinesthetic, and orthographic representations. *Consonant sounds* are introduced using mouth pictures and SLP model as cognate pairs by place/manner of articulation and grouped according to tactile–kinesthetic description (lip, tongue, air, nasal, and wind). They are introduced in the following order: lip (*p/b, f/v*), tongue (*t/d, k/g, th/th*), air (*s/z, sh/zh, ch/j*), tongue (*l/r*), nasal (*m/n/ng*), and wind (*h/w/wh*). When mastery of a consonant pair is achieved (e.g. *p/b*) in perception and production (approximately 85% accuracy), the next sound pair is introduced (e.g. *t/d*). Once a sound pair is introduced, training continues on this pair in all subsequent sessions. Once a participant can perceive and produce all consonants in isolation, corresponding graphemes are introduced using the corresponding mouth picture. *Vowel sounds* are trained according to lip and jaw placement via mouth pictures and letter tiles. Vowel sounds (*ee*, *o*, *oo*) are introduced with consonants to allow for minimal pair discrimination (e.g. *eep*, *op*, *oop*). The remaining vowels are trained after consonants.	The purpose of Stage 2 is to extend skills acquired in Stage 1 to *phoneme sequences*. Treatment tasks remain similar to Stage 1 tasks, with the exception that sounds will be produced in combinations rather than isolation. Training progresses from shorter, monosyllabic sequences to longer, multisyllabic (more complex) sequences (e.g. VC, CV, CVC, CCV, VCC, CCVC, CVCC, CCVCC, CVCV). Both real and nonwords are trained using phonologic tasks (in other words, only phonological features, *not* semantic features, are trained for real words). Nonword training is introduced before real word training to allow for emphasis on phonology; however, as treatment progresses nonwords and real words are trained simultaneously.

Introduction of sounds and sound sequences	Participant observes SLP producing a single sound (e.g. /*p*/). SLP asks participant what they observed (heard, saw) and if needed, describes what articulators are moving and how they move. For the sound /*p*/, for example, 'the lips come together and blow apart, the sound is "quiet" so the voice is turned off, the tongue is not moving.' The participant is then shown the line drawing of the mouth posture corresponding to the sound. After looking at the mouth picture and hearing the SLP's production, the participant is then asked to repeat the sound while looking in the mirror. The participant is also asked to place their hand on their throat in order to feel for vocal fold vibration ('quiet' versus 'noisy'). Following production, the SLP asks the participant what s/he saw and felt when the sound was made. Socratic questioning is used to enable the participant to 'discover' the auditory, visual, articulatory, and tactile/kinesthetic attributes of the sound (e.g. 'What do you feel when you make that sound? What moved? What did you see when you made that sound?', etc.). Within therapy progression for all levels is based on 85% accurate performance on task.	The process of 'discovering' sounds primarily occurs in Stage 1; however, knowledge of the auditory, visual, articulatory, and tactile/kinesthetic attributes of sounds can also be used later in the program as a cueing technique to identify individual phonemes within a phoneme sequence. For example, if a participant had trouble parsing the initial sound in *peef*, the SLP would use Socratic questioning (e.g. 'What do you feel when you make that first sound? What moved? Did your lips or tongue move when you made that sound?', etc.) to help identify the initial sound /*p*/. Put differently, rather than give the participant a model and tell them what the initial sound is, the SLP assists the participant in self-awareness of errors and how to repair them.
Perception tasks	Perception of sounds in isolation can be trained through various multimodal tasks. Examples: *Mouth pictures*: SLP produces a sound (e.g. *p*) and asks the participant to choose that sound from an array of mouth pictures (e.g. *p*, *b*, *t*, *d*) *Colored blocks*: SLP produces a string of individual sounds (e.g. *p*, *t*, *t*, *b*) and asks the participant to lay out blocks to demonstrate ability to discriminate sounds (e.g. blocks: red, blue, blue, green). *Verbal*: SLP produces two sounds (e.g. *p*, *p* or *p*, *b*) and asks the participant 'same or different.' *Letters*: SLP produces a sound and asks participant to point to the corresponding letter from an array of letters.	The SLP produces a real or nonword sound combination and asks the participant to depict the target through various tasks: *Mouth pictures*: If the participant heard the CVC *peef*, they would select the pictures corresponding to *p*, *ee*, and *f*. *Colored blocks*: If the participant heard the CVCV *peefee*, they would select three differently colored blocks arranged in the following order: white, black, red, black. *Verbal*: If the participant heard the CCVCs *grook* and *glook*, the SLP would ask 'same or different.' *Letters*: If the participant heard *chootee*, s/he would select the corresponding letter tiles.
Production tasks	Production of sounds in isolation can be trained through various tasks. Here are some examples: *Mouth pictures*: The SLP shows participant a mouth picture and asks the participant to produce that sound (e.g. *d*). *Motor description*: The SLP describes a sound (e.g. 'make the sound where your voice is noisy and your tongue quickly taps the roof of your mouth') and asks the participant to say the sound. *Verbal*: The SLP asks the participant to repeat a sound *p* or a string of individual sounds *p*, *p*, *s*, *d*. *Letters*: The SLP shows the participant a letter to elicit production of the sound.	The SLP elicits a real or nonword sound combination by asking the participant to produce the target through various tasks: *Mouth pictures*: The SLP lays out a series of mouth pictures and asks the participant to 'touch and say' each sound (*f-ee-p*) and then blend the sounds to produce the target (*feep*). *Verbal*: The SLP asks the participant to repeat a nonword *grook* and parse the word apart (*g-r-oo-k*). *Letters*: The SLP lays out letter tiles (or writes letters on dry erase board). The participant parses out the sounds by underlining and verbalizing each grapheme and then blends the sounds to produce the target.

Note: This appendix is meant to provide an overview and quick reference for those already familiar with the phonomotor treatment program. Readers interested in implementing this program are strongly encouraged to contact the first author of this paper for further information. Reprinted with permission from the American Speech Language Hearing Association.

Appendix B: Trained stimuli used in treatment.

Trained sounds in isolation	
IPA symbol	Graphemic representation(s)
p	p
b	b
f	f
v	v
t	t
d	d
k	k
g	g
θ	th
ð	th
s	s
z	z
ʃ	sh
ʒ	zh
ʧ	ch
ʤ	j
l	l
r	r
h	h
w	w
wh	wh
m	m
n	n
ŋ	ng
i	ee
ɪ	i
ɛ	e
eɪ	ae
æ	a
ʌ, ə	u
ɑ, ɔ	o, aw
o, oʊ	oe
ʊ	oo
u	oo
ɑɪ	ie
ju	ue
ɔɪ	oi, oy
aʊ	ow, ou
ɝ, ɚ	er, ir, ur
ɔr	or
ɑr	ar

References

ASHA. 2012. (09/25) Accent modification [retrieved 10/11/12]. Available from: http://www.asha.org/public/speech/development/Accent-Modification/.

Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., Amtmann, D. 2013. The Communicative Participation Item Bank (CPIB): item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research*, 56(4): 1190–1208.

Beeson, P., Robey, R. 2006. Evaluating single-subject treatment research: lessons learned from the aphasia literature. *Neuropsychology Review*, 16: 161–169.

Brookshire, C., Conway, T., Hunting Pompon, R., Oelke, M., Kendall, D. 2014. Effects of intensive phonomotor treatment on reading in eight individuals with aphasia and phonological alexia. *American Journal of Speech-Language Pathology*, 23(2): S300-11.

Brown, A. 1988. Functional load and the teaching of pronunciation. *TESOL Quarterly*, 22(4): 593–606.

Burgess, J., Spencer, S. 2000. Phonology and pronunciation in integrated language teaching and teacher education. *System*, 28: 191–215.

Chakraborty, R., Domsch, C., Gonzales, M.D. 2011. Articulatory behaviors of nonnative speakers: role of L2 proficiency and accent modification. *Perceptual and Motor Skills*, 113(1): 311–330.

Champagne-Muzar, C., Scheiderman, E.I., Bourdages, J.S. 1993. Second language accent: the role of the pedagogical environment. *International Review of Applied Linguistics*, 31(2): 143–160.

Couper, G. 2006. The short and long-term effects of pronunciation instruction. *Prospect*, 21(1): 46–66.

de Bot, K. 1983. Visual feedback of intonation I: effectiveness and induced practice behavior. *Language and Speech*, 26(4): 331–350.

Derwing, T., Munro, M., Wiebe, G. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3): 393–410.

Eckman, F. 2008. Typological markedness and second language phonology. In: Edwards, H., Jette, G., Zampini, M.L., (eds.) *Phonology and second language acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company, p. 95–116.

Flege, J.E. 1995. Second language speech learning: theory, findings, and problems. In: Strange W., (ed.) *Speech perception and linguistic experience: issues in cross-language research*. Baltimore, MD: York Press, p. 233–277.

Gluszek, A., Dovidio, J.F. 2010. The way they speak: a social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, 14(2): 214–237.

Hammill, D., Brown, V., Larsen, S., Wiederholt, L. 1994. *Test of Adolescent and Adult Language. Assessing linguistic aspects of listening, speaking, reading, and writing*. Austin, TX: Pro-Ed.

Kendall, D., Brookshire, C., Oelke, M., Nadeau, S. 2013. Phonomotor rehabilitation of anomia in aphasia. Proceedings from 43rd Clinical Aphasiology Conference, Tucson, AZ.

Kendall, D., del Toro, C., Nadeau, S., Johnson, J., Rosenbek, J., Velozo, C. The development of a standardized assessment of phonology in aphasia. In: Clinical Aphasiology Conference; 2010 May 23-27; Isle of Palms, SC.

Kendall, D., Rosenbek, J., Heilman, K., Conway, T., Klenberg, K., Gonzalez, L., *et al.* 2008. Phoneme-based rehabilitation of anomia in aphasia. *Brain and Language*, 105: 1–17.

Kertesz, A. 1982. *The Western Aphasia Battery.* New York: Grune and Stratton, Inc.

Kleim, J. 2008. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research*, 51: S225–S239.

Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., Berke, G.S. 1993. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research*, 36: 21–40.

Lado, R. 1957. *Linguistics across cultures: applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review*, 74(6): 431–461.

Lichtheim, L. 1885. On aphasia. *Brain*, VII: 433–484.

Lindamood, C.H., Lindamood, P.C. 1998. *The Lindamood phoneme sequencing program for reading, spelling, and speech.* Austin, TX: Pro-Ed.

Munro, M.J., Derwing, T.M. 2006. 'The functional load principle in ESL pronunciation instruction: an exploratory study. *System*, 34: 520–531.

Nadeau, S.E. 2001. Phonology: a review and proposals for a connectionist perspective. *Brain and Language*, 79: 511–579.

Nadeau, S.E. 2014. Neuroplastic mechanisms of language recovery after stroke. In: Tracy J.I., Hampstead B.M., Sathian K., (eds.) *Cognitive plasticity in neurologic disorders.* New York, NY: Oxford University Press, p. 61–84.

Portney, L.G., Watkins, M.P. 2000. *Foundations of clinical research: applications to practice.* 2nd ed. Upper Saddle River: Prentice Hall Health.

Raven, J.C. 1978. *Coloured Progressive Matrices, Sets A, AB, B.* Los Angeles, CA: Western Psychological Services.

Rothauser, E.H., Chapman, W.D., Guttman, W., Hecker, M.H.L., Nordby, K.S., Silbiger, H.R., *et al.* 1969. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3): 225–246.

Saito, K., Lyster, R. 2012. Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /r/ by Japanese learners of English. *Language Learning*, 62(2): 595–633.

Sheldon, A., Strange, W. 1982. The acquisition of R and L by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3: 243–261.

Shrout, P.E., Fleiss, J.L. 1979. Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin*, 86(2): 420–428.

Sikorski, L.D. 2005. Foreign accents: suggested competencies for improving communicative pronunciation. *Seminars in Speech and Language*, 26(2): 126–130.

Strange, W. 1995. Cross-language studies of speech perception a historical review. In: Strange W. (ed.) *Speech Perception and Linguistic Experience*. Timonium, Maryland: York Press, pp. 3–45.

Storkel, H.L., Armbruster, J., Hogan, T.P. 2006. Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, 47: 1175–1192.

Vaden, K.I., Halpin, H.R., Hickok, G.S. 2009. Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Available from: http://www.iphod.com.

Yorkston, K.M., Beukelman, D.R. 1981. *Assessment of Intelligibility of Dysarthric Speech*. Tigard, OR: CC Publications.