Journal of Academic Language & Learning
Vol. 9, No. 1, 2015, A1-A17. ISSN 1835-5196

Evidence that L2 production training can enhance perception

Gary Linebaugh

Department of English, American University of Sharjah, Sharjah, United Arab Emirates

Email: glinebaugh@aus.edu

Thomas Roche

SCU College, Southern Cross University, Lismore, NSW, Australia; and,

Faculty of English Studies, Sohar University, Sohar, Sultanate of Oman

Email: thomas.roche@scu.edu.au

(Received 10 May, 2014; Published online 6 March, 2015)

It is often readily accepted that perception precedes production in second language acquisition. According to Flege's (1995) Speech Learning Model and Broselow and Park's (1995) Split Parameter Setting Hypothesis, accurate second language (L2) perception necessarily precedes accurate L2 production. This paper examines whether, contrary to that assumption, production can inform perception, whether training in the production of problematic L2 sounds can enhance perception of those sounds. Participants were 46 Arabic speaking learners of English and took part in a between-groups experiment. They were assigned to either an articulatory training or focused exposure condition for learning three problematic English contrasts: /æ, ʌ/, /ɜ, ɔ/ and /g, ʤ/. Performance on pre-, post- and post-post-condition perceptual discrimination tests was used to assess participants' improvement in ability to perceptually discriminate the sounds after training in production or after focused aural exposure. Results point to the efficacy of the articulatory training, and thereby provide strong evidence that production can inform perception and that L2 acquisition can be facilitated through targeted training in articulation.

Key Words: L2 speech perception, L2 speech production, SLA, second language teaching, teaching pronunciation, articulatory training.

1. Introduction

The acquisition of the phonology of a second language involves the acquisition of phonetic categories in the second language. Conventionally, the establishment of phonetic categories involves a mapping between a specific set of acoustic cues and a vocal tract configuration where the mapping is mediated by linguistic representations variously conceived of as bundles of phonetic features, phonetic segments, phonemes, or words. A learner, upon hearing speech, analyzes the acoustic cues and matches them with the abstract representations; thereby creating a phonetic category that is in turn accessed in speech production.

In first language acquisition, the direction of this mapping is uncontroversial; after sufficient auditory exposure, the child has established phonetic categories and begins to produce the sounds of the language. Perception precedes production. The belief that perception is a prerequisite to accurate production in second language acquisition (SLA) is also widespread. According to Flege's (1995) well known Speech Learning Model (SLM), accurate perception of L2 sounds is a necessary, though not necessarily sufficient, condition for accurate production. Broselow and Park's (1995) Split Parameter Setting Hypothesis (SPSH) also claims that accurate L2 perception necessarily precedes L2 production. Much of the SLA literature and many contemporary second language teacher training textbooks reflect the assumption that perception is a necessary precondition for production (see International House, 2009: Krashen, 1996; Nation & Newton, 2009; Ridgway, 2000; Scrivener, 2005; Ur, 2008).

There are, however, a number of studies that raise questions about the precedence relationship between perception and production in SLA. Goto (1971) and Sheldon and Strange (1982) both found that the ability of Japanese learners of English to accurately produce the English liquids /l/ and /r/ was more advanced than their ability to perceptually distinguish the two sounds. Flege and Eefting (1987) found that the ability of Dutch speakers to produce English-like Voice Onset Timing (VOT) was more advanced than their ability to discriminate between Dutch-like VOT and English-like VOT. Mack (1989) found that the perceptual identification and discrimination abilities of French-English bilinguals lagged behind their productive abilities with respect to /d/ versus /t/ and /i/ versus /ɪ/. Tsukada et al. (2005) found that the ability of Korean children to produce English vowels (i, ɪ, e, ɛ, æ, ɑ, ʌ, u) surpassed their ability to perceptually discriminate the same vowels. Kluge, Rauber, Reis, and Bion's (2007) results indicate that speakers of Brazilian Portuguese were more accurate in production than perception of English nasal codas. Zampini and Green (2001) found that English speaking learners of Spanish were more Spanish-like in production than in perception of Spanish /p/.

These studies provide evidence that productive skills can exceed perceptual skills, and it is tempting to claim as a corollary that production can precede perception. However, positing such a corollary claim is problematic due to the difficulty in directly comparing perception and production abilities. Mack (1989, 2003) and Listerri (1995) note the difficulties in comparing ability in perception to ability in production, and Mack (2003) notes that testing the two modalities necessarily involves different methodologies, tasks, and evaluation procedures. These differences preclude drawing reliable conclusions about whether L2 perception or L2 production skills are more advanced. In addition, even if we accept that production skills can be more advanced, it is not logical to assume that that means production can precede perception. It is possible that perception and production are on separate tracks, that acquisition of one skill is at least partly independent of the other. As Broselow and Kang (2013) note, "the question of whether L2 perception and production develop in tandem" (p. 533) is an important issue in the study of second language phonology. The nature of this important relationship between the perception and production of second language speech sounds is not yet well understood, and this paper adds to the understanding of that relationship by examining the link between perception and production directly. We aim to determine if training in production of problematic L2 sounds (sounds that are not accurately perceived as different by L2 learners) leads to improvement in ability to perceptually discriminate between those sounds.

The difficulty associated with perceiving and producing L2 sounds has been observed to continue even after long-term exposure to the L2 and among advanced learners (Darcy et al. 2012; Eckman, Elreyes, & Iverson, 2003; Flege, Yeni-Komshian, & Liu, 1999; Mack, 1989; Pallier, Bosch, & Sebastián-Gallés, 1997; Pallier, Colomé, & Sebastián-Gallés, 2001). Of particular interest for our purposes are problematic phonemic contrasts that Darcy et al. (2012) call spurious homophones. Perceptual assimilation (Best, 1995; Flege, 1995) occurs when two phonemes merge into a single category due to similarity and under influences from the first language (L1). This specific type of perceptual assimilation is referred to as Single Category assimilation in Best (1995) and is quite similar to what Flege (1995) terms equivalence classification. A well-known example of this is the difficulty Japanese speakers face in distinguishing English /l/ and /r/, thereby creating the spurious homophone comprised of 'lock' and 'rock'. The experiment described here is designed to discover if articulatory training in the production of perceptually assimilated sounds can help learners more sharply define the boundaries between the assimilated phonetic categories, and if more sharply defined categories can in turn lead to improved ability in perceptually identifying the sounds. In other words, can production inform perception by shaping and defining the phonetic categories? Such a finding would contradict the notion that perception necessarily precedes production.

1.1. How production can inform perception

The notion that production can inform perception raises the question as to what mechanisms might be availed if training in production directly leads to improved perception. We cite three strands of evidence in support of this possibility. 1) There is reason to believe that babbling and pre-speech sounds inform acquisition of first language speech sounds. 2) Audio feedback from a speaker's own speech plays a role in shaping and adjusting phonetic categories. 3) There is evidence that speech perception involves the activation of motor codes associated with the production of speech.

In first language acquisition, infants perceive speech sounds and words, making a connection between sounds and meaning, before producing language sounds. It is quite possible, however, that the non-language sounds of babbling are part of the language development process in infants. Proprioception coupled with auditory feedback during babbling may provide the child with knowledge regarding the matching of particular acoustic signals with particular vocal tract configurations. This knowledge is then utilized in the establishment of phonetic categories once the child starts making sound-meaning connections. Jakobson (1968), in what is known as the discontinuity hypothesis, discounted the possibility that babbling is a precursor to speech and argued that acquisition of anything recognizable as a phonological system begins only when the child starts making the sound-meaning connection. However, in direct tests of the discontinuity hypothesis, Vihman, Macken, Miller, Simmons and Miller (1985) found evidence of continuity, evidence that the pre-language sounds of babbling and the sounds of language are part of the same developmental process. They found a striking consistency between babbling and word production in individual children across time regarding length of vocalizations, phonotactic constraints, and distribution of consonants. Their findings suggest that it is inaccurate to claim that babbling is unrelated to speech, and they speculate that the sounds of babbling are indeed precursors to speech. In more recent work, Guenther (2006) claims that babbling forms a bridge to early speech and that the two together are essential in establishing connections among articulatory, somatosensory, and acoustic information. If such is the case, children are not solely dependent on accessing innate feature geometries or phonetic categories in order to begin perceiving and producing speech. They have an inventory of vocal tract configurations matched with their auditory consequences. We suggest that second language learners may benefit in a similar way. Training in production and production practice may shape and adjust second language phonetic categories by connecting orosensory and acoustic information.

The orosensory information and audio feedback associated with babbling also pertains to mature speakers. Linguistic accommodation is a well-known process in which speakers emulate an interlocutor, and Casserly and Pisoni (2010) found that speakers also adjust their speech based on perception of their own speech. Manipulation of speakers' perception of their own speech led to immediate and robust alterations in production. Such adaptation of the sensorimotor system suggests that speech production activates a robust feedback operation. According to Casserly and Pisoni (2010), "[a] speaker's perception of his or her own speech plays a significant role in the planning and execution of future speech production" (p. 249). Levelt, Roelof, and Meyer (1999) report that we do indeed monitor our own speech and that this ability to self-monitor is exploited in the process of phonological encoding. Thus, there is further reason to believe that production can shape and inform phonetic categories. Proprioception coupled with auditory feedback from one's own speech can lead to adjustment of phonetic categories and can solidify categorical differences. It is even possible that new phonetic categories can be formed through this process.

The Motor Theory of Speech Perception (Galantucci, Fowler, & Turvey, 2006; Liberman 1957; Liberman & Mattingly 1985) claims that speech perception is mediated by reference to motor codes involved in speech production. In effect, all speech, both perceived and produced, is converted into a gestural code. In support of the Motor Theory, neuroimaging studies show that speech perception invokes motor cortical activity (Pulvermuller et al. 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). The implications for this study are clear; training in production may create or adjust the gestural codes that are essential for accurate perception.

The experiment described below tests the hypothesis that articulatory training in the production of problematic sounds in a second language can enhance the ability of learners to perceptually discriminate those sounds. The experiment involves a perceptual identification task in which participants are asked to identify which of two perceptually confusable sounds they have heard. In a preliminary study (Linebaugh & Roche, 2013), we found that articulatory training in the production of /p, b/ enhanced ability to perceptually discriminate the two sounds in a way that focused listening did not. These results were encouraging, and in order to more thoroughly test the hypothesis, we identified three more pairs of sounds that were perceptually confusable for Omanis learning English.

1.2. Problematic English sounds for Arabic speakers

English consonants with equivalents or near-equivalents in Arabic are seen in the darkly shaded boxes in Table 1. The consonants in unshaded boxes have no Modern Standard Arabic MSA equivalent or near-equivalent. There are few studies of Omani Arabic speakers' difficulties with English phonology, but Al-Beloushi's (2012) unpublished study of young Omani learners found that they struggled to produce the phonemes identified in the unshaded boxes in Table 1 as well as the voiced palatal affricate /ʤ/ which appears in the lightly shaded box in Table One. This sound, in fact, proved the most problematic of all for the young Omani learners. Holes (2004) reports that the sounds /g, ʤ/ are dialectal variants of the phoneme represented by the Arabic letter 'ji:m'. He further notes that Omani Arabic is one of the dialect areas in which /g/ is often used where /ʤ/ is used in MSA, and Smith (2001) notes that Arabic speakers in general produce one or the other of these sounds according to their local dialect in all phonetic contexts.

Table 1. Near equivalent (darkly shaded) and problematic (unshaded, lightly shaded) English consonants for Arabic L1 speakers (after Smith (2001) and Kharma and Hajjaj (1989)).

p	b	t	d	ʧ	ʤ	k	g
f	v	θ	ð	s	z	ʃ	ʒ
m	n	ŋ	h	l	r	w	y

Based on the above information and personal observations of difficulties encountered by Omani students, we selected /g, ʤ/, for investigation in this study.

There are three short vowels in MSA /i, u, a/, plus long versions of each of those three. Due to this relatively small vowel inventory, learning to perceive and produce the larger number of English vowels is often problematic, and as Smith (2001) notes, most English vowels present problems for Arabic speaking learners. Through general observation and informal assessment, we discovered that among others our Omani students had difficulty differentiating the pairs /æ, ʌ/ and /ɜ, ɔ/, so we chose these vowel pairs for investigation. The low front vowel /æ/ is a common allophone of short /a/ in Arabic, but the other three vowels are entirely missing from Arabic.

As seen in Table 2, the three sound contrasts we chose represent a cross section of possibilities in terms of the relationship between the English sound pairs and the Arabic sound inventory.

Table 2. Sound contrasts investigated in the study.

English sound pair	Exists in Arabic	Does not exist in Arabic
/g, ʤ/	/g/ or /ʤ/ depending on dialect	/g/ or /ʤ/ depending on dialect
/æ, ʌ/	/æ/ as an allophone	/ʌ/
/ɜ, ɔ/		/ɜ, ɔ/

The first language experience the participants have with the sounds is variable. In the first two pairs, the participants can be expected to have L1 experience of different types with one of the two sounds: as a dialectal variant of a phoneme and as a common contextual allophone. In the third case, participants will have had no L1 experience with either of the two sounds. The selection of these contrasts allows us to test the hypothesis that training in production enhances perception across a variety of sounds, consonants as well as vowels, and across a variety of prior experiences with the sounds.

2. The study

2.1. Participants

Participants were first year English students at Sohar University in the Sultanate of Oman. They had previously completed the English Language Foundation Program at Sohar University, nominally testing at 4.5 or better on all sub-skill areas of IELTS. All participants were native speakers of Arabic. Females comprise the vast majority of students in the English Department at Sohar University, and this participant pool reflects that. Table 3 provides information about the 6 participant groups: 3 experimental groups and 3 control groups. Convenience sampling was used; training and testing took place as part of regular class meetings. The experimental groups received articulatory training in the production of the problematic sounds. In order to ensure that any improvement in perceptual ability was not simply due to the focused exposure to the sounds of interest, we used control groups who took part in focused listening activities but received no articulatory training. If improvement is seen in perceptual skills of the experimental group but not the control group, it is reasonable to assume the improvement is due to the training and not just the additional exposure to the sounds.

Table 3. Participant Information.

Group	Males	Females	Age range	Median age
/æ, ʌ/ Experimental	2	29	17-25	20
/æ, ʌ/ Control	-	14	18-24	20
/g, ʤ/ Experimental	2	18	18-24	20
/g, ʤ/ Control	2	14	17-25	20
/ɜ, ɔ/ Experimental	3	22	18-26	20
/ɜ, ɔ/ Control	-	17	17-24	20

2.2. Materials

Stimuli consisted of sets of minimal pairs that exemplify the contrasting pairs of sounds identified as problematic for Omani learners of English. The specific sounds tested in this study were /æ, ʌ/, /g, ʤ/ and /ɜ, ɔ/; ten minimal pairs were selected for each of those pairs. A list of the minimal pairs used is provided in Appendix A. For the sound pairs /æ, ʌ / and /g, ʤ/, the first author, a native speaker of American English, was recorded reading carrier sentences (‘The man said ___ again”) that included one of the words from each minimal pair. Each sentence was read twice. There were twenty trials for each test, two for each of the ten minimal pairs, with a pause of approximately 4 seconds between trials. To avoid a familiarity effect, where participants are familiar with one of the words from the minimal pair but not the other, the words chosen for the stimuli were among the top 2000 words on a word frequency list based on the *British National*

Corpus (BNC, The British National Corpus, 2007)[1]. Due to a shortage of appropriate /g, ʤ/ minimal pairs, an exception was made, and we included minimal pairs including common English names (Joe, James, John, and Jane). The frequency list for the BNC does not include names. The second author, a native speaker of Australian English was recorded reading the stimulus sentences involving the /ɜ, ɔ/ contrast. A speaker of Australian English was required for this particular pair because /ɜ/ is not commonly found in American English. The recordings were made using *Audio Hijacker* on an Apple MacBook Pro computer. The training for each sound pair was done by the person who recorded the stimuli.

2.3. Procedure

The experiment began with a pre-test (T1); participants listened to the stimulus sentences through headphones in a computer laboratory and indicated on an answer sheet which sound they believed they had heard, 'cat' or 'cut' for example. The answer sheets were collected before training started.

After the pre-test, the experimental groups received articulatory training for the sounds of interest. The control group received focused listening practice but no articulatory practice. The treatment, articulatory training or focused listening was not extensive and lasted only about 20 minutes. Immediately after the treatment, a post-test (T2) was administered. In order to see if any improvement was longer lived, a post-post test (T3) was given one week later. Those two tests followed the same procedure as the pre-test.

2.4. Description of training

/æ, ʌ /

Training started with the trainer modeling pronunciation of minimal pairs; participants were asked to listen and repeat. After that, articulatory training in production of the two vowels began. The trainer focused attention on the position of the tongue and jaw, and utilized the University of Iowa website (University of Iowa, n.d.) animated videos depicting the production of the two vowels. The trainer then informed participants that the tongue is in a neutral position in the middle of the mouth for production of the central vowel /ʌ/, but that for /æ/, the tip of the tongue touches or nearly touches the lower teeth. It was also pointed out that the jaw is lowered in the production of this sound. Following Avery and Ehrlich (1992), /æ/ was described as a somewhat ugly sound when produced in isolation, and this was reinforced by the trainer producing the sound with an exaggerated sense of displeasure. In the final training activity, participants produced the two vowels in rapid succession, saying 'æʌ æʌ æʌ æʌ æʌ'.

Treatment for the control group began in the same way, with a listen and repeat activity. After that the trainer led the participants in a game that required them to identify which word they heard ('cut' or 'cat' for instance). The treatment for the control group ended with another listen-and-repeat drill.

/g, ʤ/

This training session also began with a listen and repeat drill and an informal assessment of participant ability to perceptually discriminate between the two sounds. The explicit articulatory training component for /g, ʤ/ focused on tongue position and the difference between affricates and stops. Participants were shown animated videos illustrating the production of these two sounds (University of Iowa, n.d.), and that was followed by a detailed description of the affricate as a combination of a stop and fricative. Placement of the tongue against the alveolar ridge for the start of the affricate was easily apprehended by the participants. Awareness of tongue placement in the velar region was less discernible, but since the voiceless velar stop is a common sound in Arabic, reference to this sound facilitated instruction in producing the voiced

[1] These lexical items fall well within the 5000-8000 word-family vocabulary knowledge generally acknowledged as necessary for tertiary study in English. These top 2000 words account for approximately 80-85% of words in spoken and written English texts (Nation and Waring, 1997).

counterpart. The articulatory training continued with the participants alternately producing the two sounds, saying /gə-ʤə- gə-ʤə - gə-ʤə - gə-ʤə/, and concluded with a listen and repeat drill. The procedure for the control group was the same as for the control group for /æ, ʌ /.

/ɜ, ɔ/

The training for production of /ɜ, ɔ/ began with display of the vowel chart from Underhill (2005, p. 15) showing both lip and tongue positions for English vowels. The trainer highlighted the different tongue positions and drew attention to the fact that /ɜ/ is a mid-central vowel and /ɔ/ is a mid-back vowel. He then drew comparisons to vowels with similar and contrasting tongue positions. Work on tongue position was followed by a self-discovery activity in which students noticed the difference between rounded and unrounded vowels. After that, there was a listen and repeat drill in which the trainer modeled pronunciation of minimal pairs, asking participants to listen and repeat. That was followed by an informal assessment of participant ability to perceptually discriminate between the two sounds in words spoken by the trainer and then by fellow students in pairs. To finish off, the participants generated lists of words containing /ɜ, ɔ/ and read from these lists, testing if fellow participants could identify which word they produced.

2.5. Reliability

In a pilot-test, we found that a 4 second pause between trials was appropriate. Participants were able to replay the sentences in their heads, but the pause was short enough that attention never wavered. Participants self-selected the volume settings on their headphones prior to the actual testing. There were no distractions; no one was permitted to enter or leave the room during the tests. In addition, the brevity of the training (about 20 minutes) and testing (about 2 minutes) guaranteed that fatigue was not an issue.

3. The results

Overall, the results support the hypothesis that explicit training in the production of problematic second language sounds can enhance the ability to perceive those sounds. With respect to the /g, ʤ/, and /æ, ʌ/ contrasts, participants given articulatory training in producing the sounds showed statistically significant improvement in performance on the post-test compared to performance on the pre-test, and improved performance carried through to the post-post test which was done a week after the training. For /æ, ʌ/, the control group, which received additional auditory exposure to the sounds but no articulatory training, showed no statistically significant improvement on either the post-test or the post-post test. As for/g, ʤ/, performance by control participants also improved from the pre-test to the post-test, and the change was statistically significant. However, unlike the experimental participants, the improvement in performance did not carry through to the post-post test, which revealed no significant change from the pre-test. As for the /ɜ, ɔ/ contrast, the experimental group's performance improved after training, but that change was not statistically significant. Neither experimental nor control groups showed statistically significant changes in ability to accurately perceive the two sounds as a result of training or additional exposure. Table 4 provides a summary of the mean scores across all three sound pairs tested.

For two of the three sound pairs tested, participants showed statistically significant improvement in ability to perceptually identify the sounds after receiving training in production of the sounds. Importantly, that improvement persisted for at least a week after training. Participants who received only additional listening practice showed statistically significant improvement in perceptual ability for only one of the three sound pairs, and that improvement had disappeared a week later. The findings provide evidence that articulatory training can improve perception in a way that additional listening exposure does not. The detail results for each pair of sounds are provided in Appendix B.

Table 4. Mean number of correct responses in 20 trials.

	/æ, ʌ/	/g, ʤ/	/ɜ, ɔ/
Experimental group			
Pre-test	14.0	15.5	15.6
Post-test	15.0*	18.4*	15.8
Post-post-test	15.0*	18.1*	16.3
Control group			
Pre-test	12.4	17.1	15.6
Post-test	12.9	18.0*	15.1
Post-post-test	12.8	17.4	15.8

*Statistically significant (p< .05 on a two-tailed paired t-test) change from score on pre-test.

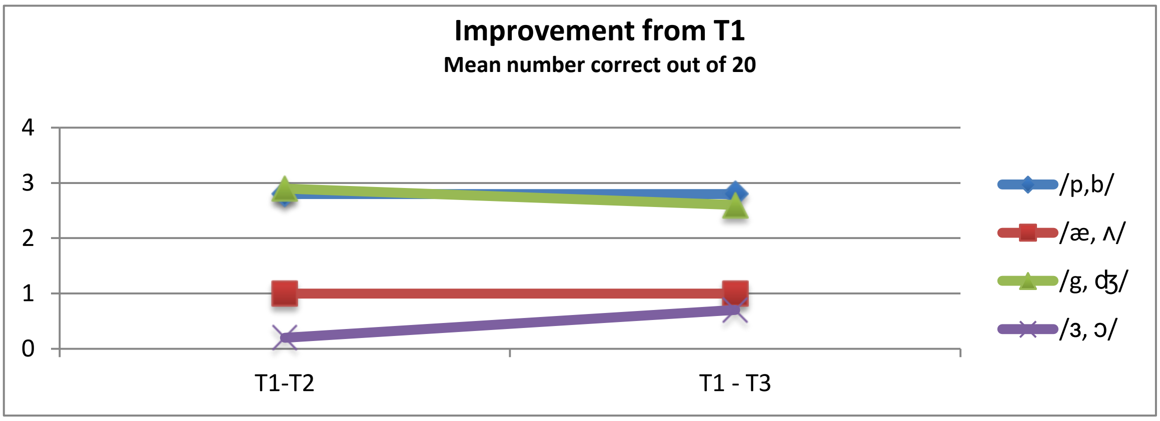

Figure 1. Improvement in experimental group performance from pre-test (T1) to post-test (T2) and from pre-test to post-post-test (T3).

Figure 1 reveals a difference between the effects of articulatory training for consonants as opposed to vowels. The improvement in mean scores for the consonant pairs /p, b/ from the preliminary study (Linebaugh & Roche, 2013) and /g, ʤ/ in this study was greater than that seen for the vowel pairs /æ, ʌ/ and /ɜ, ɔ/. It is possible that the greater orosensory awareness associated with training in production of the consonants, compared to training in the production of vowels, may be responsible for this difference. It is worth noting that for the vowel /æ/, training included drawing attention to the fact that this vowel is produced with the tongue touching (or close to) the lower teeth. The lowering of the jaw associated with this sound is also highly proprioceptive. Thus, training for this vowel involved more sensory awareness than does training for other English vowels. This fact may have been a key factor in the success of the training for this particular vowel. Lip rounding, which was part of the training for /ɜ, ɔ/, is also fairly proprioceptive, but the difference in tongue position for /ɜ/ compared to /ɔ/ is not easily discernible. This indicates that sensory awareness may be an important consideration in determining the best uses of articulatory training. It is possible that sounds that are produced with vocal tract configurations that are less discernible by learners require significantly more training than those that are more discernible. The training regimen in this study was quite short, about twenty minutes. It is possible that with more training, more positive results would have been seen for /ɜ, ɔ/.

It is also possible that differences with respect to salience in Arabic phonology played a role. As mentioned previously, for all but the last sound pair, participants would have had L1 experience with one of the two paired sounds, /g/ or / ʤ/ as a dialectal variant of a phoneme, / æ/ as an allophone, and in the preliminary study, /b/ as a phoneme. Thus, for all but the last pair, speakers essentially had to learn to produce only one of the sounds; the other already existed in their first language phonology. This provides another reason to believe that more extensive training may have produced more positive results for /ɜ, ɔ/. Alternatively, given that this sound pair was the only one involving Australian English, it is possible that the lack of positive results for this vowel pair is attributable to lesser familiarity on the part of the participants with Australian English compared to American English.

4. Discussion

The results indicate that production can inform perception in a second language in the sense that enhanced knowledge of production leads to enhanced perceptual ability. Results presented here suggest that learners are able to utilize auditory feedback from their own speech to shape, adjust, or define phonetic categories in the second language, and those more accurate phonetic categories lead to improved perceptual ability. The received wisdom that perception necessarily precedes production is in error. Further studies involving a wider range of speech sounds and with different L1 speakers would further strengthen this claim. This finding does not deny that meaningful input and sufficient exposure is key in attaining second language phonology. It does, however, suggest that for certain problematic (perceptually assimilated) sounds, exposure may not be enough. Language learners often encounter situations where, despite considerable exposure, they fail to accurately discriminate between two similar second language sounds. These findings indicate that learners can benefit from explicit articulatory training in the production of those sounds. Further, this finding does not deny that misperception is the cause of production problems. It simply shows that directly addressing production problems can be beneficial. It is not necessary in all cases to correct problems related to misperception before tackling production issues. Flege's (1995) SLM and Broselow and Park's (1995) SPSH are important and may accurately describe second language phonological acquisition processes in play for the majority of second language sounds, but we must recognize the complexity of the process and not claim that specific parameters determine the outcomes in all cases for all sounds. We find compelling evidence that any model of second language phonological acquisition must accommodate the fact that production can inform perception.

The centrality of gestures is reinforced by these findings. Acquiring L2 phonetic categories is an essential element of second language phonological acquisition, and our findings show that this process is advanced through training that focuses attention on vocal tract gestures. As learners gain knowledge of gestures, the accuracy of phonetic categories is solidified. Thus there is evidence for the hypothesis of the Motor Theory of Speech Perception that gestural knowledge is a key element of perception. Without accurate gestural knowledge, phonetic categories for perceptually problematic sounds are incomplete and overlapping. Knowledge of gestures for such sounds coincides with the establishment of accurate phonetic categories.

The Motor Theory of Speech Perception is not specifically about second language phonology, but it may be the case that the primacy of gestures is even greater for second language learners. As mentioned above, there is neurolinguistic evidence that speech perception activates motor cortices involved with speech production. Callan, Jones, Callan, and Akahane-Yamada (2004) found that activation of motor cortices was even greater among nonnative speakers when attempting to perceptually identify phonetic contrasts that were ambiguous. Native speakers by contrast, exhibit greater activation only in auditory cortices in the same situation. Thus, there is even more reason to believe that knowledge of second language articulatory configurations is key in identifying and recognizing problematic sounds.

In summary, learners use somatosensory awareness and auditory feedback from their own speech to more sharply delineate phonetic boundaries between perceptually assimilated second language sounds, in essence unassimilating the sounds. This results in more native-like phonetic categories, and that results in more accurate perception. Production can inform perception in the case of perceptually assimilated sounds.

The implications for second language teaching are clear; learners can benefit from explicit training in the production of problematic sounds. Over the last several years there has been a focus on the importance of suprasegmentals in teaching pronunciation with less attention paid to individual segments (e.g. tonic stress in Avery & Ehrlich, 1992; and sentence stress in Hahn, 2004). There is increasing evidence that the accurate production and perception of segments is essential in interactions between non-native speakers (Jenkins, 2000, 2002; Saito 2011), and given the growing recognition of the need to accommodate users of English as a Lingua Franca (ELF), SLA research and practice should pay increased attention to the importance of the production and perception of individual segments[2].

Acknowledgements

This research was supported by a grant from the Omani Research Council [Grant number ORG SU HER12 004].

Appendix A. List of minimal pairs used in stimuli.

/æ, ʌ /	/g,ʤ/	/ɜ, ɔ/
cap-cup	get-jet	walk-work*
hat-hut	go-Joe	bored-bird*
ankle-uncle	bug-budge	war-were*
track-truck	dog-dodge	short-shirt*
mad-mud	log-lodge	form-firm
ran-run	leg-ledge	store-stir
match-much	gain-Jane	warm-worm
cat-cut	games-James	torn-turn
fan-fun	gone-John	
drank-drunk	egg-edge	

There were 20 trials for each contrast using each of the 10 minimal pairs twice. Due to a shortage of minimal pairs for the /ɜ, ɔ/ contrast, the pairs with asterisks were each used 3 times.

Appendix B. Detailed results

/æ, ʌ /

The results by participant for /æ, ʌ/ can be seen in Tables B1 and B2. The mean score of experimental participants on the pre-test was 14 while the mean score on the post-test was 15. Thus there was not a large improvement, but the improvement was fairly consistent across participants and the improvement was statistically significant ($t(30) = -2.8$, $p < .01$) on a t-test. The mean score on the post-post-test administered one week after the training was also 15.0, and this difference was also statistically significant ($t(30) = -3.6$, $p < .01$). On average, control participants scored 12.4 on the pre-test, 12.9 on the post-test, and 12.8 on the post-post-test. None of the changes in average scores were statistically significant. The results support the hypothesis that explicit training in production of problematic sounds can improve the ability to perceive those sounds, and that the improvement is not attributable simply to the additional aural exposure to the sounds. (Note: One of the control participants did not take the post-post test.)

/g, ʤ/

The results by individual participant for /g, ʤ/ can be seen in Tables B3 and B4. The mean score of experimental participants on the pre-test was 15.5 while the mean score on the post-test was 18.4, and 18.1 on the post-post-test. The improvement seen after articulatory training is substantial and is statistically significant with respect to both the post-test ($t(20) = -5.8$, $p < .01$) and the post-post-test ($t(18) = -4.2$, $p < .01$). Control participants also improved on average, with a mean score of 17.1 on the pre-test and 18.0 on the post-test. This change is statistically significant ($t(13) = -2.5$, $p < .05$). This improvement, however, proved somewhat transitory as

[2] For a thorough discussion of this topic, see Linebaugh and Roche (2013).

the mean score on the post-post-test fell to 17.4, and the difference between scores on the pre-test and the post-post-test were not statistically significant (t(12) = -.6, p = .58). Providing additional listening practice focused on these two sounds provided a benefit in terms of ability to perceive them, but that benefit did not carry over in the longer term. It should be pointed out here that the additional listening practice involved listen and repeat drills and thus did involve some production. It is possible that the transitory improvement shown by the /g, ʤ/ control participants was due to the production practice coupled with the listening practice. It is, however, only with articulatory training that we see sustained improvement in perception, and the results for /g, ʤ/ also support our hypothesis. Articulatory training provides a boost to perceptual ability that lasts for at least a week while the benefit derived from additional aural exposure is more transitory, and has disappeared after a week. (Note: Two experimental participants and one control participant did not take the post-post test.)

Table B1. Number of correct answers out of 20 trials. Experimental group.

Participant	Pre-test	Post-test	Post-post-test	Participant	Pre-test	Post-test	Post-post-test
1	11	9	15	17	16	17	16
2	12	15	13	18	12	14	14
3	15	16	17	19	10	13	11
4	16	17	17	20	16	17	16
5	8	15	12	21	14	14	14
6	13	14	14	22	18	20	19
7	15	17	15	23	19	19	19
8	13	14	15	24	12	10	13
9	12	14	14	25	12	16	15
10	16	16	15	26	11	12	12
11	15	14	15	27	18	20	18
12	14	14	14	28	19	19	19
13	10	12	13	29	20	20	20
14	10	10	11	30	16	15	14
15	11	12	15	31	19	19	19
16	12	10	11				
				Mean	14.0	15.0	15.0
				SD	3.1	3.1	2.5

Table B2. Number of correct answers out of 20 trials. Control group.

Participant	Pre-test	Post-test	Post-post-test	Participant	Pre-test	Post-test	Post-post-test
1	12	10	11	8	9	11	12
2	19	19	18	9	12	10	10
3	10	8	11	10	9	11	9
4	17	17	16	11	10	12	12
5	13	14	16	12	9	10	11
6	15	16		13	13	15	13
7	16	15	18	14	9	13	10
				Mean	12.4	12.9	12.8
				SD	3.2	3.1	3.0

Table B3. Number of correct answers out of 20 trials. Experimental group.

Participant	Pre-test	Post-test	Post-post-test	Participant	Pre-test	Post-test	Post-post-test
1	12	18	16	12	17	20	20
2	16	19	17	13	11	17	
3	20	20	20	14	20	20	20
4	14	17		15	10	15	14
5	9	14	16	16	15	20	20
6	15	20	20	17	19	20	20
7	19	20	20	18	19	20	19
8	16	17	16	19	19	19	18
9	12	18	12	20	14	19	19
10	17	18	19	21	17	17	20
11	15	19	18				
				Mean	15.5	18.4	18.1
				SD	3.2	1.7	2.3

Table B4. Number of correct answers out of 20 trials. Control group.

Participant	Pre-test	Post-test	Post-post-test	Participant	Pre-test	Post-test	Post-post-test
1	19	19	19	8	18	17	17
2	17	18	18	9	19	20	19
3	14	15	13	10	16	20	15
4	15	14	14	11	15	17	14
5	16	18	20	12	16	18	
6	16	18	18	13	20	20	20
7	18	18	19	14	20	20	20
				Mean	17.1	18.0	17.4
				SD	1.9	1.8	2.4

Table B5. Number of correct answers out of 20 trials. Experimental group.

Participant	Pre-test	Post-test	Post-post-test	Participant	Pre-test	Post-test	Post-post-test
1	18	17	18	14	17	17	17
2	16	16	15	15	16	16	
3	15	11	13	16	16	16	15
4	11	16	14	17	15	10	18
5	15	15	15	18	16	16	
6	15	17	17	19	14	17	16
7	18	16	18	20	15	17	18
8	13	13	14	21	18	19	16
9	16	15	17	22	18	14	16
10	16	16	14	23	15	18	16
11	18	18	19	24	14	16	18
12	14	17	16	25	14	14	16
13	17	18	18				
				Mean	15.6	15.8	16.3
				SD	1.7	2.1	1.6

Table B6. Number of correct answers out of 20 trials. Control group.

Participant	Pre-test	Post-test	Post-post-test	Participant	Pre-test	Post-test	Post-post-test
1	13	14	13	10	15	15	16
2	17	11	13	11	16	17	17
3	19	14	16	12	17	17	
4	15	12	16	13	11	14	12
5	13	13	14	14	14	16	18
6	16	16	17	15	18	19	18
7	17	15	15	16	18	17	18
8	18	18	18	17	17	14	15
9	17	15	16				
				Mean	15.9	15.1	15.8
				SD	2.1	2.1	1.9

/ɜ, ɔ/

The results for /ɜ, ɔ/ can be seen in Tables B5 and B6. The mean scores of experimental participants were 15.6 on the pre-test, 15.8 on the post-test, and 16.3 on the post-post-test. None of the changes were statistically significant (Pre to post-test: $t(24) = -.433$, $p = .669$; pre to post-post-test: $t(22) = -1.859$, $p = .076$). For the control group, the same three scores were 15.9, 15.1, and 15.8. None of those changes were statistically significant (Pre to post-test: $t(16) = 1.412$, $p = .177$; pre to post-post-test: $t(15) = .259$, $p = .799$). The results for this sound pair do not support our hypothesis. (Note: two of the experimental participants and one of the control participants did not take the post-post test.)

Finally, it can be reported that the changes in performance from the post-test to the post-post-test was not significantly significant for any of the groups.

References

Al-Beloushi, J.M. (2012). *Investigating EFL Omani Grade 2 learners' ability to produce English phonics*. Unpublished Dissertation. M.Ed. SQU, Muscat, Oman.

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford, England: Oxford University Press.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.) *Speech perception and linguistic experience: issues in cross-language research*. (pp. 171–204). Timonium, MD: York Press.

Broselow, E., & Kang, Y. (2013). Phonology and speech. In J. Herschensohn & M. Young-Scholten (Eds.). *The Cambridge handbook of second language acquisition* (pp. 529-553). Cambridge: Cambridge University Press.

Broselow, E., & Park, H.B. (1995). Mora conservation in second language prosody. In J. Archibald (Ed.). *Phonological acquisition and phonological theory* (pp. 81-109). Hillsdale, NJ: Lawrence Erlbaum.

Callan, D.E., Jones, J.A., Callan, A.M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, *22*(3), 1182-1194.

Casserly, E.D., & Pisoni, D.B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 629-647.

Darcy, I., Dekydtspotter, L., Sprouse, R.A., Glover, J., Kaden, C., McGuire, M., & Scott, J.H.G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front round vowels in L1 English-L2 French acquisition. *Second Language Research*, *28*(1), 5-40. DOI: 10.1177/0267658311423455

Eckman, F.R., Elreyes, A., & Iverson, G.K. (2003). Some principles of second language phonology. *Second Language Research, 19*(3), 169-208.

Flege, J.E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Hueber & C. Ferguson (Eds.). *Crosscurrents in Second Language Acquisition and Linguistics Theories* (pp. 249-289). Amsterdam: John Benjamins.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.). *Speech perception and linguistic experience: issues in cross-language research* (pp. 233-77). Timonium, MD: York Press.

Flege, J.E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics,15*, 67-83.

Flege, J.E., Yeni-Komshian, G.H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78-104.

Galantucci, B., Fowler, C.A., & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, *13*(3), 361.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds. *Neuropsychologia*, *9*(3), 317-323.

Guenther, F.H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*, 350-365.

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Washington D.C.: Georgetown University Press.

International House. (2009). Unit 4, Section 3: Learning to listen by listening. *DELTA Module One Course Material*.

Jakobson, R. (1968). *Child language: aphasia and phonological universals, Vol. 72*. The Hague: Mouton De Gruyter.

Jenkins, J. (2000). *The phonology of English as an international language.* Oxford, England: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, *23*, 83-103.

Kharma, N., & Hajjaj, A. (1989). *Errors in English among Arab speakers: Analysis and Remedy*. London: Longmann.

Kluge, D.C., Rauber, A.S., Reis, M.S., & Bion, R.A.H. (2007). The relationship between the perception and production of English nasal codas by Brazilian learners of English. In *Proceedings of Interspeech 2007*, 2297-2300.

Krashen, S.D. (1996). The case for narrow listening. *System*, *24*, 97-100.

Levelt, W.J., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.

Liberman, A.M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, *29*, 117-123.

Liberman, A.M., & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Linebaugh, G., & Roche, T. (2013). Learning to hear by learning to speak: The effect of articulatory training on Arab learners' English phonemic discrimination. *Australian Review of Applied Linguistics*, 36(2).

Listerri, J. (1995). Relationships between speech perception and speech production in a second language. In K. Elenius & P. Branderud (Eds.). *Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 4* (pp. 92-99). Stockholm: Stockholm University.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception and Psychophysics*, *46*, 187-200.

Mack, M. (2003). The phonetic systems of bilinguals. In M. Banich & M. Mack (Eds.). *Mind, brain, and language* (pp. 309-349). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Nation, I.S.P., & Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking.* New York: Routledge Publisher.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.). *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge, UK: Cambridge University Press.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in vowel acquisition. *Cognition*, *64*, 9-17.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based vs. abstract lexical entries. *Psychological Sciences*, *12*, 445-49.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, USA*, *103*(20), 7865-7870.

Ridgway, T. (2000). Listening Strategies – I beg your pardon? *ELT Journal*, *54*(2), 196-197.

Saito, K. (2011). Identifying Problematic Segmental Features to Acquire Comprehensible Pronunciation in EFL Settings: The Case of Japanese Learners of English. *RELC Journal*, *42*(3), 363-378.

Scrivener, J. (2005). *Learning Teaching*. Oxford, UK: Heinemann.

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*(03), 243-261.

Smith, B. (2001). Arabic speakers. In M. Swan & B. Smith (Eds.). *Learner English: A teacher's guide to interference and other problems* (pp. 195-213). Cambridge: Cambridge University Press.

The British National Corpus. (2007). *The British National Corpus*, version 3 (BNC XML Edition). Oxford University Computing Services on behalf of the BNC Consortium. Retrieved October 12, 2012, from http://www.natcorp.ox.ac.uk/

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263-290.

Underhill, A. (2005). *Sound Foundations.* London, UK: Macmillan.

University of Iowa. (n.d.) *Phonetics: The sounds of American English.* Retrieved October 12, 2012, from http://www.uiowa.edu/~acadtech/phonetics/english/frameset.html

Ur, P. (2008). *A Course in Language Teaching: Practice and theory*. Cambridge, UK: Cambridge University Press.

Vihman, M.M., Macken, M.A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A reassessment of the continuity issue. *Language, 61*(2), 397-445.

Wilson, S.M., Saygin, A.P., Sereno, M.I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701-702.

Zampini, M. L., & Green, K.P. (2001). The voicing contrast in English and Spanish: The relationship between perception and production. In J.I. Nicol (Ed.). *One mind, two languages: Bilingual language processing* (pp. 23-48). Malden, Massachusetts: Blackwell.