Perceptual paths to accurate production of L2 vowels: The role of individual differences*

WENDY BAKER AND PAVEL TROFIMOVICH

Abstract

This study investigated whether individual differences in learners' age of arrival (AOA) and length of residence (LOR) in a country where a second language (L2) is spoken determine the relationship between L2 perception and production. In the first experiment, 40 Korean learners of English and 10 native English speakers participated in vowel perception and production tasks in English. Results demonstrated that the relationship between perception and production depended on learners' AOA, not LOR. In the second experiment, the same Korean learners of English participated in a vowel perception task in which they judged the accuracy of their own production of English vowels. Results demonstrated that self-perception, or the ability to accurately hear one's own productions, may underlie, at least for some learners, accurate L2 production. Overall, results suggested that self-perception is an important factor in determining the perception-production link and indicated that self-perception may have implications for L2 speech learning.

1. Introduction

Does accurately *perceiving* second-language (L2) vowels and consonants ("sounds") help in accurately *producing* them? In other words, is there a direct relationship between perception and production when learning an L2? Most L2 speech learning theories are based on the assumption that these two skills are related and that accurate perception precedes or is necessary for accurate production. For example, several studies have documented that L2 learners cannot produce sounds accurately without first perceiving differences between them and similar L2 and/or native-language (L1) sounds (Flege 1995; Underbakke 1993). Likewise, perceptual training (teaching learners to identify or discriminate L2 sounds correctly) is often used to improve both L2 perception and production (Borden et al. 1983; Bradlow et al. 1997; Wang et al. 2003). Although the nature of the relationship between perception and production is still

IRAL 44 (2006), 231–250
DOI 10.1515/IRAL.2006.010
0019042X/2006/044-0231
©Walter de Gruyter

unknown and heavily debated (Fowler 1996; Lindblom 1996), support for the interdependence of perception and production is more positive than negative (Fox 1982). Such a link is not necessarily obvious, however, considering that the two skills – one involving motor control and the other auditory processing – may or may not be controlled by different mechanisms (Allport et al. 1987).

The relationship between perception and production is especially difficult to determine in adult L2 acquisition because, in comparison to L1 development in children, adult L2 acquisition is variable along many dimensions (Bialystok and Miller 1999; Flege et al. 1999). Among the most salient of these dimensions are such individual differences as learners' length of residence (hereafter "LOR") and their age of arrival (hereafter "AOA") in the target country (a country where an L2 is spoken). Indeed, these two factors have been shown to profoundly affect L2 learners' perception and production abilities (e.g., Baker and Trofimovich 2005). For example, child learners, or learners with an earlier AOA, tend to perceive and produce L2 sounds more accurately than do adult learners, or learners with a later AOA (see Flege 1999 for a review of age effects in L2 phonological learning). Moreover, learners with a longer LOR, or with more L2 experience, are more likely to accurately perceive and produce at least some L2 sounds than are learners with a shorter LOR, or less L2 experience (e.g., Flege et al. 1997). If, as these findings demonstrate, L2 perception and production depend on learners' AOA and LOR, it should be possible, then, by carefully controlling these two factors, to gain insights into the L2 perception-production relationship. Understanding the perception-production relationship and the effect of individual differences on this relationship is important for both theoretical and pedagogical reasons. Theoretically, knowing whether perception abilities influence production abilities may illuminate how second languages are acquired (Rochet 1995; Wang et al. 2003). For pedagogical reasons, such an understanding may help to determine what types of L2 training will be most effective in improving both L2 perception and production abilities (Bradlow et al. 1997).

In general, three hypotheses about the relationship between speech perception and production have been advanced. In both L1 and L2 acquisition studies, the most widely supported hypothesis is that accurate perception is at least one necessary component of accurate production (Flege 1995; Flege et al. 1997; Wode 1996), which translates into perception abilities usually surpassing, and therefore "preceding," production abilities, especially for beginning L2 learners (Flege 1988). Even advanced L2 learners, whose perception and production abilities are nearly asymptotic, perceive some vowels more accurately than they produce them (Flege et al. 1999).

The results of acquisition studies, however, are not conclusive. Other researchers have proposed that production and perception are interdependent and develop simultaneously (Best 1995; Fowler 1996). This hypothesis assumes

that perception entails the ability to visually detect distal articulatory properties of speech (i.e., tongue movement, vocal tract size, etc.) and maintains that perception and production are always aligned, so that perception never surpasses, and therefore never precedes, production and vice versa. The McGurk effect (McGurk and MacDonald 1978), the phenomenon that listeners use both visual (i.e., the perception of the *articulatory* gesture) and auditory (i.e., the perception of the *auditory* signal) information to recover speech, and the results of imitation studies (Ryalls and Pisoni 1997), which illustrate that children rely on imitation when learning to speak, support the hypothesis that perception and production abilities are interdependent (Meltzoff 2002) and develop out of using both visual and auditory cues to learn speech. Speech processing models based on this hypothesis predict that speech perception and production abilities in L2 learners develop in synchrony and depend on how difficult it is to recover slight articulatory differences between L1 and L2 sounds (Best 1995). In support of this hypothesis, for example, Flege et al. (1997) demonstrated that, at least for some L2 learners, perception and production abilities were aligned.

In contrast, yet another hypothesis states that accurate production precedes accurate perception, so that some speech contrasts are maintained in production before they are actually perceived. Many L1 studies of dialect change, for example, have proposed that changes first occur in production before they are sustained in perception (Labov et al. 1991; see, however, Bowie 2000). L2 studies often demonstrate that at least some learners are able to produce differences between L2 sounds that they cannot perceive (Flege et al. 1997; Zampini and Green 2001), such as Japanese learners of English who can produce a contrast between English /ɹ/ and /l/ without being able to perceive it (Sheldon and Strange 1982; Smith, 2001). In fact, more and more L1 and L2 studies are suggesting that production abilities can and often do develop before perception abilities (Kosky and Boothroyd 2003; Parsloe 1998; Smith 2001), although no study has demonstrated this to be the case for all language learners. Indeed, even within the same study, often differences in the perception-production relationship are noted for different learners (Beach et al. 2001; Zampini and Green 2001). Possible explanations for why production may precede perception include either that the two abilities develop independently or that production is necessary for accurate perception. This "production-first" hypothesis agrees with those conceptualizations of L1 development that suggest that there are two separate underlying representations in children's linguistic systems, known as the input and the output lexicons (Locke 1988; Schwartz and Leonard 1982) and that L1 phonological development and L1 phonological disorders (Straight 1980) – and perhaps L2 phonological development as well – can be described as a gradual process of "alignment" of the phonological specifications for the output (production) and the input (perception) lexicons.

Conflicting results such as these suggest that the perception-production relationship may differ as a result of individual differences such as learners' AOA and LOR. That is, different studies may yield different results because they examine the perception-production link for learners of different ages and at different stages of learning. For example, Sheldon and Strange (1982), who examined intermediate-level Japanese learners of English, found that some learners were more accurate in their production than in their perception of English /ɹ/ and /l/. By contrast, Flege et al. (1999) showed that some advanced Italian learners of English (those who had lived in an English-speaking country for more than 18 years) perceived English vowels as accurately as they produced them. Perhaps observing the *development* of perception and production abilities (i.e., testing learners with varying AOAs and LORs) may be essential in differentiating among these three hypotheses. Thus, the objectives of the present study were (1) to determine whether and how the perception-production relationship depends on learners' AOA and LOR, and (2) to explore what other factors may influence or explain the perception-production link.

To explore the development of L2 speech perception and production abilities, 40 Korean learners of English were asked to perform perception and production tasks of English vowels. In Experiment 1, the influence of the learners' AOA and LOR on the perception-production relationship was evaluated to understand the developmental component of speech learning. This experiment also sought to determine whether the learners' production accuracy was a better predictor of their perception accuracy than was AOA or LOR. In Experiment 2, the influence of the learners' self-perception ability on their perception and production was examined to gain insights into the role of other individual differences in the perception-production relationship.

2. Experiment 1: Relationship between perception and production

The purpose of Experiment 1 was to determine whether L2 learners' AOA and LOR in the target country (a country where the learners' L2 is spoken) influence the L2 perception-production relationship.

2.1. *Participants*

The participants were 40 Korean learners of English and 10 native English speakers. Prior to testing, all participants reported that their hearing was normal. All were students at a U.S. university at the time of testing. The learners differed in their LOR in the U.S. (less than 1 year, 3 years, 10 years) and in their AOA in the U.S. (early, late). Based on AOA and LOR, the learners were assigned to four groups, with 10 participants in each group. The learners in the first three groups were adult (late) learners of English. The learners in these three groups differed in their LOR and were included in the study to exam-

Table 1. *Characteristics of participants in Experiments 1 and 2*

Group	AOA[a]	LOR[b]	Chron. Age[c]	Use of Korean[d]
Late+0	29.0 (2.7)	0.2 (0.1)	29.3 (2.7)	74 (22.2)
Late+3	24.0 (3.7)	3.0 (0.4)	27.3 (3.5)	63 (20.0)
Late+10	21.4 (2.8)	9.8 (2.3)	32.1 (2.7)	44 (22.2)
Early	9.0 (1.2)	11.1 (1.9)	20.3 (1.6)	19 (20.3)
NE			25.6 (4.5)	

a Age of arrival in the U.S., in years
b Length of U.S. residence, in years
c Chronological age, in years
d Percent of daily Korean use. Standard deviations appear in parentheses.

ine the influence of this factor on the perception-production relationship. The learners in the first of these groups, the "Late+0" group (where "+0" indicates less than 1 year of U.S. residence), had arrived in the U.S. at a mean age of 29, had resided in the U.S. for less than 3 months on average, and were on average 29 years old at the time of testing. They were not proficient in English and represented beginning-level learners. The learners in the second group, the "Late+3" group (where "+3" indicates about 3 years of U.S. residence), had arrived in the U.S. at a mean age of 24, had resided in the U.S. for an average of 3 years, and were on average 27 years old at the time of testing. These learners were intermediate-level learners. The learners in the third group, the "Late+10" group (where "+10" indicates about 10 years of U.S. residence), had arrived in the U.S. at a mean age of 21, had resided in the U.S. for an average of 10 years, and were on average 32 years old at the time of testing. They were advanced-level learners.

The learners in the fourth group, the "Early" group, were child (early) learners of English. They had arrived in the U.S. as children, at about 9 years of age, had resided in the U.S. for a mean of 11 years, and were on average 20 years old at the time of testing. This group differed from the Late+10 group only in their age of arrival (and chronological age) and was included in this study to examine the influence of AOA on the perception-production relationship. In addition, 10 native English speakers, the "NE" group, who were all residents of Illinois at the time of testing, participated as a comparison group (Table 1).

2.2. *Materials*

The materials used in this study included English vowels which are particularly difficult for Korean learners of English to both perceive and produce (Flege et al., 1997) – /i/, /ɪ/, /u/, /ʏ/, /θ/, and /ɛ/. To reduce coarticulation effects, these vowels were placed in 18 monosyllabic words with the English bilabial plosive

Table 2. *Word stimuli*

English vowel					
/i/	/ɪ/	/u/	/ʏ/	/æ/	/ɛ/
beat	bit	boot	book	bat	bet
bead	bid	booed	good	bad	bed
heed	hid	who'd	hood	had	head

/b/ (with one exception, *good*) and the glottal fricative /h/ as syllable onsets (Table 2), because these two consonants exert the least amount of coarticulatory effects on following vowels (Olive, Green, and Coleman 1993). The materials were recorded using a Shure unidimensional head-mounted microphone and Sony TCD-D8 DAT recorder.

For the speech perception task, the materials were produced by three male native English speakers (ages: 21-26) who had no or minimal experience with a foreign language. Each speaker produced six words from the larger set of 18 words, namely, *had, head, hid, heed, hood, who'd*, in a carrier phrase "I say _____ for you". Each word contained one of the six vowels examined in this study. These target words were elicited from three different speakers for two reasons. First, using multiple speakers to record tokens for a perception task allows for the creation of a task that includes a large number of unique tokens (as opposed to a small number of repeated one-speaker tokens). Second, a perception task that includes tokens produced by multiple speakers is likely to yield a speaker-independent measure of perception ability, one that does not reflect a learner's ability to perceive tokens spoken by one particular L2 speaker (Flege et al. 1999).

For the speech production task, the materials were produced by a female native English speaker (age: 31) who had minimal experience with a foreign language. A different speaker was used to elicit words for the production task in order to minimize the effect of learners' familiarity with the speaker's voice. The female speaker's voice used to record the word tokens for the production task was thus deemed to be maximally dissimilar to the voices of the three male speakers who recorded the stimuli for the perception task. The female speaker recorded the entire set of 18 words. All recordings were digitized at 16 kHz, normalized for peak intensity, and ramped off during the first and last 15 ms to prevent audible clicks. Prior to the experiment, 10 native speakers of English identified the stimuli with a 98 % accuracy in an open-choice identification task.

2.3. *Procedure*

Perception task. To obtain a perception measure, the participants were asked to perform an open-choice identification task. In this task, the participants each heard the six English words (*had, head, hid, heed, hood, who'd*) presented over the headphones one at a time and matched each word, by clicking a button on the computer screen, with one of six response alternatives labeled orthographically as the English words *had, head, hid, heed, hood, who'd,* respectively. The stimuli on each trial were played one at a time, with each new trial presented 1 sec following the participant's response to a previous trial. No feedback and no practice session were given. The participants could listen to each token as many times as they wished, although they could not listen to more than one token at a time. This was done so that they could not contrastively compare different tokens. Allowing participants to listen to the tokens multiple times ensured that their responses were not based on guesses but instead reflected their phonological encoding of the vowel in question. The six words recorded by the three speakers were presented in three randomized lists for a total of 54 one-stimulus trials per participant. The perception score for each participant was derived by averaging correct identification rates across the 54 tokens of the six words.

Production task. To explore the relationship between the learners' ability to perceive and produce the same vowels, the same participants were asked to perform a word-production task. In this task, the participants were asked to name black-and-white line drawings, whose names (or descriptions) contained the six vowels examined in this study (Table 2). For example, the word *bat* was used to name the drawing depicting a baseball bat. Because some drawings did not unambiguously depict the objects to be named, the participants were familiarized with the intended words during the study phase. In the study phase, the participants twice viewed each drawing and repeated the intended word recorded by the female native English speaker. The intended word was presented first in a sentence, which provided the necessary context, and then in citation form. These sentences (e.g., *She had a good day* describing an image of a smiling girl) provided context to help participants remember the more abstract words such as *had* and *who'd.* The participants had no less difficulty remembering these words than the more concrete words with unambiguous labels (e.g., *bat*). In the test phase, which contained two blocks of 18 randomized trials, the participants attempted to spontaneously name the drawings by labeling them with the appropriate words. When no response was given, the expected word was played over a loudspeaker and the participant repeated it. Only spontaneous productions from the test phase were used in subsequent analyses.

The version of a picture-naming task used in this study, one that included an auditory model for the learners to repeat, has been successfully used to elicit production data in previous studies of L2 phonological learning (e.g., Tsukada et al. 2004). This task allows for eliciting fluent speech while avoiding reading (a potentially confounding factor) as part of the task. This task was deemed appropriate for this study because it familiarized the participants (in both the study phase of the task and during the first block of trials) with the words to be used in describing the pictures and thus ensured that the participants produced identical and therefore maximally comparable speech samples. Because only spontaneous productions from the test phase were used in all subsequent analyses, the likelihood of the participants' direct mimicry of the auditory models played was minimized.

The production data were subsequently presented to another group of native English speakers to obtain word-production ratings. The 900 recorded word tokens (18 words $\times$ 50 participants) were randomized and re-recorded onto a high-quality audio tape, with each stimulus presented with a 4-second interval. The audio tape was played to a group of 10 native English listeners from Illinois for identification (age range: 18–27). In this task, the listeners were instructed to choose, on an answer sheet, one of four response alternatives created for each of the 18 word tokens. For example, for the word token *bit*, the response alternatives were: *bit*, *beat*, *bead*, or *bid*. (The response alternatives used for each word can be found in the Appendix.) The production score was derived by averaging the 10 listeners' correct identification rates for each participant's 18 words.

In each case but one (i.e., *goed*), real English words were used as response alternatives to make the listeners' task as natural as possible (see Appendix). For most response alternatives, it was possible to find words that ended in either a voiced or a voiceless alveolar plosive (i.e., /d/, /t/). A small number of other response alternatives were real words that either ended in other consonants (e.g., *book*) or did not end in a consonant at all (e.g., *who*). Because the listeners were instructed to choose the response alternative that contained the same vowel as that spoken by the participant, the response alternatives were assumed to have provided appropriate responses.

Although acoustic analyses of the participants' production of English vowels (e.g., those documenting differences in vowel formant frequencies) might have provided a more fine-grained measure of the participants' production accuracy (Baker and Trofimovich 2005), the listener judgments obtained here were deemed appropriate as a global measure of production accuracy. This is because accurate production of L2 vowels likely depends on a number of articulatory and acoustic dimensions (e.g., formant frequency variations, duration, lip rounding), which would be difficult to examine in separate acoustic analyses. Listener judgments of accuracy are reflective of these dimensions (e.g.,

Flege et al. 1999) and likely represent consequences of speech comprehension typical of interactions between native and non-native speakers (Bradlow and Pisoni 1999).

2.4. *Results and discussion*

Effect of AOA and LOR. The perception and production scores for each participant in the four learner groups and the native English speakers were submitted to two separate one-way analyses of variance (ANOVAs). These analyses yielded a significant group effect (defined as a combination of AOA and LOR variables) in perception, $F(4,45) = 31.30$, $p < .0001$, and in production, $F(4,45) = 46.03$, $p < .0001$. Tukey HSD post-hoc tests (in these and subsequent Tukey tests, between-group differences were deemed significant at $p \leq .05$) revealed that (1) the Early and the NE groups' perception scores and their production scores did not differ from each other, (2) the Late+0, Late+3, and Late+10 groups all had significantly lower perception and production scores than the NE and the Early groups did, and (3) among the three late learner groups (Late+0, Late+3, Late+10), the Late+10 group had a significantly higher production score than the Late+0 group did (Table 3). In other words, the early learners perceived and produced English vowels similarly to the native English speakers. The late learners with 3 and 10 years of U.S. residence did not perceive English vowels significantly more accurately than did the learners with less than 1 year of U.S. residence. Only the late learners with 10 years of U.S. residence produced English vowels significantly more accurately than did the late learners with less than 1 year of U.S. residence. These findings suggested that learners' AOA, more so than their LOR, predicted the learners' perception and production performance.

To explore within-group differences between perception and production, separate paired-samples *t*-tests were performed within each group on perception and production data. Surprisingly, however, perception and production abilities were not statistically significantly different within any group, $ts(9) < 1.37$, $ps > .20$, indicating that each group both perceived and produced L2 vowels with comparable accuracy. These results suggested that perception does not

Table 3. *Percent correct and standard deviations (in parentheses) for perception and production in Experiment 1*

	Group				
	NE	Early	Late+10	Late+3	Late+0
Perception	94.3 (4.6)	95.2 (5.6)	59.8 (10.7)	63.0 (18.5)	50.4 (13.5)
Production	96.4 (2.0)	91.9 (7.2)	64.0 (10.7)	57.7 (14.5)	50.4 (9.5)

Table 4. *Correlations between perception, production scores and participant variables*

	AOA[a]	LOR[b]	Use of Korean[c]
Perception	−.69**	.41	−.53**
Production	−.82**	.57**	−.60**

a Age of arrival in the U.S., in years
b Length of U.S. residence, in years
c Percent of daily Korean use
** $p < .001$

precede production at the stages of L2 learning examined in this study, and that these two perceptuomotor skills develop simultaneously, supporting the second hypothesis discussed above.

To further explore the perception-production link, zero-order correlations for the 40 L2 learners were computed between (1) production and perception scores and (2) participant variables (AOA, LOR, percent of daily use of Korean) and production and perception scores separately. The first analysis yielded a strong relationship between perception and production, $r(38) = .73$, $p < .001$, two-tailed. The second analysis (see Table 4) demonstrated that AOA correlated more strongly with production than with perception accuracy, although both correlations were strong and significant. The second analysis also demonstrated that LOR significantly correlated with production and that percent of daily use of Korean significantly correlated with both perception and production. Taken together, these findings indicated that there was a complex relationship between the learners' perception and production abilities on the one hand, and such factors as AOA, LOR, and amount of daily Korean use on the other. Speculative and perhaps premature, these conclusions merely confirmed that perception and production abilities are related to each other and to a number of individual-difference factors. Clearly, these analyses did not indicate whether one ability precedes the other.

Individual differences in production. The obtained group data and correlations may not have provided accurate insights into how perception and production abilities are related, nor how such a relationship develops as an L2 is learned. It is possible that our failure to detect oft-cited within-group differences between perception and production in bilinguals and L2 learners (e.g., Sheldon and Strange 1982) at the stages of L2 development examined in this study was an artifact of the grouping variables used in the analyses – AOA and LOR. For example, it may be that length of U.S. residence is not an adequate measure of amount of L2 experience, especially for late L2 learners (Flege and Liu 2001). Because the goal of this study was to examine whether the

perception-production relationship differs across several stages of L2 learning, the data were re-analyzed by grouping the participants according to one of the two skills – the learners' production ability. It was hypothesized that doing so could provide a better criterion for describing between-participant individual differences.

The following data re-analysis involved three learner groups (Early, Late+3, Late+10), that is, the groups of learners whose perception and production abilities were unlikely to be equally poor (as was the case for the learners in the Late+0 group) or equally good (as was the case for the learners in the NE group). Using these 30 learners' production accuracy as the grouping variable (instead of AOA or LOR), data re-analysis yielded 3 new participant groups: those with good (85-100%; $n = 8$; Early: 8), intermediate (70–85 %; $n = 8$; Late+3: 3, Late+10: 3, Early: 2), and poor (less than 70 %; $n = 14$; Late+3: 7, Late+10: 7) production accuracy. These three groups were then compared to the NE group, which represented the "end" state of English learning, and to the Late+0 group, which ostensibly represented the "initial" state of English learning because all participants in this group were functionally monolingual Korean speakers.

In close agreement with perceptually based accounts of the development of L2 perception and production (Flege 1995), it was hypothesized that good production would be associated with good perception, and poor production with poor perception. For intermediate learners, it was hypothesized that there may be a possible misalignment of both skills (a characteristic of an intermediate learning stage) if accurate perception precedes accurate production.

The perception and production scores were examined in two one-way ANOVAs (one comparing perception and the other comparing production data) to determine whether the newly formed groups (Good, Intermediate, Poor), the Late+0 group, and the NE group differed in these two skills and whether using this new grouping would provide more insights into the perception-production relationship. These analyses yielded a significant group effect for both perception, $F(4,45) = 23.34$, $p < .0001$, and production, $F(4,45) = 97.90$, $p < .0001$. Tukey HSD post-hoc tests revealed that (1) the learners in the good production-skill group did not differ from those in the NE group in either perception or production, (2) the learners in the poor production-skill group did not differ from those in the Late+0 group in either perception or production, and (3) the three production-skill groups (Good, Intermediate, Poor) differed from one another in both perception and production in all cases but one, that is, in the comparison of perception between the intermediate and poor production-skill groups (Table 5). However, no differences emerged when perception and production abilities were compared using paired-samples *t*-tests within each of the three production-skill groups: Good, $t(7) = .42$, $p > .05$, Intermediate, $t(7) = .91$, $p > .05$, and Poor, $t(13) = 2.14$, $p > .05$, suggesting that perception and pro-

Table 5. *Percent correct and standard deviations (in parentheses) for perception and production in Experiment 1*

	Group				
	NE	Good	Intermediate	Poor	Late+0
Perception	94.3 (4.6)	95.8 (5.8)	69.7 (23.7)	61.1 (11.3)	50.4 (13.5)
Production	96.4 (2.0)	94.9 (3.3)	76.7 (5.4)	54.5 (9.1)	50.4 (9.5)

duction abilities were aligned at each level of production mastery, again supporting the second hypothesis.

Surprisingly, however, the learners whose production accuracy was intermediate (75–80 %) displayed an exceptionally large amount of variability in their perception scores ($M = 69.7$, $SD = 23.7$, see Table 5). That is, a close examination of the individual perception and production scores in the intermediate group revealed that perception was not aligned with production. As Figure 1 depicts, for about half of the learners in this group (designated on the X-axis as *perc* > *prod*), perception accuracy surpassed production accuracy, being identical to the English speakers' perception accuracy. Assuming that there is a perceptual basis for accurate production, these learners' production accuracy may eventually align with their perception ability and the English speakers' production accuracy. However, the other half of the learners in this group (designated on the X-axis as *prod* > *perc*) attained better accuracy in production than in perception (Sheldon and Strange 1982; Zampini and Green 2001). It may be the case that either production precedes perception for these learners, thus lending support to the third hypothesis entertained earlier, or perhaps factors other than perception underlie production accuracy in these individuals.

One such factor might be the ability to perceive sound contrasts in one's own production, or what is often called internal perception, self-hearing, or self-perception (Panagos and King 1975; Sheldon and Strange 1982; Shuster 1998). Self-perception, or the ability to perceive one's own speech, is often distinguished from other-perception, or the perception of others' speech. There is evidence in L1 acquisition that in order to learn to speak, infants may attempt to match the acoustics of their own productions to the acoustics of others' speech, thereby establishing accurate perceptual and articulatory representations for L1 sounds (Meltzoff 2002; Vihman 1991). Similar evidence is marshaled in research on language-impaired children who may have an overall superior ability to understand others but are unable to understand their own speech (Panagos and King 1975), a reflection of what Van Riper and Irwin (1958) termed a "non-functioning or inefficient intra-personal auditory circuit" in children with phonological disorders (1958: 113). Findings such as these strongly implicate

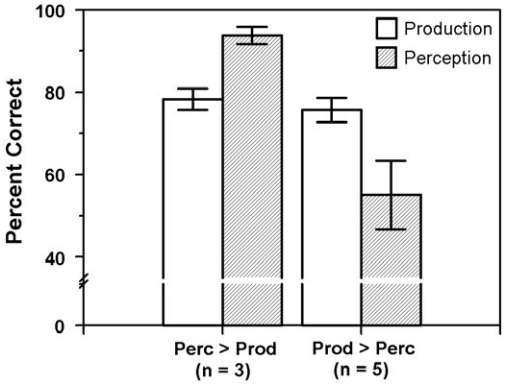

Figure 1. *Percent correct in perception and production for learners with intermediate production accuracy. Brackets enclose 1 SD.*

self-perception as a necessary and perhaps fundamental element in L1 production ability.

Extended to L2 acquisition, a similar claim may offer a feasible explanation for the somewhat surprising finding of this study. It may be that self-perception provides the perceptual basis for superior production abilities to those participants whose production abilities exceed perception abilities, allowing for the needed mapping of auditory targets onto accurate articulatory targets. Indeed, other researchers have found that self-perception may be a more accurate indicator of perceptual abilities than is other-perception (Sheldon and Strange 1982). More importantly, however, from a theoretical perspective, such an explanation would accentuate perceptual bases of accurate production, much in line with most research in the field (Flege 1995).

Beyond a doubt, however, self-perception in the absence of the ability to accurately perceive others may not in itself translate into accurate production. Perhaps what self-perception, as part of an auditory-articulatory feedback mechanism, underlies is the ability to *compare* one's own attempts at speech with others' productions. Essential to both L1 and L2 learning, self-perception cast in such terms would entail the ability to perceive differences between one's own speech and the auditory targets in others' speech, thus allowing learners to modify their speech to match these targets. Testing this hypothesis was the purpose of Experiment 2.

3. Experiment 2: Self-perception

The purpose of this experiment was to determine whether accurate production is contingent not just on accurate perception of others' speech (other-

perception), but also on accurate perception of one's own speech (self-perception, or the perception of self-produced speech). Two possibilities about the relationship among these skills were proposed. The first possibility was that, in some learners, other-perception should be better than, or equal to, self-perception. For example, earlier research has documented that some L2 learners (Sakow and McNutt 1993) and L1 children and adolescents with phonological disorders (Shuster 1998) are better at accurately perceiving others' speech than they are at perceiving their own. Other learners, however, particularly those whose production is either very good or very poor, may be equal in their ability (or inability) to perceive themselves and to perceive others. For instance, learners with good production skills will be equally good at other-perception and self-perception because they are able to match their own productions with the auditory targets in others' speech. Likewise, learners with poor production skills will also be poor at other-perception and self-perception because they are unable to match their own productions with the auditory targets in others' speech. The second possibility was that some other learners – especially those whose production accuracy outstrips perception accuracy – should perform better at self-perception than other-perception and perhaps even production because, as hypothesized earlier, the ability to perceive one's own speech in relation to the auditory targets in others' speech may be essential in developing accurate production skills. That is, even those learners who are apparently better at production than perception, may still have a perceptual basis for their production abilities.

3.1. *Procedure*

To test these hypotheses, the same learners were asked to return (after about 3 months) and perform a word identification task on their own speech to obtain a measure of self-perception. The stimuli and procedure in this experiment were identical to those used in the production-rating task in Experiment 1. In this case, the learners heard the productions of 18 English words by all 50 participants (including themselves) and were asked, as were the English listeners in Experiment 1, to choose one of four possible response alternatives for each of the 18 words (see Appendix). The learners' responses were deemed "correct" if they marked the same response to their speech as was marked by the English listeners in Experiment 1 (whether it was the intended word or not). Self-perception scores were available for 14 learners, with 6, 4, and 4 learners in the good, intermediate, and poor production groups, respectively.

3.2. *Results and discussion*

The Wilcoxon matched-pairs signed-rank test, a nonparametric test for analyzing data from two related samples, was used to compare the learners' per-

Table 6. *Percent correct and standard deviations (in parentheses) for perception, production, and self-perception in Experiment 2*

	Group		
	Good	Intermediate	Poor
Perception	95.1 (6.7)	80.1 (14.8)	59.7 (9.2)
Production	94.1 (3.4)	74.7 (3.5)	55.3 (4.7)
Self-Perception	93.5 (6.5)	83.3 (16.4)	56.9 (12.3)

ception and self-perception scores within the groups of learners with good and poor production skills, yielding the following results. Representing asymptotic performance, the learners whose productions skills were good were equally good at other-perception and self-perception, $Z = .67$, $n = 6$, $p = .50$. Furthermore, the learners whose production skills were poor were equally poor at other-perception and self-perception, $Z = 1.10$, $n = 4$, $p = .27$, performing at chance level and therefore marking the beginning of L2 development. Overall, in both of these stages of L2 learning (beginning and advanced), other-perception and self-perception were completely aligned with production, conforming to the first possibility about the relationship among production, self-perception, and other-perception skills discussed above (see Table 6).

The role of self-perception in the developing relationship between perception and production was, however, most salient in the scores of learners with intermediate production skills. Of the four learners with intermediate production skills who returned to perform this task, two performed better at other-perception (93 % correct) than self-perception (81 % correct) and production (75 % correct). Apparently, for these learners, accurate production skills were being "guided" by their ability to accurately perceive others' speech. The remaining two learners with intermediate production accuracy performed better at production (74 % correct) than at other-perception (68 % correct). As predicted, these learners' self-perception accuracy (86 % correct) appeared to exceed both their production accuracy and their accuracy in perceiving others' speech. Conceivably, for these two learners, accurate production skills were being "guided" by their ability to accurately perceive their own speech. Implicit in these results, which conformed, albeit inconclusively (as no statistical analyses were possible due to small cell sizes), to the second possibility about the relationship among production, self-perception, and other-perception, is the claim that self-perception was perhaps important in the development of L2 production abilities.

4. General discussion

Of the three conceptualizations of the perception-production relationship, the results of this study overall supported the claim that accurate L2 production is contingent on accurate L2 perception (Flege 1995). However, as results of Experiment 1 demonstrated, the perception-production relationship may differ depending on learners' LOR and AOA, two individual-difference factors of which the first typically indexes the amount of L2 experience and the second characterizes learners' age at the time L2 learning begins. First, the perception-production relationship appears to depend on learners' LOR in the target country. That is, perception and production may be aligned at initial and more advanced stages of L2 learning. However, in the intermediate stages of L2 learning (where presumably most of the learning occurs), perception and production skills are misaligned, suggesting that at least some conflicting findings in examinations of the perception-production link may be due to variations in learners' amount and type of L2 experience (Flege et al. 1999; Sheldon and Strange 1982; Smith 2001). Second, the perception-production relationship also appears to depend on learners' AOA in the target country. That is, those learners who were exposed to their L2 in childhood were the only learners who both perceived and produced L2 sounds with native-like accuracy, suggesting that the perception-production relationship develops differently in child and adult L2 learning and that children's ability to translate accurate perception into accurate production (and perhaps vice versa) may give children a potential advantage over adults (Flege et al. 1999).

More interestingly, the results of this study suggested that the ability to accurately perceive one's own speech (self-perception) is important and perhaps even essential in the development of accurate L2 production. In particular, self-perception may enable learners to gradually attune their production to their perception. As results of Experiment 2 demonstrated, those intermediate learners whose production accuracy surpassed perception accuracy (a finding which may otherwise seem surprising) were good self-perceivers and were, in fact, better at self-perception than production. Being able to perceive their own speech accurately enabled these learners to produce sounds accurately, suggesting that accurate production may nevertheless be perceptually based (Newman 1998). Conversely, other intermediate learners were better at other- and self-perception than at production and perhaps belonged in the developmental stage in which their self-perception ability was improving, enabling them, with more experience, to attain native-like L2 accuracy. Certainly, further research investigating the role of self-perception in L2 speech learning may provide insights into the relationship between ultimate attainment and learners' self-perception ability and, eventually, between perception and production. In any case, self-perception – conceptualized as both an auditory and articula-

tory feedback mechanism – may be viewed as a necessary link between speech perception and speech production. That is, without self-perception, which may occur at an auditory (Flege 1995) or articulatory (Best 1995) processing level, L2 learners may not be able to link auditory perceptual targets with their own articulatory efforts.

Within a broader context of language processing and learning, self-perception, or learners' ability to accurately perceive their own speech, may be related to individual differences in learners' phonological short-term memory capacity and, by extension, to learners' ability to store and learn phonological representations of words. Captured within the phonological loop component of Baddeley's model of working memory (Baddeley and Hitch 1974; see Baddeley, Gathercole, and Papagno 1998 for review), this view of individual differences in phonological short-term memory suggests that, in spoken language processing, incoming sound sequences are first stored in a short-term phonological memory store (phonological loop) and that those sequences that are kept "active" in it (e.g., through the process of subvocal rehearsal) are subsequently transferred for long-term storage.

Certainly, learner differences in phonological short-term memory capacity may explain why self-perception is an essential element in learning L2 sounds and, more importantly, why learners differ in their self-perception ability. It is possible, then, that learners with a larger phonological short-term memory capacity may be better able than learners with a smaller phonological short-term memory capacity to hold in their short-term memory an L2 sound sequence and to match it with their own articulation of that sequence, making required adjustments to their own production (cf. Masoura and Gathercole 1999; Papagno et al. 1991). That is, producing an L2 sound sequence accurately (and knowing when one does or does not do so) may be requisite for creating accurate long-term memory representations for L2 sounds. Learners' phonological short-term memory capacity may, in turn, determine the rate at which such representations are created.

Although suggestive, the findings of this study are somewhat speculative at this time and certainly need to be replicated with a larger number of L2 learners. This notwithstanding, the findings point to an important relationship between L2 perception and production and suggest that a number of individual differences (LOR, as a measure of L2 experience; AOA, as an index of learners' age at the time L2 learning begins; and, notably, learners' ability to accurately perceive their own speech) may underlie the development of L2 perception and production. Overall, the findings of this study suggest that there are L2 perception-production links and indicate that self-perception may exemplify one such link.

Brigham Young University, <wendy_baker@byu.edu>
Concordia University, Montréal, <pavel@education.concordia.ca>

Appendix

Table 7. *Response alternatives used in the production-rating tasks in Experiment 1 and Experiment 2*

/i/		/ɪ/		/u/	
beat:	beat, bit, bid, bead	*bit*:	bit, bid, bead, beat	*boot*:	boot, booed, bud, boo
bead:	beat, bead, bid, bit	*bid*:	bit, bid, bead, beat	*booed*:	boot, booed, bud, boo
heed:	heat, heed, hit, hid	*hid*:	heat, heed, hit, hid	*who'd*:	who'd, hood, hoot, who

/ʏ/		/æ/		/ɛ/	
book:	book, boot, booed, buck	*bat*:	bat, bad, bed, bet	*bet*:	bet, bed, bad, bat
good:	good, gooed, goed, goat	*bad*:	bad, bat, bed, bet	*bed*:	bed, bet, bad, bat
hood:	hood, who'd, hoot, who	*had*:	had, hat, head, heed	*head*:	head, had, hat, heed

Note

∗ This research was partially supported by University of Illinois at Urbana-Champaign research grants to both authors. The authors extend their sincerest gratitude to James E. Flege and Molly Mack for their support and advice throughout this research project, to Randall Halter for his invaluable statistical advice, and to Randall Halter, Elizabeth Gatbonton, and two anonymous reviewers for their comments on earlier drafts of this paper.

References

Allport, Alan, MacKay, Donald G., Prinz, Wolfgang, and Eckart Scheerer (1987). *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing.* London: Academic Press.

Baddeley, Alan and Graham J. Hitch (1974). Working memory. In *The Psychology of Learning and Motication,* G. Bower (ed.), 47–89. New York: Academic Press.

Baddeley, Alan, Gathercole, Susan, and Costanza Papagno (1998). The phonological loop as a language learning device. *Psychological Review* 105 (1): 158–173.

Beach, Elizabeth. F., Burnham, Denis and Christine Kitamura (2001). Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. *The International Journal of Bilingualism* 5 (2): 221–235.

Best, Catherine. T. (1995). A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Winifred Strange (ed.), 107–126. Baltimore: York Press.

Bialystok, Ellen and Barry Miller (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition* 2 (2): 127–145.

Borden, Gloria, Gerber, Adele, and Gary Milsark (1983). Production and perception of the /r/–/l/ contrast in Korean adults learning English. *Language Learning* 33 (4): 499–526.

Bowie, David (2001). Perception and production in a series of related mergers. In *LACUS Forum XXVII: Speaking and Comprehending*, Ruth Brend, Alan K. Melby and Arle R. Lommel (eds.), 297–306. Fullerton, CA: LACUS.

Bradlow, Ann R. and David B. Pisoni (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America* 106: 2074–2085.

Bradlow, Ann R., David. B. Pisoni, Reiko Akahane-Yamada, and Yohichi Tohkura (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101 (4): 2299–2310.

Flege, James E. (1999). Age of learning and second-language speech. In *Second Language Acquisition and the Critical Period Hypothesis*, David Birdsong (ed.), 101–132. Mahwah, N.J.: Lawrence Erlbaum Press.

— (1988). Factors affecting the degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 91: 370–389.

— (1995). Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* Winifred Strange (ed.), 233–272. Baltimore: York Press.

Flege, James E. and Serena Liu (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23: 527–552.

Flege, James E., Bohn, Ocke-Schwen and Sunyoung Jang (1997). Effects of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics* 25: 169–186.

Flege, James E., MacKay, Ian R. A. and Diane Meador (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America* 106: 2978–2987.

Fowler, Carol A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* 99: 1731–1740.

Fox, Robert A. (1982). Individual variation in the perception of vowels: Implications for a perception–production link. *Phonetica* 39: 1–22.

Kosky, Christine and Arthur Boothroyd (2003). Perception and production of sibilants by children with hearing loss: A training study. *The Volta Review* 103: 71–98.

Labov, William, Karen, Mark and Corey Miller (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3: 33–74.

Lindblom, Bjorn (1996). Role of articulation in speech perception: Clues from articulation. *Journal of the Acoustic Society of America* 99: 1683–1692.

Locke, John L. (1988). Variation in human biology and child phonology: A response to Goad and Ingram. *Journal of Child Language* 15: 663–668.

Masoura, Elvira V. and Susan E. Gathercole (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology* 34 (5–6): 383–388.

McGurk, Harry and John MacDonald (1978). Auditory-visual coordination in the first year of life. *International Journal of Behavioral Development* 1: 229–239.

Meltzoff, Andrew (2002). Elements of a developmental theory of imitation. In *The Imitative Mind: Development, Evolution, and Brain Bases,* Andrew Meltzoff and W. Prinz (eds.), 19–41. New York: Cambridge University Press.

Newman, Rachel S. (1998). Individual differences and the link between speech perception and speech production. Unpublished doctoral dissertation. State University of New York, Buffalo.

Olive, Joseph P., Greenwood, Alice and John Coleman (1993). *Acoustics of American English Speech: A Dynamic Approach.* New York: Springer Press.

Panagos, J. M. and Ruth R. King (1975). Self and mutual speech comprehension by deviant- and normal-speaking children. *Journal of Speech and Hearing Research* 18: 653–662.

Papagno, Costanza, Valentine, Tim and Alan Baddeley (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language* 30 (3): 331–347.

Parsloe, Rachael (1998). Use of the speech pattern audiometer and the electropalatograph to explore the speech production/perception relationship in a profoundly deaf child. *International Journal of Language and Communication Disorders* 33 (1): 109–121.

Rochet, Bernard L. (1995). Perception and production of second-language speech sounds by adults. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Winifred Strange (ed.), 379–410. Baltimore: York Press.

Ryalls, Brigette O. and David B. Pisoni (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology* 33 (3): 441–452.

Sakow, Margaret and James C. McNutt (1993). Perception of /r/ by native speakers of Japanese and Korean: Internal and external perception. *International Review of Applied Linguistics* 31 (1): 46–53.

Schwartz, Richard G. and Laurence B. Leonard (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language* 9: 319–336.

Sheldon, Amy and Winifred Strange (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.

Shuster, Linda I. (1998). Perception of incorrectly and correctly produced /r/. *Journal of Speech, Language, and Hearing Research* 41: 941–950.

Smith, Laura C. (2001). L2 acquisition of English liquids: Evidence for production independent from perception. In *The Past, Present, and Future of Second Language Research,* Xenia Bonch-Bruevich, William J. Crawford, John Hellermann, Christina Higgins, and Hanh Nguyen (eds.), 3–22. Somerville, MA: Cascadilla Press.

Straight, Stephen H. (1980). Structural commonalities between comprehension and production. *Revue de Phonetique Appliquee* 55–56: 313–316.

Tsukada, Kimiko, David Birdsong, Molly Mack, Hyekyung Sung, Ellen Bialystok and James E. Flege (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica* 61: 67–83.

Underbakke, Melva E. (1993). Hearing the difference: Improving Japanese students' pronunciation of a second language through listening. *Language Quarterly* 31: 67–89.

Van Riper, Charles and John V. Irwin (1958). Voice and Articulation. Oxford, England: Prentice-Hall.

Vihman, Marilyn M. (1991). Ontogeny of phonetic gestures: Speech production. In *Modularity and the Motor Theory of Speech Perception*. Ignatius Mattingly and Michael Studdert-Kennedy (eds.), 69–84. Hillsdale, NJ: Lawrence Erlbaum.

Wang, Yue, Allard Jongman, and Joan A. Sereno (2003). Phonetic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America* 113 (2): 1033–1043.

Wode, Henning (1996). Speech perception and L2 phonological acquisition. In *Investigating Second Language Acquisition*, Peter Jordens and Josine Lalleman (eds.), 321–353. Berlin: Mouton de Gruyter.

Zampini, Mary L. and Kerry P. Green (2001). The voicing contrast in English and Spanish: The relationship between perception and production. In *One Mind, Two Languages*, Janet Nicol (ed.), 23–48. Oxford: Blackwell.