Individual differences in second language speech perception across tasks and contrasts: The case of English vowel contrasts by Korean learners

Abstract

The present study examines whether individual differences in second language (L2) learners' perceptual cue weighting strategies reflect systematic abilities. We tested whether cue weights indicate proficiency in perception using a naturalistic discrimination task as well as whether cue weights are related across contrasts for individual learners. Twenty-four native Korean learners of English completed a two-alternative forced choice identification task on /ɪ/-/i/ and /ɛ/-/æ/ contrasts varying orthogonally in formant frequency and duration to determine their perceptual cue weights. They also completed a two-talker AX discrimination task on natural productions of the same vowels. In the cue-weighting task, we found that individual L2 learners varied greatly in the extent to which they relied on particular phonetic cues. However, individual learners' perceptual weighting strategies were consistent across contrasts. We also found that more native-like performance on this task – reliance on spectral differences over duration – was related to better recognition of naturally produced vowels in the discrimination task. Therefore, the present study confirms earlier reports that learners vary in the extent to which they rely on particular phonetic cues. Additionally, our results demonstrate that these individual differences reflect systematic cue use across contrasts as well as the ability to discriminate naturally produced stimuli.

Keywords

L2 speech perception, Individual differences, Identification task, Discrimination task, Vowel contrasts

Individual differences in second language speech perception across tasks and contrasts: The case of English vowel contrasts by Korean learners

1 Introduction

It is well known that second language (L2) learners have difficulties perceiving speech sound contrasts not present in their native language. Some documented difficulties include the English /l/-/ɹ/ contrast for Japanese learners (Aoyama et al. 2004; Iverson et al. 2003) and the English /ɪ/-/i/ contrast for Spanish (Escudero 2000, 2005; Morrison 2008) and Korean learners (Flege et al. 1997; Tsukada et al. 2005). This may be related to the observation that native listeners categorize speech sounds using multiple acoustic cues but that, for any given contrast, they pay more attention to certain cues over others; this is referred to as perceptual cue weighting (Dorman et al. 1977; Holt and Lotto 2006). Developmental studies of speech perception showed that native listeners learn relevant cue weighting strategies as they gain more native language experience (Nittrouer and Miller 1997). For example, English listeners attend to both formant frequency and vowel duration when categorizing vowel contrasts but they attend more to the former than the latter (Escudero 2000, 2005; Francis et al. 2008; Hillenbrand et al. 2000).

When acquiring a new contrast, part of the challenge for L2 learners is determining which cue should be weighted most strongly. Indeed, previous research has shown that perceptual difficulties like those identified above might be attributable to learners' paying attention to less relevant acoustic-phonetic information present in these speech sounds. For example, it has been reported that Japanese learners of English rely on second formant (F2) frequencies instead of relevant third formant (F3) frequencies to distinguish English /l/ and /ɹ/ (Iverson et al. 2003). Similarly, Spanish, Mandarin, and Korean learners of English attend to duration differences rather than spectral differences for English front vowel categorization (Flege et al. 1997). These difficulties observed in the use of phonetic cues by learners from different L2 backgrounds are likely to be overcome by successful learning of the relevant cue weights in the course of acquisition (Escudero et al. 2011; Iverson et al. 2005; Kondaurova and Francis 2010). However, research that has probed this issue has generally reported the challenges that learners experience at a group level while individual-level difficulties are less well studied. The overarching goal of the present study is to investigate L2 cue weighting strategies in terms of successful learning of non-native speech sound contrasts at an individual level.

It has been observed that native listeners exhibit a large amount of individual variability in cue weighting and that overall group patterns tend to mask important differences across individuals (e.g., Idemaru et al. 2012; Kong and Edwards 2011; Shultz et al. 2012). Individual differences in cue weights are especially evident, however, in L2 speech perception (Kong and Edwards 2015; Schertz et al. 2015; Wanrooij et al. 2013). For example, Schertz et al. (2015) found that some Korean learners distinguish English stop categories primarily by VOT while others use primarily f0. Furthermore, individual differences in cue weights are stable over time for both L1 and L2 listeners (Idemaru et al. 2012; Schertz et al. 2015). However, little is known about whether individual differences in cue weighting strategies are related across speech sound contrasts.

Moreover, there is a lack of evidence on whether individual differences in L2 cue weights are directly related to success in perceiving L2 contrasts. A number of studies have found some evidence that individual differences in sensitivity to sound contrasts predict success in L2 training studies (Chandrasekaran et al. 2010; Lengeris and Hazan 2010; Perrachione et al. 2011) and in identification ability after naturalistic learning (Hattori and Iverson 2009).

The present study aims to expand our understanding of individual differences in L2 speech perception by examining (1) how individual cue weights are related across two difficult front vowel contrasts for Korean learners of English: /ɪ/-/i/ and /ɛ/-/æ/ (Tsukada et al. 2005), and (2) whether learners' perceptual cue weighting strategies (attention to formant frequency vs. attention to duration) are linked to their ability to discriminate naturally produced words. We predict that cue weights are correlated across contrasts within an individual, that is, that learners who attend more to, for example, duration in one vowel contrast will do the same for the other contrast. Furthermore, we predict that learners who rely more on formant frequency cues to distinguish the vowel contrasts in a controlled cue weighting task – consistent with native listeners' strategies (Kondaurova and Francis 2008, 2010; Liu and Holt 2015) – will show better discrimination abilities with naturally varying vowels than those who rely more on duration. In other words, we will test whether individual cue weighting strategies predict degree of successful perception of L2 speech sound contrasts.

The participants in this study were part of a larger longitudinal study that included both child and adult learners. However, the effect of age on perceptual cue weighting is not the main research question in the present study, which focuses on individual variability rather than on group-level differences. Indeed, we do not predict that overall cue weights will differ significantly between adults and children in a controlled immersion environment although the rate of perceptual development may, of course, differ depending on age.

2 Methods

2.1 Participants

Twenty-four native Korean learners of English, 11 adults (KA, mean = 40.3 years, range = 36–46) and 13 children (KC, mean = 9.9 years, range = 7–13), participated in an identification task and a discrimination task after slightly over one year of immersion in Canada (mean length of residence = 14.3 months, range = 13–17). All participants were highly-motivated learners, who had come to Canada for the purpose of learning English, and none had lived in an English-speaking country prior to their arrival. In Canada, KA and KC learners received relatively comparable L2 input by taking full-time English language courses in language schools (mean = 22.7 hours a week, range = 15–40) or by attending English-medium schools (mean = 30.3 hours a week, range = 20–40), respectively. None of the participants reported speech or hearing impairments and all participants had little or no knowledge of second languages other than English. Ten native listeners of Canadian English (NE, mean = 23.6, range = 18–30) also participated in the experiment as controls. Participants were compensated for their participation in the experiment.

2.2 Stimuli

There were two tasks in this study, an identification task and a discrimination task, to be discussed in section 2.3. Stimuli in the identification task were created by resynthesizing speech continua based on natural recordings of the endpoints of both /bɪt/-/bit/ and /bɛt/-/bæt/ contrasts produced by a male native speaker of Canadian English in his 20s. Five spectral continua for each vowel contrast from *bit* to *beat* and from *bet* to *bat* were generated by *TANDEM-STRAIGHT* (Kawahara et al. 2009) and were fully crossed with 5 duration steps using the PSOLA method in *Praat* (ver. 5.3.55, Boersma and Weenick 2013), creating 25 tokens for each /ɪ/-/i/ and /ɛ/-/æ/ continuum, respectively. Table 1 provides the 5 steps of F1 and F2 values and the 5 steps of duration values for each continuum.

Table 1 F1, F2, and duration values in the /ɪ/-/i/ and /ɛ/-/æ/ continua for the identification task.

	/ɪ/-/i/			/ɛ/-/æ/		
Step	F1 (Hz)	F2 (Hz)	Duration (ms)	F1 (Hz)	F2 (Hz)	Duration (ms)
1	437	1938	70	654	1710	100
2	366	2202	110	670	1684	140
3	333	2323	150	697	1677	180
4	312	2409	190	756	1674	220
5	269	2463	230	850	1704	260

To test learners' ability to discriminate between L2 vowel contrasts, a set of 18 minimal pairs of monosyllabic English words (6 pairs for /ɪ/-/i/, 6 pairs for /ɛ/-/æ/, and 6 pairs of fillers) were recorded by two female native speakers of Canadian English in their 20s.[1] Only frequently used English words were selected to ensure that they are recognizable by all participants, as shown in Table 2.

[1] An anonymous reviewer pointed out that /æ/ is raised and sometimes diphthongized as a result of a sound change in some North American dialects of English. The raising and fronting of /æ/ are particularly associated with the Northern Cities Shift, but Canada is the region with the least raising of /æ/ (Labov et al. 2006). The two speakers who recorded the stimuli in the discrimination task are from Eastern Canadian regions and their /æ/ was not raised or diphthongized.

Table 2 Stimuli in the discrimination task.

/ɪ/-/i/	/ɛ/-/æ/	Fillers
sit-seat	*set-sat*	*fool-full*
dip-deep	*beg-bag*	*pool-pull*
chick-cheek	*dead-dad*	*Luke-look*
live-leave	*said-sad*	*bowl-ball*
chip-cheap	*pen-pan*	*coat-caught*
fit-feet	*bed-bad*	*low-law*

2.3 Procedure

2.3.1 Identification task

First, a two-alternative forced choice identification task was conducted to probe learners' cue weighting strategies. Pictures that represent the response words were used to avoid orthographic bias. Participants were instructed to press the left (←) or right (→) arrow on the laptop keyboard to identify the picture that corresponds to the word they heard. Twenty-five stimuli were repeated 5 times for each contrast, trials were blocked by contrast (i.e., 25 stimuli × 5 repetitions × 2 vowel contrasts = 250 trials), and all trials within a block were randomized.

2.3.2 Discrimination task

After completing the identification task, listeners participated in a two-talker AX discrimination task to test learners' ability to discriminate L2 vowels. In this task, participants heard natural recordings of isolated words produced by two different talkers and judged whether they heard "same" (e.g., $[\text{bɛd}]_1$-$[\text{bɛd}]_2$) or "different" (e.g., $[\text{bɛd}]_1$-$[\text{bæd}]_2$) stimuli. "Same" words in each pair indicated the same lexical item produced by two different talkers and "different" words in each pair indicated different lexical items produced by two different talkers. The inter-stimulus interval between the two words in each pair was 1,200 ms to prevent participants from discriminating the stimuli acoustically and to ensure phonological processing (Escudero et al. 2009; Werker and Logan 1985).

Participants were informed that they would hear two words in a pair spoken by two different talkers. They were instructed that they should click "same" on the computer screen if they heard the same word and "different" if they heard two different words; they were told not to attend to subtle acoustic differences reflecting indexical properties of the two talkers. Four stimuli (2 same: $[\text{bɛd}]_1$-$[\text{bɛd}]_2$, $[\text{bɛd}]_2$-$[\text{bɛd}]_1$ and 2 different: $[\text{bɛd}]_1$-$[\text{bæd}]_2$, $[\text{bæd}]_2$-$[\text{bɛd}]_1$) for each minimal pair were repeated 2 times in each block (i.e., 4 stimuli × 18 minimal pairs × 2 repetitions = 144 trials). Participants were permitted to take a short break after the first block (i.e., 72 trials), and all trials within a block were randomized.

3 Results

3.1 Perceptual cue weighting

To examine the extent to which each acoustic dimension (i.e., vowel formants vs. duration) contributes to learners' vowel categorization, the participants' responses in the identification task were first analyzed with a mixed-effects logistic regression model, using the *glmer()* function from the *lme4* package (ver.1.1-11) in *R* (R Development Core Team 2008). The model included fixed effects for GROUP (adults vs. children), CONTRAST (/ɪ/-/i/ vs. /ɛ/-/æ/), SPECTRUM (spectral steps), DURATION (duration steps), as well as interactions between CONTRAST and SPECTRUM and between CONTRAST and DURATION. All predictors were standardized prior to analysis to reduce collinearity and make the intercept interpretable as an overall mean. Categorical variables, namely GROUP and CONTRAST, were centered (–0.5 and 0.5), and continuous variables, namely SPECTRUM and DURATION, were standardized by centering and dividing by 2 standard deviations. By-participant random intercepts and random slopes for each participant were also included in the model. To measure individual differences in L2 learners' perceptual cue weighting strategies, we built a series of separate logistic regression analyses fitted to each learner's vowel categorization responses to obtain individual learners' perceptual weights (β) for each acoustic dimension (Morrison 2005; Morrison and Kondaurova 2009). The beta-coefficients from the individual models served as measures of the perceptual weights, and were used to test whether individuals' cue weighting strategies can predict their discrimination abilities (see section 3.2).[2]

Figure 1 shows responses to each acoustic dimension for each contrast, as well as the overall responses to spectral and duration cues averaged across the vowel contrasts in order to focus on the use of each acoustic dimension by L2 learners. Overall, the cue weighting strategies show that the learners used both acoustic dimensions relatively similarly in their categorization of English vowels. However, these strategies differ from those of native English listeners in that native listeners predominantly use spectral quality while duration has a much weaker effect on their vowel categorization, consistent with previous studies (Kondaurova and Francis 2008, 2010; Liu and Holt 2015).

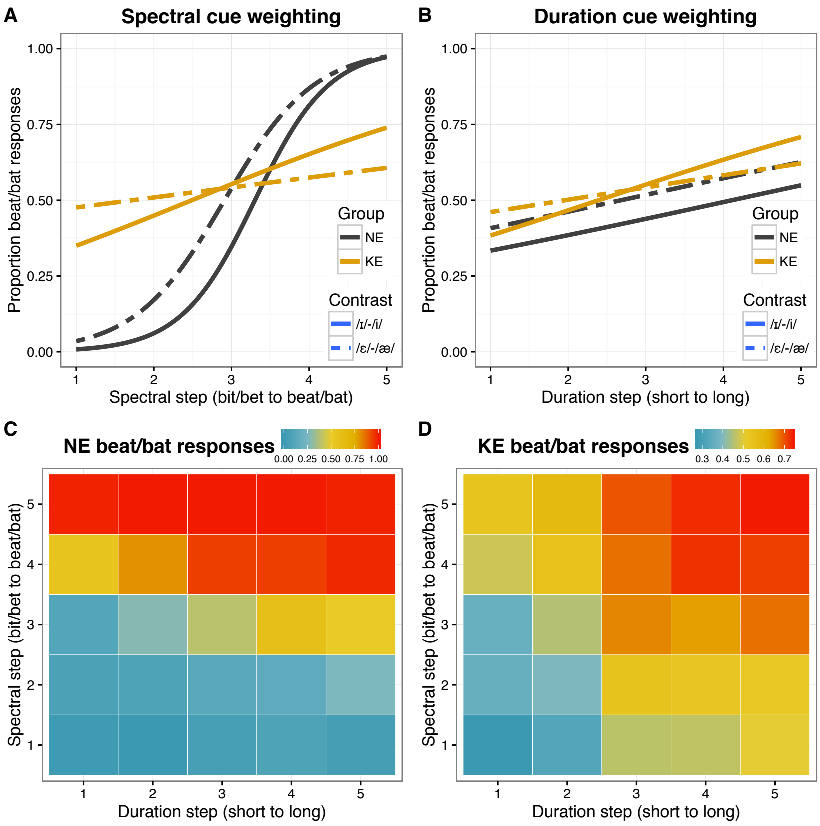

Figure 1 Proportion of *beat* (solid lines) and *bat* (dashed lines) responses by native English listeners (NE, black) and Korean learners of English (KE, yellow) along the vowel spectral continuum (A) and vowel duration continuum (B), as well as heatmap plots of overall responses averaged across the vowel contrasts for each combination of stimulus steps by the native listeners (C) and the learners (D).

[2] There were a few participants who showed negative values for cue weights. However, the present study used the magnitude of cue weights, rather than the direction, to assess the weight of a particular acoustic dimension that contributes to L2 learners' vowel categorization.

The results of the overall mixed-effects logistic regression model are summarized in Table 3. The results show that both spectral (β = 1.67, p = 0.003) and duration dimensions (β = 1.58, p = 0.001) significantly contribute to participants' categorization responses. A corresponding model for the native listeners (Table A1 in the Appendix) also found significant effects of both dimensions on vowel categorization, but a much stronger influence of spectral quality than duration (β = 8.45, $p < 0.001$ vs. β = 2.07, $p < 0.001$). There was also a significant interaction between CONTRAST and SPECTRUM (β = 1.36, p = 0.01), suggesting that the L2 learners weighted spectral quality more for /ɪ/-/i/ than for /ɛ/-/æ/.

Table 3 Summary of fixed effects in a mixed-effects logistic regression model in the identification task by the learners. Model coefficient estimates (β), standard errors, corresponding z-values, and p-values. Reference level is provided in parentheses.

Predictor	Estimate (β)	Std. Error	z	p
Intercept	0.38	0.09	3.85	< 0.001
GROUP (KC)	–0.28	0.19	–1.43	0.15
CONTRAST (/ɪ/-/i/)	0.18	0.25	0.73	0.46
SPECTRUM	1.67	0.57	2.90	0.003
DURATION	1.58	0.50	3.13	0.001
CONTRAST × SPECTRUM	1.36	0.57	2.35	0.01
CONTRAST × DURATION	0.90	0.92	0.97	0.32

Figure 2 shows the categorization responses of three L2 learners employing different strategies averaged across the two contrasts. The plots display considerable differences in individual learners' cue weighting strategies, including the largest differences observed between cue weights for vowel quality and duration (i.e., Learners 106 and 107), as well as a learner who gave relatively equal weight to both acoustic dimensions (i.e., Learner 114). That is, to categorize the vowel contrasts, some learners weighted spectral quality to a greater degree, while others predominantly relied on vowel duration; still others used both dimensions. Clearly, the overall group patterns displayed in Figure 1D mask substantial individual differences in learners' use of spectral quality and duration, much like what was observed in Schertz et al. (2015). Since native listeners rely mostly on spectral cues, these individual cue weighting strategies are expected to lead to differential abilities in L2 vowel discrimination, as will be discussed in section 3.2.

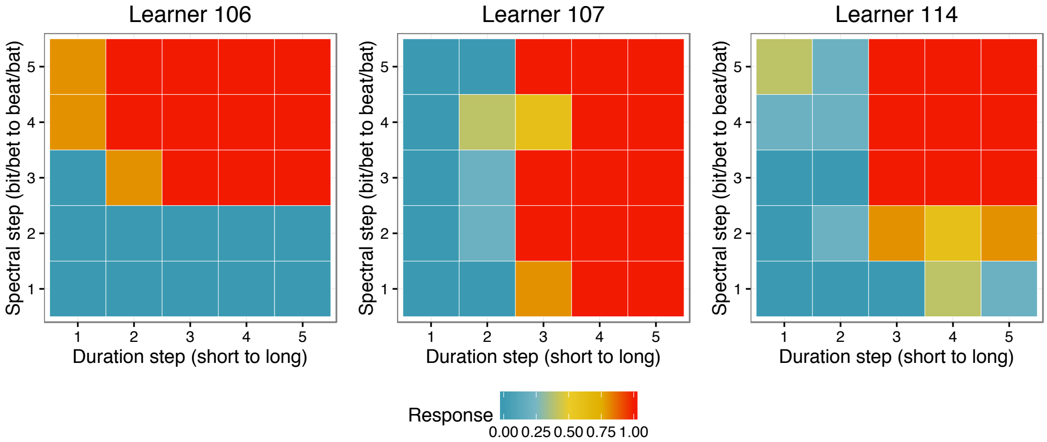

Figure 2 Heatmap plots of overall responses averaged across the vowel contrasts of two learners with the largest reliance on a particular acoustic dimension (spectrum for Learner 106 and duration for Learner 107), and one learner who weighted both dimensions relatively equally in vowel categorization (Learner 114).

Compared to the individual patterns in Figure 2, the overall cue weighting patterns shown in Figure 1 may suggest that most L2 learners use both spectral and duration cues to relatively similar degrees in vowel categorization in English, that is, that most learners have a profile similar to Learner 114 in Figure 2. In order to examine whether the overall group pattern also holds true for individual learners, we performed a correlation analysis of individual weights between spectral and duration cues. As shown in Figure 3, the analysis revealed that there was no significant correlation between the two dimensions ($r = -0.15$, $p = 0.29$). This indicates that most L2 learners attended to one acoustic dimension or the other, but not to both simultaneously.

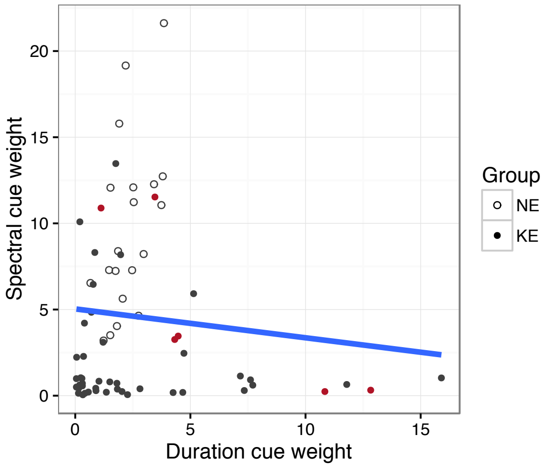

Figure 3 Correlation between individual L2 learners' perceptual weights for spectral quality and duration (KE, filled black). Individual learners from Figure 2 (filled red) and cue weights from NE listeners (hollow) are also plotted for reference.

The results from individual cue weights exhibit considerable variability within individual L2 learners, indicating differential cue weighting strategies at the level of the individual. To determine whether these differential cue weighting strategies for individual learners indicate consistent reliance on a particular cue, correlation analyses were conducted to examine the relation between individual learners' cue weights across different contrasts.

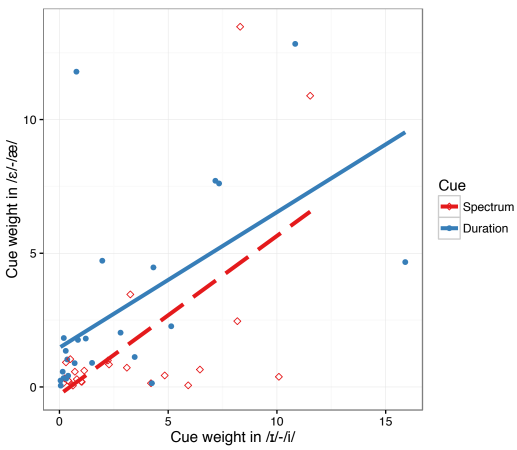

Figure 4 The relation between individual L2 learners' perceptual weights across contrasts.

Figure 4 shows correlations between cue weights across vowel contrasts for each of the cues. The correlation analysis revealed that the learners' cue weights were correlated across contrasts for both cues (Spectrum: $r = 0.60$, $p = 0.001$, Duration: $r = 0.55$, $p = 0.004$). That is, learners who showed greater spectral reliance on the /ɪ/-/i/ contrast also relied more on that same cue for the /ɛ/-/æ/ contrast. The same pattern was also observed for the duration cue. This suggests that individual learners' perceptual weighting strategies are stable across contrasts.

3.2 Relation between cue weights and discrimination abilities

To examine whether individual learners' discrimination abilities in the discrimination task are predicted by their cue weighting strategies, linear mixed-effects models were built using the *lmer()* function from the *lme4* package (ver.1.1-11) in *R*. Individual learners' discrimination abilities were computed based on perceptual sensitivity as indexed by d-prime (d′), which is a measure of discrimination ability (Macmillan and Creelman 2005). As in the previous model for perceptual cue weighting, the model included fixed effects for GROUP (adults vs. children), CONTRAST (/ɪ/-/i/ vs. /ɛ/-/æ/), SPECTRUM (spectral weights), and DURATION (duration weights), as well as interactions between CONTRAST and SPECTRUM and between CONTRAST and DURATION. GROUP and CONTRAST were centered, and SPECTRUM and DURATION were standardized by centering and dividing by 2 standard deviations to reduce collinearity and make the intercept interpretable as mean perceptual sensitivity. Random intercepts for each participant were included in the model. Figure 5 illustrates relations between cue weights and discrimination performance. Native English listeners performed at ceiling on the discrimination task (/ɪ/-/i/: 97% correct, /ɛ/-/æ/: 94.1% correct).

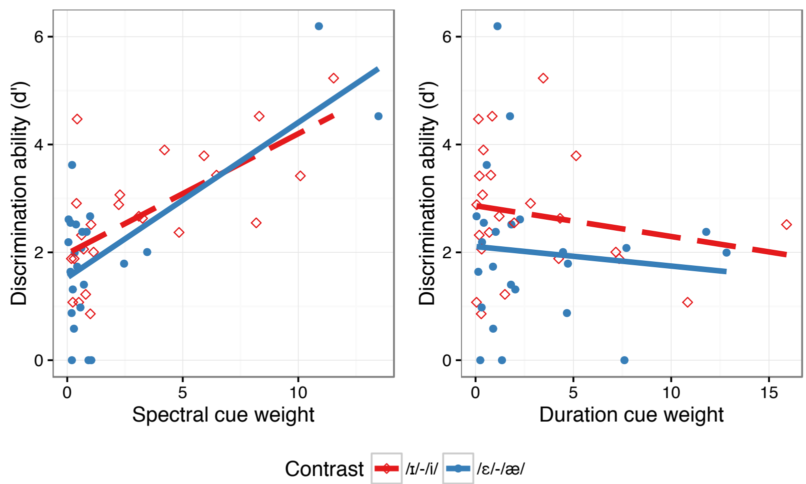

Figure 5 The relation between individual L2 learners' cue weights and discrimination abilities across contrasts.

The results of the linear mixed-effects model are summarized in Table 4. There was a significant effect of CONTRAST ($\beta = 0.45$, $p = 0.01$), suggesting that the learners' discrimination abilities were overall better for the /ɪ/-/i/ contrast than the /ɛ/-/æ/ contrast. A significant effect of SPECTRUM was found ($\beta = 1.03$, $p < 0.001$); that is, learners who attended more to spectral information in categorizing vowels were more sensitive to the phonological contrast. However, individual learners' vowel duration weight was not correlated with discrimination performance ($\beta = 0.13$, $p = 0.61$), suggesting that attending to vowel duration (which is a non-primary cue for native listeners) did not affect discrimination performance. A marginal interaction effect of CONTRAST and SPECTRUM suggests that the learners' spectral weight may have had a greater influence on their discrimination performance for the /ɛ/-/æ/ contrast than for the /ɪ/-/i/ contrast ($\beta = –0.62$, $p = 0.08$), but this effect may have been due to a few exceptional learners.

Table 4 Summary of fixed effects in a linear mixed-effects regression model in the discrimination task. Model coefficient estimates (β), standard errors, corresponding t-values, and p-values. Reference level is provided in parentheses.

Predictor	Estimate (β)	Std. Error	t	p
Intercept	2.38	0.20	11.64	< 0.001
GROUP (KC)	0.16	0.42	0.39	0.69
CONTRAST (/ɪ/-/i/)	0.45	0.16	2.83	0.01
SPECTRUM	1.03	0.28	3.61	< 0.001
DURATION	0.13	0.27	0.50	0.61
CONTRAST × SPECTRUM	–0.62	0.34	–1.82	0.08
CONTRAST × DURATION	–0.17	0.33	–0.52	0.60

4 Discussion and conclusion

As has previously been observed, our group analyses masked a large amount of individual variability in learners' use of phonetic cues in L2 speech perception. Although there were no group-level differences between adult and child learners in their overall cue weights, individual learners' cue weighting strategies varied considerably such that some learners, both adults and children, used spectral differences to distinguish the vowel contrasts while others used duration differences. The magnitude of cue weights also differed substantially across L2 learners. However, individual learners' use of spectral and duration differences as cues to the vowel contrasts were systematic in two ways.

First, cue weights were correlated across contrasts. Our results showed that learners who weighted spectral differences for one vowel contrast tended to rely more on spectral differences for the other vowel contrast. The same pattern also held true for the duration dimension. These findings extend results from previous studies which observed that individual cue weights are stable over testing sessions in native listeners (Idemaru et al. 2012) and in L2 learners (Schertz et al. 2015) to show that cue weights are also stable across contrasts.

The second type of systematic variability we observed was the correlation between cue weights and performance on the natural vowel discrimination task at an individual level. In other words, more native-like performance in the cue weighting task, namely the use of spectral differences over duration, was associated with better recognition of naturally produced vowels. In contrast, duration cue weight, which is not a primary cue to the English vowel contrasts, did not predict learners' discrimination performance. This means that the ability to attend to phonetic dimensions in general does not improve discrimination. Rather, only increased attention to the relevant acoustic dimensions of a phonological contrast entails better discrimination. Surprisingly, however, a few individuals achieved good discrimination without having strong cue weights for either dimension. Either these individuals were able to make use of some other source of information not captured by the two dimensions manipulated in the cue weighting task, or their performance in the cue weighting task did not adequately reflect their abilities to use the cues we varied.

Remarkably, most L2 learners in the present study attended to one acoustic dimension or the other, but not to both simultaneously. This mirrors the patterns observed by Schertz et al. (2015) for Korean learners of the English stop voicing contrast and by Escudero (2000) and Morrison (2008) for Spanish learners of the English /ɪ/-/i/ contrast. This might suggest that L2 learners differ from native listeners in terms of patterns of multiple cue use. That is, most L2 learners in this study did not exhibit the combined use of cues, which is commonly described for native listeners (e.g., Coleman 2003; Hillenbrand et al. 2000; Lisker 1986). Note, however, that there is relatively little evidence as to whether or not all native listeners use multiple cues for a given contrast. There is some evidence that for the English voicing contrast, only a subset of native listeners use both cues (VOT and f0) while all use the primary cue (VOT) (Kong and Edwards 2011; Llanos et al. 2013). Llanos et al. (2013) also found that all L2 learners used the primary cue while the present study and Schertz et al. (2015) found that many did not. More research is required to explore the effects of using a single cue or multiple cues and their roles in speech perception at the level of the individual.

The findings of this study indicate the importance of learning relevant cue weighting strategies in L2 speech perception. In other words, our results suggest that successful learning of the relevant acoustic dimension for target categories predicts L2 learners' abilities to perceive naturally produced contrasts, in line with Hattori and Iverson (2009). However, the current study also opens up possibilities for further research on the source of individual differences in L2 learners' perceptual abilities. Notably, this study controlled many factors that have previously been found to correlate with individual variability in L2 speech perception including motivation, amount of input, and length of residence in an L2-speaking environment (see Piske et al. 2001). Future work should consider how cognitive and social factors interact with linguistic factors in individual differences in L2 speech perception.

In conclusion, the present findings confirm earlier reports that individual L2 learners vary greatly in the extent to which they rely on particular acoustic-phonetic cues in the perception of L2 sound contrasts (Kong and Edwards 2015; Schertz et al. 2015, 2016). However, we further demonstrated that these individual differences are not random. Instead, they are systematically associated with how well L2 learners discriminate contrasts and reveal that successful learners use a stable cue weighting strategy across contrasts.

Appendix

Table A1 Summary of fixed effects in a mixed-effects logistic regression model in the identification task by the native listeners. Model coefficient estimates (β), standard errors, corresponding *z*-values, and *p*-values. Reference level is provided in parentheses.

Predictor	Estimate (β)	Std. Error	*z*	*p*
Intercept	–0.56	0.29	–1.91	0.05
CONTRAST (/ɪ/-/i/)	–1.35	0.65	–2.05	0.04
SPECTRUM	8.45	0.71	11.79	< 0.001
DURATION	2.07	0.17	12.11	< 0.001
CONTRAST × SPECTRUM	3.18	1.53	2.07	0.03
CONTRAST × DURATION	0.64	0.34	1.87	0.06

References

Aoyama, Katsura, James E. Flege, Susan Guion-Anderson, Reiko R. Akahane-Yamada & Tsuneo Yamada. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32(2). 233–250.

Boersma, Paul & David Weenink. 2013. *Praat: Doing phonetics by computer* (Version 5.3.55).

Chandrasekaran, Bharath, Padma D. Sampath & Patrick C. M. Wong. 2010. Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America* 128(1). 456–465.

Coleman, John. 2003. Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics* 31. 351–372.

Díaz, Begoña, Cristina Baus, Carles Escera, Albert Costa & Núria Sebastián-Gallés. 2008. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences* 105(42). 16083–16088.

Dorman, Michael F., Michael Studdert-Kennedy, & Lawrence J. Raphael. 1977. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics* 22(2). 109–122.

Dörnyei, Zoltán & Peter Skehan. 2003. Individual differences in second language learning. In Catherine J. Doughty & Michael H. Long (eds.), *The Handbook of Second Language Acquisition*, 589–630. Malden, MA: Blackwell Publishing.

Escudero, Paola. 2000. *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speakers*. Edinburgh: University of Edinburgh MA thesis.

Escudero, Paola. 2005. *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization.* Utrecht: Utrecht University dissertation.

Escudero, Paola, Titia Benders & Silvia Lipski. 2009. Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal* *of Phonetics* 37. 452–465.

Escudero, Paola, Titia Benders, & Karin Wanrooij. 2011. Enhanced bimodal distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America* 130(4). EL206–EL212.

Flege, James E., Ocke-Schwen Bohn & Sunyoung Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25. 437–470.

Francis, Alexander L., Natalya Kaganovich & Courtney Driscoll-Huber. 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America* 124(2). 1234–1251.

Hattori, Kota, & Paul Iverson. 2009. English/r/-/l/category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America* 125(1). 469-479.

Hillenbrand, James M., Michael J. Clark, & Robert A. Houde. 2000. Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America* 108(6). 3013–3022.

Holt, L. Lori & Andrew J. Lotto. 2006. Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America* 119(5). 3059–3071.

Idemaru, Kaori, Lori L. Holt & Howard Seltman. 2012. Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *Journal of the Acoustical Society of America* 132(6). 3950–3964.

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yoh'ich Tohkura, Andreas Kettermann & Claudia Siebert. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87(1). B47–B57.

Iverson, Paul, Valerie Hazan, & Kerry Bannister. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America* 118(5). 3267–3278.

Kawahara, Hideki, Toru Takahashi, Masanori Morise & Hideki Banno. 2009. Development of exploratory research tools based on TANDEM-STRAIGHT. *Proceedings of Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference.* 111–120.

Kondaurova, Maria V. & Alexander L. Francis. 2008. The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America* 124(6). 3959–3971.

Kondaurova, Maria V. & Alexander L. Francis. 2010. The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners Comparison of three training methods. *Journal of Phonetics* 38(4). 569–587.

Kong, Eun Jong & Jan R. Edwards. 2011. Individual differences in speech perception: Evidence from visual analogue scaling and eye-tracking. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 1126–1129.

Kong, Eun Jong & Jan R. Edwards. 2015. Individual differences in L2 learners' perceptual cue weighting patterns. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Walter de Gruyter.

Lengeris, Angelos & Valerie Hazan. 2010. The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America* 128(6). 3757-3768.

Levy, Erika S. 2009. Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *Journal of the Acoustical Society of America* 125(2). 1138–1152.

Levy, Erika S. & Winifred Strange. 2008. Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics* 36(1). 141–

157.

Lisker, Leigh. 1986. “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees, *Language and Speech* 29(1). 3–11.

Liu, Ran & Lori L. Holt. 2015. Dimension-based statistical learning of vowels. *Journal of Experimental Psychology: Human Perception and Performance* 41(6). 1783–1798.

Llanos, Fernando, Olga Dmitrieva, Amanda A. Shultz & Alexander L. Francis. 2013. Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *Journal of the Acoustical Society of America* 134(3). 2213–2224.

Macmillan, Neil A. & C. Douglas Creelman. 2005. *Detection theory: A user's guide* (2nd Edition). Mahwah, NJ: Lawrence Erlbaum.

Morrison, Geoffrey S. 2005. An appropriate metric for cue weighting in L2 speech perception: Response to Escudero and Boersma (2004). *Studies in Second Language Acquisition* 27(4). 597–606.

Morrison, Geoffrey S. 2008. L1-Spanish speakers' acquisition of the English /i/-/ɪ/ contrast: Duration-based perception is not the initial developmental stage. *Language and Speech* 51(4). 285–315.

Morrison, Geoffrey S. & Maria V. Kondaurova. 2009. Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis (L). *Journal of the Acoustical Society of America* 126(5). 2159–2162.

Nittrouer, Susan & Marnie E. Miller. 1997. Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, *101*(4), 2253–2266.

Perrachione, Tyler K, Ji-Yeon Lee, Louisa Y. Y. Ha & Patrick C. M. Wong. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America* 130(1). 461–472.

Piske, Thorsten T., Ian R. A. MacKay & James E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29(2). 191–215.

R Development Core Team. 2008. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, URL http://www.R-project.org, ISBN 3-900051-07-0.

Schertz, Jessamyn, Taehong Cho, Andrew Lotto & Natasha Warner. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52. 183–204.

Schertz, Jessamyn, Taehong Cho, Andrew Lotto & Natasha Warner. 2016. Individual differences in perceptual adaptability of foreign sound categories. *Attention, Perception, & Psychophysics* 78. 355–367.

Shultz, Amanda A., Alexander L. Francis & Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America* 132(2). EL95–EL101.

Tsukada, Kimiko, David Birdsong, Ellen Bialystok, Molly M. Mack, Hyekyung Sung & James E. Flege. 2005. A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics* 33(3). 263–290.

Wanrooij, Karin, Paola Escudero & Maartje E. J. Raijmakers. 2013. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics* 41(5). 307–319.

Werker, Janet F. & John S. Logan. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37(1). 35–44.