pISSN 2005-8063
eISSN 2586-5854
2026. 6. 30.
Vol.18 No.2
pp. 1-12

말소리와 음성과학
Phonetics and Speech Sciences

한국음성학회지

https://doi.org/10.13064/KSSS.2026.18.2.001

English vowel perception hierarchy of Korean learners: Effects of Residential Context*

Dong-Jin Shin**

Department of English Language and Literature, Jeonju University

Abstract

The current study investigates the perceptual hierarchy of vowels among Korean speakers learning English to determine the influence of the residential context, (between Korea versus the United Kingdom), on this hierarchy. Sixty-five Korean participants (30 Korea-based, 35 UK-based) performed a vowel identification task and a speech-in-noise recognition task. Despite the two groups yielding highly comparable hierarchies (ρ=.883, ρ<.001), differences arose between the two groups in the form of the identification accuracy of /ɒ/, /u/, /aɪ/, and /aʊ , all of which tended to be higher in the Korea-based group. The poor performance of the UK-based group in identifying /ɒ/ suggests that the influence of British English plays a role here. A generalized linear mixed-effects logistic regression analysis found residential context to be a statistically significant correlate of identification accuracy even after controlling for individual differences. While native accent speech-in-noise recognition was a reliable predictor of participant performance, Korean accent speech-in-noise recognition was not. Confusion matrix analysis revealed three common trends: systematic confusion between the back vowels, bidirectional confusion between /i/ and /ɪ/, and dispersed confusion with /ʌ/. Together, these results indicate a significant L1 influence on Korean learners' vowel perception, with residential context exerting selective rather than uniform effects on individual vowels.

Keywords: vowel perception, perception hierarchy, Korean English learners, residential context, L2 speech perception

1. Introduction

In second language (L2) speech perception research, vowel perception has been one of the core topics of investigation. The influence of a learner's first language (L1) vowel system on L2 vowel perception has been extensively discussed within theoretical frameworks such as Flege's (1995) Speech Learning Model (SLM), Best's (1995) and Best & Tyler's (2007) Perceptual Assimilation Model (PAM/PAM-L2), and EscuderoNeyra's (2005) Second Language Linguistic Perception Model (L2LP).

The Korean vowel system consists of seven monophthongs

* I would like to thank the three anonymous reviewers for their helpful comments and valuable suggestions.
** djshin@jj.ac.kr, Corresponding author
Received 13 May 2026; Revised 8 June 2026; Accepted 14 June 2026

(/i, e, ɛ, a, o, u, ɯ/), whereas English has 11 to 15 vowels depending on the dialect. This asymmetry leads Korean learners of English to experience systematic difficulties in perceiving certain English vowel contrasts, as consistently reported in previous research (e.g., Ingram& Park, 1997; Tsukadaet al., 2005). Specifically, vowel pairs that lack phonemic contrasts in Korean, such as /ɪ/-/iː/, /ɛ/-/æ/, /ɑː/-/ɔː/, and /ʌ/-/ɑː/, are predicted to be perceptually confusable under SLM and PAM frameworks.

However, relatively few studies have systematically investigated the perception hierarchy of a comprehensive set of English vowels. That is, which vowels are most easily identified and which are most difficult for Korean learners. Understanding this hierarchy can provide valuable insights for both theoretical modeling and pronunciation pedagogy.

Meanwhile, the effect of target language environment on L2 speech perception is another important research topic. According to Flege's (2009) revised SLM (SLM-r), the quantity and quality of exposure to the target language plays a crucial role in forming L2 phonetic categories. Shin (2018) examined the effect of language experience on speech-in-noise recognition for Korean L2 speakers, finding that while overall recognition accuracy did not differ between Korea-resident and UK-resident groups, the two groups showed different patterns of interlanguage benefit, suggesting that residential context influences perceptual processing in qualitative rather than purely quantitative ways. The vowel identification data collected in that study provides a unique opportunity to examine whether these qualitative differences extend to the level of individual vowel perception. The following section reviews the theoretical frameworks and empirical findings that bear on this question, before identifying the specific gaps that motivate the present study.

Against this backdrop, two gaps remain in the existing literature. First, while individual vowel contrasts have been studied for Korean learners, no research to date has systematically constructed a perception hierarchy covering the full inventory of English vowels, nor examined how such a hierarchy is modulated by residential context. Second, the vowel-level perceptual patterns underlying the group differences reported by Shin (2018) have not been analyzed, leaving open the question of whether Korea-resident and UK-resident learners differ not only in broader speech processing strategies but also in the internal structure of their vowel perception hierarchies. The present study makes three distinct contributions to the literature. First, it provides the first systematic perception hierarchy covering the inventory of 14 English vowels for Korean learners, allowing simultaneous evaluation of theoretical predictions from the SLM-r, PAM-L2, and L2LP across the entire vowel system rather than for isolated contrasts. Second, it examines how residential context modulates this hierarchy at the level of individual vowels, going beyond the group-level comparisons of previous research to reveal the vowel-specific nature of immersion effects. Third,by incorporating speech-in-noise (SIN) recognition accuracy as a participant-level covariate in the analytic model, it isolates group and vowel-specific effects from variation attributable to general L2 perceptual ability, thereby providing a more conservative estimate of residential context effects than analyses that treat all participants as exchangeable. Drawing on the vowel identification data collected in Shin (2018), the present study addresses these gaps through the following research questions:

- RQ1: What is the vowel perception hierarchy for Korean learners of English?
- RQ2: Does the vowel perception hierarchy differ between Korea-resident (KO) and UK-resident (UK) learners?
- RQ3: What confusion patterns emerge, and how do they reflect the influence of the Korean vowel system?

Speech-in-noise (SIN) recognition accuracy is incorporated as a participant-level covariate in the analyses rather than framed as an independent research question, on the grounds that general L2 speech processing ability is a known correlate of vowel identification performance (Shin, 2018) and must be partialled out before group and vowel-specific effects can be properly interpreted. The descriptive predictive performance of this covariate is reported in the Results section in support of the primary analyses, but no theoretical claim is staked on SIN itself.

2. Theoretical Background

2.1. Speech Learning Model (SLM/SLM-r)

The SLM (Flege, 1995) and SLM-r (Flege & Bohn, 2021) treat perceived phonetic similarity between L1 and L2 sounds as the primary determinant of L2 category formation. A central mechanism of the model is equivalence classification, whereby an L2 sound acoustically similar to an existing L1 category fails to trigger the formation of a new phonetic category and is instead assimilated to the L1 category, resulting in persistent non-native perception. Conversely, L2 sounds sufficiently distinct from any L1 category are more likely to trigger new category formation. This generates the well-known prediction that 'new' L2 sounds are ultimately acquired more accurately than 'similar' ones, because the latter are more readily assimilated to existing L1 categories.

The SLM-r (Flege & Bohn, 2021) extends this framework in three respects relevant to the present study: (a) L1 acquisition mechanisms remain accessible for L2 learning across the lifespan; (b) the formation of new L2 categories depends on the precision of L1 categories at the time of L2 learning onset; and (c) the quantity and quality of L2 input—rather than age of acquisition or length of residence per se—are critical determinants of phonetic category development. Crucially for the present study, the SLM-r predicts that learners in immersion contexts who receive higher-quality input may develop more refined L2 categories, though this improvement is expected to be selective rather than uniform across the vowel inventory.

For Korean learners of English, SLM/SLM-r predicts that English /uː/ should be relatively easy to perceive due to its close correspondence with Korean /u/, whereas /ʌ/ and /ɒ/, which lack direct Korean counterparts, should pose greater difficulty. Empirical work with Korean EFL learners (Lee & Shin, 2015) has shown, however, that cross-language vowel labelling alone accounts for the identification accuracy of only some English vowels, indicating that perceived phonetic distance must be considered alongside learners' broader interlanguage phonological system and degree of target-variety exposure.

2.2. Perceptual Assimilation Model (PAM/PAM-L2)

The PAM (Best, 1995) and PAM-L2 (Best & Tyler, 2007) approach L2 perceptual difficulty from a phonological standpoint, classifying L2 contrasts by how they assimilate to L1 categories. Two-category and uncategorized–categorized assimilations predict relatively easy discrimination, category-goodness assimilation predicts moderate difficulty, and single-category assimilation—where two L2 sounds are both assimilated to a single L1 category—predicts the greatest difficulty.

For Korean learners of English, these patterns generate specific predictions. The English vowels /ɒ/, /ɔː/, and /ʌ/ are all likely to be assimilated to the Korean /a/ category, constituting a three-way single-category assimilation that predicts extensive mutual confusion. Similarly, the English /iː/–/ɪ/ contrast is predicted to undergo single-category assimilation to Korean /i/, as Korean lacks a tense–lax vowel distinction (Ingram & Park, 1997). Empirical work supports these predictions: Lee & Suh (2016) found that English vowel pairs /iː/–/ɪ/, /e/–/ɛ/, /uː/–/ʊ/, and /ʌ/–/ɔː/ were each assimilated to a single Korean vowel category, with /ʌ/–/ɔː/ in particular showing assimilation patterns more strongly modulated by English–Korean orthographic correspondence than by acoustic similarity alone—suggesting that factors beyond acoustic distance contribute to PAM-L2 predictions for Korean learners. Systematic /ɒ/–/ɔː/–/ʌ/ confusion has been documented across both General American and British English input conditions (Lee & Cho, 2020; Lee & Shin, 2015), indicating that this confusion pattern is robust across input varieties rather than being an artifact of exposure to a particular accent.

In contrast, English diphthongs such as /eɪ/, /aɪ/, and /aʊ/ possess dynamic spectral properties that may facilitate two-category or uncategorized–categorized assimilation, as their spectral trajectories are sufficiently distinct from any single Korean vowel category (Escudero Neyra, 2005). Diphthongs are therefore predicted to be among the easiest English vowels for Korean learners to identify—a prediction consistent with the identification data reported by Lee & Shin (2015), in which /aɪ/ and /aʊ/ were among the most accurately identified vowels by Korean EFL learners.

2.3. Second Language Linguistic Perception Model (L2LP)

A third theoretical perspective is provided by Escudero Neyra's (2005) L2LP, which extends beyond the static predictions of the SLM and PAM-L2 by offering an explicitly developmental account of how L2 perceptual categories emerge and evolve over time. The model distinguishes between the initial state of L2 perception—which is a full copy of the L1 perception grammar—and subsequent developmental stages during which learners may create new perceptual categories or adjust existing boundaries. Escudero & Boersma (2004) demonstrated this framework with Spanish learners of English, who failed to discriminate the English /i/-/ɪ/ contrast because both vowels were assimilated to Spanish /i/. This parallel is directly relevant to Korean learners, who face the same challenge with the English /i/-/ɪ/ pair (Ingram & Park, 1997; Yang, 1996).

The L2LP model also predicts that the degree of perceptual difficulty depends on the specific mapping pattern: 'new' types (where an L1 category must split into two L2 categories) are more difficult than 'similar' ones (where an L2 category corresponds roughly to an L1 category but differs in phonetic detail). For Korean learners, the English back vowel system presents a classic 'new' mapping, requiring them to ultimately differentiate the single Korean /a/ category into multiple distinct English categories.

2.4. The Korean Vowel System and L2 English Vowel Perception

The Korean vowel system has undergone significant changes in recent decades that are relevant to understanding L2 English vowel perception. Traditionally described as having ten monophthongs, the contemporary Korean vowel system of younger speakers is widely recognized as consisting of seven monophthongs: /i/, /e/, /ɛ/, /a/, /o/, /u/, /ɯ/ (Yang, 1996). Critically, the ongoing merger of /e/ and /ɛ/ among younger Korean speakers (Ingram & Park, 1997; Yun, 2014) means that the functional vowel inventory may be even smaller, with only six phonemically distinct monophthongs for many speakers. This reduced vowel inventory creates a substantial asymmetry with the English vowel system, which contains 11 to 15 vowels depending on the dialect. The asymmetry is particularly pronounced in the back vowel region: Korean has only /a/, /o/, and /u/ in the low-to-back area, whereas English vowels distinguish among /ɑ/, /ɔ/, /ʌ/, /o/, /ʊ/, and /u/. Previous research has consistently identified the back vowel region as the locus of greatest perceptual difficulty for Korean learners (Cho & Jeong, 2013; Tsukada et al., 2005).

Furthermore, Korean lacks the tense-lax distinction that characterizes several English vowel pairs (/i/-/ɪ/, /u/-/ʊ/). Although Korean traditionally possessed a phonemic vowel length distinction, this contrast has largely collapsed among younger Seoul speakers and is no longer considered functionally operative in the contemporary vowel system (Magen & Blumstein, 1993; Silva, 2006). Consequently, Korean does not employ vowel duration as a reliable phonemic cue in the way that English does, where the tense-lax distinction involves differences in both spectral quality and duration (Ingram & Park, 1997). Yun (2014) found that Korean speakers tend to classify the English /u/-/ʊ/ contrast primarily by duration rather than vowel quality, suggesting an incomplete perceptual representation of the relevant spectral cues.

Park et al. (2010) further documented how L1 Korean phonological processes systematically interfere with Korean learners' English sound production, demonstrating that such L1 interference is modulated by both the type of phonological process and the morphological structure of target words. While that study focused on consonant clusters, its findings underscore the broader principle that L1 phonological knowledge exerts a pervasive and patterned influence on L2 sound acquisition at multiple levels of linguistic structure.

2.5. Effects of Residential Context on L2 Speech Perception

The role of target language environment in L2 speech acquisition has been investigated extensively. Flege et al. (1997) reported that length of residence (LOR) in the target language environment was positively correlated with L2 vowel perception accuracy for non-native speakers of English. However, subsequent research has shown that immersion does not guarantee uniform improvement across all learners or all vowel contrasts. For example, Flege & Liu (2001) demonstrated that LOR provides a useful estimate of L2 input only for learners who have regular opportunity and need to use the L2. Chinese adults with longer residence in the United States showed significantly higher scores in English consonant identification and listening comprehension, but this effect was modulated by the quality and frequency of L2 interaction. Baker & Trofimovich (2005) further showed that the effect of immersion is not uniform across all vowel contrasts; selective improvement was observed for specific contrasts, suggesting that immersion benefits particular aspects of the vowel system rather than producing wholesale improvement.

Aoyama et al. (2004) found that Japanese children showed significant improvement in English /r/-/l/ perception after 0.5 to 1.5 years of residence in the United States, whereas adults did not, suggesting that age-related factors may interact with residential context. More recently, the SLM-r (Flege& Bohn, 2021) has emphasized that the quantity and quality of L2 input — rather than LOR per se — are the critical determinants of L2 phonetic category development.

In the specific context of Korean learners, Shin (2018) examined the effect of language experience on SIN recognition for Korean L2 speakers, comparing Korea-resident and UK-resident groups. While overall recognition accuracy did not differ significantly between the two groups, they showed different patterns of interlanguage benefit: the UK group showed greater sensitivity to Korean and Southern British English accent type variation, suggesting that residential context influences perceptual processing in qualitative rather than purely quantitative ways. The vowel identification data collected in that study provides a unique opportunity to examine whether these qualitative differences in perceptual processing are reflected in vowel-level perception patterns.

2.6. Vowel Perception Hierarchy Research

Despite extensive research on specific vowel contrasts, relatively few studies have systematically investigated the perception hierarchy of a comprehensive set of L2 vowels—that is, the ordering of vowels from easiest to most difficult to identify—for any learner population. Such hierarchies are valuable because they reveal the cumulative effect of L1-L2 phonological relationships across the entire vowel system, rather than focusing on isolated contrasts. Cebrian (2006) examined Catalan speakers' categorization of Canadian English vowels and found that perceptual assimilation patterns predicted categorization accuracy: vowels that mapped clearly onto L1 categories were categorized more accurately than those with ambiguous mappings. Escudero & Boersma (2004) similarly demonstrated that the full pattern of cross-linguistic vowel mappings—not just individual contrasts—determines the overall landscape of perceptual difficulty for L2 learners. For Korean learners specifically, Cho & Jeong (2013) reported overall vowel perception accuracy of approximately 60%, with monophthongs and short vowels being more difficult to perceive than their diphthong and long counterparts. However, that study did not construct a complete perception hierarchy or examine the effect of residential context.

Given the limitations of existing research, understanding the full perception hierarchy is important for two reasons. Theoretically, a hierarchy based on a comprehensive set of vowels makes it possible to test whether the predictions of the SLM-r, PAM-L2, and L2LP hold simultaneously across the entire vowel system, rather than for isolated contrasts examined in previous studies. Pedagogically, identifying which vowels are most resistant to perceptual learning—and whether residential context alters this ordering—can directly inform the design of perceptual training programs by indicating where instructional resources should be concentrated.

3. Method

The data analyzed in the present study were drawn from Shin (2018), in which 65 Korean speakers completed both a SIN recognition task and a vowel identification (VI) task as part of a single experimental session. In Shin (2018), the VI task functioned as a predictor variable within the SIN recognition analysis: VI accuracy was entered as a fixed-effect predictor in a linear mixed-effects model of SIN recognition, and its correlation with SIN accuracy was reported for the pooled sample (r=.36 with Korean-accented stimuli; r=.58 with English native-accented stimuli). However, the VI data were not analyzed as a primary outcome in their own right: no per-vowel identification accuracies were reported, no perception hierarchy was constructed, no group-level comparison of VI performance was conducted, no confusion matrix analysis was performed, and no statistical model was fitted to the VI data as a dependent variable. The present study therefore constitutes an independent analysis of the VI dataset, motivated by a set of research questions that were not addressed in Shin (2018): namely, the structure of the vowel perception hierarchy, the modulation of that hierarchy by residential context at the level of individual vowels, and the nature of the systematic confusion patterns underlying group differences. This reanalysis was conducted with appropriate institutional approval, and all data were collected with participants' informed consent as reported in Shin (2018).

3.1. Participants

Sixty-five native Korean speakers participated in this study: 30 in the Korea-resident group (KO group) and 35 in the UK-resident group (UK group). The KO group consisted of Korean learners of English living in Seoul, Korea, and the UK group consisted of Korean learners of English living in London, UK. The same participants were also tested for a speech-in-noise recognition task as in Shin (2018). Table 1 presents the descriptive statistics of the two groups. The two groups did not differ significantly in their speech-in-noise recognition accuracy for either the native English accent or the Korean-accented conditions.

Table 1. Participant information

	KO Group (n=30)	UK Group (n=35)	*t*-value	*p*-value
SIN accuracy (native)	.584 (.124)	.572 (.111)	0.41	.686
SIN accuracy (Korean)	.614 (.072)	.581 (.075)	1.75	.085
VI accuracy	.642 (.174)	.569 (.141)	1.87	.067

Data are presented as mean (SD).
SIN accuracy represents the proportion correct in recognizing BKB sentences embedded in noise (see Shin, 2018).
SIN, speech-in-noise recognition; VI, vowel identification; KO group, Korea-resident group; UK group, UK-resident group.

3.2. Stimuli

The stimuli for the vowel identification task were 14 monosyllabic English words in a bVt context, recorded by one female speaker of Southern British English (see Shin, 2018 for details). It should be noted that the use of a single Southern British English speaker introduces a potential stimulus–listener dialect alignment confound: the UK-resident group may have had greater familiarity with this speaker's dialect, which could have differentially affected identification accuracy independently of any residential context effect. This design limitation should be borne in mind when interpreting group differences in the results that follow. The 14 target words were: beat /i/, bit /ɪ/, bet /ɛ/, bat /æ/, bait /eɪ/, bite /aɪ/, bout /aʊ/, boat /əʊ/, boot /u/, bot /ɒ/, bought /ɔ/, bart /ɑ/, bert /ɜ/, and but /ʌ/. Each participant heard each word four times, yielding a total of 56 trials per participant.

3.3. Procedure

Participants listened to each stimulus and selected the word they heard from a 14-alternative forced choice display on a computer screen. The presentation order was randomized for each participant. Responses were coded as correct (1) or incorrect (0) for each trial. The vowel identification task was administered alongside the speech-in-noise recognition task reported in Shin (2018).

3.4. Data Analysis

Vowel identification accuracy was calculated as the proportion of correct responses for each vowel and each group to construct the perception hierarchy. The correlation between the two groups' hierarchies was assessed using Spearman's rank correlation coefficient, which was chosen over Pearson's correlation because the hierarchy data are ordinal in nature and do not assume a linear relationship or normal distribution.

Per-vowel group differences were tested using chi-square tests of independence. Because the dependent variable for each trial was binary (correct=1, incorrect=0) and each vowel was presented four times per participant, the chi-square test was selected as a nonparametric test appropriate for comparing response proportions between two independent groups. These chi-square analyses were conducted as exploratory descriptive tests to characterize the pattern of per-vowel group differences, rather than as confirmatory hypothesis tests. Accordingly, no correction for multiple comparisons was applied to the chi-square results; confirmatory inference regarding group and vowel-specific effects was instead based on the generalized linear mixed model (GLMM) reported below, which simultaneously models all vowels within a single inferential framework and is not subject to the same multiple comparison concerns.

To examine the simultaneous effects of multiple predictors on vowel identification accuracy, a mixed-effects logistic regression model (GLMM) was fitted. A GLMM was chosen because the dependent variable is binary (correct vs. incorrect on each trial) and the data have a hierarchical structure in which multiple trials are nested within participants, violating the independence assumption of standard logistic regression. The model included a random intercept for each participant to account for individual differences in overall vowel identification ability. Random slopes for vowel were considered but were not included in the final model for two reasons: first, the ratio of model parameters to participant count (n=65) rendered a full random slope structure for 14 vowel levels computationally intractable and susceptible to non-convergence; second, the implementation used for model estimation (see below) does not support complex crossed random effects structures. The random intercept-only specification was therefore adopted as the most parsimonious structure that adequately accounts for the non-independence of observations within participants.

The model was estimated using variational Bayes (VB) approximation as implemented in the BinomialBayesMixedGLM function in Python's statsmodels library. VB approximation was selected over standard Laplace approximation for two reasons: its superior convergence stability in high-dimensional random-effect models with binary outcomes, and its computational efficiency relative to Markov chain Monte Carlo methods for datasets of the present scale. It is acknowledged that VB approximation tends to underestimate posterior variance, which can result in narrower credible intervals and potentially anti-conservative p-values for fixed-effect estimates; however, this limitation does not affect the direction or rank ordering of fixed-effect estimates, and the substantive group × vowel patterns reported below are robust to this caveat (cf. Blei et al., 2017).

The model included residential group (KO vs. UK), vowel (14 levels), speech-in-noise recognition accuracy in the native English accent condition (n_se), speech-in-noise recognition accuracy in the Korean-accented condition (n_ko), and the group × vowel interaction as fixed-effect predictors, with participant as a random intercept. This specification allowed us to assess whether group differences in vowel identification varied across individual vowels while controlling for individual differences in general L2 speech perception ability and accounting for the non-independence of observations within participants. To facilitate interpretation of the magnitude of fixed effects, odds ratios (OR=e^β) are reported alongside β coefficients for all GLMM predictors.

Confusion matrices were computed for each group by cross-tabulating stimulus vowels against response vowels. Each row was normalized by dividing cell frequencies by the row total, yielding proportions that represent the conditional probability of each response given a particular stimulus.

4. Results

4.1. Vowel Perception Hierarchy

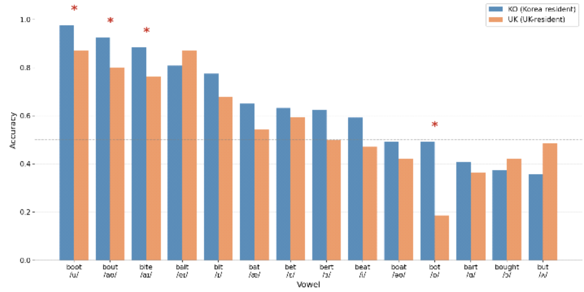
This section addresses RQ1 by presenting the overall vowel identification accuracy and the perception hierarchy for both groups. The overall vowel identification accuracy was .642 (*SD*=.174) for the KO group and .569 (*SD*=.141) for the UK group. Group and vowel-specific effects were examined using a mixed-effects logistic regression model (GLMM) while controlling for individual differences. The full GLMM results, including fixed-effect coefficients, group × vowel interaction terms, and predictor effects, are reported in Section 4.5.

In the GLMM, residential group was a significant fixed-effect factor of vowel identification accuracy (β=0.30, *SE*=0.05, *z*=5.72, $p<.001$), indicating that after individual differences in overall ability were accounted for through the random intercept, group membership was significantly associated with identification accuracy. Speech-in-noise recognition accuracy in the native English accent condition was also a significant predictor (β=3.69, *SE*=0.34, *z*=10.80, $p<.001$), whereas SIN accuracy in the Korean-accented condition was not (β=0.82, *SE*=0.52, *z*=1.57, *p*=.116). The random intercept for participants had a standard deviation of 0.716, confirming substantial individual variation in overall vowel identification ability and justifying the use of a mixed-effects model.

Table 2 presents the vowel perception hierarchy for each group, ranked by identification accuracy. The Spearman rank correlation between the two groups' hierarchies was strong and significant (ρ=.883, $p<.001$), indicating that the overall ordering of vowel difficulty was largely consistent across groups. Both groups showed high accuracy for diphthongs (/eɪ/, /aɪ/, /aʊ/) and the high back vowel /u/, while back vowels (/ɒ/, /ɔ/), rhotacized vowels (/ɑ/), and /ʌ/ were consistently difficult. Figure 1 displays the per-vowel identification accuracy for both groups, with asterisks indicating vowels for which significant group differences were found.

Table 2. Vowel perception hierarchy by group

Rank	KO Group	Accuracy	UK Group	Accuracy
1	boot /u/	.975	bait /eɪ/	.871
2	bout /əʊ/	.925	boot /u/	.871
3	bite /aɪ/	.883	bout /əʊ/	.800
4	bait /eɪ/	.808	bite /aɪ/	.764
5	bit /ɪ/	.775	bit /ɪ/	.679
6	bat /æ/	.650	bet /ɛ/	.593
7	bet /ɛ/	.633	bat /æ/	.543
8	bert /ɜ/	.625	bert /ɜ/	.500
9	beat /i/	.592	but /ʌ/	.486
10	boat /əʊ/	.492	beat /i/	.471
11	bot /ɒ/	.492	bought /ɔ/	.421
12	bart /ɑ/	.408	boat /əʊ/	.421
13	bought /ɔ/	.375	bart /ɑ/	.364
14	but /ʌ/	.358	bot /ɒ/	.186



*$p<.05$.

Figure 1. Vowel perception accuracy by group.

4.2. Effect of Residential Context

Addressing RQ2, this section examines whether the perception hierarchy differs between the KO and UK groups at the level of individual vowels. Table 3 presents the chi-square test results for per-vowel group differences. Significant group differences were found for bot /ɒ/ (χ^2=26.11, $p<.001$), boot /u/ (χ^2=7.99, *p*=.005), bout /aʊ/ (χ^2=7.28, *p*=.007), and bite /aɪ/ (χ^2=5.41, *p*=.020). Notably, for all four vowels, the KO group showed higher accuracy than the UK group. In contrast, for but /ʌ/, the UK group (.486) tended to outperform the KO group (.358), although this difference was only marginally significant (χ^2=3.78, *p*=.052). The most striking group difference was observed for bot /ɒ/: the KO group achieved .492 accuracy, whereas the UK group was at .186. These chi-square tests were conducted at the trial level and therefore should be interpreted as descriptive screening statistics rather than as formal tests of group differences; the GLMM reported in Section 4.5, which explicitly models participant-level variation through random intercepts, provides the primary inferential basis for the group × vowel interactions discussed throughout this section. The confusion patterns underlying this group difference are examined in detail in Section 4.3.

Table 3. Chi-square tests for group differences by vowel

Vowel	KO Group	UK Group	χ^2	*p*-value	Sig.
bait /eɪ/	.808	.871	1.49	.222	
bart /ɑ/	.408	.364	0.36	.549	
bat /æ/	.650	.543	2.64	.104	
beat /i/	.592	.471	3.28	.070	
bert /ɜ/	.625	.500	3.60	.058	
bet /ɛ/	.633	.593	0.29	.589	
bit /ɪ/	.775	.679	2.54	.111	
bite /aɪ/	.883	.764	5.41	.020	*
boat /əʊ/	.492	.421	1.02	.313	
boot /u/	.975	.871	7.99	.005	**
bot /ɒ/	.492	.186	26.11	<.001	***
bought /ɔ/	.375	.421	0.40	.526	
bout /aʊ/	.925	.800	7.28	.007	**
but /ʌ/	.358	.486	3.78	.052	

*$p<.05$, **$p<.01$, ***$p<.001$.

4.3. Confusion Patterns and L1 Influence

Figure 2 presents the row-normalized confusion matrices for the KO and UK groups. Addressing RQ3, the confusion matrix analysis examines what error patterns emerge and how they reflect the influence of the Korean L1 vowel system. Three major confusion patterns emerged from the confusion matrices. First, extensive confusion was observed among back vowels. Bot /ɒ/, bought /ɔ/, and boat /əʊ/ were systematically confused with each other in both groups, reflecting the absence of corresponding phonemic contrasts in Korean. The bought → boat confusion was particularly prominent (KO: 50.0%, UK: 39.3%). Furthermore, the UK group's notably low accuracy for bot /ɒ/ (.186) can be attributed to its predominant confusion with /əʊ/ (boat, 38.6%) and /ɔ/ (bought, 17.9%), likely reflecting the influence of the British English vowel system in which LOT /ɒ/ differs acoustically from American English /ɑ/.

Second, bidirectional confusion between beat /i/ and bit /ɪ/ was observed in both groups. In the KO group, beat /i/ was misidentified as bit /ɪ/ in 17.5% of trials and bit /ɪ/ as beat /i/ in 20.8%. In the UK group, this confusion was even stronger: beat /i/ → bit /ɪ/ at 43.6% and bit /ɪ/ → beat /i/ at 27.1%. This pattern is consistent with Single Category assimilation to Korean /i/.

Third, but /ʌ/ showed dispersed confusion across multiple categories. The KO group confused but primarily with bot (33.3%) and bart (12.5%), while the UK group confused it with bat (17.1%) and bart (15.0%). This dispersed pattern suggests that /ʌ/ has not established a stable perceptual category for Korean learners.

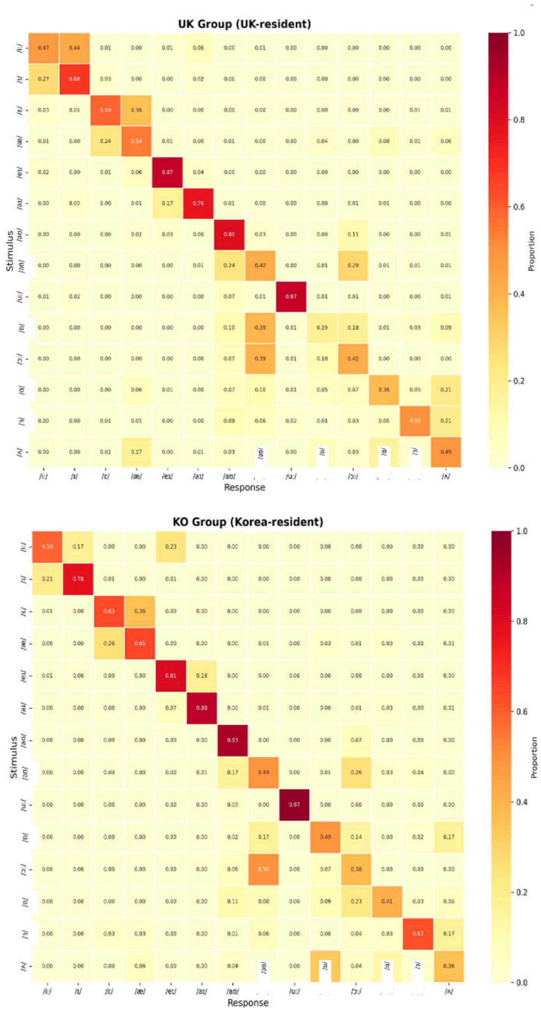


Figure 2. Confusion matrices for vowel identification (row-normalized proportions).

4.4. Speech-in-Noise Recognition as a Predictor

As a supplementary descriptive analysis of the covariate included in the primary model, this section examines the relationship between speech-in-noise recognition ability and vowel identification accuracy. SIN recognition accuracy in the native English accent condition was significantly correlated with vowel identification accuracy in both groups (KO: r=.563, p=.001; UK: r=.610, p<.001). SIN accuracy in the Korean-accented condition showed a significant correlation only in the KO group (r=.380, p=.038) but not in the UK group (r=.273, p=.112). The GLMM confirmed these bivariate patterns at the trial level: SIN accuracy in the native English accent condition was the strongest participant-level predictor (β=3.69, p<.001), whereas SIN accuracy in the Korean-accented condition did not reach significance (β=0.82, p=.116). These results indicate that the ability to process native-accented English speech in noise is a robust predictor of vowel identification accuracy, even after controlling for group membership and vowel-specific effects. Figure 3 illustrates the relationship between SIN recognition accuracy and vowel identification accuracy for each group.

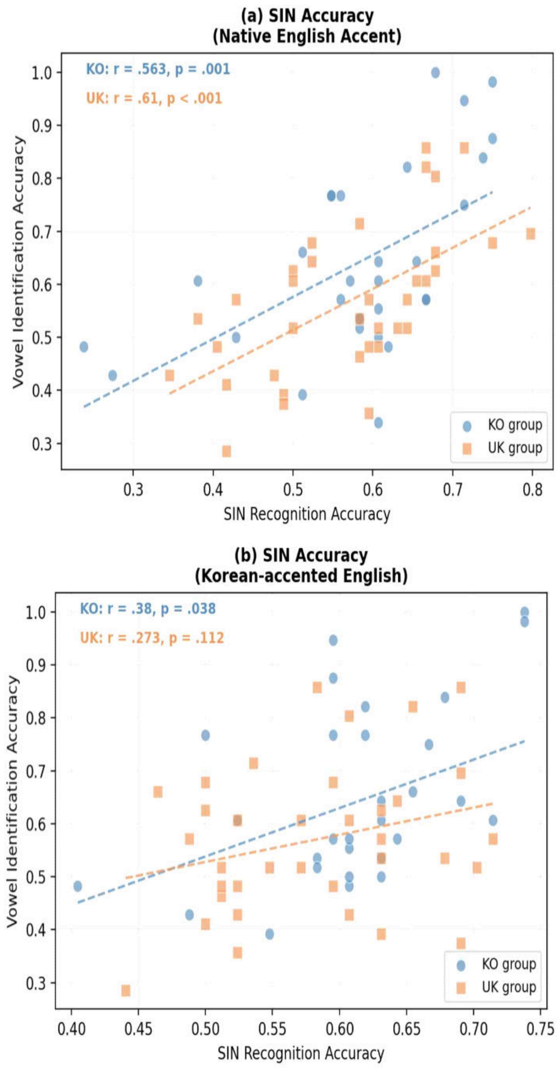


Figure 3. Relationship between Speech-in-Noise (SIN) recognition accuracy and vowel identification accuracy.

4.5. Predictors of Vowel Identification Accuracy

This section presents the full results of the mixed-effects logistic regression model (GLMM), integrating the fixed-effect structure, group × vowel interaction terms, and participant-level predictors. Residential group was a significant fixed-effect factor of vowel identification accuracy (β=0.30, SE=0.05, z=5.72, p<.001, OR=1.35), indicating that group membership was significantly associated with accuracy after individual differences were accounted for through the random intercept. The random intercept for participants had a standard deviation of 0.716, confirming substantial individual variation in overall vowel identification ability and justifying the mixed-effects structure.

The GLMM detected significant group × vowel interactions for 12 out of 13 vowels (relative to the reference vowel bait), with the majority of interaction coefficients showing negative values, indicating vowel-specific differences in identification accuracy in the UK group. The largest interaction was observed for bot /ɒ/ (β=−1.87, p<.001, OR=0.15), and the only positive interaction was for but /ʌ/ (β=+0.35, p=.048, OR=1.42), indicating a selective group-related advantage for this vowel in the UK group.

Turning to participant-level predictors, SIN recognition accuracy in the native English accent condition was significantly correlated with vowel identification accuracy in both groups (KO: r=.563, p=.001; UK: r=.610, p<.001) and was the strongest predictor in the GLMM (β=3.69, SE=0.34, z=10.80, p<.001, OR=40.0). SIN accuracy in the Korean-accented condition showed a significant bivariate correlation only in the KO group (r=.380, p=.038) but not in the UK group (r=.273, p=.112), and did not reach significance in the GLMM (β=0.82, SE=0.52, z=1.57, p=.116, OR=2.27).

5. Discussion

5.1. Robustness of the L1-Based Vowel Perception Hierarchy

The most notable finding of the present study is the strong Spearman rank correlation between the two groups' vowel perception hierarchies (ρ=.883, p<.001). This finding provides compelling evidence that the perception hierarchy is primarily determined by the structural relationship between the L1 and L2 vowel systems, and that extended residence in the target language environment does not fundamentally restructure this hierarchy.

Before interpreting group-level patterns, it is important to acknowledge the substantial individual variation captured in the GLMM. The random intercept for participants (SD=0.716) indicated that individual differences in overall vowel identification ability were considerable—variation that was independent of group membership. This finding suggests that residential context alone does not determine vowel perception outcomes, and that learner-internal factors such as general L2 processing ability play a significant role alongside L1-L2 structural relationships. The GLMM's ability to partial out this individual variation thus provides a cleaner estimate of group and vowel-specific effects than would otherwise be possible.

This robustness is consistent with predictions from multiple theoretical frameworks. Under the SLM-r (Flege & Bohn, 2021), the hierarchy reflects the perceived phonetic distance between English vowels and their closest Korean counterparts: vowels with clear L1 mappings (e.g., English /u/ → Korean /u/ '우') are perceived more accurately than those without direct counterparts (e.g., /ʌ/). Under PAM-L2 (Best & Tyler, 2007), the hierarchy reflects assimilation patterns, with two category contrasts being easier than single category ones.

The vowels ranked highest in both groups—/u/, /aʊ/, /aɪ/, /eɪ/—share two key characteristics. First, /u/ benefits from a relatively clear cross-linguistic category mapping to Korean /u/, and its peripheral position in the vowel space may contribute to its spectral distinctiveness. Second, the diphthongs possess dynamic spectral properties—significant formant movement—that may serve as salient perceptual cues even when static spectral properties overlap with those of other vowels. This interpretation is consistent with Escudero Neyra's (2005) L2LP framework, which predicts that dynamic spectral cues can facilitate perceptual categorization when static cues are ambiguous, though direct acoustic measurement of the stimuli would be needed to confirm this account.

In contrast, the lowest-ranked vowels included two low-back vowels, /ɒ/ and /ɔ/, characterized by relatively similar formant structures and assimilated to the Korean /a/ region, constituting a case of multiple single category assimilation under PAM-L2. The vowel /ʌ/, while also among the lowest-ranked, occupies a different perceptual position: as elaborated in Section 5.3, its identification is characterized by dispersed confusion across multiple L1 categories rather than systematic assimilation to a single one, consistent with Uncategorized assimilation under PAM-L2. The GLMM results support these distinctions, with the fixed-effect coefficients for bot, bought, and but among the largest negative values, confirming their difficulty even after controlling for individual differences in general perceptual ability.

The intermediate position of the tense-lax pairs (/i/-/ɪ/, /ɛ/-/æ/) in the hierarchy is also theoretically informative. These pairs underwent single category assimilation to their respective Korean counterparts, but their bidirectional confusion rates were lower than those for the back vowel triad. This suggests that some degree of perceptual differentiation has occurred for front vowel pairs, possibly facilitated by the greater acoustic distance between tense and lax front vowels in English compared to the compact back vowel space. This is consistent with the L2LP model's prediction that perceptual categories can gradually differentiate through extended L2 exposure (Escudero & Boersma, 2004).

5.2. Selective Effects of Residential Context

Before interpreting any group-level patterns in the present study, it is necessary to confront directly a structural limitation of the experimental design. Because all vowel identification stimuli were produced by a single Southern British English speaker, the observed differences between the KO and UK groups are necessarily confounded with differential familiarity with the specific phonetic properties of the Southern British English variety. This confound cannot be resolved post hoc; rather, it defines the interpretive boundary within which all group-level findings must be situated. Specifically, two categories of inference can be drawn from the present results: those that remain valid under the confound, and those that require future research to disentangle.

The finding that is most robustly interpretable under this constraint is the selective advantage of the UK group for /ʌ/ (but). This vowel showed a significant positive group × vowel interaction in the GLMM (β=+0.35, p=.048), with the UK group outperforming the KO group (.486 vs. .358). Crucially, a stimulus–listener dialect alignment account does not predict this direction of group difference. In Southern British English, /ʌ/ is realized as a fronted, near-open central vowel that approaches [ɐ] and is therefore acoustically closer to /æ/ than its General American counterpart (Roach, 2009; Wells, 1982). If UK-residents' familiarity with SBE /ʌ/ were the operative factor, one would expect more frequent /ʌ/→/æ/ confusions in the UK group — a pattern that is indeed observed (but→bat: 17.1% in the UK group). Yet despite this dialect-driven confusability with an adjacent category, the UK group still identified /ʌ/ more accurately overall than the KO group. Dialect alignment, in other words, predicts a UK *disadvantage*for /ʌ/, not the observed advantage. A more principled account, consistent with the SLM-r's emphasis on input quantity as a driver of phonetic category development (Flege & Bohn, 2021), is that the UK group's more frequent real-world exposure to /ʌ/ in naturalistic English input facilitated the formation of a more stable perceptual category for this vowel — robust enough to withstand the spectral proximity to /æ/ in the SBE stimuli. This interpretation, while still inferential, is not reducible to the stimulus dialect confound and thus represents the strongest evidence in the present dataset for a genuine residential context effect on L2 vowel perception.

In contrast, the findings for /ɒ/ (bot) and the /i/-/ɪ/ contrast must be interpreted with greater caution. The most striking group difference in the present study was the UK group's markedly lower accuracy for /ɒ/ (.186 vs. .492 in the KO group), yielding the largest group × vowel interaction coefficient in the GLMM (β=−1.87, p<.001). A theoretically motivated account is that the UK group's developing representations were shaped by the British English LOT vowel /ɒ/, which is realized as a rounded, slightly raised vowel acoustically distinct from the General American /ɑ/ represented in the stimulus. If so, the UK group's lower accuracy would reflect not a deficit in vowel perception per se, but rather a category-level reorganization driven by dialect-specific input—a pattern that is in principle consistent with SLM-r predictions about input quality effects. However, since this account and the stimulus dialect confound make identical predictions for this vowel, the two explanations cannot be empirically distinguished within the present design. The finding is therefore reported as a theoretically motivated observation rather than evidence for a residential context effect, and its interpretation must await future studies employing stimuli from multiple dialect varieties.

Similarly, the finding that /i/-/ɪ/ confusion was stronger in the UK group (beat → bit: 43.6%; bit → beat: 27.1%) than in the KO group (17.5% and 20.8%) is consistent with the possibility that British English input, in which the spectral distance between /i/ and /ɪ/ may be smaller than in some other varieties, reinforced single category assimilation rather than promoting differentiation. This interpretation is plausible, but again the stimulus dialect confound prevents causal attribution to residential context. It is noted, however, that the strengthened confusion in the UK group runs counter to the simple prediction that immersion facilitates category differentiation, and therefore carries theoretical interest regardless of its ultimate cause.

Taken together, the pattern of group × vowel interactions reveals that the effect of residential context on L2 vowel perception is not uniform. The /ʌ/ finding suggests a genuine frequency-driven benefit of immersion for at least one vowel. The /ɒ/ and /i/-/ɪ/ findings suggest that immersion may introduce dialect-specific category reorganization that does not straightforwardly improve identification accuracy relative to a non-immersion baseline.It bears emphasis, however, that length of residence, cumulative L2 input quantity, and frequency of English use were not measured at the participant level in the present design; the immersion-based interpretations offered above therefore remain inferential associations rather than quantitatively validated effects. The GLMM's ability to simultaneously model all vowel-specific effects while accounting for individual differences provides a more statistically complete picture than would otherwise be possible, and the present findings collectively motivate a research agenda in whichlength of residence, individual usage patterns, and dialect exposure are directly measured and experimentally controlled.

5.3. Theoretical and Pedagogical Implications of the Confusion Patterns

The confusion matrix analysis provides rich empirical evidence for evaluating predictions from PAM-L2 and related models. Three major patterns emerged, each with distinct theoretical implications.

First, the triangular confusion among /ɒ/, /ɔ/, and /əʊ/ in the back vowel region represents a classic case of multiple Single Category assimilation. The particularly prominent bought → boat confusion (KO: 50.0%, UK: 39.3%) is consistent with the possibility that Korean learners may place greater weight on lip rounding as a cue for distinguishing back vowels, potentially associating the rounded quality of /ɔ/ with the Korean /o/ category. However, as cue-weighting was not directly measured in this study, this account remains speculative and would require acoustic or perceptual cue manipulation experiments to verify.

Second, the bidirectional /i/-/ɪ/ confusion, discussed in Section 5.2 as a finding constrained by the stimulus dialect confound, also bears on PAM-L2 predictions when viewed from the confusion-pattern perspective. The intensified confusion in the UK group runs counter to the simple SLM-r prediction that increased L2 exposure facilitates separate category formation, suggesting that the trajectory of category differentiation may depend on the spectral properties of the specific input variety rather than on exposure quantity alone.

Third, the dispersed confusion pattern of /ʌ/ across multiple categories suggests Uncategorized assimilation under PAM-L2, where the L2 vowel does not find a stable L1 perceptual anchor. The different error patterns between groups (KO: primarily bot and bart; UK: primarily bat and bart) suggest that the two groups have developed different, but equally unstable, perceptual strategies for this vowel.

These findings have clear pedagogical implications. Korean English learners would benefit most from targeted perceptual training on the back vowel contrasts (/ɒ/-/ɔ/-/ʌ/), which represent the most persistent perceptual difficulty regardless of residential context. For UK-resident learners specifically, training should also address the /i/-/ɪ/ contrast. High-variability phonetic training using multiple speakers and dialect varieties may be particularly effective in promoting robust L2 category formation.

With respect to training methodology, a body of research on perceptual training has demonstrated that high-variability training protocols—in which stimuli are drawn from multiple talkers rather than a single speaker—promote more robust and generalizable L2 phonetic category formation than single-talker training (Bradlow et al., 1997; Lively et al., 1993; Logan et al., 1991). The present study's finding that single-speaker stimuli were sufficient to produce strong ceiling effects for some vowels (e.g., /eɪ/, /aʊ/) but not others (e.g., /ɒ/, /ʌ/) is consistent with the view that the most perceptually resistant categories—those subject to multiple single category assimilation—may require the broader acoustic sampling that multi-talker, multi-dialect training provides. These observations suggest that training programs for Korean EFL learners should prioritize back vowel contrasts and, where possible, incorporate stimuli from multiple English varieties to promote category robustness across dialect variation.

5.4. Speech-in-Noise Recognition as a Predictor

Both the bivariate correlations and the GLMM confirmed that SIN recognition accuracy in the native English accent condition is a significant predictor of vowel identification accuracy (bivariate: KO r=.563, p=.001; UK r=.610, p<.001; GLMM: β=3.69, p<.001). In contrast, SIN accuracy in the Korean-accented condition did not reach significance in the GLMM (β=0.82, p=.116), although it showed a significant bivariate correlation in the KO group only (r=.380, p=.038).

It is important to note that the SIN recognition scores represent participants' accuracy in recognizing BKB sentences embedded in noise under two accent conditions, not language proficiency measures per-se. The native English accent SIN score captures the listener's ability to process L2 speech under degraded conditions, drawing upon both segmental and suprasegmental skills. Its stronger predictive power suggests that general native-accented L2 speech processing ability is a more reliable index of underlying vowel perception ability than processing of non-native-accented speech.

The asymmetric pattern for Korean-accented SIN scores—significant bivariately in the KO group but not in the UK group, and non-significant in the GLMM—is theoretically interesting. One possible account is that the KO group, who are likely to encounter Korean-accented English more frequently in their immediate environment, may be better positioned to leverage their vowel knowledge when processing Korean-accented speech in noise. However, since frequency of exposure to Korean-accented English was not directly measured, this interpretation remains inferential. The pattern is nonetheless consistent with Shin's (2018) finding that the two groups showed different patterns of interlanguage benefit.

5.5. Methodological Consideration of the GLMM Approach

The present study employed a mixed-effects logistic regression model (GLMM), which represents a methodological improvement over the standard logistic regression commonly used in L2 vowel perception research. The GLMM revealed several findings that were not apparent in the standard analysis.

First, the random intercept for participants (SD=0.716) confirmed that individual differences in overall vowel identification ability were substantial. In designs with this degree of individual variation, failing to account for within-participant clustering through a mixed-effects structure risks underestimating standard errors and inflating Type I error rates, a well-documented concern in the psycholinguistic literature (cf. Baayen et al., 2008). Second, the GLMM detected significant group × vowel interactions for 12 out of 13 vowels (relative to the reference vowel bait), whereas the chi-square tests identified significant differences for only 4 vowels. Third, the effect of /ʌ/ (but), which was only marginally significant in the chi-square analysis (p=.052), reached significance in the GLMM (p=.048), a finding with substantive theoretical implications for understanding the selective effects of immersion.

These results underscore the importance of employing appropriate statistical models that respect the hierarchical structure of experimental data. Researchers conducting forced-choice identification tasks with repeated measures should consider mixed-effects models as the default analytical framework.

6. Conclusion

The present study investigated the English vowel perception hierarchy of 65 Korean learners and compared Korea-resident and UK-resident groups. The main findings can be summarized as follows.

First, the vowel perception hierarchy is strongly constrained by the L1 Korean vowel system, with diphthongs and /u/ at the top and /ɑ/, /ɔ/, and /ʌ/ at the bottom. Second, the two groups' hierarchies were highly correlated (ρ=.883), but significant per-vowel differences emerged, particularly for /ɒ/ (bot), possibly reflecting the influence of the British English variety. Third, confusion matrix analyses revealed systematic L1-driven error patterns, including back vowel triangular confusion, /i/-/ɪ/ confusion, and dispersed /ʌ/ confusion. Fourth, speech-in-noise recognition accuracy was significantly correlated with vowel identification accuracy, suggesting a shared perceptual foundation.

As discussed in detail in Section 5.2, the principal limitation of the present study concerns the confounding of residential context with stimulus dialect, given that all stimuli were produced by a single Southern British English speaker. Compounding this, length of residence, frequency of English use, and degree of exposure to specific dialect features were not measured for the UK group, and the construct of residential context was not operationally defined with sufficient precision. The present findings should therefore be interpreted as group-level associations rather than as evidence of a causal residential context effect, and future research should employ stimuli from multiple dialect varieties together with detailed measures of individual exposure history.

A second design feature warrants acknowledgement. The 14-alternative response set used orthographic forms (bot, bart, bert) anchored in General American grapheme–phoneme conventions; for UK-residents whose ambient input maps these graphemes to non-rhotic and rounded variants, the display may have introduced a mismatch absent for the KO group. Because this orthographic confound and the dialect alignment confound (Section 5.2) make indistinguishable predictions for /ɒ/, /ɑ/, and /ɜ/ in the present design, the two cannot be separated post hoc. Response formats minimizing orthographic mediation (e.g., picture-pointing) would be needed to adjudicate.

Despite these limitations, the present study advances our understanding of L2 vowel perception in three ways. Theoretically, the findings engage the three frameworks in complementary ways: the L1-driven assimilation patterns observed in the confusion matrices instantiate PAM-L2's typology, with back vowel triangular confusion exemplifying multiple Single Category assimilation and dispersed /ʌ/ errors exemplifying Uncategorized assimilation; the selective UK-group advantage for /ʌ/ is consistent with SLM-r's prediction that input quantity drives phonetic category development; and the intermediate position of tense-lax pairs in the hierarchy aligns with L2LP's account of gradual perceptual differentiation under sustained L2 exposure. Methodologically, the application of GLMM to forced-choice identification data demonstrates the value of mixed-effects modeling in detecting group- and vowel-specific patterns that aggregate analyses would obscure. Pedagogically, the stable difficulty hierarchy—anchored by back vowel confusion regardless of residential context—provides a principled basis for prioritizing instructional targets in Korean EFL pronunciation programs.

Acknowledgements

I would like to thank the three anonymous reviewers for their helpful comments and valuable suggestions.

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech, 48*(1), 1-27.

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

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Fleg*e (pp. 13-34). Amsterdam, Netherlands: John Benjamins.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association, 112*(518), 859-877.

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

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34*(3), 372-387.

Cho, M. H., & Jeong, S. (2013). Perception and production of English vowels by Korean learners: A case study. *Studies in Phonetics, Phonology and Morphology, 19*(1), 155-177.

EscuderoNeyra, P. R. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Utrecht University, Utrecht, Netherlands.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(4), 551-585.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore, MD: York Press.

Flege, J. E. (2009). Give input a chance! In T. Piske, & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175-190). Bristol: Multilingual Matters.

Flege, J. E., & Bohn, O. S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge: Cambridge University Press.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437-470.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527-552.

Ingram, J. C., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25*(3), 343-370.

Lee, S., & Cho, M. H. (2020). A comparison of vowel perception between American English and British English by Korean University students. *The Journal of the Korea Contents Association, 20*(7), 203-211.

Lee, S., & Shin, D. J. (2015). Estimating Korean EFL listeners' perception of English vowels with reference to cross-language labelling. *Studies in Phonetics, Phonology and Morphology, 21*(2), 297-321.

Lee, S., & Suh, H. (2016). Perceptual assimilation between Korean and English vowels by Korean learners of English. *Journal of Language Sciences, 23*(4), 83-101.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training

Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242-1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-886.

Magen, H. S., & Blumstein, S. E. (1993). Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics, 21*(4), 387-409.

Park, D., Lee, S., & Cho, M. H. (2010). Interference of L1 phonological processes in English learning. *Journal of English Linguistics, 16*(3), 187-215.

Shin, D. J. (2018). Effect of experience to language on speech-in-noise recognition for Korean L2 speakers. *English Language and Literature Studies, 44*(2), 301-319.

Silva, D. J. (2006). Variation in voice onset time for Korean stops: A case for recent sound change. *Korean Linguistics, 13*(1), 1-16.

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

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24*(2), 245-261.

Yun, G. (2014). Korean listeners' perception of L2 English phoneme contrast. *Studies in Phonetics, Phonology and Morphology, 20*(2), 161-185.

• **Dong-Jin Shin,** Corresponding author
Associate Professor, Dept. of English Language and Literature
Jeonju University
303 Cheonjam-ro, Wansan-gu, Jeonju 55069, Korea
Tel: +82-63-220-3091
Email: djshin@jj.ac.kr
Areas of interest: Acoustic phonetics, L2 acquisition