Korean Journal of English Language and Linguistics, **Vol 26, February 2026, pp. 215-234**
DOI: 10.15738/kjell.26..202602.215

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

Effects of Fluency on the Consistency and Contrast of English Vowel Production in Korean Elementary School Students

Soyoung Yi · Hyunsong Chung (Korea National University of Education)

Received: December 26, 2025
Revised: January 15, 2026
Accepted: January 16, 2026

Yi, Soyoung (First author)
MA Student, Department of English Education
Korea National University of Education
250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju-si, Chungbuk, Korea
Tel: +82-43-230-3554
Email: 200811019@daum.net

Chung, Hyunsong (Corresponding author)
Professor, Department of English Education
Korea National University of Education
250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju-si, Chungbuk, Korea
Tel: +82-43-230-3554
Email: hchung@knue.ac.kr

ABSTRACT

Yi, Soyoung and Hyunsong Chung. 2026. Effects of fluency on the consistency and contrast of English vowel production in Korean elementary school students. *Korean Journal of English Language and Linguistics* 26, 215-234.

The present study examines how fluency relates to (i) within-category consistency and (ii) between-category contrast in English vowel production by Korean elementary school learners. We analysed both reading and spontaneous utterances produced by 45 male and 49 female students across four fluency levels. Vowel consistency was assessed using compactness scores (ellipse area in the F1–F2 space), and vowel contrast was assessed using Pillai scores derived from MANOVA for six vowel contrasts. The results indicate that vowel compactness is strongly structured by vowel category and shows no reliable effect of fluency for either gender, whereas vowel-contrast distinctness increases with fluency at the highest level (F05). These findings suggest that fluency-related improvement in vowel production is expressed primarily through enhanced contrastive organisation rather than through reduced within-category dispersion. Accordingly, pronunciation instruction may be more effective when it emphasises contrastive learning rather than general articulatory ‘accuracy’.

KEYWORDS

vowel formant, compactness score, Pillai score, pronunciation

1. Introduction

Acquiring English vowel contrasts is widely recognised as a persistent challenge for Korean learners of English. One major reason is that, unlike consonants, English vowels are less reliably distinguished by categorical articulatory positions and are instead defined by relative positioning within the vowel space (Celce-Murcia et al. 2010). In addition, L1 phonological structure strongly shapes L2 vowel acquisition. Korean lacks tense–lax contrasts, whereas English contains several such oppositions; consequently, Korean learners frequently assimilate new English vowel categories into existing L1 vowel categories, impeding the formation of distinct L2 categories (Flege 1995). This difficulty is particularly evident in contrasts such as /i/–/ɪ/ and /u/–/ʊ/, but it also extends to other contrasts such as /ɛ/–/æ/, /oʊ/–/ɔ/, and /ɔ/–/ɑ/ (Flege 1995, Lee and Cho 2013). When such distinctions are not acquired early—particularly before puberty—pronunciation outcomes are more likely to fossilise, and subsequent acquisition often requires substantially greater effort.

The importance of early exposure is, however, constrained by institutional realities. Under the 2022 national curriculum, annual English instructional time is limited to between 109 and 163 class hours for each of 3rd and 4th graders and between 164 and 244 class hours for each of 5th and 6th graders (Ministry of Education 2022) in Korea. These figures are markedly lower than those allotted to other core subjects, which makes it difficult to implement systematic pronunciation instruction within regular class hours. As a result, elementary learners may receive insufficient opportunities to engage with difficult vowel contrasts in a sustained and structured manner.

Research on vowel production has traditionally concentrated on adult learners' pronunciation (Cho and Jeong 2013, Hwang 2015, Lee and Rhee 2019, Park and Lee 2024). In contrast, child L2 speech research suggests that although children may have developmental advantages in acquiring new phonetic categories, their productions remain strongly shaped by L1 vowel inventories and limited L2 experience. Within this framework, the Speech Learning Model (SLM/SLM-r) argues that L2 pronunciation development depends on how learners perceptually relate L2 sounds to existing L1 categories: when L2 vowels are perceived as similar to L1 categories, learners are less likely to establish new phonetic categories, resulting in slower progress in production (Flege 1995, Flege and Bohn 2021). Within the Speech Learning Model (SLM/SLM-r), the relative ease of acquiring L2 vowel contrasts depends on how they are mapped onto existing L1 categories. For Korean learners of English, contrasts based on front–back vowel quality may be relatively easier because they align with perceptually salient spectral differences and clearer L1-based organisation. In contrast, high–mid distinctions often involve crowded regions of the English vowel space, making category separation less robust, particularly in early stages of learning. Tense–lax contrasts are frequently the most difficult, as Korean lacks an equivalent phonological distinction, increasing the likelihood of equivalence classification and compressed contrast in production. Importantly, the model also highlights the role of sustained exposure and cumulative experience in supporting phonetic learning over time, while recognising that learning efficiency may gradually decline with age. These assumptions further imply that higher L2 proficiency or fluency is closely associated with the conditions that facilitate phonetic learning, particularly accumulated experience and increased perceptual attunement. Building on this perspective, the present study investigates the development of vowel production in Korean elementary school learners to examine how early-stage L2 experience may shape emerging phonetic patterns in speech production.

Although children have traditionally been assumed to hold a biological advantage over adults in second language (L2) speech acquisition (Lenneberg 1967), empirical studies based on fine-grained analyses of learner speech suggest that child learners also require substantial time and sustained exposure to establish stable L2 vowel categories (Flege 1995, Tsukada et al. 2005). Relatedly, models such as the Speech Learning Model (SLM; Flege 1995) and the Perceptual Assimilation Model (PAM; Best 1993) explain L2 vowel category acquisition as being strongly constrained by how learners perceptually relate L2 vowels to existing L1 categories. In SLM, L2 vowels that are perceived as sufficiently distinct from L1 vowels are more likely to trigger the formation of new phonetic categories, whereas L2 vowels that are interpreted as equivalent to an L1 vowel may undergo equivalence classification, limiting category differentiation and slowing the development of target-like production. PAM similarly predicts that L2 vowel perception and discrimination will depend on learners' assimilation patterns, with greatest difficulty expected when two L2 vowels are mapped onto a single L1 category. Together, these models provide a principled account of why vowel learning often requires extended experience and why certain contrasts—particularly those lacking clear counterparts in the L1—remain persistently challenging.

Longitudinal evidence further indicates that developmental gains are neither immediate nor uniform across vowel contrasts. For example, Tsukada et al. (2005) showed that Korean children residing in North America generally outperformed adults in both vowel discrimination and production, and that children's perceptual accuracy improved significantly as length of residence increased. In production, child learners were also more likely than adults to achieve native-child-like acoustic distances, suggesting that prolonged immersion can support the emergence of clearer phonetic category separation. However, even these child learners showed persistent difficulty during early stages of exposure, particularly for contrasts absent from Korean, such as /ɛ/–/æ/, highlighting that young age alone does not guarantee rapid or target-like development.

Research conducted in Korean classroom contexts reports similar patterns. While both child and adult learners may successfully differentiate relatively salient contrasts (eg /ɑ/–/ɔ/), finer distinctions such as /ɛ/–/æ/ or /ʊ/–/u/ remain problematic even after extended instruction (Kim 2016). Moreover, Korean child learners have repeatedly been shown to produce relatively compressed vowel spaces and ambiguous category boundaries, often exhibiting merger-like tendencies towards nearby Korean vowel categories (Ahn 2019, Jung 2006). Recent acoustic evidence using Euclidean distance further confirms that Korean children's tense–lax vowel contrasts may be significantly smaller than those of native speakers, indicating limited acoustic distinctiveness in production (Lim 2020).

Longitudinal evidence outside Korea likewise supports the view that vowel contrasts are refined gradually rather than undergoing abrupt shifts. Oh et al. (2011), for instance, showed that Japanese elementary learners in immersion settings developed increased spectral separation and more target-like formant values over time, particularly for vowels with no close L1 equivalent. In the Korean context, tense–lax difficulties are well documented. Korean learners often compress multiple English vowels into fewer acoustic categories, leading to overlapping vowel spaces in production (Lee and Cho 2013). Moreover, while age of onset may facilitate development of temporal cues (eg duration), improvements in vowel quality (F1/F2 patterns) appear slower and require extensive exposure and practice (Park and Lee 2024). High-variability phonetic training has been shown to yield modest production improvements in Korean elementary learners, with stronger gains in discrimination than in articulation, suggesting partial decoupling of perception and production at this developmental stage (Hwang and Lee 2015).

Against this background, the present study investigates whether individual differences in fluency relate to how Korean elementary learners organise English vowel categories acoustically. Specifically, we examine whether fluency predicts (i) greater within-category consistency and (ii) clearer between-category contrast in vowel production. This study addresses the following research questions:

1) How does fluency affect the consistency of vowel production in Korean elementary school students' English speech?
2) How does fluency affect the distinctness of vowel contrasts in Korean elementary school students' English production?

2. Methods

2.1 Data

This study used “Children’s English Speech Data for Learning” corpus, developed by and publicly available through AI Hub. We selected speech data from 94 elementary school students who each produced more than 100 utterances. The dataset contains WAV files of both reading and spontaneous speech, along with corresponding labelling files in JSON format, including metadata such as speaker information, recording environment and text labels.

Using a Python script, we converted the JSON files into CSV format. Forced alignment was carried out using the Montreal Forced Aligner to align the transcriptions with the audio recordings. The alignment output was then manually checked and adjusted before formant measurement. We extracted Bark-transformed F1 (F1Bk) and F2 (F2Bk) values at the mid-point of each stressed monophthong using Praat’s “To Formant (burg)” function. Five formants were extracted with a ceiling of 5,500 Hz.

For pre-processing, we used the variables “SpeakerName,” “Gender,” “Fluency” and “LabelText.” The dataset provides five fluency levels (F01–F05), with F05 representing the most fluent speakers. The fluency ratings were obtained using SpeechPro, an English pronunciation assessment tool (MediaZen 2020). SpeechPro categorises fluency into five levels: F01 (Nice), F02 (Good), F03 (Great), F04 (Excellent), and F05 (Awesome). The tool provides automated comparisons and analyses of learners’ speech against native-speaker pronunciation at the sentence, word, and phoneme levels, with particular emphasis on phoneme-level pronunciation accuracy. Because F01 included relatively few speakers, we merged F01 with F02 and therefore analysed four fluency levels (F02–F05). Table 1 summarises the distribution of speakers by fluency level and gender. Table 2 summarises vowel token counts by fluency level and gender.

Table 1. Speaker Counts by Fluency Level and Gender

Fluency	Male	Female	Total
F02	11	13	24
F03	12	12	24
F04	10	12	22
F05	12	12	24
Total	45	49	94

Table 2. Token Counts by Vowel, Fluency and Gender

Vowel	Male				Female			
	F02	F03	F04	F05	F02	F03	F04	F05
i	323	350	245	376	364	334	373	392
ɪ	469	624	506	637	613	616	587	639
ɛ	299	330	279	352	334	327	356	364
æ	554	701	554	808	706	810	728	791
ɑ	321	449	367	436	450	388	533	485
ɔ	159	235	190	264	233	224	232	216
ʊ	94	126	125	136	95	146	128	121
u	309	332	307	339	330	289	353	271

2.2 Measuring Vowel Consistency: Compactness Score

Research on L2 vowel production increasingly emphasises that phonetic learning involves not only movement towards target-like category means but also reduced within-category variability, reflecting greater production consistency and more stable phonetic representations. Kartushina and Frauenfelder (2014) argue that variability reduction is a key indicator of phonetic category formation, consistent with the Speech Learning Model framework (Flege 1995, Flege and Bohn 2021). From this perspective, learning is reflected in tighter clustering of vowel tokens in acoustic space, even when category centres remain non-native-like.

Empirical work has operationalised such consistency using compactness or dispersion measures in the F1–F2 space (eg Bradlow 1995). Oh (2024) further demonstrates the usefulness of compactness-based approaches in examining L1 category precision and its relation to L2 speech learning. Oh (2024) suggests that vowel compactness (i.e., low variability / high precision in vowel production) in a learner's Korean (L1) vowel categories is closely linked to how accurately they can produce English (L2) vowels. Specifically, learners who produced Korean vowels with less dispersion (more compact, stable categories) tended to produce English vowels in a way that more closely approximated the target acoustic space. In other words, greater Korean vowel compactness predicted better English vowel production accuracy across individuals. Oh interprets this relationship as reflecting differences in learners' acoustic sensitivity: speakers with more compact Korean categories may be better able to detect fine-grained acoustic differences and form clearer phonetic boundaries, which supports both consistent L1 production and more target-like L2 vowel realisation. Compactness measures are particularly appropriate for early-stage L2 learners and for children, whose productions often show heightened token-to-token variability due to developing motor control and attentional constraints.

In the present study, vowel compactness is calculated as the area of a one-standard deviation ellipse in the F1–F2 space:

$$CS = \sigma F1 \sigma F2 \pi \quad (1)$$

where $\sigma F1$ and $\sigma F2$ represent one standard deviations of F1 and F2 values for a given vowel category (Kartushina and Frauenfelder 2014). Smaller CS values indicate tighter clustering and greater consistency, whereas larger values indicate greater dispersion and reduced stability.

For each speaker, we computed the centroid (mean) of F1Bk and F2Bk for each vowel and then calculated the compactness score for each vowel category (CSV) using Equation (1). Statistical analysis was conducted in R (version 4.5.2) using linear mixed-effects modelling with the lme4 and lmerTest packages. CSV was modelled as the dependent variable, with Fluency and Vowel as fixed effects. Because each speaker contributed one observation per vowel category, Speaker was included as a random intercept.

```
model <- lmer (CSV ~ Fluency + Vowel + (1 | Speaker), data = df)
```

Separate models were fitted for each gender. Statistical significance for fixed effects was evaluated using Satterthwaite-approximated degrees of freedom as implemented in lmerTest.

2.3 Measuring Vowel Overlap: Pillai Score

Quantifying vowel overlap is central to sociolinguistic and laboratory phonological research. Multiple metrics have been proposed for evaluating overlap, including Euclidean distance, regression-based methods, spectral overlap indices, and the Pillai–Bartlett trace (Mairano et al. 2019, Nycz and Hall-Lew 2014). Pillai scores have gained particular prominence because they provide a multivariate measure of distinction while allowing the inclusion of relevant predictors or controls (Hay et al. 2006, Nycz and Hall-Lew 2014).

The Pillai–Bartlett trace (hereafter Pillai score) originates in the work of Bartlett (1939) and Pillai (1955). In typical phonetic applications, a MANOVA tests whether variation in acoustic measures (eg F1 and F2) is predicted by a categorical factor representing two vowel categories. The statistic ranges from 0 to 1, with lower values indicating greater overlap and higher values indicating clearer separation. Small Pillai values are therefore consistent with merger-like overlap, although overlap alone is not sufficient for a categorical merger diagnosis.

Pillai scores have been applied widely across languages and contrast types (Stanley and Sneller 2023). They have been particularly influential in work on low back vowel overlap in North American English (Hall-Lew 2013, Kendall and Fridland 2017), and Becker (2019) explicitly recommends Pillai scores for quantifying such overlap. Accordingly, Pillai scores are now a standard tool for evaluating whether phonological distinctions are maintained or weakened in production.

Because Pillai values are influenced by sample size, Stanley and Sneller (2023) proposed a practical benchmark for interpreting observed Pillai scores. They derived a sample-size-dependent threshold corresponding to the 95th percentile of Pillai scores expected if two vowel categories were drawn from the same distribution. Denoting the average group size (*n*/2) as *m*, the cutoff is defined as:

$$p_{95} = \frac{e}{m} \quad (2)$$

where e is Euler’s number ($e \approx 2.718281828459045$). For example, when $n = 20$ ($m = 10$), the cutoff is approximately 0.2718. Stanley and Sneller (2023) therefore emphasise that token counts and p-values should be reported alongside Pillai scores to support interpretation.

More recently, Han and Lee (2025a) conducted a MANOVA of six American English monophthong pairs using z-scored F1, F2, and duration as dependent measures. They found that all pairs were acoustically distinct, although /u/–/ʊ/ showed minor and statistically non-significant overlap. In a related study, Han and Lee (2025b) applied a threshold-based Pillai approach to expert-rated pronunciation assessments of Korean learners of English and reported that several contrasts, including /u/–/ʊ/, were often realised non-contrastively. Importantly, expert ratings aligned closely with Pillai-based contrast measures, supporting the use of the metric for evaluating L2 contrast acquisition.

In the present study, we computed Pillai scores separately for each speaker and each vowel contrast: /i/ vs. /ɪ/, /ɪ/ vs. /ɛ/, /ɛ/ vs. /æ/, /u/ vs. /ʊ/, /ʊ/ vs. /ɔ/ and /ɔ/ vs. /ɑ/. For each contrast, vowel category served as the independent variable and F1Bk and F2Bk served as dependent variables in a MANOVA.

Figures 1 and 2 illustrate three vowel contrasts produced by one low-fluency speaker (F02) and one high-fluency speaker (F05). The ellipses represent 2.5 standard-deviation contours for each vowel category. Visual inspection indicates substantial overlap for /i/–/ɪ/ and /ɛ/–/æ/ across fluency levels, whereas /ɪ/–/ɛ/ remains relatively distinct. Combining Pillai values with contrast-specific thresholds derived from token counts provides a more principled basis for evaluating whether contrasts are robustly maintained.

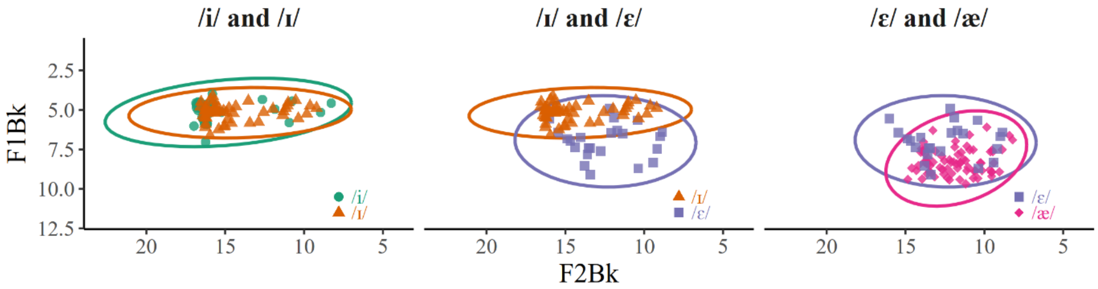

Figure 1. Distributions of /i/-/ɪ/, /ɪ/-/ɛ/ and /ɛ/-/æ/ Tokens in F1Bk-F2Bk Space for Speaker JCS (F02)

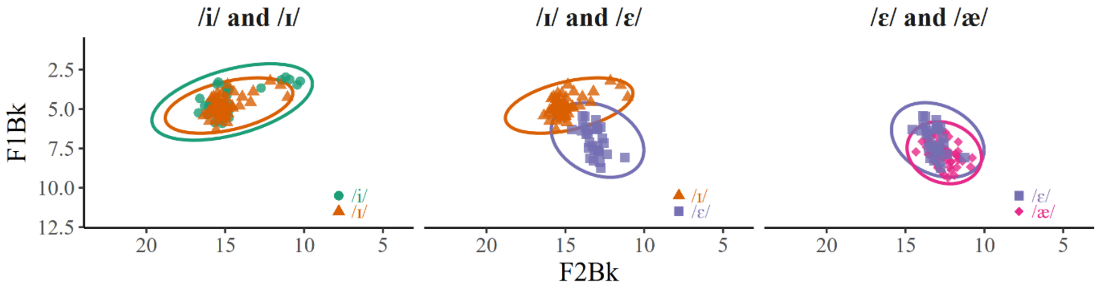

Figure 2. Distributions of /i/-/ɪ/, /ɪ/-/ɛ/ and /ɛ/-/æ/ Tokens in F1Bk-F2Bk Space for Speaker GSC (F05)

All analyses were implemented in Python. For each speaker and contrast, the function removed missing values, verified adequate sample size, fitted a MANOVA model (MANOVA.from_formula (“F1Bk + F2Bk ~ Group”)), and extracted Pillai scores and *p*-values. We then calculated contrast-specific thresholds following Stanley and Sneller (2023). If a Pillai value exceeded the threshold, the contrast was classified as distinct; otherwise, it was classified as merged. We subsequently modelled distinctness using logistic regression with Distinct as the dependent variable and Fluency, Contrast and their interaction as predictors:

```
model <- glm (Distinct ~ Fluency * Contrast, data = df, family = binomial)
```

Fluency and Contrast were treated as categorical predictors, so coefficients are interpreted as log-odds relative to reference categories.

3. Results

3.1 Vowel Consistency

Figure 3 illustrates the vowel ellipses for male speakers across fluency levels, showing the distribution of each vowel category within one standard deviation of the mean F1Bk and F2Bk values. Figure 4 presents the corresponding distributions for female speakers. In both figures, the ellipses represent dispersion in the acoustic space and therefore correspond directly to the compactness score (CSV), with larger ellipses indicating greater within-category variability and smaller ellipses indicating tighter clustering. Table 3 reports the mean CSV values (and standard deviations) for each vowel across fluency levels, separately for male and female speakers.

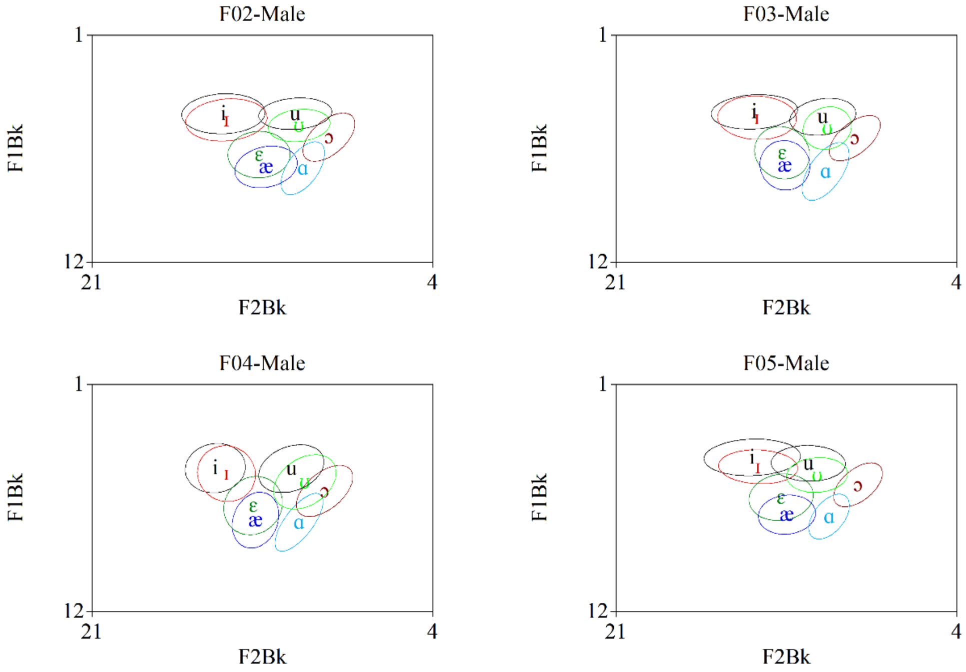
F02-Male
F03-Male
F04-Male
F05-Male

Figure 3. Vowel Ellipses for Male Students Within One Standard Deviation

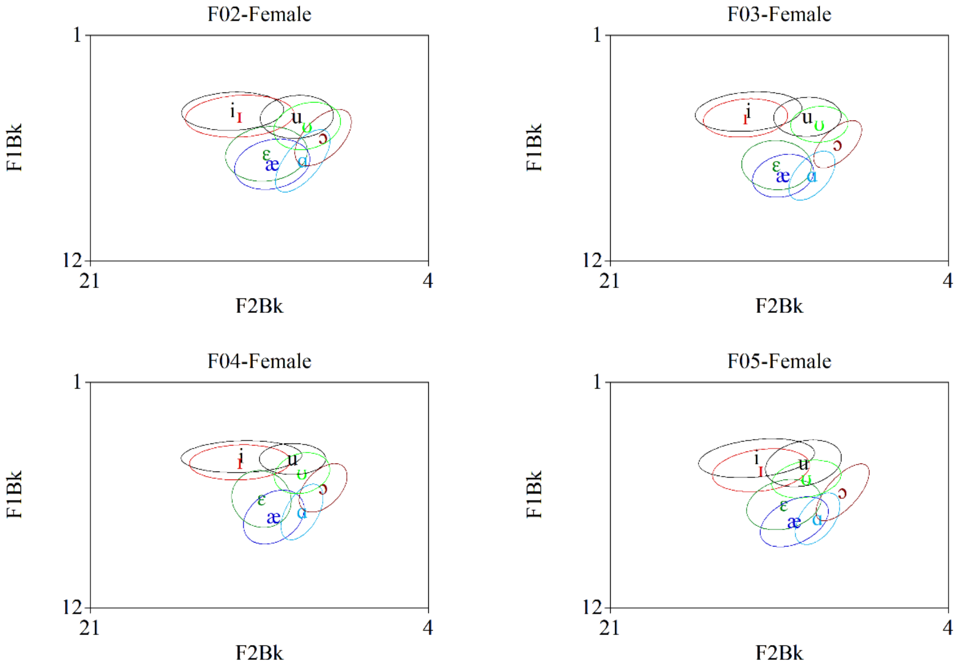
F02-Female
F03-Female
F04-Female
F05-Female

Figure 4. Vowel Ellipses for Female Students Within One Standard Deviation

Table 3. Mean CSV Values by Vowel, Fluency and Gender (Standard Deviation in Parentheses)

Vowel	M				F			
	F02	F03	F04	F05	F02	F03	F04	F05
i	5.01	4.12	3.81	3.91	5.07	5.19	5.80	6.74
	(2.47)	(2.22)	(2.50)	(2.80)	(2.34)	(3.61)	(3.00)	(3.80)
ɪ	4.40	4.40	3.70	3.04	4.94	4.92	4.66	5.26
	(2.67)	(3.28)	(1.90)	(1.53)	(2.50)	(2.89)	(2.72)	(2.80)
ɛ	4.22	3.18	4.30	3.45	5.54	5.18	5.23	4.36
	(2.02)	(1.65)	(2.55)	(2.15)	(1.92)	(2.69)	(2.21)	(2.90)
æ	3.48	2.89	2.81	3.14	5.19	3.48	4.02	4.08
	(1.53)	(1.42)	(1.12)	(1.64)	(1.59)	(1.30)	(1.31)	(2.46)
u	2.63	3.43	3.59	3.00	4.34	3.58	3.07	3.65
	(1.01)	(2.08)	(1.41)	(2.43)	(2.33)	(1.22)	(1.15)	(1.79)
ʊ	2.42	2.08	3.58	2.18	3.70	2.68	3.32	2.94
	(1.14)	(1.71)	(2.35)	(1.06)	(2.11)	(1.44)	(2.93)	(2.62)
ɔ	3.95	2.93	3.29	3.80	4.33	3.59	3.09	4.69
	(1.68)	(.95)	(1.62)	(2.00)	(2.05)	(2.06)	(1.51)	(2.15)
ɑ	3.62	3.66	3.11	2.65	5.38	3.88	3.30	3.20
	(.77)	(1.18)	(1.19)	(.91)	(2.63)	(1.14)	(.90)	(1.15)

Figures 5 and 6 display the relationship between fluency level (F02–F05) and vowel compactness scores (CSV) for eight vowel categories (/i, ɪ, ɛ, æ, u, ʊ, ɔ, ɑ/), presented separately for male and female speakers. The x-axis represents fluency level and the y-axis represents CSV values. For each vowel, individual speaker values are shown alongside the group mean and associated confidence intervals. These figures allow for visual comparison of how vowel compactness varies across fluency levels and across vowel categories within each gender.

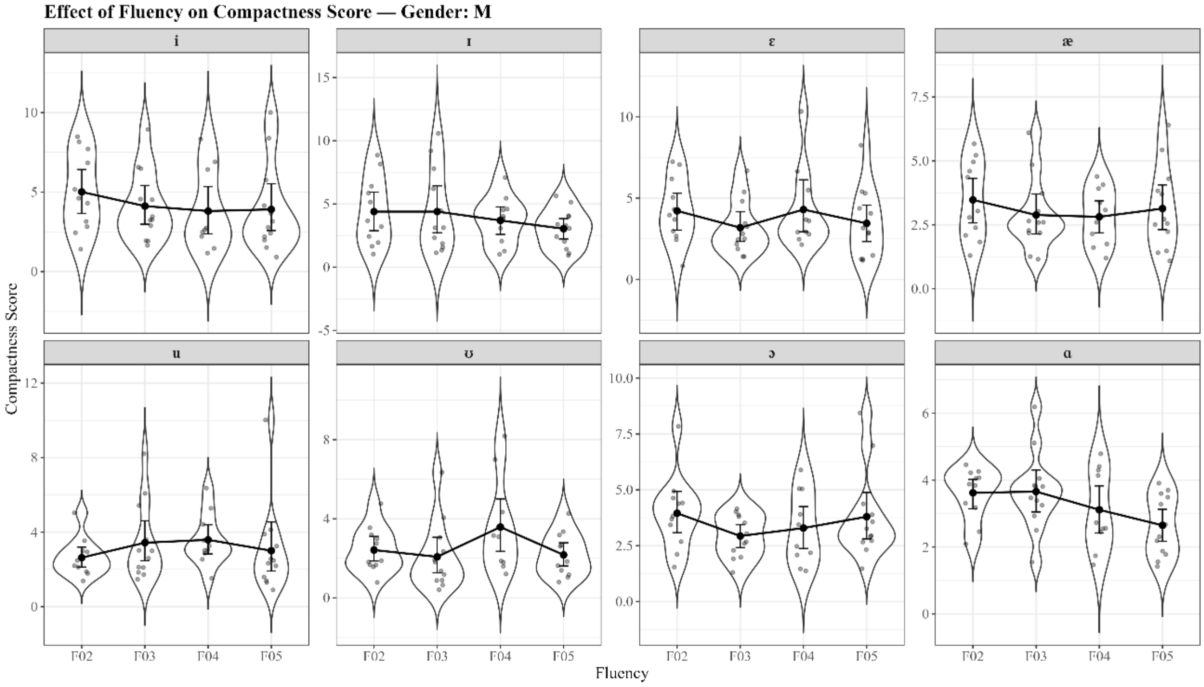

Figure 5. Effect of Fluency on Vowel Compactness Score (CSV) for Male Speakers

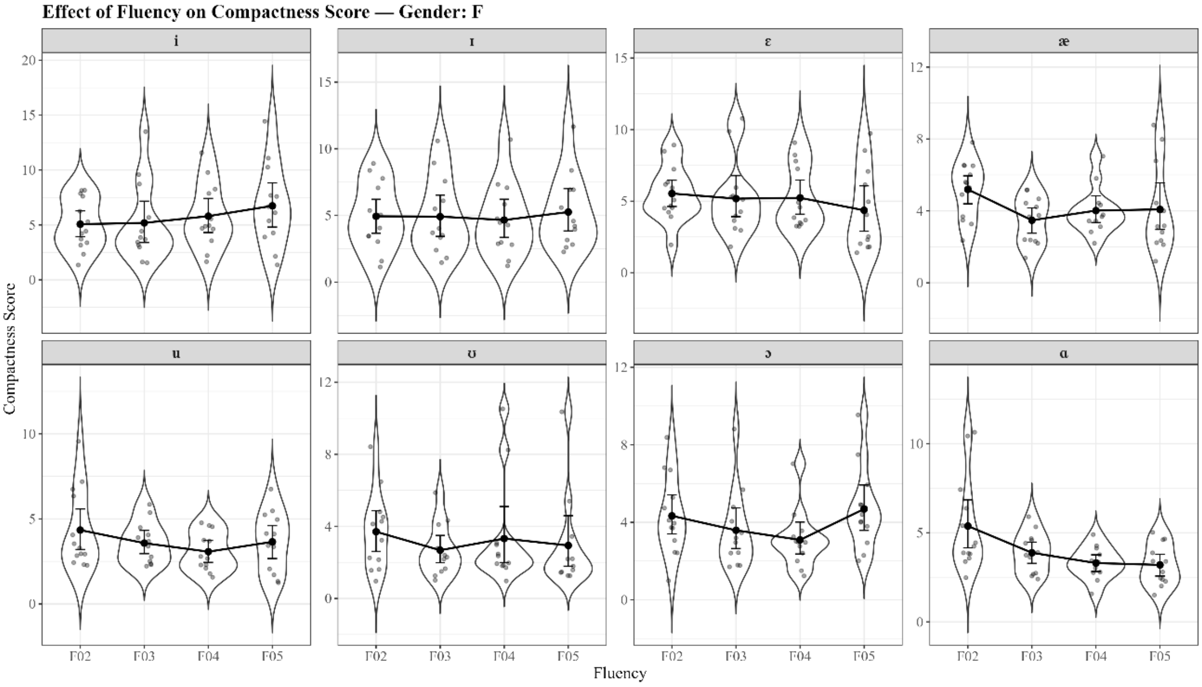

Figure 6. Effect of Fluency on Vowel Compactness Score (CSV) for Female Speakers

Descriptive statistics reveal that for male speakers, mean CSV values generally fall between approximately 2.0 and 5.0 across vowels and fluency levels. At the lowest fluency level (F02), front vowels such as /i/, /ɪ/, and /ɛ/ exhibit relatively high mean CSV values compared to back vowels such as /u/ and /ʊ/, suggesting greater dispersion in the acoustic realisations of front vowels at low fluency. Across fluency levels, however, changes in compactness do not display a consistent monotonic pattern. Instead, compactness appears to vary in a vowel-dependent manner: certain vowels show slight reductions in mean CSV values with increasing fluency (eg /i/, /ɪ/), whereas other vowels remain relatively stable across fluency groups (eg /ɑ/). Standard deviations remain moderate across conditions, indicating substantial inter-speaker variation within each fluency level.

For female speakers, mean CSV values are generally higher than those observed for male speakers across most vowels and fluency levels. At lower fluency (F02), front vowels (/i/, /ɪ/, /ɛ/, /æ/) show particularly large CSV means, whereas back vowels (/u/, /ʊ/) display lower values. Similar to the male data, compactness does not shift uniformly across fluency levels. Instead, the pattern is again vowel-specific. Notably, at the highest fluency level (F05), female speakers show especially high mean CSV values for /i/ and /ɪ/, whereas compactness for other vowels remains comparatively lower. Overall, these descriptive results suggest that within-category variability is more strongly structured by vowel category and gender than by fluency level.

To evaluate these observations statistically, separate linear mixed-effects models were fitted for male and female speakers, with CSV as the dependent variable and fluency and vowel category as predictors (Table 4). In both models, the intercept represents the estimated mean CSV for the reference vowel /i/ at the baseline fluency level (F02). For neither gender did fluency level significantly predict CSV values: for male speakers, none of the fluency contrasts (F03–F05 relative to F02) reached statistical significance, and an equivalent pattern was observed for female speakers. This indicates that, after controlling for vowel category, increases in fluency are not associated with systematic changes in vowel compactness.

Table 4. Results of Linear-Mixed Effects Models for Male and Female Speakers

	Male				Female			
term	estimate	SE	*t*-value	*p*-value	estimate	SE	*t*-value	*p*-value
(Intercept)	4.51	0.42	10.64	0.000	6.16	0.47	13.12	.000
FluencyF03	-0.38	0.50	-0.76	0.454	-0.75	0.57	-1.33	.191
FluencyF04	-0.19	0.53	-0.36	0.717	-0.75	0.57	-1.32	.193
FluencyF05	-0.57	0.50	-1.13	0.263	-0.45	0.57	-0.79	.435
Vowel ɪ	-0.33	0.33	-1.01	0.314	-0.74	0.39	-1.91	.057
Vowel ɛ	-0.46	0.33	-1.39	0.167	-0.60	0.39	-1.54	.124
Vowel æ	-1.13	0.33	-3.43	0.001	-1.47	0.39	-3.79	.000
Vowel u	-1.05	0.33	-3.20	.004	-2.01	0.39	-5.18	.000
Vowel ʊ	-1.69	0.33	-5.13	.029	-2.51	0.39	-6.47	.000
Vowel ɔ	-0.72	0.33	-2.19	.000	-1.75	0.39	-4.51	.000
Vowel ɑ	-0.95	0.33	-2.89	.002	-1.72	0.39	-4.43	.000

In contrast, vowel category exerted strong and consistent effects on CSV for both genders. Relative to the reference vowel /i/, several vowel categories exhibited significantly lower compactness scores. For male speakers, vowels such as /æ/, /ɑ/, /ɔ/, /ʊ/, and /u/ showed significant negative coefficients, with /ʊ/ displaying the largest reduction in CSV. Female speakers showed a comparable but more pronounced pattern, where multiple vowels—including /æ/, /ɑ/, /ɔ/, /ʊ/, and /u/—were associated with significantly lower CSV values than /i/. These results indicate that vowel compactness is shaped primarily by vowel category and that the magnitude of vowel-category effects is generally larger among female speakers.

Taken together, the results suggest that within-category vowel consistency (as measured by CSV) is structured mainly by vowel identity rather than by fluency, and that fluency-related gains in elementary learners may not necessarily be reflected in reduced within-category dispersion.

3.2 Vowel Overlap

Table 5 presents representative Pillai scores derived from MANOVA for two speakers, illustrating how vowel overlap and contrast distinctness are quantified within this framework.

Table 5. Examples of Pillai Score Results for Two Speakers Following MANOVA

Speaker	Gender	Fluency	Contrast	nA	nB	n_total	Pillai	*p*-value	Threshold95	Distinct
CMJ	M	F05	/i/-/ɪ/	36	60	96	0.58173	2.5E-18	0.057265	1
			/ɪ/-/ɛ/	60	64	124	0.069143	0.013105	0.045667	1
			/ɛ/-/æ/	64	76	140	0.528527	4.28E-23	0.039868	1
			/u/-/ʊ/	19	26	45	0.415298	1.28E-05	0.122323	1
			/ʊ/-/ɔ/	26	24	50	0.414474	3.45E-06	0.108731	1
			/ɔ/-/ɑ/	24	23	47	0.400088	1.31E-05	0.116886	1
BYH	F	F03	/i/-/ɪ/	63	86	149	0.06125	0.009912	0.036606	1
			/ɪ/-/ɛ/	86	39	125	0.746393	4.51E-37	0.045305	1
			/ɛ/-/æ/	39	94	133	0.144599	3.9E-05	0.042405	1
			/u/-/ʊ/	23	18	41	0.112904	0.102671	0.133196	0
			/ʊ/-/ɔ/	18	29	47	0.771241	8.06E-15	0.116886	1
			/ɔ/-/ɑ/	29	37	66	0.27012	4.93E-05	0.085535	1

In Table 5, nA and nB represent token counts for the first and second vowel categories in each contrast, while n_total indicates the combined number of observations. The Pillai column reports the multivariate statistic derived from F1Bk and F2Bk values. Threshold95 values were calculated using the sample-size-dependent equation described in Section 2.3. For example, for speaker CMJ's /i/–/ɪ/ contrast (n_total = 96), the mean group size is 48, yielding a threshold of approximately 0.0573. If the Pillai score exceeds this threshold, the contrast is classified as distinct; if not, it is classified as merged. As shown in the table, BYH's /u/–/ʊ/ contrast at fluency level F03 falls below the threshold and is therefore classified as merged, whereas the remaining contrasts for this speaker are classified as distinct.

To examine whether fluency and contrast type predict the likelihood of producing distinct contrasts, separate logistic regression models were fitted for male and female speakers, with Distinct as the dependent variable. Odds ratios (OR) were calculated to facilitate interpretation.

Table 6. Results of the Logistic Regression Model of Pillai Scores for Male Speakers

term	estimate	SE	statistic	p-value	OR
(Intercept)	-2.30	1.05	-2.20	0.03	0.10
FluencyF03	1.61	1.21	1.33	0.19	5.00
FluencyF04	1.90	1.23	1.54	0.12	6.67
FluencyF05	3.40	1.24	2.74	0.01	30.00
Contrast_/ɪ/-/ɛ/	4.61	1.48	3.10	0.00	100.00
Contrast_/ɛ/-/æ/	2.12	1.21	1.75	0.08	8.33
Contrast_/u/-/ʊ/	1.74	1.22	1.43	0.15	5.71
Contrast_/ʊ/-/ɔ/	3.81	1.31	2.91	0.00	45.00
Contrast_/ɔ/-/ɑ/	3.81	1.31	2.91	0.00	45.00
FluencyF03: Contrast_/ɪ/-/ɛ/	14.65	1882.92	0.01	0.99	2312975.86
FluencyF04: Contrast_/ɪ/-/ɛ/	14.37	2062.64	0.01	0.99	1734731.89
FluencyF05: Contrast_/ɪ/-/ɛ/	12.86	1882.92	0.01	0.99	385495.98
FluencyF03: Contrast_/ɛ/-/æ/	-1.43	1.47	-0.97	0.33	0.24
FluencyF04: Contrast_/ɛ/-/æ/	-0.87	1.54	-0.56	0.57	0.42
FluencyF05: Contrast_/ɛ/-/æ/	-1.61	1.58	-1.02	0.31	0.20
FluencyF03: Contrast_/u/-/ʊ/	-0.36	1.50	-0.24	0.81	0.70
FluencyF04: Contrast_/u/-/ʊ/	-0.49	1.54	-0.32	0.75	0.61
FluencyF05: Contrast_/u/-/ʊ/	-2.51	1.51	-1.66	0.10	0.08
FluencyF03 Contrast_/ʊ/-/ɔ/	-2.01	1.59	-1.27	0.21	0.13
FluencyF04: Contrast_/ʊ/-/ɔ/	-2.01	1.66	-1.21	0.22	0.13
FluencyF05: Contrast_/ʊ/-/ɔ/	13.66	1882.92	0.01	0.99	856657.73
FluencyF03: Contrast_/ɔ/-/ɑ/	-1.50	1.64	-0.92	0.36	0.22
FluencyF04: Contrast_/ɔ/-/ɑ/	-1.20	1.80	-0.67	0.50	0.30
FluencyF05: Contrast_/ɔ/-/ɑ/	13.66	1882.92	0.01	0.99	856657.72

The male-speaker model (Table 6) indicates that fluency exerts a strong positive influence on the probability of producing distinct vowel contrasts, although this effect is statistically reliable only at the highest fluency level. Specifically, F05 significantly increases the odds of producing a distinct contrast relative to F02 ($\beta = 3.40$, $p = .006$, OR = 30). This suggests that highly fluent male speakers are approximately 30 times more likely to produce distinct contrasts than those in the lowest fluency group. Fluency levels F03 and F04 show positive coefficients but do not reach statistical significance, indicating that although distinctness may gradually increase, reliable improvement is observed primarily at the highest fluency level.

Contrast type also strongly predicts distinctness. Relative to the reference contrast, the /ɪ/–/ɛ/ contrast greatly increases the odds of distinct production (β = 4.61, *p* = .0019, OR = 100). Similarly, /ʊ/–/ɔ/ and /ɔ/–/ɑ/ also show large and significant effects (β = 3.81, *p* = .0036, OR = 45), indicating that these contrasts are robustly maintained across speakers. By contrast, /ɛ/–/æ/ shows a marginal main effect (β = 2.12, *p* = .080), suggesting a tendency towards clearer distinction but without strong statistical support. The /u/–/ʊ/ contrast does not show a significant main effect, consistent with its reputation as a difficult contrast for Korean learners.

Although interaction terms between Fluency and Contrast were included, none reached statistical significance. Several interaction estimates show extremely large standard errors and implausible OR values, indicating numerical instability likely due to sparse observations or quasi-separation in certain fluency-by-contrast cells. These patterns suggest that the dataset does not provide reliable evidence that fluency systematically moderates the effect of contrast type among male speakers.

Table 7. Results of the Logistic Regression Model of Pillai Scores for Female Speakers

term	estimate	SE	statistic	*p*-value	OR
(Intercept)	-1.20	0.66	-1.83	0.07	0.30
FluencyF03	0.51	0.90	0.57	0.57	1.67
FluencyF04	0.51	0.90	0.57	0.57	1.67
FluencyF05	1.90	0.90	2.11	0.03	6.67
Contrast_/ɪ/-/ɛ/	19.77	1809.05	0.01	0.99	385495977.26
Contrast_/ɛ/-/æ/	1.67	0.87	1.92	0.05	5.33
Contrast_/u/-/ʊ/	0.39	0.89	0.44	0.66	1.48
Contrast_/ʊ/-/ɔ/	1.05	0.86	1.22	0.22	2.86
Contrast_/ɔ/-/ɑ/	2.41	0.93	2.59	0.01	11.11
FluencyF03: Contrast_/ɪ/-/ɛ/	-0.51	2611.15	0.00	1.00	0.60
FluencyF04: Contrast_/ɪ/-/ɛ/	-0.51	2611.15	0.00	1.00	0.60
FluencyF05: Contrast_/ɪ/-/ɛ/	-1.90	2611.15	0.00	1.00	0.15
FluencyF03: Contrast_/ɛ/-/æ/	0.12	1.26	0.09	0.93	1.12
FluencyF04: Contrast_/ɛ/-/æ/	17.59	1882.92	0.01	0.99	43368297.31
FluencyF05: Contrast_/ɛ/-/æ/	-0.76	1.32	-0.58	0.56	0.47
FluencyF03: Contrast_/u/-/ʊ/	0.64	1.23	0.52	0.60	1.89
FluencyF04: Contrast_/u/-/ʊ/	0.30	1.23	0.24	0.81	1.35
FluencyF05: Contrast_/u/-/ʊ/	-0.75	1.23	-0.61	0.54	0.47
FluencyF03 Contrast_/ʊ/-/ɔ/	2.04	1.49	1.37	0.17	7.70
FluencyF04: Contrast_/ʊ/-/ɔ/	0.34	1.22	0.28	0.78	1.40
FluencyF05: Contrast_/ʊ/-/ɔ/	0.65	1.49	0.44	0.66	1.92
FluencyF03: Contrast_/ɔ/-/ɑ/	0.68	1.53	0.45	0.65	1.98
FluencyF04: Contrast_/ɔ/-/ɑ/	16.85	1882.92	0.01	0.99	20816782.79
FluencyF05: Contrast_/ɔ/-/ɑ/	-0.70	1.53	-0.46	0.65	0.49

The female-speaker model (Table 7) reveals a broadly similar pattern, though fluency effects are weaker than in the male data. The highest fluency level (F05) significantly increases the odds of distinct production relative to F02 ($\beta = 1.90$, $p = .035$, OR ≈ 6.67), indicating that highly fluent female speakers are over six times more likely to produce distinct contrasts than low-fluency speakers. Fluency levels F03 and F04 show small positive coefficients but do not approach significance, again suggesting that meaningful fluency-related gains emerge primarily at the highest level.

Contrast effects show a somewhat different profile. The /ɔ/–/ɑ/ contrast exhibits the clearest and strongest effect ($\beta = 2.41$, $p = .0097$, OR ≈ 11.11), indicating that this contrast is reliably distinct among female speakers. The /ɛ/–/æ/ contrast shows a marginal effect ($\beta = 1.67$, $p = .055$), suggesting a tendency towards increased distinctness but not at conventional significance levels. Other contrasts do not show significant effects.

As in the male model, interaction terms are uniformly non-significant and numerically unstable. Several interactions are associated with extremely large standard errors and inflated OR estimates, which is consistent with sparse cell counts and quasi-separation rather than meaningful interaction effects. Therefore, there is no reliable evidence that fluency moderates contrast type effects for female speakers.

To ensure comparability across speakers, we equalised sample sizes using contrast- and gender-specific downsampling, following Stanley and Sneller (2023). For each contrast within each gender, the minimum token count across speakers was used as the target downsample size (Table 8). We then generated 1,000 bootstrap samples for each speaker and computed Pillai scores for each sample. The mean of the 1,000 Pillai scores was used as the adjusted Pillai value for that speaker. Correlations between full-dataset Pillai scores and bootstrapped means were high (males: .9830, females: .9477), indicating that the downsampled estimates closely approximate the original values while reducing sample-size bias.

Table 8. Downsample Sizes by Contrast and Gender

Vowel contrast	Male	Female
/i/-/ɪ/	49	16
/ɪ/-/ɛ/	38	21
/ɛ/-/æ/	34	26
/u/-/ʊ/	14	9
/ʊ/-/ɔ/	13	10
/ɔ/-/ɑ/	17	18

Finally, log-transformed Pillai scores were modelled as a function of fluency for each vowel contrast (Figures 7 and 8).

Figures 7 and 8 display log-transformed Pillai scores for six vowel contrasts (/i/–/ɪ/, /ɪ/–/ɛ/, /ɛ/–/æ/, /u/–/ʊ/, /ʊ/–/ɔ/, and /ɔ/–/ɑ/) across four fluency groups (F02–F05). Each point represents a speaker’s mean Pillai value derived from 1,000 bootstrap samples using the contrast- and gender-specific downsample sizes. The dotted horizontal line marks the 95% null threshold expected for a fully merged vowel pair given the relevant sample size. Values above this line indicate statistically reliable vowel separation, whereas values below the line may reflect substantial overlap but do not necessarily imply a categorical merger. For example, in the /i/–/ɪ/ contrast in Figure 7, the fitted regression line shows a positive trend, indicating that distinctiveness increases with greater fluency. The dotted horizontal line corresponds to $\log(.111) = -2.199$, which is used here as a benchmark for contrast maintenance. Pillai values above this threshold suggest that a speaker’s productions exhibit meaningful multivariate separation between the two vowels, whereas values falling below the threshold indicate weaker or collapsed contrast realization. Although considerable variability is observed at each fluency level, the rising regression line indicates that, as fluency increases, a larger proportion of speakers produce Pillai scores above the threshold, reflecting more consistent and robust differentiation of the /i/–/ɪ/ contrast.

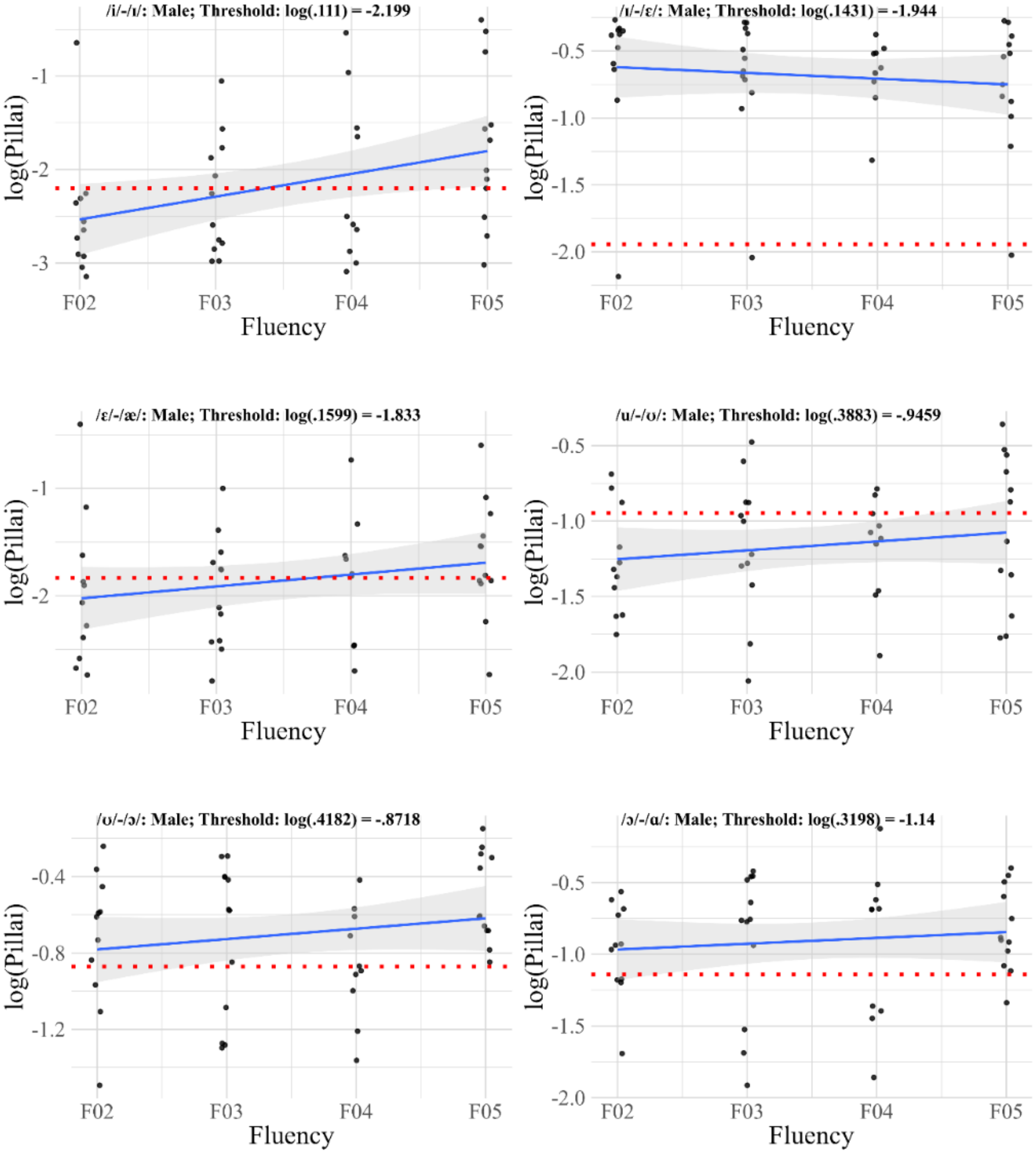

Figure 7. Log-transformed Pillai Scores by Fluency Level with Predicted Regression Lines for Each Vowel Contrast (Male Speakers)

Among male speakers, for the /ɪ/–/ɛ/, /ʊ/–/ɔ/, and /ɔ/–/ɑ/ contrasts, Pillai values exceed the threshold across fluency levels, indicating broadly maintained vowel distinctions. The /ɪ/–/ɛ/ contrast displays a slightly downward slope, suggesting that the degree of distinction may diminish as fluency increases, even though it remains above the threshold. Pillai values cross the threshold upward as fluency increases for the /i/–/ɪ/ and /ɛ/–/æ/ contrasts, indicating that these contrasts become more clearly distinct at higher fluency levels. In the /u/–/ʊ/ contrast, the fitted line lies below the threshold across fluencies, meaning that this contrast does not reach distinct status and remains relatively weakly distinguished for male speakers.

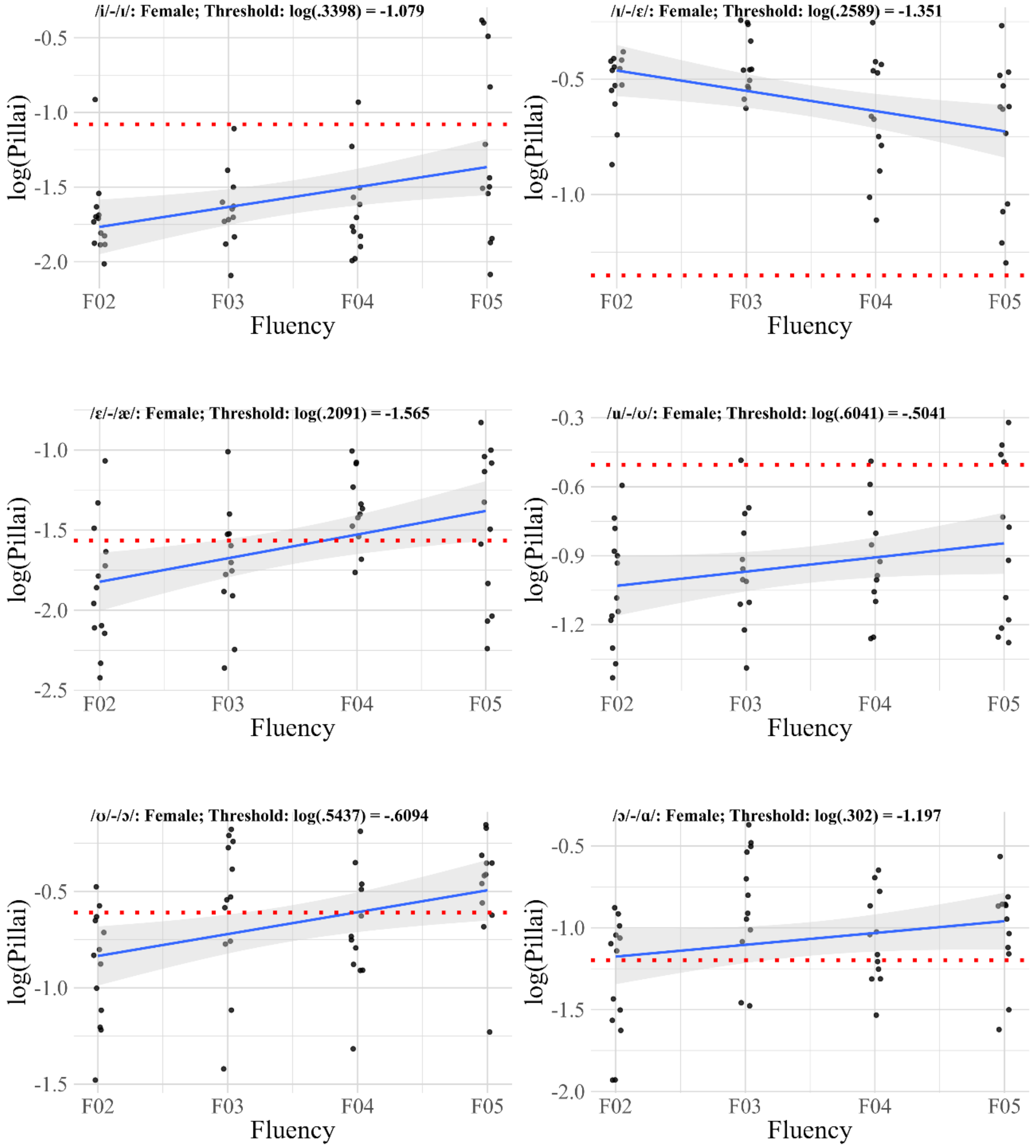

Figure 8. Log-transformed Pillai Scores by Fluency Level with Predicted Regression Lines for Each Vowel Contrast (Female Speakers)

Among female speakers, the /i/–/ɛ/ contrast lies well above the threshold, meaning that most speakers have little difficulty maintaining a clear distinction in this contrast. However, as in the male data, the regression slope is slightly downward and variability is quite large in F05, suggesting that there are no additional gains in distinctiveness for this contrast as fluency increases. The fitted line for the /ɔ/–/ɑ/ contrast is just above the threshold, indicating that the distinction in this contrast is also successfully maintained. In contrast, for the /i/–/ɪ/ and /u/–/ʊ/ contrasts, the fitted lines fall well below the null threshold, indicating comparatively close vowel categories and a weaker separation between the two vowels. For the /ɛ/–/æ/ and /ʊ/–/ɔ/ contrasts, the fitted lines cross the threshold upward as fluency increases, suggesting that these contrasts become more distinct at higher fluency levels.

Overall, while both male and female speakers maintain clear distinctions in the /ɪ/–/ɛ/ and /ɔ/–/ɑ/ contrasts, the /u/–/ʊ/ contrast appears consistently weak across genders, and the /i/–/ɛ/ contrast shows a moderate reduction in acoustic distance.

4. Discussion and Conclusion

This study examined whether fluency is associated with (i) within-category consistency of English vowels and (ii) between-category distinctness of vowel contrasts in Korean elementary school learners. Using distribution-based measures of vowel compactness (CSV) and Pillai-based overlap, the results reveal a notable dissociation between these dimensions of phonetic learning. Compactness was strongly structured by vowel category and exhibited no reliable effect of fluency for either gender. In contrast, vowel-contrast distinctness increased with fluency at the highest level (F05). This pattern accords with the Speech Learning Model (SLM/SLM-r), which posits that successful L2 phonetic acquisition depends on establishing categories sufficiently distinct from L1 categories, and that progress may occur unevenly across different aspects of phonetic organisation (Flege 1995, Flege and Bohn 2021). In the present data, fluency appears to be more strongly related to contrast maintenance than to the stabilisation of vowel categories.

The absence of a significant fluency effect on CSV suggests that increased fluency does not necessarily entail reduced within-category dispersion for elementary learners. Compactness reflects within-category variability, which in child speech may be influenced not only by phonological representations but also by developmental constraints such as immature articulatory control, attentional variability, and limitations in motor planning (Kartushina and Frauenfelder 2014). Previous work similarly indicates that children may show improvements in temporal measures and global fluency before demonstrating adult-like refinement of spectral precision (Park and Lee 2024). Moreover, Korean learners' production is shaped by L1 phonological structure, particularly the absence of tense–lax distinctions, which may contribute to persistent dispersion in L2 vowel categories even with increasing experience (Flege 1995, Lee and Cho 2013). From this perspective, the robust vowel-category effects observed in the CSV models likely reflect stable differences in category learnability and articulatory demands rather than fluency-driven improvement in consistency.

In contrast, the Pillai-based analyses demonstrate that fluency is meaningfully related to vowel contrast distinctness, particularly at the highest fluency level. For both genders, F05 significantly increased the likelihood of producing distinct contrasts. This suggests that when learners reach a sufficiently advanced level of fluency, they are more able to maintain and implement contrastive organisation in multidimensional acoustic space. Importantly, these fluency-related gains were contrast-specific. Contrasts known to be challenging for Korean learners—especially /u/–/ʊ/ and /i/–/ɪ/—exhibited weaker separation and more delayed improvement, consistent with extensive evidence that such contrasts are perceptually and articulatorily assimilated into single L1 categories (Flege 1995, Han and Lee 2025b, Lee and Cho 2013). Furthermore, the use of sample-size-adjusted Pillai thresholds following Stanley and Sneller (2023) strengthens interpretation by distinguishing genuine acoustic overlap from overlap that may be exaggerated by small or imbalanced token counts. Collectively, these results indicate that fluency contributes to contrast maintenance, but primarily at the highest fluency level and most clearly for contrasts that are acoustically tractable.

These findings have direct implications for pronunciation instruction under the constraints of the Korean national curriculum, which allocates relatively limited instructional hours to English compared with other subjects (Ministry of Education 2022). Because fluency-related gains in the present study were expressed primarily in between-category distinctness rather than within-category compactness, instructional approaches may be more effective when they prioritise contrastive learning, such as minimal-pair instruction and targeted contrast training, rather than a general emphasis on articulatory "accuracy." High-variability phonetic training has been shown to facilitate perceptual improvement and modest production gains in young learners, particularly when training targets contrasts that are strongly susceptible to L1 interference (Hwang and Lee 2015). The persistent weakness of the /u/–/ʊ/ contrast across fluency levels further suggests that tense–lax contrasts require sustained and explicit instructional attention rather than reliance on exposure or general fluency development.

Several limitations should be acknowledged. Although the dataset is large for child speech research, some fluency–contrast combinations were sparse, resulting in numerical instability and quasi-separation in interaction terms in the logistic models. This is a recognised issue in sociophonetic modelling of categorical outcomes (Nycz and Hall-Lew 2014). Future research could address this limitation through more balanced elicitation designs or by applying modelling techniques better suited to sparse data. In addition, although compactness and Pillai scores capture complementary aspects of production, they do not directly measure perceptual accuracy. Integrating perception-based measures would allow future studies to more directly examine the relationship between discrimination and production in child L2 phonological development (Flege and Bohn 2021).

Despite these limitations, the present study demonstrates that in Korean elementary learners, fluency is more closely related to between-category phonological organisation than to within-category stabilisation. This finding reinforces the view that early L2 vowel acquisition involves gradual, contrast-dependent restructuring under strong L1 influence and underscores the importance of systematic, sustained, and contrast-focused pronunciation instruction during the elementary school years, before phonological patterns become fossilised (Flege 1995).

References

Ahn, S. 2019. *A Study on English Vowel Production and Awareness of Korean Elementary School Students*. Unpublished MA thesis, Pusan National University, Korea.

Bartlett, M. S. 1939. A note on tests of significance in multivariate analysis. *Mathematical Proceedings of the Cambridge Philosophical Society* 35(2), 180-185.

Becker, K. 2019. Introduction. In K. Becker ed., *The Low-Back-Merger Shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and Short Front Vowel Shifts Across North America*, 1-30. Publication of the American Dialect Society 104. Durham, NC: Duke University Press.

Best, C. T. 1993. Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage and J. Morton, eds., *Developmental Neurocognition: Speech and Face Processing in the First Year*, 289-304. Dordrecht: Kluwer Academic.

Bradlow, A. R. 1995. A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America* 97, 1916-1924.

Celce-Murcia, M., D. M. Brinton and J. M. Goodwin. 2010. *Teaching Pronunciation: A Course Book and Reference Guide* (2nd edition). Cambridge: Cambridge University Press.

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

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*, 233-277. Timonium, MD: York Press.

Flege, J. E. and O.-S. Bohn. 2021. The revised Speech Learning Model (SLM-r). In R. Wayland, ed., *Second Language Speech Learning: Theoretical and Empirical Progress*, 3-83. Cambridge: Cambridge University Press.

Hall-Lew, L. 2013. ‘Flip-flop’ and mergers-in-progress. *English Language & Linguistics* 17(2), 359-390.

Han, Y. and J.-K. Lee. 2025a. Measuring vowel contrasts revisited. In *Proceedings of the 2025 Spring Conference of the Korean Society of Speech Sciences*, 38-39.

Han, Y. and J.-K. Lee. 2025b. Testing vowel contrasts in L2 English: A threshold-based Pillai score approach. In *Proceedings of the 2025 Spring Conference of the Korean Society of Speech Sciences*, 75-77.

Hay, J., P. Warren and K. Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4), 458-484.

Hwang, H. and H.-Y. Lee. 2015. The effect of high variability phonetic training on the production of English vowels and consonants. In *Proceedings of 18th International Congress of Phonetic Sciences*. Available online at https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0466.pdf

Hwang, I. 2015. An acoustic analysis of English vowels produced by Korean university students. *The Linguistic Association of Korea Journal* 23(1), 65-90.

Jung, S. J. 2006. *A Comparative Analysis on English Vowels by Measuring Formant Frequencies: A Case of Korean Elementary School Students*. Unpublished MA thesis, Seoul National University of Education, Korea.

Kartushina, N. and U. H. Frauenfelder. 2014. On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology* 5, 1-17.

Kendall, T. and V. Fridland. 2017. Regional relationships among the low vowels of U.S. English: Evidence from production and perception. *Language Variation and Change* 29(2), 245-271.

Kim, M. 2016. *A Comparative Study of English Vowels Produced by Korean Child and Adult Learners*. Unpublished MA thesis, Chosun University, Korea.

Lee, K.-Y. and M.-H. Cho. 2013. Production of English vowels by Korean learners. *The Journal of the Korea Contents Association* 13(9), 495-503.

Lee, S. and S. C. Rhee. 2019. The relationship between vowel production and proficiency levels in L2 English produced by Korean EFL learners. *Phonetics and Speech Sciences* 11(2), 1-13.

Lenneberg, E. H. 1967. *Biological Foundations of Language*. New York, NY: Wiley.

Lim, Y.-S. 2020. Production of English tense and lax vowels by young speakers of Korean dialects. *The Journal of Education* 40(1), 157-172.

Mairano, P., C. Bouzon, M. Capliez and V. de Iacovo. 2019. Acoustic distances, Pillai scores and LDA classification scores as metrics of L2 comprehensibility and nativelikeness. In *Proceedings of 19th International Congress of Phonetic Science*. ffhal-03046802.

MediaZen. 2020. *Children’s English Speech Data for Learning: Data Documentation and Category Definitions* (in Korean). National Information Society Agency. Available online at https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=541

Ministry of Education. 2022. *2022 Revised National Curriculum: English*. Sejong: Republic of Korea.

Nycz, J. and L. Hall-Lew. 2014. Best practices in measuring vowel merger. In *the Proceedings of Meetings on Acoustics: The Acoustical Society of America through the American Institute of Physics*. 1st ed. 20, 1-19.

Oh, G. E., S. Guion-Anderson, K. Aoyama, J. E. Flege, R. Akahane-Yamada and T. Yamada, T. 2011. A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics* 39(2), 156-167.

Oh, S. (2024). *Individual Differences of L1 Category Precision in L2 Speech Learning*. Unpublished doctoral dissertation, The University of Wisconsin-Milwaukee, USA.

Park, M. S. and J. K. Lee. 2024. Korean ESL learners' production of English vowel contrasts: Developmental variations in L2 sound learning. *Korean Journal of English Language and Linguistics* 24, 1318-1332.

Pillai, K. C. S. 1955. Some new test criteria in multivariate analysis. *The Annals of Mathematical Statistics* 26(1), 117-121.

Stanley, J. A. and B. Sneller. 2023. Sample size matters in calculating Pillai scores. *The Journal of the Acoustical Society of America* 153, 45-67.

Tsukada, K., D. Birdsong, E. Bialystok, M. Mack, H. Sung and J. 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.

Examples in: English
Applicable Languages: English
Applicable Level: Elementary