



Visual identity cues in L2 Korean pronunciation assessment^{*}

Shiyu Zhang^{**a} · Drew Crosby^{***b} · Jiyoung Shin^a
(Korea University^a · The University of Suwon^b)

Zhang, Shiyu, Drew Crosby, and Jiyoung Shin. 2026. Visual identity cues in L2 Korean pronunciation assessment. *Linguistic Research* 43(Special Edition): 1-40. Pronunciation assessment is generally expected to reflect learners' speech accuracy, yet growing evidence suggests that listeners' judgments may also be shaped by social information about the speaker. This study examines whether visual identity cues influence pronunciation evaluation in Korean language education and whether their effects vary across learner voices from different national backgrounds. Using a matched-guise design, speech produced by advanced Korean learners from China, Japan, Vietnam, and the United States was presented under two visual identity conditions: the same learner speech was paired either with a Korean visual identity cue or with a visual identity cue matching the speaker's nationality. Crucially, the acoustic signal was identical across the two conditions, so only the visual identity information differed. Twenty-one Korean language education evaluators rated the accuracy of the learners' pronunciation. A cumulative link mixed-effects model revealed a significant interaction between visual identity cue and learner nationality, showing that the relationship between visual identity cues and pronunciation ratings differed across the four learner voices rather than producing a uniform overall effect. This indicates that visual identity cues may shape the evaluation of identical speech differently depending on the particular speaker-image pairing. These findings suggest that pronunciation evaluation is not determined solely by the acoustic signal and highlight the potential role of socially mediated expectations in professional pronunciation assessment. (Korea University · The University of Suwon)

Keywords pronunciation assessment, visual identity cues, matched-guise, social cognition, reverse linguistic stereotyping, Korean language education, L2 Korean

* We sincerely thank all participants, Jeffrey Holliday, Dongsu Choi, Jeongho Ko, Jihye Yoon, Yoonjeong Seo, Rok Sim, Jong Hyuk Lee, and the two anonymous reviewers for their valuable contributions and constructive feedback on this study.

** First author

*** Corresponding author

1. Introduction

In second language speech research, learners' speech quality is typically evaluated in two ways. The first relies on objective acoustic analyses, in which learners' speech is assessed by measuring acoustic properties and comparing them with those of native speakers. For example, vowel studies typically examine acoustic measures such as formant frequencies (F1/F2) or vowel-distance measures (Kim 2018; Choe and Ahn 2026), whereas stop production is commonly evaluated using acoustic parameters such as voice onset time (VOT) and fundamental frequency (F0) (Oh 2018). The second approach relies on perceptual evaluations by listeners, in which native speakers, language teachers, or other evaluators assess learners' speech using measures such as pronunciation accuracy, comprehensibility, intelligibility, and accentedness (Munro and Derwing 1995, 1999; Munro 2008; Lim and Seo 2011; Kim and Lee 2019; Jung and Holliday 2026). However, an accumulating body of research suggests that speech perception is not a purely objective acoustic process but is also influenced by listener perceptions and expectations regarding the speaker's social identity (e.g., Rubin 1992; Strand 1999; Kang and Rubin 2009; McGowan 2015; Ko 2026). In other words, pronunciation evaluation may reflect not only the quality of the speech itself, but also listeners' social impressions of and expectations toward particular groups.

Although previous research has demonstrated that speech perception possesses clear social-cognitive properties, this issue has received relatively limited attention in the field of L2 pronunciation evaluation, particularly within Korean language education contexts. Meanwhile, existing Korean pronunciation research has largely focused on the pronunciation accuracy of learners or the acoustic characteristics of their speech (e.g., Cho 2004; Zhao 2006; Kim and Kim 2015; Ha and Lee 2016; Lee and Jung 2020; Cho 2024), leaving the potential influence of social cognition on listeners' evaluations underexplored.

Building on this line of research, the present study focuses specifically on visual identity cues (specifically visual nationality cues), which we define as visually presented identity-related information that may shape listeners' perception of a speaker's identity. This operationalization is narrower than the broader concept of social identity discussed in previous studies. Accordingly, the present study aims to investigate whether changing only these visual identity cues, while keeping the speech signal constant, affects native Korean language education evaluators' evaluations of the same

pronunciation. More broadly, the study examines whether pronunciation evaluation may be influenced by socially meaningful visual information about learners, even when the acoustic signal remains unchanged.

2. Literature review

2.1 Social cognition and speech perception

The theoretical basis for this study lies in research showing that speech perception is shaped not only by acoustic signals but also by social cognition. According to social cognition theory (Fiske and Taylor 1991), individuals rely on pre-existing categorical knowledge and stereotypes when processing information and making rapid judgments about others. This cognitive mechanism also operates in language use and perception, such that listeners' interpretation and evaluation of speech are influenced not only by the acoustic signal itself but also by the speaker's perceived social identity and the associated social knowledge. Within this broader social-cognitive framework, previous studies have approached the influence of social information on speech perception from several related perspectives, including reverse linguistic stereotyping (Rubin 1992; Kang and Rubin 2009), raciolinguistic ideologies (Flores and Rosa 2015; Hughes 2020, 2023), and visual–auditory congruence (McGowan 2015; Gnevsheva 2018).

Before reviewing these lines of research in detail, we briefly discuss several recurring perceptual constructs, since previous studies have used different listener-based measures of L2 speech perception. *Accentedness* refers to “[t]he extent of difference perceived by speakers of one linguistic variety when listening to speakers of other varieties” (Derwing and Munro 2015: 175), whereas *intelligibility* concerns “the extent to which a speaker’s message is actually understood by a listener” (Munro and Derwing 1995: 76). *Comprehensibility*, in contrast, concerns “the ease or difficulty a listener experiences in understanding an utterance” (Derwing and Munro 2015: 5). Although these dimensions are related, they capture different aspects of listeners’ responses to L2 speech and are therefore not interchangeable.

Early evidence for the influence of social identity on language evaluation comes from the matched-guise experiments of Lambert et al. (1960). Their study

demonstrated that listeners' evaluations changed even when the speech remained identical, provided that the perceived social identity of the speaker changed. This finding laid the foundation for subsequent research examining the relationship between language attitudes, social identity, and speech perception.

Building on this work, Rubin (1992) demonstrated that even when listeners heard identical English speech, simply manipulating the speaker's racial cues influenced not only perceptions of accentedness but also speech comprehension itself. Subsequently, Kang and Rubin (2009) introduced the concept of Reverse Linguistic Stereotyping (RLS), arguing that listeners' expectations about a speaker's social identity can distort the perception of speech itself. In other words, listeners do not always perceive non-native speech accurately, but instead often "hear" what they expect to hear.

More recently, Hughes (2020, 2023) examined socially differentiated pronunciation evaluation from the perspective of raciolinguistic ideologies and the concept of the "white listening subject" (Flores and Rosa 2015). Hughes argues that speech evaluation is not merely an immediate response to the acoustic signal but is also shaped by socially constructed language ideologies. In educational contexts, even teachers with linguistic training and extensive assessment experience may be influenced by non-linguistic identity cues, such as race, gender, and clothing, when evaluating pronunciation. This perspective suggests that bias in pronunciation evaluation cannot be attributed solely to learners' speech itself but should also be understood in relation to evaluators' social cognition and institutionalized language ideologies.

A related but distinct line of research focuses on the congruence between visual and auditory information rather than on evaluative bias itself. McGowan (2015) and Gnevsheva (2018) demonstrate that listeners' judgments are also shaped by the degree of congruence between visual identity cues and auditory information. Specifically, McGowan (2015) found that listeners transcribed Chinese-accented English in noise more accurately when the speech was paired with a congruent Chinese face than with an incongruent Caucasian face. Similarly, Gnevsheva (2018) found that when listeners saw a Caucasian speaker, they generally did not expect to hear a strong foreign accent; consequently, when foreign-accented speech was paired with a Caucasian face, the mismatch between expectation and speech made the accent more salient. These findings suggest that visual identity cues do not simply produce uniformly positive or negative evaluations; rather, they influence speech perception by shaping listeners' expectations about how particular speakers are expected to sound.

These studies used different perceptual measures, including accentedness judgments, speech comprehension, and transcription accuracy; nevertheless, they collectively demonstrate that non-acoustic social information can influence how listeners perceive and evaluate L2 speech. Although the studies discussed above have primarily been conducted in English-speaking contexts, similar issues have recently begun to receive attention in Korean. Korean perception studies provide emerging, though not entirely uniform, evidence that visual identity cues can influence listeners' evaluations of accented Korean. Preliminary evidence from Jung and Idemaru (2025), reported in a conference abstract, investigated how social information influences Korean listeners' perception of English-accented Korean. They paired identical English-accented Korean speech with three different identity cues: a Caucasian guise described as an American learner of Korean, an Asian guise described as a Korean American, and a silhouette described as a foreign learner of Korean. Their preliminary findings indicated that both intelligibility and accentedness judgments varied depending on the social identity associated with the same speech, with the Caucasian American learner guise receiving more favorable evaluations than the Korean American guise. However, Graillot and Oh (2025) suggest that visual-cue effects in the Korean context may be conditional rather than uniform. In their study, Korean listeners rated Chinese- and English-accented Korean under matching-ethnicity, mismatching-ethnicity, and no-visual identity cue conditions. Speaker proficiency emerged as the strongest predictor of accentedness ratings, and English-accented Korean was rated as more foreign-accented than Chinese-accented Korean, whereas visual condition did not show a significant main effect. Meanwhile, a significant interaction between visual mismatch and low proficiency suggested that visual-auditory mismatch may still amplify accentedness judgments under some conditions. Taken together, these studies suggest that visual identity cues can affect Korean accented-speech perception, but that their effects may vary depending on task type, perceived speaker identity, speaker proficiency, and the degree of congruence between visual and auditory information.

Related Korean evidence also suggests that visual identity cues may affect not only listeners' evaluations of L2 Korean, but also native Korean speakers' speech behavior toward learners: In a matched-guise study of foreigner-directed speech (FDS), Kim (2026) paired L1-English Korean learner voices differing in proficiency with photographs of either white¹ or East Asian people and examined how native Korean

speakers modified their speech in a route-description task. The results showed that speakers generally produced more exaggerated FDS toward lower-proficiency learners, but that racialized visual cues also played a role: white learner guises elicited larger vowel spaces, slower speech, and, for some measures, more pause-related adjustment than East Asian learner guises. These findings suggest that Korean speakers' expectations about learners' linguistic needs may be shaped not only by actual proficiency cues, but also by racialized visual cues. Importantly, if visual identity cues associated with an L2 learner can shape how Korean speakers modify their own speech when addressing that learner, the same type of visual identity cues may also plausibly influence how they evaluate the learner's speech.

Beyond pronunciation itself, Korean language education has also been shown to reflect ideologies about who Korean learners are and how they are expected to participate in Korean-speaking communities. Kiaer et al. (2024), for example, show that Korean textbooks present address terms in simplified ways that position L2 learners as relatively peripheral to Korean relational hierarchies, while Korean speakers' expectations about address-term use differ depending on whether the interlocutor is Korean or non-Korean, and further depending on perceived ethno-national identity. Although their study concerns pragmatics rather than pronunciation, it suggests that Korean speakers' expectations toward L2 learners are socially differentiated rather than neutral.

This body of work, therefore, suggests that L2 speech is not perceived, evaluated, or responded to solely on the basis of acoustic input. Matched-guise research, reverse linguistic stereotyping, expectation-mismatch studies, visual–auditory congruence studies, and raciolinguistic approaches provide different but complementary evidence that social information can shape speech perception and evaluation. Recent research on L2 Korean learners further suggests that the influence of social information may extend beyond perceptual judgments. As Kim (2026) demonstrates, racialized visual cues can shape how Korean speakers modify their own speech toward L2 learners, while research in Korean language education more broadly suggests that expectations toward learners may also be socially differentiated. Thus, these findings raise the possibility that visual identity cues may influence not only how Korean speakers

1 Following Hughes (2023: 16), we use lowercase white when referring to racialized whiteness, while preserving capitalization in direct quotations, titles, and source material to avoid the appearance of participating in white supremacist ideologies.

respond to learners in interaction, but also how they evaluate learners' speech. However, existing Korean research has primarily examined accentedness, intelligibility, speech accommodation, and pragmatic expectations, while relatively little work has tested whether Korean language education evaluators' pronunciation evaluations are influenced by visual identity cues when the acoustic signal is held constant. Accordingly, the present study employs a matched-guise design to investigate whether Korean language education evaluators' evaluations of identical L2 Korean speech vary according to visual identity cues designed to signal learners' nationality.

2.2 Bias in pronunciation evaluation

In the present study, pronunciation evaluation refers specifically to listeners' judgments of how accurate or correct a learner's pronunciation sounds. This differs from accentedness, which concerns the perceived degree of difference between linguistic varieties. Accordingly, speech may be perceived as accented without necessarily being judged inaccurate, and the two constructs are not treated as synonymous in the present study. Pronunciation evaluation is particularly relevant in educational contexts because teachers routinely make judgments about the quality and accuracy of learners' pronunciation. Such evaluations may nevertheless be influenced by expectations arising from social cognition. Particularly in second language teaching, teachers not only function as evaluators, but their judgments may directly affect learner feedback, placement, and learners' perceptions of their own language abilities. Despite this, as Hughes (2020: 544) points out, "teachers' perception of students is largely unstudied." This is especially the case in Korean pronunciation research, where researchers commonly employ Korean language teachers, graduate students in Korean language education, or native speakers as evaluators, assuming that their ratings objectively reflect learners' pronunciation ability. This practice effectively positions language educator evaluations as professional and normative standards of pronunciation judgment. However, relatively little research has examined whether evaluators with professional linguistic knowledge and teaching experience remain susceptible to learners' identity cues during pronunciation evaluation. Existing evidence comes primarily from ESL contexts (e.g., Hughes 2020, 2023), where teacher evaluations have been shown to be shaped by both teacher and perceived student characteristics. Hughes

(2020) found that one ESL teacher's pronunciation ratings of identical Standardized American English speech varied according to the supposed speaker's race, gender, and heritage clothing, with no main effects but complex interactions among these cues. Hughes (2023) extended this work in a mixed-methods study of ESL teachers, showing that pronunciation ratings were associated with both rater characteristics and perceived student characteristics. She interprets these patterns as evidence that pronunciation judgments may be shaped by reverse linguistic stereotyping, raciolinguistic ideologies, and expectation downshift, whereby teachers rate some learners more favorably because they hold lower baseline expectations for those learners' pronunciation. Whether similar effects extend to Korean pronunciation assessment remains largely unknown.

In addition to perceived learner characteristics, rater-related factors may also influence judgments of L2 speech. From a language-assessment perspective, Kang (2012) distinguishes between trait-relevant factors, which reflect characteristics of the speaker's performance that are relevant to the ability being assessed, and trait-irrelevant factors, which are unrelated to the target ability but may nevertheless introduce variation into ratings. Kang examined the relative contributions of these two sources by comparing instrumentally measured prosodic features, such as articulation rate, overall pitch range, pausing, and stress-related measures, with five rater background variables: native-speaker status, linguistic sophistication, teaching/tutoring experience, contact with non-native speakers, and negative past experience in courses taught by international teaching assistants (ITAs). The results showed that evaluations of ITAs' oral performance were influenced not only by properties of the speech itself but also by rater background. Notably, negative prior experiences in ITA-taught courses predicted lower ratings of both oral proficiency and instructional competence. Kang (2012) suggested that such experiences may shape listeners' expectations of ITA speech, which may subsequently influence their evaluations. Importantly, although trait-relevant prosodic factors made a greater relative contribution to the ratings, the trait-irrelevant variance associated with rater background was nevertheless "not insubstantial."

Taken together, these findings suggest that pronunciation and L2 speech evaluations are not determined solely by properties of the speech signal. Rather, judgments may also be shaped by rater characteristics, prior social experiences, and expectations associated with the perceived identity of the speaker. This raises an

important question for Korean pronunciation assessment: whether visual identity cues signaling a learner's perceived nationality can similarly influence language educators' evaluations even when the speech itself remains unchanged.

2.3 Korean pronunciation features across learner groups and listener expectations

If visual identity cues signaling learners' perceived nationality can influence teachers' evaluations, it is also worth considering whether such evaluations vary across learners from different national backgrounds. Previous studies have shown that Korean learners from different L1 backgrounds exhibit recurring pronunciation patterns, especially in the production of Korean monophthongs and certain consonant contrasts. Chinese-,² Japanese-,³ Vietnamese-,⁴ and English-speaking⁵ learners have all been reported to show systematic deviations in Korean vowel production and/or consonant production/perception, which are often attributed to L1 transfer.⁶ These recurring patterns are relevant because listeners who are familiar with Korean spoken by learners from particular L1 backgrounds may also become familiar with characteristic pronunciation patterns associated with those learner groups. Such accent-specific familiarity may, in turn, shape how listeners interpret subsequent learner speech and the expectations they bring to pronunciation evaluation.

In addition to familiarity with L1-related pronunciation patterns, previous research has also suggested that social preference toward different learner groups may be related to native listeners' perceptions of foreign accentedness. Jung (2020) found significant cross-national differences in both perceived foreign accent strength and social preference among native Korean listeners. Chinese learners were perceived as having the strongest foreign accent, whereas Americans received the highest social preference ratings. In Jung's (2020) study, social preference was operationalized as participants' self-reported preference toward people of different nationalities rather than as a separate theoretical construct. Thirty native Korean undergraduate and graduate students rated Chinese, American, Japanese, and Vietnamese people on two 7-point

2 See Zhao (2006), Yun (2018), Wen (2019), Wen (2021) for examples.

3 See Cho (2004), Ha and Lee (2016) and Lee and Ha (2021) for examples.

4 See Kim (2016), Jung and Lee (2017), Jang (2018), Lee and Jung (2020), and Cho (2024) for examples.

5 See Cho (2005), Kim, Park, and Chun (2006), Kim and Kim (2015), and Roh and Song (2024) for examples.

6 See Yoo and Kang (2018) for a comparison among learner groups.

scales measuring social distance and social preference. Americans received the highest preference ratings, followed by Japanese, Vietnamese, and Chinese people, while social distance showed the reverse pattern overall. Jung (2020) further noted that Vietnamese people received relatively high preference ratings despite greater perceived social distance, suggesting that limited prior knowledge about Vietnam and Vietnamese people may have led participants to respond more neutrally or with curiosity rather than on the basis of firmly established attitudes. More recently, Korean foreign accent perception research has begun to focus more directly on how native Korean listeners perceive and evaluate non-native Korean speech. Jung and Holliday (2026) found that native Korean listeners' ratings of foreign accent and non-nativeness were associated with certain acoustic properties, particularly deviations in F0 and F1. However, they also pointed out that concepts such as "foreign accent" and "non-nativeness" themselves rely on listeners' subjective interpretations, and that different listeners may understand "foreignness" and "nativeness" differently. This suggests that speech evaluation in Korean may involve not only acoustic characteristics of speech itself, but also listeners' social perceptions of foreignness and nativeness.

Therefore, in actual pronunciation evaluation contexts, listeners' judgments may reflect not only genuine acoustic differences in learner speech, but also social expectations associated with learners' perceived nationality and language background. Nevertheless, existing studies have focused primarily on the phonetic characteristics of learner pronunciation differences, while paying relatively little attention to whether listeners are evaluating actual acoustic differences or are simultaneously influenced by pre-existing expectations toward particular learner groups.

2.4 Research gap and contribution of the present study

Although previous studies have separately examined Korean learners' pronunciation differences and the role of social cognition in speech perception, several limitations remain. First, most studies have not clearly distinguished the independent effects of acoustic speech properties and social identity information on pronunciation evaluation. Second, although previous research suggests that listeners may hold different foreign accent perceptions and social expectations toward different learner groups, relatively few studies have experimentally investigated how visual identity cues influence Korean

pronunciation evaluation. Finally, as discussed in Section 2.2, it is also important to examine whether language education evaluators, despite their professional linguistic knowledge and expertise, may still be influenced by cues beyond the speech signal itself, such as visual identity cues, when evaluating Korean learners' pronunciation.

To address these issues, the present study adopts a matched-guise experimental design in which the speech signal remains constant while learners' visual identity cues are systematically manipulated. Specifically, this study investigates whether visual identity cues influence native Korean language education evaluators' pronunciation evaluations when the speech itself remains identical, and whether such evaluation patterns differ according to learner nationality.

The findings of this study contribute to a deeper understanding of the relationship between acoustic information and social expectations in second language pronunciation evaluation. Furthermore, this study reconsiders whether pronunciation evaluation can truly be regarded as an objective judgment based solely on pronunciation accuracy, while also providing a new empirical perspective on pronunciation evaluation mechanisms in Korean language education contexts.

3. Methodology

To examine the role of visual and auditory cues in pronunciation evaluation, this study adopted a matched-guise paradigm and constructed two audiovisual conditions. By varying the visual identity cues presented with the speech, the study investigates whether language education evaluators' evaluations of the same pronunciation differ across audiovisual conditions.

3.1 Experimental conditions

Two types of cues were manipulated in the experiment:

- (i) *Visual identity cues*: operationalized as facial appearance, heritage clothing and a nationality label presented on screen that indexed the speaker's nationality category (Korean vs. American vs. Chinese vs. Japanese vs. Vietnamese); and

- (ii) *Acoustic cues*: operationalized as Korean speech produced by speakers with different language backgrounds (Korean vs. English vs. Chinese vs. Japanese vs. Vietnamese).

By combining these two types of cues, two experimental conditions were constructed as visualized in Figure 1:

(1) Matched Condition:

In this condition, participants viewed an image of a purported L2 learner and listened to Korean speech produced by an L2 learner from the same national background. Thus, the visual identity cue and the speaker's language background were nationality-congruent. The visual identity cues in this condition are expected to activate language education evaluators' expectations regarding the pronunciation of learners from different national backgrounds.

(2) Korean Condition:

In this condition, participants viewed an image of a purported Korean person while listening to the same L2 Korean speech used in the matched condition. Thus, the visual identity cue and the speaker's language background were nationality-incongruent. Rather than treating this condition as an expectation-free baseline, we use it as a contrasting visual-identity cue condition. Importantly, this condition may itself activate expectations associated with Korean or native-speaker identity, rather than eliminating social expectations altogether. In this condition, four different Korean images were used, with each image consistently paired with one learner voice. The Korean images were not crossed across learner voices; that is, each learner voice was always presented with the same Korean image to all participants assigned to this condition. This fixed pairing was used to avoid presenting multiple distinct voices as if they belonged to the same depicted individual.

Based on the conditions described above, the present study focuses on comparing the effect of visual identity cues on pronunciation ratings while holding the auditory input constant (i.e., learner speech in both conditions). If visual identity cues influence

pronunciation evaluation, ratings were expected to differ between the two conditions; however, no uniform directional difference was assumed across learner groups.

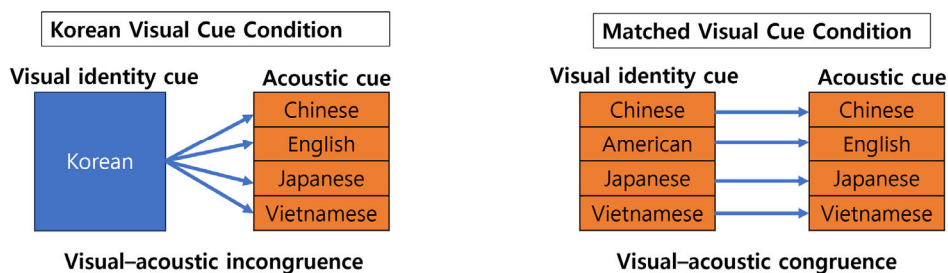


Figure 1. Experimental design illustrating visual-acoustic identity congruence across the two visual identity cue conditions

3.2 Stimuli

3.2.1 Speech stimuli

Speech stimuli were recorded by four non-native speakers of Korean from the United States, China, Japan, and Vietnam. All L2 speakers had advanced Korean proficiency, having achieved Level 5 or 6 on the Test of Proficiency in Korean (TOPIK), a standardized Korean language proficiency test for non-native speakers. Specifically, the speakers from China, Japan, and Vietnam had achieved Level 6, whereas the speaker from the United States had achieved Level 5.

The stimuli consisted of 14 nonwords consisting of three phones each. Short pseudowords were used rather than sentence-level materials in order to minimize potential effects of higher-level prosodic variation, such as intonation. Real lexical items were also avoided because lexical familiarity and difficulty may influence L2 production and, in turn, listeners' perception of foreign accentedness (Jung 2020). The stimuli were therefore constructed in a controlled V.CV format. The phones selected were each of the seven Korean monophthongs followed by the consonants /k/ and /p/ and ending in the vowel /a/ (7 vowels x 2 consonants). The consonants /k/ and /p/ were selected because both belong to the Korean lenis stop category and allowed the construction of a sufficiently large and balanced set of pronounceable

pseudowords across the target vowels. It should be noted that /aka/ is a real word in Korean meaning “child”; therefore, it was replaced with the nonword /ata/ in order to avoid potential lexical-semantic influences on perceptual evaluation. A complete list of the stimulus words is provided in Table 1.

Table 1. Monophthongs and corresponding stimulus words

Monophthong	Stimulus Words (IPA)	Stimulus Words (Hangul)
/i/	/ika/, /ipa/	이 가, 이 바
/i/	/ika/, /ipa/	으 가, 으 바
/u/	/uka/, /upa/	우 가, 우 바
/o/	/oka/, /opa/	오 가, 오 바
/ʌ/	/ʌka/, /ʌpa/	어 가, 어 바
/a/	/ata/, /apa/	아 다, 아 바
/ɛ/	/ɛka/, /ɛpa/	애 가, 애 바

Each stimulus word was recorded twice by each speaker, and each version was presented once during the experiment. Each of the two repetitions was randomly put into the first or the second half of the experiment. Then each half was placed in a pseudo-random order such that no two stimulus words nor speaker occurred twice in a row. Recordings were conducted in a meeting room of the Graduate Library at Korea University using a TASCAM DR-05X recorder, with a sampling rate of 44.1 kHz and a 24-bit resolution.⁷

Following recording, all speech materials were post-processed in Praat. All spoken tokens were shorter than 1000 ms and placed within a 1000-ms audio clip, with the complete speech signal retained in all cases. The total duration was standardized by adjusting only the surrounding silence, and no portion of the spoken material was removed. Intensity was normalized to 55 dB to ensure comparability across stimuli during experimental presentation. Across experimental conditions, each speech stimulus remained acoustically identical and was paired only with different visual stimuli. Prior to the perception experiment, all participants completed an online pure-tone perception task as a preliminary screening procedure. The task was intended to confirm that participants were attending to the auditory stimuli and could reliably

⁷ Stimuli from a native Korean speaker were also recorded and included in the experiment with corresponding Korean visual identity cues. However, for ease of interpretability these are not included in the analysis here.

distinguish the basic differences presented in the task before proceeding to the main experiment. Only participants who achieved 100% accuracy were allowed to take part in the experiment. The pure-tone stimuli were generated in Praat and consisted of six stimuli manipulating two acoustic parameters, intensity and frequency, each set at three levels (low, middle, and high). Because the experiment was conducted online and participants' listening levels were not calibrated, the task was not treated as a calibrated hearing test. The specific parameters of the pure-tone stimuli are provided in Table 2.

Table 2. Acoustic parameters of pure tone stimuli

Condition type	Frequency (Hz)	Intensity (dB)	Description
Intensity level 1	640	50	Fixed frequency, low intensity
Intensity level 2	640	60	Fixed frequency, medium intensity
Intensity level 3	640	70	Fixed frequency, high intensity
Frequency level 1	440	60	Fixed intensity, low frequency
Frequency level 2	640	60	Fixed intensity, mid frequency
Frequency level 3	840	60	Fixed intensity, high frequency

3.2.2 Visual stimuli

To provide visual identity cues related to the speaker's perceived nationality, purported speaker images were presented in the experiment. All individuals depicted in the images were dressed in traditional clothing⁸ associated with the intended nationality, and the corresponding nationality was explicitly labeled above each image. To minimize visual variability unrelated to the research objectives, all images were standardized in terms of size, resolution, and composition.

The visual stimuli were selected from copyright-free image websites: Pexels (<https://www.pexels.com/>) and Wikimedia Commons (<https://commons.wikimedia.org/>) or were purchased from iStock (<https://www.istockphoto.com/kr>) or Shutterstock (<https://www.shutterstock.com/>).

8 Because the United States does not have traditional national clothing comparable to those associated with China, Korea, Japan, or Vietnam, we selected an image of a racially ambiguous woman wearing clothing featuring the American flag as the visual stimulus intended to represent an American English speaker.

To ensure that the selected pictures were reliably perceived as representing the intended nationality, twenty native speakers of Korean rated all candidate images on a 7-point scale according to their perceived match with the target nationality. Based on these ratings, images with the highest mean scores and a concentrated rating distribution (within ± 2 standard deviations) were selected as the final visual stimuli. The final set of visual stimuli consisted of four images for the Korean condition and one image each for the American, Japanese, Chinese, and Vietnamese nationality conditions. The complete set of visual stimuli is provided in the Appendix, along with source and licensing information.

3.3 Participants

Before formal data analysis, participant records were screened to distinguish completed experimental sessions from pilot/test entries. According to the Gorilla recruitment record, 43 participants accessed the experiment using the purchased participant tokens, and 37 completed the experiment. No participants were recorded by Gorilla as rejected or deleted. Of the 37 completed sessions, three were determined to correspond to pilot/test versions of the experiment based on experiment version number. Consequently, these sessions were removed from the downloaded dataset prior to analysis. Meanwhile, thirteen corresponded to a third condition currently under analysis and whose results are, accordingly, not discussed here. The present analysis, therefore, includes the 21 completed participants assigned to the two visual identity cue conditions relevant to the research questions: the *Matched Condition* and the *Korean Condition*.

Thus, the final dataset included 21 participants with backgrounds in Korean language education, including Korean language teachers, pre-service teachers, current graduate students in Korean language education, and graduates in Korean language education. In the present study, we refer to this group as “Korean language education evaluators.” Teaching experience was defined on the basis of participants’ self-reported prior experience teaching Korean. Participants reporting any prior Korean-language teaching experience were classified as having teaching experience. Of the 21 participants included in the present analysis, 14 reported prior teaching experience and 7 reported none. In the *Matched Condition*, 8 participants had teaching experience

and 3 had none; in the *Korean Condition*, 6 had teaching experience and 4 had none. These and additional participant demographics are reported in Table 3.

Table 3. Summary of participant demographics

Characteristic	Matched Condition	Korean Condition	Total
Participants (N)	11	10	21
Women	9	10	19
Men	2	0	2
With Teaching Experience	8	6	14
Without Teaching Experience	3	4	7
Age Range (years)	25–44	23–42	23–44
Age, M (SD)	30.73 (6.96)	29.60 (5.34)	30.19 (6.11)

3.4 Procedure

The experiment was conducted using the Gorilla online experimental platform. After providing informed consent, participants completed a brief demographic questionnaire and were then randomly assigned to one of the visual conditions. Participants read standardized instructions prior to the experiment. Before the main experiment, participants completed three practice trials to familiarize themselves with the procedure and rating task.⁹

During the main experiment, in each trial, participants first viewed a purported speaker image displayed at the center of the screen while listening to a Korean speech stimulus. Each speech stimulus could be replayed at most once. Participants then rated the accuracy of the speaker's pronunciation using a 7-point Likert scale ranging from 1 (매우 정확하지 않다, “very inaccurate”) to 7 (매우 정확하다, “very accurate”). Consistent with the definition of pronunciation evaluation adopted in this study, the scale was intended to capture listeners' perceived correctness of the pronunciation rather than an objective acoustic measure of phonetic accuracy.¹⁰ After completing

9 To minimize pre-experiment exposure to the stimuli, the practice trials were designed to differ sufficiently from the main experimental trials. We used audio from an L1 Korean man and an L1 English Korean-learning white man (the second author). The visual cues presented were of an Asian man dressed in a Hanbok and an actual photograph of the second author dressed in a business suit.

10 Previous research (Jung 2020) suggests that native Korean listeners' evaluations of foreign accented

the rating, participants proceeded to the next screen, where they listened to the same speech stimulus again and were asked to type the word they had just heard. The stimulus again could be replayed at most one additional time.¹¹ An example of the experimental interface is shown in Figure 2.



Figure 2. Example of an experimental trial interface

During the main experimental phase, all speech stimuli were presented in a pseudo-randomized order as described above. The entire experimental session lasted approximately 25–35 minutes.

3.5 Statistical analysis

The pronunciation rating data in this study were analyzed using the open-source statistical software R (R Core Team 2026).

Because the dependent variable was a 7-point ordinal pronunciation rating, the primary analysis was a cumulative link mixed-effects model fit with the *clmm()* function in the *ordinal* package (Christensen 2023) using the Laplace approximation.

The dependent variable was the pronunciation rating (*Score*) ordered from 1 to 7. Fixed effects included *Visual identity cue* and speaker *Nationality*, as well as their interaction.

Visual identity cue consisted of two levels: Korean and Matched. In the Korean

Korean reflect their overall perception of the speaker’s pronunciation, with perceived acceptability as Korean playing an important role in such judgments.

¹¹ The results for the second screen are currently under analysis and not discussed further here.

visual identity cue condition, L2 speech was paired with a Korean visual identity cue; in the matched visual identity cue condition, L2 speech was paired with a visual identity cue corresponding to the speaker's nationality. *Nationality* refers to the language background of the learner voice: Chinese, English, Japanese, or Vietnamese.

All categorical predictors were coded using treatment coding. The reference levels were set to Chinese for *Nationality* and Korean for *Visual identity cue*. Accordingly, model coefficients represent effects on the latent ordinal scale relative to the Chinese learner voice in the Korean visual identity cue condition.

The random-effects structure included random intercepts for *Item* and *Participant* to account for variability across stimulus items and participants. Accordingly, the model structure was specified as follows: $\text{Score} \sim \text{VisualIdentityCue} * \text{Nationality} + (1 | \text{Item}) + (1 | \text{Participant})$.

Fixed effects were evaluated using Type III Wald chi-square tests with the *Anova()* function in the *car* package (Fox and Weisberg 2019). To further interpret the interaction, post hoc comparisons were conducted using the *emmeans* package (Lenth 2025). Specifically, an estimated marginal means comparison of the two Visual identity cue conditions was conducted after averaging equally across the four learner voices and simple effects of *Visual identity cue* were examined within each level of *Nationality*, with *p*-values adjusted across the four comparisons using the Holm-Bonferroni method.

4. Results

The results of this study are organized around the following two research questions: First, does the manipulation of visual identity cues influence Korean language education evaluators' perceptual evaluation of learners' pronunciation? Second, does this visual identity cue effect vary depending on the learners' nationality?

4.1 Descriptive results

Figure 3 presents the overall proportion of pronunciation ratings given by Korean language education evaluators under different visual identity cue conditions (Korean vs. Matched). This figure is based on all learner data and illustrates rating differences

across visual identity cue conditions collapsed across learner nationality. As shown in the figure, ratings cluster in the 5-7 range, although the Korean condition has a higher proportion of 6s and the Matched condition has a higher proportion of 5s, suggesting a tendency to give voices in the Korean condition slightly higher ratings on average.

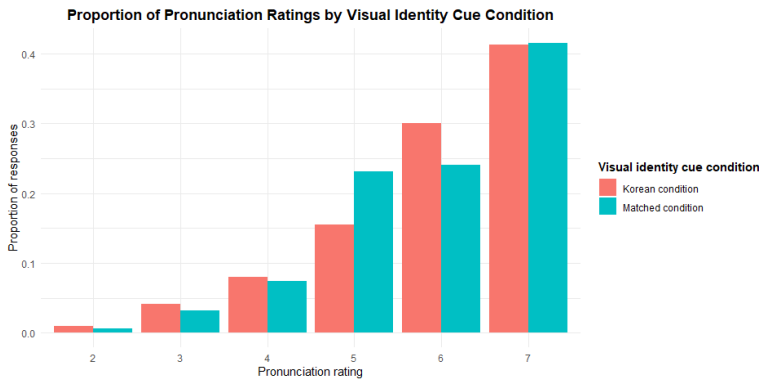


Figure 3. Proportion of pronunciation ratings by visual identity cue condition

Figure 4 shows the proportion of pronunciation ratings divided for each learner nationality under the two visual identity cue conditions (Korean vs. Matched). For the Japanese and Vietnamese learners, the proportion of 6s and 7s is higher for the Korean condition than the Matched condition while the proportion of 5s is lower. The American learner follows a similar pattern, but there is little difference in the proportion of 7s between the two conditions and the conditional difference in 5s is larger than the two aforementioned nationalities. The ratings for the Chinese learner follow a different pattern: the proportion of 7s is much higher for the matched visual identity cue condition than the Korean condition and the proportion of 5s and 6s is similar between the two conditions. This seems to suggest that for the Chinese learner, the matched condition is actually receiving higher pronunciation ratings on average than the Korean condition.

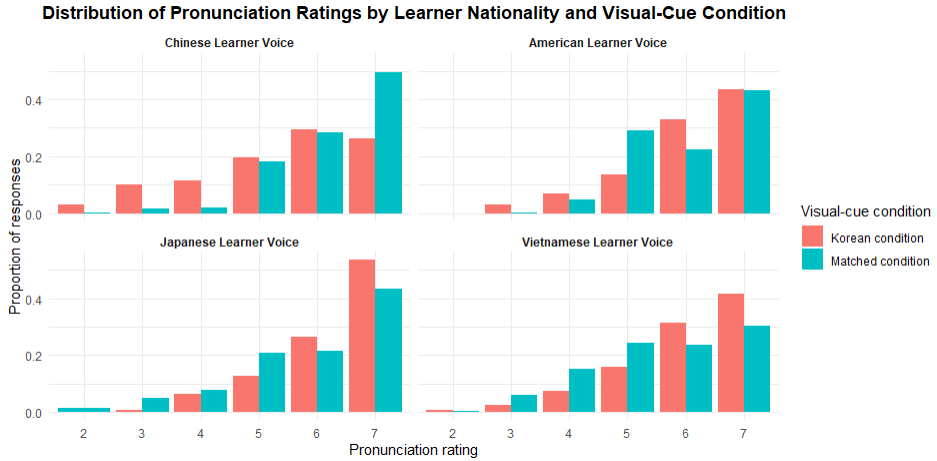


Figure 4. Proportion of pronunciation ratings by learner nationality and visual identity cue condition

4.2 Inferential statistical models

The summary of the ordinal mixed-effects model is given in Table 4 while the results of Type III Wald chi-square tests performed on the model are given in Table 5. The results of said tests indicate that the simple effect of *Visual identity cue* is significant ($\chi^2(1) = 5.100$, $p = .024$), the simple effect of *Nationality* is significant ($\chi^2(3) = 125.525$, $p < .001$), and the interaction between *Visual identity cue* and *Nationality* is also significant ($\chi^2(3) = 158.558$, $p < .001$). This indicates that the effect of *Visual identity cue* condition on ratings varies significantly depending on learner nationality.

Table 4. Fixed-effect estimates from the cumulative link mixed-effects model predicting ordinal pronunciation ratings

Predictor	β	SE	z	p
Visual identity cue (Matched)	1.873	0.830	2.258	.024 *
Nationality (American)	1.321	0.168	7.845	< .001 ***
Nationality (Japanese)	1.885	0.175	10.785	< .001 ***
Nationality (Vietnamese)	1.179	0.166	7.088	< .001 ***
Visual identity cue $\times$ American	-1.805	0.240	-7.520	< .001 ***
Visual identity cue $\times$ Japanese	-2.597	0.247	-10.530	< .001 ***
Visual identity cue $\times$ Vietnamese	-2.758	0.240	-11.497	< .001 ***

Note. Signif. codes: *** $p < .001$; ** $p < .01$; * $p < .05$; . $p < .10$.

At the level of fixed effects, with the Chinese learner in the Korean condition as the reference group, the difference between Matched and Korean conditions is estimated at log odds of 1.873 ($SE = 0.830$, $z = 2.258$, $p = .024$). In addition, all interaction terms between *Nationality* and *Visual identity cue* are significant (all $p < .001$), further indicating that the effect of *Visual identity cue* differs across learner voices.

Table 5. Type III Wald chi-square tests for the cumulative link mixed-effects model predicting ordinal pronunciation ratings

Effect	DF	χ^2	p
Visual identity cue	1	5.100	.024 *
Nationality	3	125.525	< .001 ***
Visual identity cue $\times$ Nationality	3	158.558	< .001 ***

Note. Signif. codes: *** $p < .001$; ** $p < .01$; * $p < .05$; . $p < .10$.

After confirming a significant interaction between *Visual identity cue* and *Nationality*, post-hoc comparisons based on estimated marginal means (with Holm correction) were conducted to examine rating differences across visual identity cues for each nationality.¹²

12 The apparent discrepancy between the results displayed in Tables 4 and 5 and those displayed in Table 6 is due to the fact that the model includes a *Visual identity cue* $\times$ *Nationality* interaction and uses Chinese as the reference level for *Nationality*. Accordingly, the *Visual identity cue* coefficient in Table 4 represents the Matched–Korean contrast for the Chinese learner voice, not an average contrast across all learner voices. The corresponding one-degree-of-freedom Type III Wald test for *Visual identity cue* in Table 5 tests this reference-level contrast within the model parameterization. By contrast, Table 6 reports an estimated marginal means comparison averaged across the four learner voices. Because the simple contrasts differed in direction, with the Chinese learner showing a positive Matched–Korean contrast, the Japanese and Vietnamese learners showing negative contrasts, and the American learner

Table 6. Overall pairwise comparison of visual identity cue averaged across learner voices

Contrast	Estimate	SE	z	p
Matched – Korean	0.083	0.815	0.102	.919

In terms of individual nationalities, the post-hoc results are summarized in Table 7. None of the simple contrasts reached significance after Holm correction. However, the Chinese learner was marginally significant, with ratings trending higher in the Matched condition than in the Korean condition (estimate = 1.873, $SE = 0.830$, $z = 2.26$, $p = .096$). The Japanese and Vietnamese learners showed the opposite numerical pattern, with lower ratings in the Matched condition, although these differences were not statistically significant. In contrast, the estimated difference for the American condition was quite small relative to the estimated difference for the other learner groups.

Table 7. Pairwise comparisons of visual identity cue within each nationality

Nationality	Contrast	Estimate	SE	z	p
Chinese	Matched – Korean	1.873	0.830	2.258	.096
American	Matched – Korean	0.068	0.828	0.082	.935
Japanese	Matched – Korean	-0.724	0.829	-0.873	.853
Vietnamese	Matched – Korean	-0.885	0.826	-1.071	.853

Note. p-values are Holm-adjusted.

These findings suggest that the relationship between visual identity cues and pronunciation ratings varied across the four learner voices. Although the simple contrasts for individual learner nationalities did not reach significance after Holm correction, the significant *Visual identity cue* $\times$ *Nationality* interaction suggests that the relationship between *Visual identity cue* and pronunciation *rating* differed across the four learner voices. Descriptively, the Chinese learner showed a different pattern from the Japanese and Vietnamese learners: ratings for the Chinese learner tended to be higher in the Matched condition, whereas ratings for the Japanese and Vietnamese learners tended to be higher in the Korean condition, while there was relatively little descriptive difference between the two conditions for the American learner. These patterns suggest that visual identity cues may have shaped pronunciation ratings differently depending on the learner voice, although the results should be

showing little difference, the averaged comparison in Table 6 showed little overall difference between the two conditions.

interpreted cautiously because the individual follow-up contrasts were not statistically significant, and each nationality was represented by one speaker.

5. Discussion

5.1 The influence of visual identity cues on Korean pronunciation evaluation

The results of the present study suggest that Korean language education evaluators' pronunciation evaluations were not determined by the acoustic signal alone. Although the overall comparison between the Korean and Matched visual identity cue conditions was not significant when averaged across learner nationalities, the cumulative link mixed-effects model showed a significant interaction between *Visual identity cue* and *Nationality*. This indicates that the relationship between visual information and pronunciation rating differed across the four learner voices. In other words, visual identity cues did not produce a uniform increase or decrease in pronunciation ratings; rather, their effect depended on which learner voice was being evaluated. This finding is important because the acoustic stimuli were held constant across the two visual identity cue conditions. Therefore, the observed interaction cannot be attributed simply to differences in the speech signal itself. Instead, the results suggest that Korean language education evaluators' evaluations of identical L2 Korean speech were shaped, at least in part, by the visual information presented with that speech. This pattern is consistent with previous work showing that language education evaluators' and listeners' evaluations of L2 speech can be influenced by non-acoustic information about the speaker, including race, gender, clothing, and perceived identity (e.g., Rubin 1992; Kang and Rubin 2009; McGowan 2015; Gnevsheva 2018; Hughes 2020, 2023; Jung and Idemaru 2025).

The present findings also extend previous Korean research on audiovisual incongruence. Graillot and Oh (2025) found that incongruent visual and auditory information could increase Korean listeners' perception of foreign accentedness, particularly under some proficiency conditions. The present study likewise shows that visual and auditory information jointly shaped listeners' responses. However, the present results suggest that audiovisual incongruence does not necessarily produce a single, uniform evaluative outcome. Rather, the direction of the visual identity cue

pattern differed across learner voices: descriptively, the Chinese learner tended to receive higher ratings in the Matched condition, whereas the Japanese and Vietnamese learners showed the opposite numerical pattern. Because the individual simple contrasts did not reach significance after Holm correction, these learner-specific patterns should be interpreted cautiously. Nevertheless, the significant interaction indicates that the effect of visual identity cues cannot be reduced to a simple distinction between native-speaker and non-native-speaker visual identity cues.

We should also note that the present study cannot determine whether listeners became more aware of foreign accentedness in either visual identity cue condition, nor can it establish whether perceived accentedness directly shaped pronunciation ratings. What can be concluded is more limited but still important: when the same L2 Korean speech was paired with different visual identity cues, language education evaluators' pronunciation ratings varied differently across the four learner voices. Thus, visual identity cues appear to participate in pronunciation evaluation alongside acoustic information, although the specific social expectations activated by those cues require further investigation.

The present findings are also consistent with those reported by Kim (2026). Kim demonstrated that Korean native speakers adjust their speech rate, vowel space, and pausing patterns according to learners' visual identity during interaction, suggesting that visual identity cues influence not only listeners' perceptions but also native speakers' own speech production strategies.

5.2 Variation in visual identity cue patterns across learner voices

Having established that the relationship between visual identity cues and pronunciation ratings differed across learner voices, a more important question concerns why this visual identity cue pattern did not operate in the same direction for all speakers. If visual identity simply induced a general negative bias toward L2 learners, learners from all national backgrounds should have received uniformly lower ratings in the Matched condition. However, this was not the pattern observed. The Chinese learner's ratings trended higher in the Matched condition than in the Korean condition, whereas the Japanese and Vietnamese learners showed a descriptive tendency to be lower in the matched condition than in the Korean condition while there was little difference

for the American learner. These findings suggest that the relationship between visual identity cues and pronunciation ratings did not follow a uniform direction across the four learner voices. These observed visual identity cue patterns did, however, differ across the four learner voices, a pattern that is compatible with differences in listeners' expectations toward particular learner groups.

One possible interpretation of this pattern comes from expectation-based accounts of speech perception. McGowan (2015) and Gnevsheva (2018) argued that the effect of visual identity cues depends on listeners' expectations regarding how members of a particular social group are expected to sound and on the degree of consistency between those expectations and the actual speech signal. In other words, within these theoretical accounts, visual identity does not inherently increase or decrease pronunciation ratings. Instead, it is the social knowledge activated by a particular identity that shapes how listeners interpret identical speech. This observation is further supported by Hughes (2023: 149) who found a trend such that Afghan students were rated as more native-like than white students. She attributes this as “likely due to bias manifesting in an expectation downshift, wherein teachers have lower expectations for Afghan students’ accents, and so overcompensate in rating them higher than they do other [sic] the other students.”¹³ From this perspective, visual identity cues do not serve to change the speech itself; rather, they alter the evaluative framework through which identical speech is interpreted.

Applied cautiously to the present findings, one possible interpretation is that the different visual identity cue patterns reflect variation in the expectations associated with different learner groups rather than a single generalized response to non-native visual identity cues. However, the present study did not directly measure such expectations, and the observed patterns therefore cannot establish an expectation-activation mechanism. These expectations, if they contributed to the present results, may develop through accumulated social, educational, and professional exposure to different learner groups. For Korean language education evaluators with classroom experience, repeated interaction with learners from different linguistic backgrounds may contribute to experience-based knowledge about the pronunciation characteristics associated with particular learner groups.

¹³ Hughes herself draws on the work of Dupree and Fiske (2019) who find that white liberals lower their level of displayed linguistic competence when speaking to Black participants, something that white conservatives do not do. They term this a “competence downshift.”

As Kang (2012) demonstrates, raters' prior contact, teaching experience, and other background factors may also contribute to the evaluative framework they bring to L2 speech assessment. Notably, in Kang's analysis of ITA oral proficiency ratings, raters who reported more frequent contact with non-native speakers tended to assign lower ratings. Kang regarded this as a counterintuitive finding and noted that the contact measure did not distinguish whether raters had previously interacted with speakers from the same language background as the ITAs being evaluated or with non-native speakers from other backgrounds. The finding therefore suggests that the relationship between prior contact and L2 speech evaluation may be more complex than a simple assumption that greater contact or familiarity necessarily leads to more favorable evaluations. In the present study, evaluators' prior experiences with different learner groups may likewise constitute one possible source of variation in the observed visual identity cue patterns, although such experience was not directly measured.

Beyond direct contact and prior experience, professional knowledge available within Korean language education may constitute another potential source of such expectations. Yun (2019) demonstrated that pronunciation has consistently been one of the primary research topics concerning Chinese, Japanese, Vietnamese, and Mongolian learners in Korean language education, whereas pronunciation-related research on American learners ranked only nineteenth. This imbalance in the research literature may contribute to differences in the pronunciation-related knowledge available within Korean language education. Together, accumulated social contact, teaching experience, and professional knowledge could contribute to differences in evaluators' expectations concerning particular learner groups, although the present study did not measure these sources of experience directly. Within this framework, the descriptive trend observed for the Chinese learner can be considered in relation to Hughes' (2023) expectation downshift account. Jung (2020) reported that, among the four learner voices examined, Korean native speakers perceived Chinese learners as having the strongest foreign accent. If Korean language education evaluators similarly hold relatively low expectations regarding the pronunciation of Chinese learners, encountering a Chinese learner could establish a lower evaluative baseline. When the actual pronunciation exceeds that expectation, listeners might provide more favorable evaluations, even though the speech itself remains unchanged. The direction of the Chinese learner's descriptive pattern is compatible with this possibility, as ratings tended to be higher in the Matched condition than in the Korean condition.

Nevertheless, because this contrast did not reach significance after Holm correction and the present study did not directly measure language education evaluators' pronunciation expectations toward Chinese learners, this interpretation remains tentative.

In contrast, the Japanese and Vietnamese learners received descriptively lower ratings in the Matched condition and the interaction terms indicate that these patterns differed significantly from the Chinese learner's pattern. However, this pattern should not necessarily be interpreted as reflecting more negative attitudes toward these learner voices. One possibility is that evaluators held different or more differentiated expectations concerning these learner groups. If so, the corresponding visual identity cues may have altered how listeners attended to or interpreted the speech. This possibility remains speculative, however, because such expectations were not directly measured. Notably, the present descriptive patterns do not straightforwardly correspond to the social preference ranking reported by Jung (2020). Jung (2020) found that Korean native speakers reported more favorable social attitudes toward Japanese people than toward Chinese or Vietnamese people. Yet evaluations of Japanese learner speech did not pattern opposite to these groups nor did Chinese or Vietnamese learner speech pattern similarly. This suggests that general social preference alone is unlikely to provide a sufficient explanation for the observed pattern.

Results for the American learner showed a different pattern, with relatively little descriptive difference between the two visual identity cue conditions in the present dataset. This pattern should be interpreted as inconclusive rather than as evidence that American or non-East-Asian visual identity cues do not affect pronunciation evaluation. Several explanations are possible, but the present data do not allow us to distinguish among them. Previous work nevertheless suggests that visible speaker characteristics can shape native Korean speakers' behavior toward L2 Korean learners. For example, Kim (2026) found that Korean native speakers adopted more prototypical Foreigner-Directed Speech (FDS) when interacting with learners who appeared white than those who appeared East-Asian. This finding suggests that visible speaker characteristics can influence interactional behavior, but it does not establish what expectations Korean language education evaluators hold specifically toward American learners in pronunciation evaluation.

Overall, the present findings are compatible with an expectation-based interpretation, but they do not establish it. Because the study did not directly measure

language education evaluators' expectations and each nationality was represented by one speaker and one matched image, the present results should be interpreted as preliminary evidence from four speaker–image pairings rather than as confirmed evidence of stable nationality-level effects. The observed patterns suggest that the relationship between visual identity cues and pronunciation ratings may vary across particular speaker–image pairings. Future research with multiple speakers per nationality and direct measures of listener expectations will be necessary to determine whether such variation reflects nationality-related pronunciation expectations or other characteristics of the individual speakers and images.

5.3 Implications for research on Korean pronunciation

The findings of this study also carry important methodological implications for research on Korean pronunciation. Native-speaker ratings and teacher evaluations have long served as primary measures of learners' pronunciation proficiency in Korean pronunciation research (e.g., Cho 2004, 2005; Kim et al. 2006; Kim and Kim 2015; Ha and Lee 2016; Yoo and Kang 2018; Roh and Song 2024). Such ratings are therefore often interpreted as reflecting learners' underlying pronunciation ability. The present findings, however, suggest that even evaluators with professional training in Korean language education may be influenced by learners' visual identity cues. Consequently, when native-speaker ratings are used as dependent measures of pronunciation proficiency, they may reflect not only learners' actual phonetic performance but also the evaluators' social cognitive interpretations of that performance.

This does not mean that the evaluations of Korean language education evaluators are unreliable or uninformative. Rather, the present findings highlight that pronunciation evaluation is a complex cognitive process in which language education evaluators interpret acoustic information in relation to accumulated teaching experience, social experience, and prior knowledge of different learner groups. Accordingly, greater caution is warranted when interpreting pronunciation ratings, particularly in distinguishing learners' acoustic performance from evaluators' socially mediated expectations about learner identity.

5.4 Limitations and directions for future research

Although the present study found that the relationship between visual identity cues and pronunciation ratings differed across learner voices, more precisely across image–voice pairings, several limitations should be acknowledged.

First, the present study did not include a No-visual identity cue condition. Instead, the experiment directly compared the Korean and Matched conditions. This design effectively examined whether different visual identity cues altered language education evaluators' evaluations of identical speech; however, it does not allow us to determine how either condition differs from evaluations made in the absence of visual information. For example, the descriptive results showed somewhat different rating distributions across the Korean and Matched conditions, but it remains unclear whether these differences resulted from the Korean visual identity cue increasing ratings, the learner visual identity cues decreasing ratings, or a combination of both effects. In addition, because the experiment relied on visually presented speaker identities, it is possible that some participants questioned the authenticity of the speaker–image pairings, especially in the Korean visual identity cue condition; the present study did not directly assess such doubts. Future studies incorporating a No-visual identity cue baseline would provide a more complete understanding of how visual identity cues influence Korean pronunciation evaluation and would also allow more direct comparisons with the findings of Graillot and Oh (2025).

Second, although the present study interpreted the findings within the frameworks of Social Cognition Theory, Reverse Linguistic Stereotyping (RLS), and previous matched-guise research, it did not directly measure the social expectations formed by evaluators during pronunciation assessment. Consequently, the explanations proposed for why different learner groups elicited different evaluation patterns remain theoretical interpretations rather than direct empirical evidence. Future research could combine pronunciation evaluation with questionnaires, interviews, ethnography, eye-tracking, or other psycholinguistic methods to investigate Korean language education evaluators' social impressions, language attitudes, and pronunciation expectations toward different learner groups, thereby clarifying how these social cognitive factors shape pronunciation evaluation.

Third, each learner nationality was represented by only one speaker. Because

speaker identity was therefore fully confounded with nationality, some of the observed differences may reflect individual speaker characteristics, including speaker-specific pronunciation features, rather than nationality alone. Further, although all speakers had advanced Korean proficiency as measured by TOPIK Level 5 or 6, TOPIK scores do not guarantee comparable pronunciation across speakers, and the American speaker had a different TOPIK level from the other speakers. These limitations restrict the extent to which the present findings can be generalized to Korean language education evaluators' evaluations of particular learner groups more broadly. Future research should include multiple speakers from each nationality to separate nationality-related effects from speaker-specific variation and to determine whether the observed evaluation patterns generalize across individuals within the same learner group.

Fourth, the photographs themselves differed in several respects, including pose, facial expression, clothing, and composition. These differences may have affected listeners' evaluations independently of the intended visual nationality cue. This issue is compounded by the fact that different Korean images were paired with different learner voices in the Korean visual identity cue condition. Future research should use more tightly controlled visual stimuli, such as images selected from standardized face databases or digitally manipulated images in which clothing, nationality labels, and other identity cues are systematically varied while facial and compositional features are held constant. The American visual identity cue also differed from the Chinese, Japanese and Vietnamese visual identity cues because it used contemporary clothing and an American flag pin rather than traditional or heritage clothing, reflecting the difficulty of selecting a comparable "traditional" American visual identity cue. Future work might pretest whether there are comparable American visual identity cues that reliably index Americanness for Korean listeners.

Fifth, because the experiment was conducted online, participants' listening environments and playback volume could not be fully controlled. The pure-tone task functioned as a basic screening procedure rather than as a calibrated hearing test. Future laboratory-based studies could provide more precise control over listening conditions.

Sixth, the present study employed disyllabic nonwords as experimental stimuli in order to minimize the influence of lexical familiarity, semantic information, and contextual cues, thereby isolating the effects of visual identity cues as much as possible. Although this highly controlled design strengthens internal validity, it inevitably limits

the generalizability of the findings to authentic communication. In natural interactions, listeners simultaneously process syntactic, semantic, discourse, and contextual information in addition to visual identity cues. Whether visual identity exerts similar effects on pronunciation evaluation under more natural communicative conditions therefore remains an open question. Future studies may employ sentence reading, spontaneous conversation, or classroom interaction to examine the role of social identity cues in more naturalistic Korean language contexts.

Seventh, the present study did not collect detailed information about evaluators' prior contact with learners from different language or nationality backgrounds. As Kang (2012) notes, general measures of contact with non-native speakers may not capture important differences in whom raters have previously interacted with. Because the present study did not distinguish the frequency or nature of evaluators' contact with Chinese, Japanese, Vietnamese, or American learners, we cannot determine whether such prior experience contributed to the different visual identity cue patterns observed across learner voices. Future research should therefore include more detailed measures of rater experience, particularly contact with specific learner groups.

Eighth, all evaluators in the present study had backgrounds in Korean language education, including both teachers and pre-service teachers. The evaluator sample was also highly gender-imbalanced, with 19 women and only 2 men, and the Korean visual identity cue condition included no male evaluators. Accordingly, the findings primarily reflect the evaluation patterns of a predominantly female sample with training or professional preparation in Korean language education, rather than naïve Korean listeners more broadly. The present study therefore cannot determine whether evaluator gender contributed to the observed rating patterns. Future research should recruit a more gender-balanced sample and compare Korean language teachers, naïve Korean native listeners, and Korean language learners to determine whether similar visual identity cue effects emerge across different listener groups.

Finally, the relatively small number of evaluators is also a limitation, particularly because visual identity cue condition was manipulated between participants while learner voice varied within participants. Future studies should include larger evaluator samples to obtain more stable estimates of visual identity cue effects and their interaction with learner voice.

6. Conclusion

The present study employed a matched-guise paradigm to investigate whether visual identity cues were associated with differences in Korean language education evaluators' pronunciation evaluations across four learner voices representing different national backgrounds. The results did not show a uniform overall visual identity cue effect across the four learner voices; however, a significant interaction indicated that the relationship between visual identity cue condition and pronunciation ratings differed across the four image–voice pairings. The findings therefore support a more cautious conclusion that the relationship between visual identity cue and pronunciation evaluation may vary depending on the particular learner voice, rather than reflecting a uniform visual identity cue effect in a single direction.

These findings contribute to a broader understanding of pronunciation evaluation in Korean language education. Language education evaluator ratings are often treated as indicators of learners' pronunciation proficiency, but the present results suggest that even when the speech signal is held constant, rating patterns may still be associated with the visual identity cue presented alongside the speech. Accordingly, pronunciation ratings should be interpreted with appropriate caution, particularly when they are used as direct indicators of learners' pronunciation performance.

Methodologically, the study demonstrates the value of matched-guise designs for separating the effects of acoustic information from those of visual identity cues in Korean pronunciation research.

References

- Cho, Minha. 2005. A study on types of errors in the pronunciation of Korean word-initial stops by Japanese and English native speakers: Focusing on lenis, fortis and aspirated pronunciation. *Korean Linguistics* 29: 237–256.
- Cho, Nam-hyun. 2024. A comparative study on the perception and production error patterns of Korean vowels for Vietnamese learners. *Journal of Education & Culture* 30(5): 903–929. <https://doi.org/10.24159/joec.2024.30.5.903>.
- Cho, Sung-Moon. 2004. An acoustic study of the pronunciation of Korean vowels uttered by Japanese speakers. *Speech Sciences* 11(3): 69–81.

- Choe, Seyeon and Hyunkee Ahn. 2026. Intra-speaker variability and perceptual sensitivity of English vowels in advanced Korean EFL learners. *Korean Journal of English Language and Linguistics* 26: 651–670. <https://doi.org/10.15738/kjell.26..202605.651>.
- Christensen, Rune H. B. 2023. Ordinal: Regression models for ordinal data. R package version 2023.12-4.1. <https://CRAN.R-project.org/package=ordinal>.
- Cottonbro Studio. 2020. Portrait of a young woman with curly hair, wearing a white sweatshirt and USA flag pin, smiling against a vibrant yellow background [Photograph]. Pexels. <https://www.pexels.com/photo/smiling-woman-wearing-an-american-flag-pin-4676440/>.
- Derwing, Tracey M. and Murray J. Munro. 2015. *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/llt.42>.
- Dupree, Cydney H. and Susan T. Fiske. 2019. Self-presentation in interracial settings: The competence downshift by White liberals. *Journal of Personality and Social Psychology* 117(3): 579–604. <https://doi.org/10.1037/pspi0000166>.
- Fiske, Susan T. and Shelley E. Taylor. 1991. *Social cognition*. New York, NY: McGraw-Hill.
- Flores, Nelson and Jonathan Rosa. 2015. Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review* 85(2): 149–171. <https://doi.org/10.17763/0017-8055.85.2.149>.
- Fox, John and Sanford Weisberg. 2019. *An R companion to applied regression*. 3rd ed. Thousand Oaks, CA: Sage. <https://www.john-fox.ca/Companion/>.
- Geraldshields11. 2018. Stylish person at Fukuoka City Hall [Photograph]. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Stylish_person_at_Fukuoka_City_Hall.jpg.
- Gnevsheva, Ksenia. 2018. The expectation mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English. *Linguistics* 56(3): 581–598. <https://doi.org/10.1515/ling-2018-0006>.
- Graillot, Kevin Sebastien and Eunhae Oh. 2025. Role of proficiency and visual cue in Korean listeners' perception of accented speech. *Phonetics and Speech Sciences* 17(2): 1–8. <https://doi.org/10.13064/KSSS.2025.17.2.001>.
- Ha, Hobin and Hwajin Lee. 2016. Analysis of Japanese learner's Korean vowel pronunciation based on vowel system and vowel feature. *The Review of Korean Cultural Studies* 55: 37–58. <https://doi.org/10.17329/kcbook.2016.55.55.002>.
- Hughes, Ruthanne. 2020. Worse for the wear: Effects of raciolinguistic ideologies, gender ideologies, and clothing on ESL pronunciation perception. In Patrick Farrell (ed.), *Proceedings of the Linguistic Society of America*, 544–557. Washington, DC: Linguistic Society of America. <https://doi.org/10.3765/plsa.v5i1.4729>.
- Hughes, Ruthanne J. W. 2023. *Ideologies of race, gender, and religion in pronunciation perception at Evangelical ESL programs*. PhD Dissertation. University of South Carolina.
- Jang, Hyejin. 2018. Experimental study on the Korean monophthongs by Vietnamese advanced Korean learners. *Korean Linguistics* 80: 211–234. <https://doi.org/10.20405/kl.2018.08.80.211>.

- Jeon, Han. 2016. Hanbok Day 2016 37 [Photograph]. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Hanbok_Day_2016_37_%2830205236080%29.jpg.
- Jung, Hwanmin. 2020. *An investigation of foreign accent in the word productions of non-native speakers of Korean*. PhD Dissertation. Korea University.
- Jung, Hwanmin and Jeffrey J. Holliday. 2026. The perception of foreign accent in isolated Korean words. *International Journal of Applied Linguistics* 36(1): 667–676. <https://doi.org/10.1111/ijal.12803>.
- Jung, Hwanmin and Nara Lee. 2017. An acoustic study on Korean monophthongs produced by Vietnamese novice learners of Korean: Focusing on female speakers. *Bilingual Research* 67: 285–307.
- Jung, Yeojin and Kaori Idemaru. 2025. Social expectations in speech perception: English-accented Korean in South Korea. *The Journal of the Acoustical Society of America* 157(4 Supplement): A122–A122. <https://doi.org/10.1121/10.0037611>.
- Kang, Okim. 2012. Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly* 9(3): 249–269. <https://doi.org/10.1080/15434303.2011.642631>.
- Kang, Okim and Donald L. Rubin. 2009. Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology* 28(4): 441–456. <https://doi.org/10.1177/0261927X09341950>.
- Kiaer, Jieun, Derek Driggs, Lucien Brown, and Naya Choi. 2024. Ideologies in second language learning: The case of Korean address terms. *Journal of Language, Identity & Education* 23(6): 858–878. <https://doi.org/10.1080/15348458.2022.2060229>.
- Kim, Choon Hak. 2016. *An experimental study on Korean pronunciation education for Vietnamese speakers*. Master's Thesis. Kunsan National University.
- Kim, Eun Jung and Seon-Jung Kim. 2015. Teaching methods of alveolar fricatives in Korean for native English speakers. *Journal of Education & Culture* 21(3): 129–147. <https://doi.org/10.24159/joec.2015.21.3.129>.
- Kim, Eun Kyung. 2018. Accuracy analysis of the Korean vowels of Chinese Korean learners. *Korean Education* 116: 315–337. <https://doi.org/10.15734/koed.116.201809.315>.
- Kim, Kee-Ho, Yoonjin Park, and Yoonsil Chun. 2006. The production and perception of the Korean stops by English learners. *The Korean Journal of Speech Sciences* 13(4): 51–67.
- Kim, Ryeo-Eun and Seok-Chae Lee. 2019. Liquid production according to English pronunciation proficiency ratings of Korean learners of English. *Korean Journal of English Language and Linguistics* 19(1): 53–75.
- Kim, Seula. 2026. *The effects of addressee race and fluency on Korean native speakers' speech accommodation and perception*. Master's Thesis. Korea University.
- Ko, Jeongho. 2026. *Shifts in inferred social information influence acoustic cue use in speech perception*. Master's Thesis. Korea University.

- Kuzmanovic, Mirko. 2023. Pretty Korean girl wearing traditional hanbok dress in Seoul, South Korea [Photograph]. iStock. <https://www.istockphoto.com/photo/pretty-korean-girl-wearing-traditional-hanbok-dress-in-seoul-south-korea-gm1564641890-527443692?searchscope=image%2Cfilm>.
- Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner, and Stanley Fillenbaum. 1960. Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology* 60(1): 44–51. <https://doi.org/10.1037/h0044430>.
- Lee, Hwajin and Hobin Ha. 2021. A longitudinal study of Korean pronunciation errors of Japanese learners: Focusing on the analysis of spontaneous speech data. *Korean Language Education Research* 56(2): 183–206. <https://doi.org/10.20880/kler.2021.56.2.183>.
- Lee, Nara and Hwanmin Jung. 2020. A perception study on Korean monophthongs produced by Vietnamese novice learners of Korean. *Bilingual Research* 78: 87–116. <https://doi.org/10.17296/korbil.2020..78.87>.
- Lee, Young Jin. 2019. Korean woman's lifestyle [Photograph]. iStock. <https://www.istockphoto.com/photo/korean-womans-lifestyle-gm1175718390-327509676>.
- Lenth, Russell V. 2025. Emmeans: Estimated marginal means, aka least-squares means. R package version 1.11.1.
- Lim, Jayeon and Misun Seo. 2011. Intelligibility and comprehensibility in the production of English sibilants by Korean learners. *Korean Journal of Applied Linguistics* 27(3): 189–212.
- McGowan, Kevin B. 2015. Social expectation improves speech perception in noise. *Language and Speech* 58(4): 502–521. <https://doi.org/10.1177/0023830914565191>.
- Munro, Murray J. 2008. Foreign accent and speech intelligibility. In Jette G. Hansen Edwards and Mary L. Zampini (eds.), *Phonology and second language acquisition*, 193–218. Amsterdam: John Benjamins Publishing Company.
- Munro, Murray J. and Tracey M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1): 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>.
- Munro, Murray J. and Tracey M. Derwing. 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 49(S1): 285–310. <https://doi.org/10.1111/0023-8333.49.s1.8>.
- Oh, Eunhae. 2018. Effects of L2 experience on the production of Korean stop contrasts by Mandarin Chinese learners. *Linguistic Research* 35(1): 233–251. <https://doi.org/10.17250/khisli.35.1.201803.009>.
- Phol_66. 2019. Asian girls wearing hanbok which is a Korean national dress in the village of Bukchon Hanok in South Korea [Photograph]. Shutterstock. <https://www.shutterstock.com/image-photo/asian-girls-wearing-hanbok-which-korean-1352528342>.
- R Core Team. 2026. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Roh, Mi-sook and Hyun-jung Song. 2024. A study on teaching Korean pronunciation using romanization: Focusing on beginner English-speaking learners. *The Journal of Yeongju Language & Literature* 57: 315–339.
- Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4): 511–531. <https://doi.org/10.1007/BF00973770>.
- Strand, Elizabeth A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1): 86–100. <https://doi.org/10.1177/0261927X99018001006>.
- Tarazevich, Anna. 2021. Elegant Asian woman wearing an ao dai and conical hat, smiling outdoors [Photograph]. Pexels. <https://www.pexels.com/photo/a-woman-in-white-ao-dai-wearing-conical-hat-7216917/>.
- Tran, Sang. 2021. Asian woman in a red qipao stands by a vintage door amidst festive red lanterns [Photograph]. Pexels. <https://www.pexels.com/photo/a-pretty-woman-in-red-dress-6724117/>.
- Wen, Xin. 2019. The pronunciation of Korean monophthongs by Chinese learners: Focus on the articulation reordering influences of the first language. *Journal of Korean Language Education* 44: 211–240. <https://doi.org/10.17313/jkorle.2019..44.211>.
- Wen, Xin. 2021. A study of Chinese female Korean learners' pronunciation of Korean monophthongs /ㅏ/: Focused on advanced Korean learners. *Korean Education* 126: 401–431. <https://doi.org/10.15734/koed..126.202103.401>.
- Yoo, Soyoung and Hyoun-Hwa Kang. 2018. Analysis of phonemic errors of Korean learners according to language and proficiency. *Language Facts and Perspectives* 44: 357–397. <https://doi.org/10.20988/lfp.2018.44..357>.
- Yun, Eunkyung. 2018. A study of L1 monophthongs perceived as continua of L2 vowels. *The Language and Culture* 14(4): 243–265. <https://doi.org/10.18842/klaces.2018.14.4.010>.
- Yun, Teresa. 2019. *A study on analysis of research trends in Korean language education focused on learners: Using language network analysis*. Master's Thesis. Kyung Hee Cyber University.
- Zhao, Wenyu. 2006. An acoustic study of the pronunciation of Korean monophthong by Chinese speakers. *The Journal of Chinese Language and Literature* 21: 365–388.

Appendix: Visual stimuli and source information

Table 8. Visual stimuli and source information

Stimulus	Visual identity cue	Source website	Creator/ Uploader	License
Korean 1		iStock	Lee (2019)	iStock License
Korean 2		Wikimedia Commons	Jeon (2016)	Wikimedia Commons License
Korean 3		Shutterstock	Phol_66 (2019)	Shutterstock License
Korean 4		iStock	Kuzmanovic (2023)	iStock License

Stimulus	Visual identity cue	Source website	Creator/ Uploader	License
American		Pexels	Cottonbro Studio (2020)	Pexels License
Chinese		Pexels	Tran (2021)	Pexels License
Japanese		Wikimedia Commons	Geraldshiel ds11 (2018)	Wikimedia Commons License
Vietnamese		Pexels	Tarazevich (2021)	Pexels License

Shiyu Zhang

Ph.D. Candidate

Department of Korean Language and Literature

Korea University, 145 Anam-ro, Seongbuk-gu,

Seoul, 02841, Korea

E-mail: saeokzhang@gmail.com

Drew Crosby

Assistant Professor

Department of English Language and Culture

The University of Suwon,

17, Wauan-gil, Bongdam-eup,

Hwaseong-si, Gyeonggi-do, 18323, Korea

E-mail: dmcrosby@suwon.ac.kr

Jiyoung Shin

Professor

Department of Korean Language and Literature

Korea University, 145 Anam-ro, Seongbuk-gu,

Seoul, 02841, Korea

E-mail: shinjy@korea.ac.kr

Received: 2026. 07. 20.

Revised: 2026. 09. 08.

Accepted: 2026. 09. 12.