

Role of proficiency and visual cue in Korean listeners' perception of accented speech

Kevin Sebastien Grailot · Eunhae Oh*

Department of English Language & Literature, Konkuk University, Seoul, Korea

Abstract

This study investigates how visual cues, speaker proficiency, and native language determine native Korean speakers' perception of accented Korean. Thirty participants rated the accentedness of L2 Korean speakers under three visual conditions, i.e., matching ethnicity, mismatching ethnicity, and no visual cues. Proficiency emerged as the strongest predictor of accentedness, with low-proficiency speakers rated significantly higher in accentedness than high-proficiency speakers. English-accented Korean was generally perceived as more foreign than Chinese-accented Korean. Visual conditions marginally affected the ratings but influenced the reaction times. Unlike earlier studies conducted primarily with English-speaking participants, our findings show the limited role of visual cues and indicate proficiency and phonetic similarity as the primary determinants of foreign-accentedness perception. Korean listeners rely more significantly on auditory than visual cues when evaluating foreign-accented speech, thus challenging assumptions about the universality of visual biases and suggesting that such biases may be culturally or linguistically specific instead of generalizable.

Keywords: accentedness perception, second language Korean, speaker proficiency, face–accent mismatch, listener familiarity

1. Introduction

Human language is inherently adaptable, enabling effective communication despite variations in pronunciation, rhythm, or intonation. This adaptability is especially evident when listeners encounter foreign-accented speech, a natural outcome of second language (L2) learning. In second language acquisition (SLA), foreign accentedness refers to how listeners judge the degree to which L2 speech deviates from native speaker norms (Munro & Derwing, 1998). While accented speech can influence judgments about a learner's proficiency, it does not necessarily hinder actual

communication. For example, heavily accented speech can still be intelligible and easy to understand (Derwing & Munro, 1997).

Although proficiency and accentedness are related, they are not the same. Proficiency refers to a speaker's overall ability in a second language, including grammar, vocabulary, fluency, and pronunciation (Kang, 2010; Tergujeff, 2021). Accentedness, on the other hand, specifically refers to how much a speaker's pronunciation differs from native speaker's norms. A person can be highly proficient, using complex grammar and speaking fluently, but still sound noticeably accented due to differences in pronunciation. Likewise, someone with a mild accent may still make

* gracekonkuk@gmail.com, Corresponding author

Received 21 February 2025; Revised 17 April 2025; Accepted 19 May 2025

© Copyright 2025 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

frequent grammatical errors if their overall proficiency is lower. Namely, accentedness is one aspect of proficiency, and it does not always reflect a speaker's full language ability.

Furthermore, perceptions of accentedness are not solely shaped by the acoustic signal. In addition to auditory features, visual cues, especially those related to perceived ethnicity, can also influence how accented a speaker is judged to be. Studies have shown that the way a speaker looks can affect how their speech is perceived, even when the audio remains constant. For example, Rubin (1992) found that identical audio recordings were rated as more accented and less intelligible when paired with a photograph of an Asian woman rather than a White woman. He attributed this effect not to audiovisual mismatch, but to stereotype-driven assumptions about who is considered a competent speaker of English, which were linked to a reduced willingness to engage with the speaker.

Later studies, such as Babel & Russell (2015), reported similar effects: Chinese-Canadian speakers were transcribed less accurately and rated as more accented when their photo was shown, while White-Canadian speakers were unaffected. They interpreted these results as evidence of a mismatch between social expectations and the speech signal, suggesting that indexical cues (like perceived ethnicity) may hinder perception when expectations are violated. Zheng & Samuel (2017) further explored whether seeing an Asian face makes speech sound more accented and found that while static images increased accentedness ratings, the effect weakened or disappeared when using dynamic, audiovisual stimuli. Crucially, their selective adaptation experiments showed no evidence that visual-ethnic cues altered low-level auditory perception, indicating that such effects are likely to operate at the level of interpretation rather than perception.

A growing body of research also shows that the influence of visual cues on accent perception is shaped by a combination of speaker characteristics, listener experience, and sociocultural context. Gnevshva (2018) demonstrated that German-accented English was rated as less accented in video-only conditions, but more accented when both audio and video were presented. This pattern suggests that visual-auditory integration is not automatic, but guided by listeners' expectations about how a speaker should sound based on their appearance. Listener-related factors also play a significant role. Hanulíková (2021) found age-related differences in visual bias: older adults rated speech as more accented when paired with an Asian face, while younger adults showed reduced bias. Furthermore, comprehension improved when visual and auditory cues aligned, particularly for younger listeners, indicating that age and experience shape the extent to which social expectations influence multimodal speech perception. The broader sociocultural context also modulates these effects. Kutlu et al. (2022) reported that static facial images affected accent ratings among listeners in Gainesville, Florida, but not in Montreal, Canada, suggesting that such effects are shaped by sociocultural context and exposure, rather than by language alone. This finding resonates with Babel & Russell (2015), who argued that mismatches between social-indexical cues and speech signals can impose a perceptual cost, but that the magnitude of this cost may vary with the listener's environment and expectations. Taken together, these studies emphasize that visual bias in accent perception is not universal, but shaped by speaker traits, listener characteristics, and the surrounding sociocultural context.

At the same time, not all research supports a consistent biasing

effect of ethnic visual cues. Eisenclas & Michael (2019) found no significant differences in comprehension or evaluation when participants were shown Asian versus Caucasian faces, suggesting that exposure to multicultural environments may reduce the impact of visual biases. Similarly, Algana (2021) reported that visual cues did not consistently enhance comprehension or reduce perceived accentedness. In some cases, they interfered with intelligibility when the visual information contradicted listeners' expectations. These findings point to the contextual and variable nature of visual cue effects, which may depend on factors such as listener background, speaker-listener familiarity, and the nature of the task.

While prior research has offered valuable insights into visual bias in English-speaking contexts, these findings stem from a wide range of methodologies and sociolinguistic settings. Some studies have examined native speech paired with manipulated ethnic cues in multicultural societies (e.g., Babel & Russell, 2015; Rubin, 1992), while others have focused on the perception of non-native speech in multilingual environments (e.g., Kutlu et al., 2022). However, it remains unclear whether these findings generalize to more culturally homogeneous contexts or to less commonly studied target languages such as Korean. The current study addresses this gap by investigating how native Korean listeners perceive accented Korean produced by speakers from different L1 backgrounds, across varying visual conditions. In addition, although speaker proficiency is well established as a predictor of perceived accentedness, its potential interaction with visual cues has received limited empirical attention. This study explores whether higher proficiency reduces susceptibility to visual bias and whether lower proficiency heightens sensitivity to socially driven expectations based on a speaker's appearance.

Finally, this study has practical implications for SLA pedagogy and cross-cultural communication. Understanding how linguistic and social factors shape accent perception in non-English contexts can inform language teaching strategies designed to reduce communicative barriers. For example, emphasizing phonetic familiarity with less common or less frequently encountered L2 accents (such as English-accented Korean) could help listeners build robust perceptual representations of diverse phonetic patterns. Ultimately, broadening learners' exposure to a wider variety of accents might encourage listeners to rely less on visual cues, which could trigger biases, and place greater weights on auditory signals. Such an approach would enhance listeners' perceptual adaptability and communicative competence in multilingual interactions (Figure 1).

2. Methods

This study explored how native Korean listeners rate foreign-accented Korean under different visual conditions (matching ethnicity, mismatching ethnicity, and no visual cues) and varying speaker proficiency levels. The methodology was inspired by Gnevshva (2018), who examined the effects of visual cues on accent perception by native English listeners. Adapting this framework, this study introduced three distinct visual conditions while controlling for auditory input to investigate the interplay between proficiency and visual bias in a non-English context.

2.1. Participants

Two groups of participants were involved in the study. The first

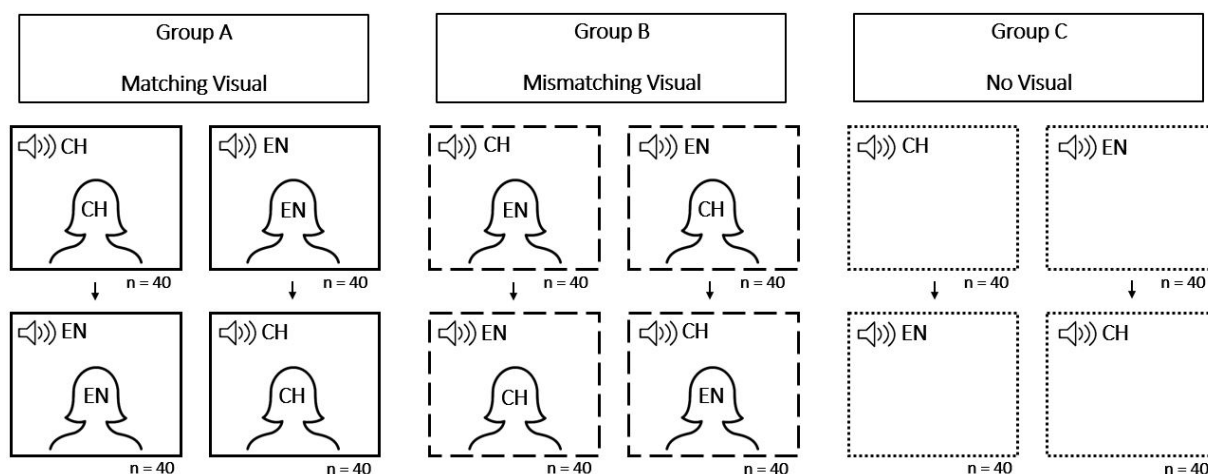


Figure 1. Overview of participant groups, visual conditions, and utterances per block.

group consisted of five native Korean speakers who served as proficiency assessment judges. Their role was to evaluate the Korean language proficiency of non-native speakers (L2 learners) based on recorded audio samples (see Appendix 1). Proficiency was assessed holistically based on general language competence, not accent alone. This group included four women and one man, aged 21 to 25 years (mean age=23.4; variance=2.64). They were all native speakers of Seoul Korean.

The second group comprised 33 native Korean speakers who participated in the main accent rating experiment. Their ages ranged from 19 to 29 years (mean age=22.15; variance=2.03), with 20 females and 13 males. All participants were university students compensated with 10,000 KRW. Data from three participants were excluded due to

ineligibility: one participant was too old (34 years old), one did not follow the instructions regarding visual cues, and one was a non-Korean. The final sample of 30 participants was randomly distributed across visual conditions. No one had hearing impairments, prior experience with accent-rating tasks, or specialized linguistic training.

2.2. Stimuli

The stimuli were selected from the AI-Hub Open Dataset¹, which contains spontaneous speech from Korean learners of various L1 backgrounds. All selected speakers were female, and their recordings were matched in duration, recording quality, and content type. Twenty speakers were initially selected: ten Chinese and ten English speakers, with each group split evenly between high and low proficiency levels. Only female speakers were included to avoid potential gender-based biases in accent perception. Each speaker provided three Korean sentences, which were combined into a 30-second audio file for proficiency assessment (see Appendix 1).

Using ePrime 2.0 software, these sentences were evaluated by five native speakers of Seoul Korean, who served as the proficiency assessment group, to identify four target speakers: one high-proficiency and one low-proficiency speaker from each

language group. These judges were instructed to focus on overall proficiency rather than accent alone and they rated the speakers on a scale going from 1 (low proficiency) to 9 (high proficiency). Due to the limited availability of highly proficient English-accented Korean speakers in the dataset, the high-proficiency English speaker received a slightly lower average rating than the high-proficiency Chinese speaker. This discrepancy is acknowledged as a potential influence and is discussed accordingly. Selected speakers included a high-proficiency Chinese speaker ($M=8.8$, $SD=0.45$), a high-proficiency English speaker ($M=7.6$, $SD=2.19$), a low-proficiency Chinese speaker ($M=1.6$, $SD=0.89$), and a low-proficiency English speaker ($M=1.8$, $SD=0.84$). These four speakers were selected as they represented the highest and lowest proficiency ratings in their respective language groups, ensuring a clear contrast in proficiency effects. Ratings were consistent across judges, with a standard deviation of 0.82.

For the accent rating task, 20 sentences were chosen from each of the selected speakers. These sentences were brief and neutral in content, avoiding any cues that might reveal the speakers' native language backgrounds. In addition, visual stimuli were prepared, consisting of images corresponding to the speaker's ethnicity. The facial images used in the visual stimuli were sourced from publicly available websites and used under fair use for research purposes. Each image was selected to represent a prototypical East Asian or Caucasian face with neutral expressions. Original sources are listed in Appendix 2. These images were paired with audio in matching and mismatching conditions, while no images were used in the audio-only condition.

2.3. Procedure

The accent rating task involved 30 participants listening to audio samples from the four selected speakers (one high- and one low-proficiency speaker from each language group) under three visual conditions: matching ethnicity, mismatching ethnicity, and no visual cue. Participants were informed prior to the experiment that they would be rating Korean speech produced by nonnative

¹ This research (paper) used datasets from 'The Open AI Dataset Project (AI-Hub, S. Korea)'. All data information can be accessed through 'AI-Hub (www.aihub.or.kr)'.

speakers. They completed a brief practice session using recordings not included in the main task to familiarize themselves with the rating scale and procedure. A post-experiment questionnaire collected information about participants' previous exposure to foreign-accented Korean and personal experience interacting with nonnative speakers. Participants rated each speaker's accent on a 9-point scale (1=no foreign accent, 9=strong foreign accent). The experiment was divided into two blocks, one featuring Chinese speakers and the other featuring English speakers. The order of blocks was counterbalanced across participants to minimize order effects.

Participants were randomly assigned to one of the three visual conditions. In the matching condition, the speaker's ethnicity corresponded to the visual cue (e.g., a Chinese speaker paired with a Chinese face). In the mismatching condition, the ethnicity of the visual cue differed from that of the speaker. In the audio-only condition, no images were shown. Participants listened to 80 utterances in total (40 per block) and were instructed to rate how strongly accented each sample sounded as quickly as possible after listening. Reaction times were recorded alongside the ratings.

2.4. Analysis

Two separate linear mixed-effects models were used to analyze accent ratings and reaction times, respectively. Both models included fixed effects for Visual Condition (Match, MisMatch, NoVisual), Proficiency (High, Low), and Language Background (Chinese, English), as well as all 2-way and 3-way interactions among these factors. Random intercepts were specified for both Participants and Sentences. All categorical predictors were contrast-coded using treatment coding, allowing for the interpretation of main effects as deviations from the grand mean. Model estimation was conducted using the lmer function from the lme4 package in R (Bates et al., 2015). Complete model summaries, including fixed effect estimates, standard errors, t-values, degrees of freedom, and *p*-values, for accent ratings are provided in Appendix 3.

3. Results

3.1. Accent Rating

The results of the linear mixed-effects model for accent indicated that speaker proficiency played a dominant role in determining perceived accentedness, while visual cues had minimal impact. The effects of visual condition (Match, MisMatch, and NoVisual) were not significant. The MisMatch condition ($\beta = -0.059$, $p = 0.436$) and the NoVisual condition ($\beta = -0.043$, $p = 0.587$) did not differ significantly from the baseline Match condition. This suggests that participants relied primarily on auditory information rather than visual cues when evaluating accentedness. Although Figure 2 suggests a slight trend toward lower perceived accentedness for English speakers in the MisMatch condition (where English-accented speech was paired with Asian faces), this result was not statistically significant. Thus, while this pattern may indicate a potential influence of participant expectations in cases of mismatched visual and auditory cues, the current data do not provide sufficient evidence to confirm such an effect. Proficiency emerged as the strongest predictor of accentedness ratings. Low-proficiency speakers were rated significantly higher in accentedness than high-proficiency speakers ($\beta = 2.026$, $p < .001$).

This effect was consistent across all visual conditions and both language groups, with high-proficiency speakers always perceived as less accented. Language background also significantly influenced accent ratings. English speakers were rated as having stronger foreign accents than Chinese speakers overall ($\beta = 0.351$, $p < .001$). However, for low-proficiency speakers, the Chinese speaker received the highest accentedness ratings, while for high-proficiency speakers, the English speaker was rated as more accented. These differences align with the proficiency judgments made by the initial assessment group.

The interaction between visual condition (MisMatch) and proficiency (Low) was significant ($\beta = 0.087$, $p = 0.044$), indicating that the MisMatch condition slightly amplified accentedness perceptions for low-proficiency speakers. However, no significant interaction was found between visual condition (NoVisual) and proficiency (Low) ($\beta = 0.042$; $p = 0.355$), which suggested that the absence of visual cues did not affect perceived accentedness. The interaction between visual condition (MisMatch) and language background (English) was also not significant ($\beta = -0.035$, $p = 0.420$), indicating that mismatch effects did not vary substantially across language groups. Similarly, the interaction between visual condition (NoVisual) and language (English) failed to reach significance ($\beta = -0.060$, $p = 0.182$).

The interaction between proficiency (Low) and language (English) was significant ($\beta = -0.484$, $p < .001$), showing that the perceived accentedness difference between high and low proficiency speakers was smaller for English speakers than for Chinese speakers. Additionally, although the interaction between Condition (NoVisual) and Proficiency (Low) was not significant, the direction of the effect was similar to the MisMatch condition, suggesting that visual absence does not systematically reduce or amplify accent ratings. None of the three-way interactions reached significance, indicating that the combined influence of Condition, Proficiency, and Language did not produce more complex effects beyond those captured by the significant two-way interactions.

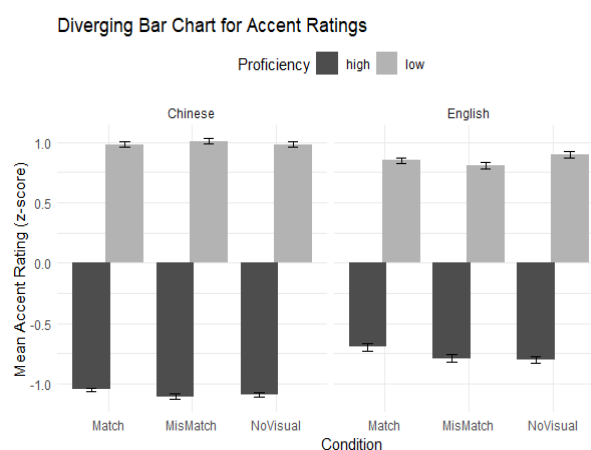


Figure 2. Diverging bar chart of accent ratings by condition (Match, MisMatch, NoVisual), speaker group (Chinese-accented, English-accented Korean), and proficiency (high, low).

3.2. Reaction Time

The mixed-effects model analyzing log-transformed reaction times (RTs) found no significant differences across visual conditions. The MisMatch condition ($\beta = -0.083$, $p = 0.687$) and

NoVisual condition ($\beta=-0.276$, $p=0.205$) showed slightly faster RTs compared to the Match condition, but these differences were not statistically significant. Low-proficiency speakers elicited slightly faster RTs than high-proficiency speakers ($\beta=-0.031$, $p=0.729$), though the difference was minimal and not significant. This trend may suggest that participants found it easier to identify accentedness in low-proficiency speakers due to their more noticeable speech deviations.

Participants responded slightly slower to English speakers than to Chinese speakers ($\beta=0.055$, $p=0.538$), though this difference was not significant. As shown in Figure 3, the slower RTs for English speakers may reflect greater cognitive processing demands, possibly due to lower familiarity with English-accented Korean compared to Chinese-accented Korean. The interaction between visual condition (MisMatch) and language (English) approached but did not reach statistical significance ($\beta=0.219$, $p=0.061$). Thus, although mismatched visual cues appeared to slightly slow reaction times for English speakers, interpretations of this finding should remain cautious given that it was not statistically significant.

No other interactions, including Condition \times Proficiency or three-way interactions, were significant ($p>0.1$). Furthermore, none of the interactions involving the NoVisual condition showed any notable trends, reinforcing the conclusion that the absence of visual cues does not meaningfully impact processing speed. Overall, these findings suggest that auditory factors, particularly speaker proficiency, are more salient than visual cues in influencing both perception and processing of foreign-accented Korean.

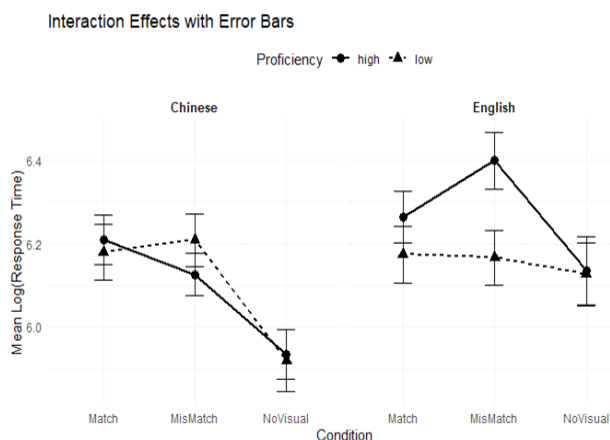


Figure 3. Effects of condition (Match, MisMatch, NoVisual), speaker group (Chinese-accented, English-accented Korean) and proficiency (high, low) on mean log-transformed reaction time.

4. Discussion

This study investigated how speaker proficiency and native language influence accentedness ratings and reaction times among native Korean listeners when evaluating foreign-accented Korean. It also examined the role of visual cues by testing three conditions: congruent ethnicity (Match), incongruent ethnicity (Mismatch), and no visual cues (NoVisual). Thirty participants rated L2 Korean speakers with Chinese and English native language backgrounds.

First, speaker proficiency was the most significant factor influencing accentedness ratings. High-proficiency speakers were consistently rated as less accented than low-proficiency speakers,

regardless of visual conditions or the speakers' native language. This finding aligns with prior research (e.g., Kang, 2010; Munro & Derwing, 1995), supporting the claim that linguistic competence plays a central role in reducing perceived foreignness. In addition, English-accented Korean was rated as more foreign than Chinese-accented Korean. This pattern likely reflects both phonological familiarity and the linguistic distance between the speaker's native language and Korean. Chinese-accented Korean is more frequently encountered in Korean contexts, which may contribute to greater listener familiarity. Phonetically, Mandarin and Korean share relatively simple syllable structures and lack the consonant clusters and vowel reductions common in English, potentially making Mandarin-accented Korean sound less foreign. As a result, participants tended to show more consistent ratings for Chinese speakers, while ratings for English-accented Korean were more variable. This pattern may also be linked to a significant interaction between visual mismatch and speaker proficiency, whereby mismatch conditions slightly increased accentedness ratings for low-proficiency speakers. Further acoustic analysis is needed to confirm whether these perceived differences are reflected in measurable phonetic features.

However, comparisons across language groups must be interpreted with caution, as the English high-proficiency speaker received lower overall proficiency ratings than the Chinese high-proficiency speaker. Independent-sample t-tests indicated a significant proficiency rating difference between high-proficiency English and Chinese speakers. This difference in proficiency levels may partly explain why the English high-proficiency speaker received higher accentedness ratings, potentially amplifying perceived differences between the language groups. Interestingly, the model also revealed a significant interaction between language and proficiency, indicating that low-proficiency English speakers were rated as less accented than expected based on the additive effects of language and proficiency alone. This may reflect a perceptual ceiling or floor effect, or that listeners' expectations shift when encountering multiple overlapping cues. It is also possible that listeners' unfamiliarity with English-accented Korean influenced their perception of proficiency, contributing to the lower ratings. Given the differences in speaker proficiency and the limited number of participants per condition, some variability in the results may be attributable to these constraints. Future research with more balanced proficiency levels and larger sample sizes is needed to validate these findings.

Although reaction time differences were not statistically significant, the overall pattern offers some tentative insights into listeners' processing of foreign-accented Korean. Responses to English-accented speech were slightly slower than those to Chinese-accented speech, particularly in the Mismatch visual condition, possibly reflecting increased cognitive effort when processing less familiar accents. A similar trend was observed for visual-auditory mismatches, where incongruent facial cues appeared to modestly delay responses, especially for English speakers. Interestingly, reaction times were also slightly faster for low-proficiency speakers, which may suggest that more noticeable deviations in speech made judgments easier, although this effect was minimal. Greater variability in response times for English-accented speakers could reflect listeners' lack of consistent perceptual expectations for this accent type. In contrast, more stable reaction patterns for Chinese-accented Korean may be associated

with greater exposure and familiarity. While these trends did not reach significance, they suggest that familiarity and expectation may play a role in the processing of accented speech, particularly when visual cues are involved.

Unlike prior studies in English-speaking contexts, which examined visual bias using native speakers and multicultural samples, the present study found no significant effect of visual cues on accentedness ratings. These differences likely reflect methodological and sociocultural variation: previous studies used L1 English speakers and diverse Western populations, whereas our study involved L2 Korean speakers assessed by native Korean listeners in a relatively homogeneous context. The results suggested that, although incongruent visual cues tended to delay processing, Korean listeners primarily relied on auditory information. This may reflect limited exposure to diverse foreign-accented Korean speech, particularly English-accented Korean, reducing the salience of visual-auditory mismatches.

This interpretation aligns with McGowan (2015), who found that transcription accuracy improved when a Chinese-accented voice was paired with a congruent Asian face and declined with a mismatched White face. Although his study used speech-in-noise rather than accentedness ratings, the findings support the broader claim that visual-auditory congruence facilitates speech processing. In the present L2–L2 context, mismatched cues may amplify perceived accentedness when listeners associate certain ethnicities with expected fluency. Chinese-accented Korean, more commonly encountered in Korea, may be associated with higher expected proficiency due to regional and cultural familiarity, whereas English-accented Korean is less familiar and perceived as more foreign. Thus, when an Asian face is paired with English-accented speech, listeners may detect a stronger mismatch, heightening perceived accentedness.

It should be noted, however, that our results contrast with findings from Kutlu et al. (2022), who found that listeners in Gainesville, Florida, a community with lower exposure to multilingualism, showed greater sensitivity to visual ethnic cues than listeners in Montreal. While both studies show the role of listener experience, the discrepancy may stem from differences in sociolinguistic context, the racialization of English-accented speech, or culturally specific expectations surrounding accentedness. It is possible that the ethnic visual stimuli used in the present study may not have elicited social expectations as strong as those observed in the Florida context examined by Kutlu et al. (2022). Moreover, Korean listeners may be less accustomed to associating facial cues with specific patterns of Korean proficiency, particularly when both groups are perceived as foreign but not racialized in contrasting ways.

Alternatively, the lack of visual effects may reflect the dominant role of phonological familiarity in accent perception. Chinese-accented Korean elicited lower accentedness ratings and faster reaction times, suggesting that listeners relied on more established auditory templates. In contrast, unfamiliarity with English-accented Korean may have increased cognitive effort, especially under mismatched conditions. It is also possible that the controlled laboratory setting—with brief utterances and static facial images—limited natural audiovisual integration. In real-world contexts, where dynamic visual cues such as facial expression and gesture are present, visual information may play a stronger role in shaping perception. Future research should investigate these possibilities

using more naturalistic and ecologically valid audiovisual stimuli.

References

- Algana, M. S. A. (2021). *L2 English listeners' comprehension: An examination of the effects of visual cues and nonnative accented speech* (Doctoral dissertation). Michigan State University, East Lansing, MI.
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, 137(5), 2823-2833.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1-16.
- Eisenclas, S. A., & Michael, R. B. (2019). What's in a face? The impact of nonlinguistic 'ethnic' facial features on accent perception. *Journal of Multilingual and Multicultural Development*, 40(10), 879-891.
- Gnevshva, K. (2018). The expectation mismatch effect in accentedness perception of Asian and Caucasian non-native speakers of English. *Linguistics*, 56(3), 581-598.
- Hanulíková, A. (2021). Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups. *PLoS ONE*, 16(10), e0259230.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301-315.
- Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, 7(1), 7.
- McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, 58(4), 502-521.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97.
- Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, 48(2), 159-182.
- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, 33(5), 511-531.
- Tergujeff, E. (2021). Second language comprehensibility and accentedness across oral proficiency levels: A comparison of two L1s. *System*, 100, 102567.
- Wesołek, S., Gulowski, P., Błaszczak, J., & Żygis, M. (2023, August). What influences foreign accent strength? Phonological and grammatical errors in the perception of accentedness (pp. 3098-3102). *Proceedings of INTERSPEECH 2023*. Dublin, Ireland.
- Zheng, Y., & Samuel, A. G. (2017). Does seeing an Asian face make speech sound more accented? *Attention, Perception, & Psychophysics*, 79, 1841-1859.

• **Kevin Sebastien Grailot**

MA student, Dept. of Language and Literature, Konkuk University

120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Korea

Tel: +82-10-9762-7088

Email: graillok@gmail.com

Areas of interest: Foreign accent perception, Sociolinguistics, Language acquisition

• **Eunhae Oh**, Corresponding author

Professor, Dept. of Language and Literature, Konkuk University

120 Neungdong-ro, Gwangjin-gu, Seoul 05029, Korea

Tel: +82-2-450-3046

Email: gracekonkuk@gmail.com

Areas of interest: Phonetics, Language acquisition, Sociophonetics

Appendix

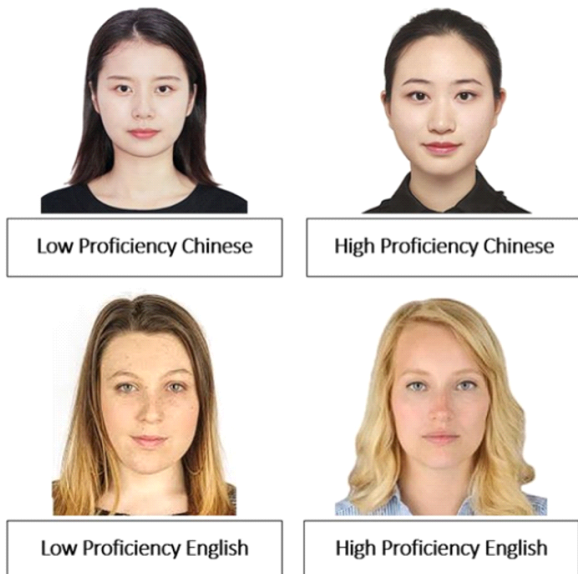
Appendix 1. Samples of the sentences used for the proficiency rating experiment, followed by samples of the sentences used for the accent rating experiment

Speaker	Proficiency	Language	Sample sentence
CN0070	High	Chinese	조부모가 손주를 돌보는 경우, 부모님과 조부모의 요가 방식의 차이가 나, 서로에게 마음의 상처를 주기도 한다.
CN0048	Low	Chinese	저 식당에서 혼자 밥을 먹어본 적이 없어요.
EN0008	High	English	회사의 일 때문에 바빠서 못 올 것 같다고 연락한 친구도 다행히 시간이 돼서 올 수 있다네요.
EN0035	Low	English	숨쉬기가 답답해요. 화장이 다 묻어서 불편해요.
CHHIGH	High	Chinese	그러면 저희는 방으로 예약하겠습니다. 방을 이용할 때 주의해야 할 점 같은 것이 있나요?
CHLOW	Low	Chinese	요즘 헬스장에 가서 운동을 많이 해요. 몸에 좋기 때문이에요.
ENHIGH	High	English	카페에는 여러 가지 음료 말고도 맛있어 보이는 달콤한 디저트들이 진열되어 있었어요.
ENLOW	Low	English	저는 내성적이에요. 그런데 저는 꼼꼼하고 그리고 친절해요.

Appendix 3. Accent rating results

Fixed effect	Estimate	Std. Error	df	t-value	p-value
(Intercept)	-1.045	0.058	49.66	-17.91	<.001***
Condition: MisMatch	-0.059	0.076	38.83	-0.79	0.436
Condition: NoVisual	-0.043	0.079	38.81	-0.55	0.587
Proficiency: Low	2.026	0.043	180.67	47.63	<.001***
Language: English	0.351	0.043	180.67	8.26	<.001***
MisMatch × Proficiency: Low	0.087	0.043	2,521.06	2.02	0.044*
NoVisual × Proficiency: Low	0.042	0.045	2,521.02	0.93	0.355
MisMatch × Language: English	-0.035	0.043	2,521.06	-0.81	0.420
NoVisual × Language: English	-0.060	0.045	2,521.02	-1.33	0.182
Proficiency: Low × Language: English	-0.484	0.060	180.67	-8.04	<.001***
MisMatch × Proficiency: Low × Language: English	-0.032	0.061	2,521.04	-0.53	0.595
NoVisual × Proficiency: Low × Language: English	0.108	0.064	2,521.02	1.70	0.089 .

p<.1, *p<.05, **p<.01, ***p<.001



Appendix 2. Visual cues used for the accent rating experiment.