Applied Psycholinguistics (2026), **47**, e11, 1–34
doi:10.1017/S0142716425100416

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

The effect of talker identity and language experience in perceiving foreign-accented speech

Jeonghwa Cho[1] and Harim Kwon[2]

[1]Department of English Language and Literature, Hongik University, Seoul, Republic of Korea and [2]Department of English Language and Literature, Seoul National University, Seoul, Republic of Korea
Corresponding author: Harim Kwon; Email: harimkwon@snu.ac.kr

(Received 6 January 2025; revised 29 October 2025; accepted 2 December 2025)

Abstract

The current study examines the effect of talker identity and linguistic experience on the perception of novel speech patterns by English speakers, focusing on vowel insertion in Korean-accented English. Experiment 1 shows that English speakers with no experience of living in Korea identify English words with vowel insertion as valid words more frequently throughout the experiment only when the talker is described as Korean, but not when the talker is American or Mexican. In Experiment 2, we find similar results with English speakers living in Korea, who provide more word responses to vowel-inserted English words in the Korean talker condition but not the American talker condition. Comparing Experiment 1 and Experiment 2, participants living in Korea show a greater preference for the inserted vowel that is similar to the one found in Korean-accented English ([ʊ]) over the control vowel ([ɪ]), as well as faster adaptation to this type. These results suggest that both talker identity and previous exposure to an accent influence how listeners perceive and adapt to foreign-accented speech, consistent with exemplar models of speech perception.

Keywords: Korean-accented English; speech adaptation; talker identity; vowel insertion

Introduction

Languages have different phonotactic constraints, allowing or prohibiting specific sound sequences. As a result, foreign sequences not permitted in speakers' or listeners' native languages are often "repaired" (e.g., Hallé et al., 2008; Kang, 2011). For example, onset consonant clusters in English are phonotactically illicit in Korean. Consequently, Korean speakers tend to perceive an illusory vowel that does not have acoustic correlates in the signal (e.g., Berent et al., 2008) or produce a vowel between consonants, either phonologically epenthesized or excrescent, to break up the illicit onset clusters (e.g., Kang, 2012; Shin and Iverson, 2011, 2014). The vowel

The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

insertion is often present in the speech of Korean learners of English and can introduce a challenge to English speakers communicating with the learners. To correctly identify English words containing onset clusters produced with the vowel insertion, English-speaking listeners have to adjust their perception strategies.

How listeners perceive foreign-accented speech may not always be identical. In the framework of the exemplar-based models (Clapp et al., 2023; Goldinger, 1998; Johnson, 1997; Nygaard and Pisoni, 1998; Palmeri et al., 1993; Sumner et al., 2014), listeners store episodic memory traces for the speech they are exposed to and use them when perceiving new speech inputs. The episodic memory not only includes phonetic details but also non-linguistic information, such as who the talker is, what the situation is, and so on. Therefore, listening to a familiar accent and a novel accent are two very different tasks. As listeners may have stored numerous exemplars for the familiar accent, they would rely on these memory traces when listening to a (purported) speaker of the accent. Thus, exemplar-based models predict that listeners' perception strategies are affected by both the purported talker identity and their linguistic experience (e.g., McGowan, 2015; Melguy and Johnson, 2021; Vaughn, 2019). This study aims to systematically investigate how speech perception is shaped by those two factors by expanding on Darcy and Thomas (2019), where English native speakers with little experience with Korean-accented English reject English words produced with vowel insertion. Here, we compare responses of English speakers with varying familiarity with Korean-accented English on a lexical decision task for English words with the vowel insertion while manipulating the talker identity.

Background

The effect of talker identity on speech perception

Earlier studies of speech perception proposed that any variations in speech, including talker details, are irrelevant to listeners' task of retrieving the linguistic messages and thus abstracted away (e.g., Jakobson et al., 1952; Kiparsky, 1973; Stampe, 1979). More recently, however, it has been suggested that listeners use such details when perceiving speech sounds and retrieving lexical information rather than filtering them out (the exemplar-based models of speech perception; e.g., Clapp et al., 2023; Goldinger, 1998; Johnson, 1997; Nygaard and Pisoni, 1998; Palmeri et al., 1993; Sumner et al., 2014). For example, the information on talkers' gender affects how listeners distinguish /s/ and /ʃ/ (Bouavichith et al., 2019; Johnson, 1991; Munson et al., 2006). In these studies, English-speaking participants reported that they were hearing /ʃ/ more often than /s/ in a /s-ʃ/ continuum when they were presented with an image of a female talker. This effect was evident in the opposite direction as well, such that the participants thought the talker was a female more often when they were primed with words starting with /ʃ/ (e.g., *shack*). This indicates that both the perception of lexical or phonological categories and the perception of speaker information are malleable and that they interact with each other.

Talkers' age, social class, and accent also affect speech perception. In an event-related potentials study, Van Berkum et al. (2008) measured listeners' neural

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

responses when they heard sentences that did not match the talker's age or social class (e.g., *Every evening I drink some wine before I go to sleep* spoken with a young child-like voice, or *I have a large tattoo on my back* with an upper-class accent). Listeners show an N400 effect in both cases, indicating that they consider such sentences semantically anomalous, although the sentences themselves were semantically natural. Furthermore, Cai et al. (2017) show that regional accents modulate speech perception—in particular, auditory word recognition. They report that when a word such as "bonnet," which has a different meaning in American English ("a hat") and in British English ("a hood of a car"), is pronounced in a British accent, participants interpret its meaning to be a car part more often than a hat. Based on these findings, Cai et al. (2017) propose that spoken word recognition involves two pathways that are activated in parallel: the lexical-semantic pathway that maps the acoustic-auditory input onto linguistic representations such as phonemes, syllables, and wordforms, and the indexical pathway that encodes the acoustic details to infer information about the talker (c.f., Pierrehumbert, 2016; Sumner et al., 2014). These two pathways interact such that the "speaker model," formed through the indexical pathway, can modulate lexical access in the lexical-semantic pathway.

Information about the talker's identity may also facilitate the processing of foreign-accented speech and adaptation to it when that information matches the phonetic characteristics of the speech sounds, including segmental and prosodic information (McGowan, 2015; Melguy and Johnson, 2021; Vaughn, 2019). While some earlier accounts have suggested otherwise, such that the non-native status of the talker inhibits foreign-accented speech perception (Kang and Rubin, 2009; Rubin, 1992), McGowan (2015) reports that English-speaking participants who see an Asian face are more accurate in transcribing Chinese-accented speech than those who see a Caucasian face. This result indicates that the alignment between the given information about the speaker and the speech signals makes it easier to process foreign-accented speech. This facilitatory effect is driven by the listeners' expectations based on the information about the talker provided along with the speech materials. Listeners understand the signals better when their expectations align with the input than when they do not, as the mismatch induces greater cognitive efforts (Van Engen and Peelle, 2014).

The effect of linguistic experience on speech perception

The aforementioned studies that show talker identity effects in speech perception suggest that listeners make use of fine-grained phonetic details to perceive the talker identity and adjust to various talkers to optimize their comprehension of the linguistic messages (e.g., Cai et al., 2017; McGowan, 2015; Vaughn, 2019). The underlying assumption is that listeners are already equipped with knowledge about the fine-grained phonetic differences that different talker identities may induce. This raises an important question about the role of prior experience with specific talkers: do listeners who are familiar with certain talkers and their speech perceive foreign-accented speech differently than those who are not?

Studies on the perception of foreign-accented speech have shown that, in general, native speakers experience more difficulties processing non-native accents than

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

native accents (Floccia et al., 2009; Munro, 1998). For example, when English-speaking participants hear English utterances (e.g., *Most people wear hats on their feet* [false]) with and without a Mandarin accent, it takes them longer to determine the truth or falsehood of the utterance when produced with a Mandarin accent than a native accent (Munro, 1998). Nevertheless, previous experience with non-native accents may influence the process of making various linguistic decisions during speech perception (e.g., Caffarra and Martin, 2019; Cai et al., 2017; McGowan, 2015). That is, listeners with prior exposure to a foreign accent can adjust their processing strategies based on relatively robust representations compared to those without such exposure. Cai et al. (2017) report that the general pattern of the accent effect on disambiguating lexical ambiguity is limited to listeners with prior exposure to the accent in question. Specifically, when British participants and American participants are asked to write a word that comes to their mind upon hearing words such as "bonnet" spoken with a British or American accent, the accent effect is not evident for American participants with little experience with British English. Caffarra and Martin (2019) show that Spanish speakers who are more familiar with English-accented Spanish experience less difficulty processing Spanish agreement errors spoken with an English accent than those with less familiarity, as evidenced by their reduced P600 response. In McGowan (2015), English speakers who have grown up in Chinese-speaking households are overall more accurate in transcribing Chinese-accented English than their monolingual counterparts.

Put together, although listeners who have been exposed to foreign-accented speech may still have difficulties with the accents (e.g., Munro, 1998), prior experience with certain accents seems to have some processing advantages, as the listeners are able to adjust their processing strategies based on the experience. Listeners who have not had such an experience, on the other hand, may not be able to readily make well-informed adjustments as they are less likely to be equipped with the knowledge about the accent in question. In the current study, we examine whether prior experience with Korean-accented English affects the accurate recognition of a specific feature of the accent, namely, vowel insertion to break up onset clusters. Specifically, we examine the proportion and speed at which words spoken with that feature are recognized by English listeners who are familiar or unfamiliar with Korean-accented English.

The role of talker identity and linguistic experience in speech adaptation

It has been repeatedly shown that listeners can quickly adapt[1] to foreign-accented speech even in just one minute (e.g., Bradlow and Bent, 2008; Clarke and Garrett, 2004; Sidaras et al., 2008; but see Adank and McQueen, 2007; Floccia et al., 2009). This kind of quick online adaptation seems to interrelate with the longer-lasting expectation-based benefits of the talker information (e.g., McGowan, 2015; Vaughn, 2019). For instance, Vaughn (2019) shows that listeners better adapt to the speech signals when provided with talker information than when not. Listeners in Vaughn (2019) are asked to transcribe English sentences spoken with a Spanish accent and are told that the talker is either a Latinx English speaker, a Spanish learner of English, or provided with no information. Not only did the former two groups show overall greater accuracy than the other group, but they also demonstrated a greater increase

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

in accuracy throughout the experiment, as revealed by a quartile-based analysis of the trials. These findings suggest that the talker information has both the expectation-based benefits (i.e., overall better performance when the talker background is provided than not) and online perceptual learning (i.e., better adaptation). Such findings seem to corroborate the claim that these two processes are not separate but have a common neural mechanism (e.g., predictive coding account; Sohoglu and Davis, 2016). When listeners are provided with prior information that matches the speech signals, the perceptual learning of, or adaptation to, the novel speech patterns is enhanced as the prediction error from the top-down predictive processes (i.e., a mismatch between the prediction and the sensory input) is reduced. These online adjustments, over time, can lead to more accurate predictions, in turn, leading to longer-term benefits of prior knowledge. While the predictive coding account makes specific predictions about the advantages (vs. disadvantages) of the matching (vs. mismatching) prior information, in Vaughn (2019), both types of prior information matched the speech signals, as the speech could be interpreted as recorded by either a Latinx English talker or a Spanish learner of English. The impact of talker information, either matching or mismatching the speech signals during online adaptation, is not clear, which is one of the research questions in the current study. Specifically, we examine English listeners' adaptation to a phonetic feature typical of Korean-accented English with matching (Korean) or mismatching (American and Mexican) talker information. Comparing Korean and Mexican talker conditions allows us to test whether listeners respond in the same manner for non-native talkers regardless of their purported native language or adapt different strategies based on the talkers' language background. Additionally, we also test whether the previous exposure to Korean accents enhances adaptation to the phonetic feature during the experiment, which has not previously been investigated.

Phonotactic repairs in L2: vowel insertion in Korean-accented English

Phonotactic constraints are language-specific. Therefore, what is allowed in one language may not be permissible in another language. Models of L2 speech learning (e.g., Flege, 1995; Best and Tyler, 2007) state that learners perceive or produce L2 sounds based on their L1. This often results in modifying L2 sound sequences by deleting, inserting, or changing sounds to comply with the learners' L1 phonotactics, a process known as phonotactic repair (e.g., Hallé et al., 2008).

One example of this kind of phonotactic repair is vowel insertion to repair phonotactically illicit consonant clusters. Spanish speakers insert [e] before word-initial [s]-consonant clusters in English, since those clusters are not allowed in Spanish (e.g., *snob* becomes *esnob*, Cuetos et al., 2011). As Korean does not allow onset consonant clusters altogether, when Korean speakers perceive or produce consonant clusters in English, they tend to insert a vowel similar to /ɨ/ inside, not before, the cluster (e.g., *blue* is perceived or pronounced as /bɨlu/, Berent et al., 2008; Kang, 2012; Shin and Iverson, 2011, 2014). Whether these inserted vowels are excrescent or phonological is controversial and beyond the scope of the current investigation. However, Darcy and Thomas's (2019) findings suggest that these phonotactic repairs may lead to L2 learners having different representations than native speakers of the target language. Darcy and Thomas (2019) illustrate such representational differences between English speakers and Korean learners of

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

English by comparing the two groups of participants in an auditory lexical decision task with English stimuli with word-onset CC clusters. The stimuli include English words produced in three forms: a native-like pronunciation (e.g., /blu/ for *blue*), a form with /ʊ/ inserted between the two consonants (e.g., /bʊlu/), and a form with /ɪ/ inserted (e.g., /bɪlu/). All stimuli are recorded by a phonetically-trained male native speaker of English. The vowel /ʊ/ sounds similar to the vowel typically inserted by Korean speakers to break up the consonant cluster, whereas /ɪ/ does not, hence serving as a control condition. English speakers in Darcy and Thomas (2019) reject the /ʊ/-inserted and the /ɪ/-inserted forms at similar rates and with comparable response times, treating both as nonwords. In contrast, Korean participants reject the /ʊ/-insertion less frequently and more slowly than the /ɪ/-insertion, suggesting that they may have the /ʊ/-insertion as well as the well-formed form in their lexicon. This outcome, according to Darcy and Thomas (2019), indicates that Korean learners of English, unlike English native speakers, have dual representation (e.g., both /blu/ and /bʊlu/ for *blue*) for English words in their mental lexicon.

In Darcy and Thomas (2019), /ʊ/ is used to approximate the Korean epenthetic vowel, which the authors transcribe as /ɯ/. While the Korean epenthetic vowel, in our view, is closer to a high central unrounded vowel /ɨ/ than /ɯ/, American English /ʊ/ would be a reasonable counterpart to this Korean epenthetic vowel due to its acoustic and perceptual similarities, regardless of the transcription convention. For example, according to Yang (1996), Korean /ɨ/ is close to American English /ʊ/ and /u/ in its formants, but is clearly distinct from /ɪ/ or /ʌ/. Similarly, Ryu (2018) shows that Korean /ɨ/ is most similar to Canadian English /ʊ/ based on a linear discriminant analysis. Furthermore, Heo and Park (2012) report that native English listeners most frequently map Korean /ɨ/ to English /ʊ/ in perceptual assimilation tasks. These findings suggest that English /ʊ/ is the most likely perceptual counterpart of Korean /ɨ/ for North American English listeners, despite their differences in lip rounding.

The difference between English speakers and Korean learners of English in Darcy and Thomas (2019) is intriguing. However, the English-speaking participants in their study would presumably have little, if any, exposure to Korean or Korean-accented English, given that they are monolinguals who have not lived in any other country than the United States. It remains to be tested whether English speakers who are familiar with Korean-accented English will show patterns more similar to Korean speakers owing to their previous exposure to the vowel insertion. Also, Darcy and Thomas (2019) have not examined whether the English monolinguals show any evidence for adaptation over the course of the experiment, although online adaptation is likely to occur, considering the aforementioned findings on rapid adaptation to foreign-accented speech (e.g., Bradlow and Bent, 2008; Clarke and Garrett, 2004; Sidaras et al., 2008).

This study

The current study aims to investigate the role of (1) purported talker identity and (2) listeners' prior linguistic experience in recognizing and adapting to foreign-accented speech, focusing on a common feature of Korean-accented English, namely the vowel insertion to break up onset clusters. We compare English speakers living in

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

the US (Experiment 1) and those living in Korea (Experiment 2) on an auditory lexical decision task on words beginning on CC clusters with or without the phonotactic repair (i.e., vowel insertion), with the talker described as American, Korean, or Mexican (Experiment 2 tested only American and Korean talker conditions). Motivated by Darcy and Thomas (2019), we use two different types of vowel insertion, [ʊ] and [ɪ]. The vowel [ʊ] is used to approximate the epenthetic [ɨ] commonly produced by Korean speakers, following Darcy and Thomas (2019). Word-initial CC clusters broken up by inserted [ʊ], but not [ɪ], could be interpreted as a feature of Korean-accented English. We ask whether the talker identity and the prior experience modulate the participants' tendency to accept tokens with [ʊ] insertion and [ɪ] insertion as English words and overall speech of doing so. In addition, we investigate whether the participants' behaviors change over the course of the experiment, showing evidence for online adaptation. The outcomes suggest that talker identity affects adaptation patterns, while previous exposure facilitates both the overall speed of recognizing those words and adaptation.

Experiment 1

Experiment 1 examines the effect of alleged talker identity (American, Mexican, and Korean) on perceiving onset clusters broken up with vowel [ʊ] or [ɪ] insertion by English speakers living in the US.

Methods

Participants

Fifty-one English speakers (20 males, 28 females, 3 other responses; age: mean = 26.26, SD = 8.36) were recruited at a university in the midwestern US (n = 26) and an online recruitment platform Prolific (n = 25). All participants who were recruited from Prolific were residing in the United States at the time of participating in the study. One participant recruited in person who reported having lived in Tokyo, Japan, was excluded from data analysis. As the Japanese also repair English consonant clusters with similar vowel insertion (e.g., Shoji and Shoji, 2013; Yazawa et al., 2015), the experience of living in Japan would have increased exposure to the vowel insertion that is tested in the current study. The rest of the participants reported neither experience of living in Korea nor other Asian countries nor knowledge of Korean. Participants were assigned to one of the three talker conditions (American: n = 17; Korean: n = 17; Mexican: n = 16). For all participants, the mean self-reported familiarity with Korean-accented English was 3.20 (SD = 1.74) on a 1-7 Likert scale. Additionally, those who were assigned to the Mexican talker rated their familiarity with Spanish-accented English as 4.50 (SD = 2.13) on average.[2]

Materials

A total of 72 experimental stimuli and 120 filler words were used. The experimental stimuli consisted of 24 English monosyllabic words containing an onset cluster consisting of an obstruent C1 followed by a liquid C2 (e.g., *blue*, *cream*) and 48

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Table 1. Examples of experimental stimuli and the mean duration, f0, F1, and F2 of the inserted vowels (SD in parenthesis)

Type	Example	Mean duration (ms)	Mean f0 (Hz)	Mean F1 (Hz)	Mean F2 (Hz)
(a) Word	*sleep* [sliːp]	–	–	–	–
(b) [ʊ]-type (Test)	[sʊliːp]	98.9 (16.9)	125.4 (14.6)	394.9 (27.5)	959.6 (122.3)
(c) [ɪ]-type (Control)	[sɪliːp]	81.3 (11.4)	121.5 (10.8)	387.5 (27.6)	1437.4 (206.8)

pseudowords that were formed with a vowel inserted between the cluster. Twenty-four of the pseudowords had the vowel [ʊ] inserted ("[ʊ]-type"; e.g., [bʊˈluː]), which sounds similar to the inserted vowel in Korean-accented English, and the other 24 contained the vowel [ɪ] inserted ("[ɪ]-type"; e.g., [bɪˈluː]). The obstruents in onset clusters had an equal number of places of articulation (eight labials, eight alveolars, and eight velars), half of which were followed by [l] and the other half followed by [ɹ].

The filler items consisted of 72 English words and 48 pseudowords to match the total number of words and pseudowords (e.g., *coob*, *begut*) throughout the experiment. In addition, as the 48 pseudowords in the experimental stimuli were disyllabic (if the listeners perceived the inserted vowels as a syllabic nucleus), the total proportion of monosyllabic (e.g., *air*) and disyllabic words (e.g., *corrupt*) and pseudowords was matched by constructing the filler items to consist of 24 monosyllabic and 48 disyllabic words, and 24 monosyllabic and 24 disyllabic pseudowords. None of the filler items contained onset clusters. As a result, a participant listened to 72 monosyllabic stimuli and 120 disyllabic stimuli.

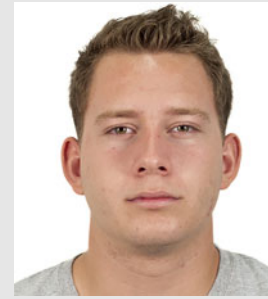
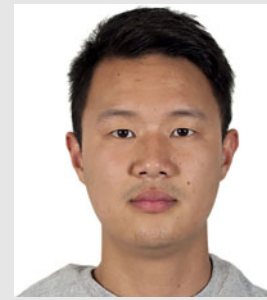
The stimuli were recorded by a linguistically trained male talker in his 20s who was born and grew up in Indiana, US, speaking English as his first language. The recording was conducted in a sound-attenuated room at Seoul National University, using Praat software (Boersma & Weenink, 2016) and a Shure BETA 87A microphone attached to a laptop computer through a Sound Devices USBPre 2 preamplifier, with the sampling rate of 44,100 Hz. The talker was presented with the stimuli in IPA (International Phonetic Alphabet) symbols and produced each item three times. From the three repetitions, one token with the most natural intonation and, for experimental pseudowords, with a clear inserted vowel, was selected for inclusion. See Table 1 for sample tokens and the mean duration, fundamental proportion (f0), and formant frequencies (F1 and F2) of the inserted vowels. A full list of the experimental stimuli with the acoustic details for each token is presented in Appendix A.

Procedure

Participants performed a lexical decision task using the Psychopy software (version 2020.1.3; Peirce, 2007) either in person or online. They listened to auditory stimuli through headphones (the AKG K271 MKII headphone was used for in-person participants, and online participants used their own) and decided whether each item was an English word or not. Online participants were instructed to complete the task

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Table 2. Descriptions of the talker's language background and face image

Talker identity	Language background	Face image
American	Next, you will hear speech from Daniel, whose face you can see in the image below. He was born and raised in Ohio, US, and English is his native language.	
Korean	Next, you will hear speech from Jihoon, whose face you can see in the image below. He was born and raised in Seoul, Korea, and speaks English as a second language.	
Mexican	Next, you will hear speech from Pedro, whose face you can see in the image below. He was born and raised in Mexico City, Mexico, and speaks English as a second language.	

in a quiet space while wearing headphones. All participants had a practice session of twelve trials consisting of six words and six pseudowords. The procedure was similar to the main experiment, but the auditory stimuli were recorded by a female English speaker, and no prior information about the talker was provided. Participants heard words and pseudowords in random order and indicated whether they heard real English words by pressing "z" or "m." The association of the keys and yes/no responses was counterbalanced, such that half of the participants pressed "z" for "yes" and "m" for "no," and the other half pressed "z" for "no" and "m" for "yes."

Before the main experiment, participants saw a screen with one of three descriptions of the talker's language background, with a corresponding image of their face selected from the Chicago Face Database (Ma et al., 2015) (Table 2).

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Then they performed a lexical decision task. The alleged talker's face remained on the screen throughout the main experiment. The main experiment consisted of three blocks with a three-minute long break between the blocks. Each block had 24 experimental items (8 items per condition) and 40 filler items. Since each word had three different forms (e.g., *blue* [bluː], [bʊluː], and [bɪˈluː]), each block contained only one of those forms per word to minimize any repetition effect. The order of the blocks was randomized for each participant. Participants' responses and reaction times were recorded during the experiment.

Data analysis

Trials with excessively short or long reaction times (less than or greater than 2.5 SD from the mean per participant) were removed, which accounted for 3.0% of the whole data. Also, data from one participant recruited online was excluded due to low accuracy on filler items (< 70%), as the low accuracy can be interpreted as a lack of sufficient effort in performing the task. For statistical analysis, participants' responses (word or not, coded as 1 or 0) were analyzed with a generalized mixed-effects regression model using the *glmer* function (Bates et al., 2015) in R (R Core Team, 2021) with a binomial family. This model had Stimuli type (word, [ʊ]-type, and [ɪ]-type), Talker (Korean, American, and Mexican), Order (Block 1, Block 2, and Block 3), and all possible interactions as fixed effects. Dummy coding was used for Stimuli type with [ʊ]-type as reference and simple effect regression coding (simple-effect contrasts with sum-to-zero weights: C1 = c(-1/3, 2/3,-1/3), C2 = c(-1/3,-1/3, 2/3)) for Talker (reference = Korean) and Order (reference = Block 1). Under this specification, the intercept equals the [ʊ]-type grand mean across Talker and Order, and the coefficients compare each non-reference level to its reference. The effects of Platform (online vs. in-person), C1 Place of articulation (Labial, Alveolar, or Velar), and C2 (r vs. l) were not statistically significant, nor did those factors increase the model fit, so they were excluded from the final model[3]. As for random effects, results from the most complex model that reached convergence are reported, which includes a random slope for Stimuli type for item and random intercepts for participant and item.

Reaction times were measured from the offset of each stimulus. As the reaction time values are positively skewed, those for word responses were transformed according to the Box-Cox test ((y^{lambda}-1)/lambda) (Box and Cox, 1964)[4] and were analyzed with a linear mixed-effects regression with the *lmer* function (Bates et al., 2015). The fixed effects were Stimuli type (word, [ʊ]-type, and [ɪ]-type), Talker (Korean, American, and Mexican), Order (Block 1, Block 2, and Block 3), Platform (coded as −0.5 for in-person, and 0.5 for online), and a four-way interaction of these factors. Stimuli type, Talker, and Order were coded in the same manner as in the previous model. C1 Place of articulation (Labial, Alveolar, or Velar) and C2 (r vs. l) were excluded from the final model as they neither increased the model fit nor showed statistically significant effects. The random effect structure included a random slope for Stimuli type for item and random intercepts for participant and item.

Alpha was set to 0.05. The in-text results focus on the predictors that are directly related to the research questions (Stimuli type, Talker, Order). See Appendix B and Appendix C for the full results of the reported models.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Results

Proportion of word responses

Figure 1 shows the by-subject mean proportion of word responses for each stimulus type and talker condition.

The statistical model for responses (Appendix B) showed that across the three blocks and three talker conditions, [ʊ]-type was responded to as a word less frequently than word type ($\beta = 3.53$, $CI = [2.87, 4.19]$, $p < 0.001$). On the other hand, [ʊ]-type was responded to as a word more frequently than [ɪ]-type ($\beta = -0.94$, $CI = [-1.33, -0.55]$, $p < 0.001$). The effects of Talker did not yield statistically significant main effects or interaction with Stimuli (all $p > 0.062$), which suggests that participants' responses for [ʊ]-type did not differ by Talker identity when the three blocks were collapsed together.

The [ʊ]-type stimuli showed evidence for adaptation, as the main effects of Order 1 (Block 1 vs. Block 2: $\beta = 0.59$, $CI = [0.22, 0.96]$, $p = 0.002$) and Order 2 (Block 1 vs. Block 3: $\beta = 0.58$, $CI = [0.21, 0.95]$, $p = 0.002$) were statistically significant. These effects indicate that the word responses for [ʊ]-type increased in Block 2 and Block 3 compared to Block 1. These order effects also interacted with Stimuli 1 ([ʊ]-type vs. [ɪ]-type) (Stimuli 1 × Order 2: $\beta = 0.78$, $CI = [0.25, 1.30]$, $p = 0.004$), Talker 1 (Korean vs. American) (Talker 1× Order 1: $\beta = -1.05$, $CI = [-1.94, -0.15]$, $p = 0.022$, Talker 1× Order 2: $\beta = -1.01$, $CI = [-1.91, -0.10]$, $p = 0.029$) and Talker 2 (Korean vs. Mexican) (Talker 2× Order 2: $\beta = -1.18$, $CI = [-2.09, -0.26]$, $p = 0.011$). Three-way interactions of Stimuli 2 ([ʊ]-type vs. word), Talker 1 (Korean vs. American), and Order 2 (Block 1 vs. Block 3) ($\beta = 4.00$, $CI = [1.20, 6.80]$, $p = 0.005$) and Stimuli 2 ([ʊ]-type vs. word), Talker 2 (Korean vs. Mexican), and Order 2 (Block 1 vs. Block 3) ($\beta = 2.42$, $CI = [0.41, 4.44]$, $p = 0.018$) were also observed, indicating that the adaptation patterns differ by stimulus type and talker identity. These interactions were further investigated with post-hoc analyses using the *emmeans*() function in the emmeans package (Lenth et al., 2018). The interaction between Stimuli 1 and Order 2 was due to a larger increase in word responses for [ɪ]-type ($\beta = 1.36$, $CI = [0.91, 1.81]$, $p < 0.001$) than [ʊ]-type from Block 1 to Block 3. The latter interactions between Stimuli 2, Talker, and Order were due to lack of increase in word responses for [ʊ]-type in American and Mexican talker conditions from Block 1 to later blocks (all $p > 0.794$) in contrast to Korean talker condition that showed a significant increase in Block 2 ($\beta = 1.13$, $CI = [0.14, 2.11]$, $p = 0.012$) and Block 3 ($\beta = 1.31$, $CI = [0.31, 2.31]$, $p = 0.002$) than Block 1. The effect of Talker identity was not significant in Block 1 (all $p > 0.354$). Finally, word responses for word type did not increase in Block 2 and Block 3 in any of the talker conditions (all $p > 0.291$).

Reaction times

Figure 2 shows reaction times (computed only for word responses) for each stimulus type and talker condition, collapsed over the three blocks (top) and per block (bottom).

In the model with reaction times as the dependent variable, reaction times were slower for [ʊ]-type than word type ($\beta = -0.45$, $CI = [-0.63, -0.28]$, $p < 0.001$)

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

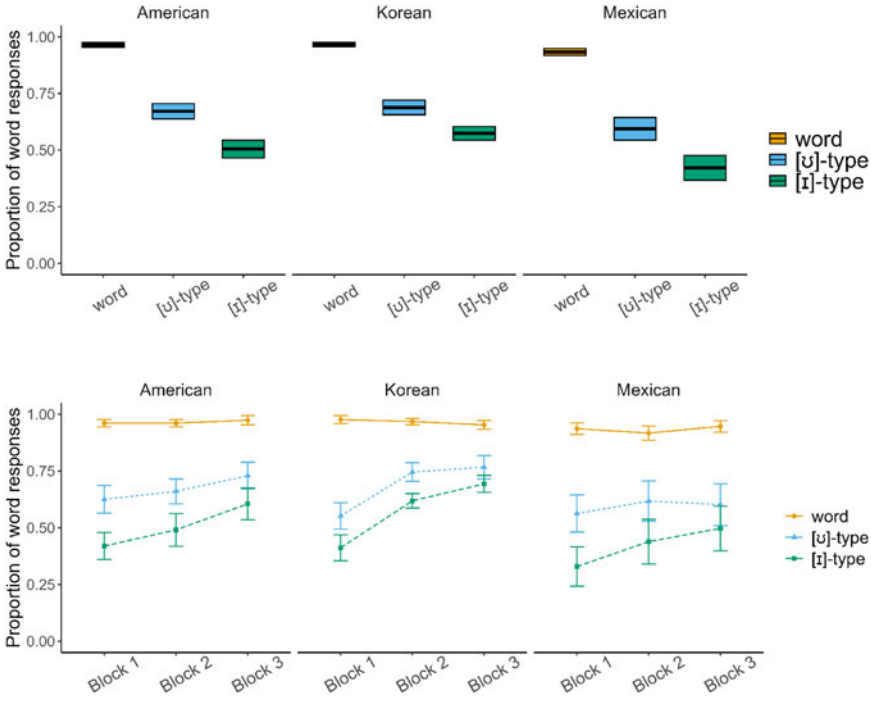


Figure 1. Proportion of word responses across blocks (top) and in each block (bottom) in Experiment 1.

when participants gave "word" responses across the talker conditions and blocks. The difference in the reaction times for [ʊ]-type and [ɪ]-type was not statistically reliable ($\beta = 0.16$, $CI = [-0.01, 0.32]$, $p = 0.068$). The main effect of Platform ($\beta = -0.56$, $CI = [-0.89, -0.24]$, $p < 0.001$) indicates generally slower reaction times for in-person participants.

When comparing blocks, the reaction times for [ʊ]-type decreased over time across three talker conditions, as indicated by the main effects of both Order 1 (Block 1 vs. Block 2) ($\beta = -0.28$, $CI = [-0.45, -0.12]$, $p < 0.001$) and Order 2 (Block 1 vs. Block 3) ($\beta = -0.47$, $CI = [-0.64, -0.30]$, $p < 0.001$). In addition, Order 2 interacted with Stimuli 2 ([ʊ]-type vs. word) ($\beta = 0.33$, $CI = [0.11, 0.55]$, $p = 0.003$), as the decrease in reaction times for [ʊ]-type is statistically larger than word type ($\beta = -0.14$, $CI = [-0.26, -0.02]$, $p = 0.020$) from Block 1 to Block 3. On the other hand, Order 1 interacted with Talker 2 (Korean vs. Mexican) ($\beta = -0.71$, $CI = [-1.13, -0.29]$, $p < 0.001$) and showed a three-way interaction with Talker 2 and Stimuli 1 ([ʊ]-type vs. [ɪ]-type) ($\beta = 0.57$, $CI = [0.03, 1.10]$, $p = 0.039$). Post-hoc analyses indicated that the effect of Order 1 was greater for Mexican ($\beta = -0.41$, $CI = [-0.71, -0.11]$, $p < 0.001$) than for Korean talker ($\beta = -0.36$, $CI = [-0.70, -0.02]$, $p = 0.040$), while the reaction times did not decrease for word type in both talker conditions.

Additionally, Platform showed a two-way interaction with Stimuli 2 ([ʊ]-type vs. word) ($\beta = -0.20$, $CI = [-0.38, -0.03]$, $p = 0.022$), and three-way interactions with Stimuli 2 and Talker 1 (Korean vs. American) ($\beta = 0.49$, $CI = [0.08, 0.90]$,

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

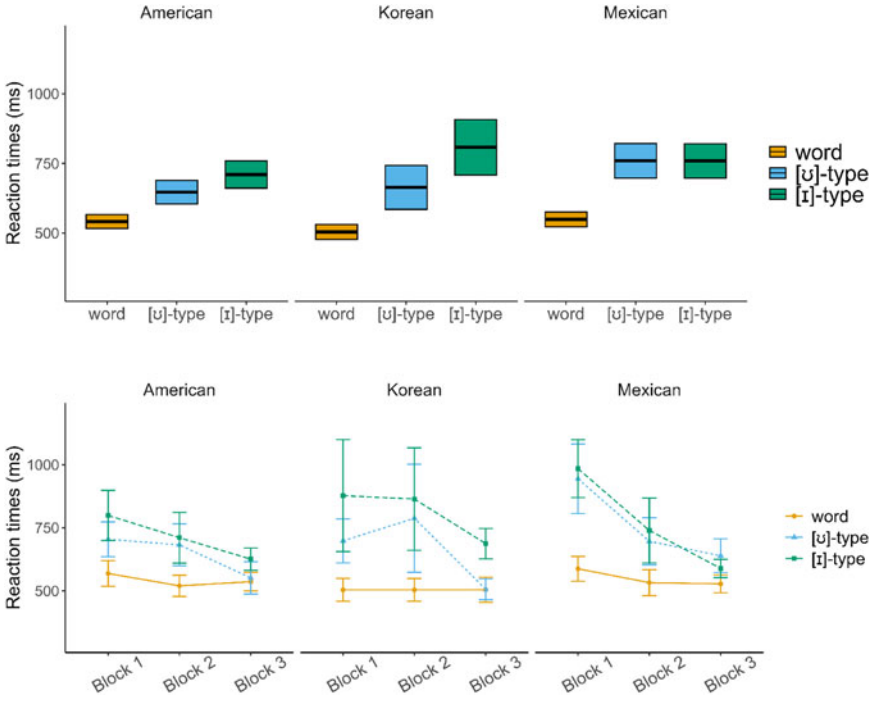


Figure 2. Reaction times for word responses across blocks (top) and in each block (bottom) in Experiment 1.

$p = 0.020$), and Stimuli 2 and Talker 2 (Korean vs. Mexican) ($\beta = 0.54$, $CI = [0.10, 0.97]$, $p = 0.017$). Post-hoc tests revealed that the effect of Stimuli 2 is greater among online participants ($\beta = -0.56$, $CI = [-0.79, -0.33]$, $p < 0.001$) than in-person participants ($\beta = -0.34$, $CI = [-0.53, -0.16]$, $p < 0.001$) in the Korea talker condition. Also, only online participants showed a statistically significant interaction between Stimuli 2 and Talker 1 ($\beta = 0.41$, $CI = [0.06, 0.76]$, $p = 0.020$). In order to assess whether the observed platform effect is consistent across all participants, we examined individual participants' responses. This follow-up inspection indicated that the interactions were largely driven by one participant in the online platform assigned the Korean talker condition, who showed an atypical pattern with extremely slow reaction times for [ʊ]-type (mean RT = 2238.48 ms) and [ɪ]-type (mean RT = 2988.56 ms) than word type (mean RT = 634.18 ms)[5]

Interim discussion

Experiment 1 examined the effect of talker identity on the perception of and adaptation to onset clusters produced with an inserted vowel [ʊ] or [ɪ] by English listeners. The listeners had limited exposure to Korean-accented English, which inserts /ɨ/, similar to [ʊ], within onset clusters. Overall responses were similar regardless of whether the talker was described as American, Korean, or Mexican; participants preferred word type to [ʊ]-type, followed by [ɪ]-type. Importantly,

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

when these measures were compared across the three blocks of the experiment, the proportion of word responses increased only for the Korean talker, whereas reaction times became faster over time, with the reduction being greater for the Mexican talker than the Korean talker.

While the general preference for word type over [ʊ]-type is as expected, the higher proportion of word responses for [ʊ]-type than [ɪ]-type contrasts with the findings in Darcy and Thomas (2019), in which the proportion of word responses for the two types was comparable among English listeners. We suspect this discrepancy may have stemmed from the stimuli. The stimuli in the current study differ from those in Darcy and Thomas (2019) in several notable ways. First, the vowels ([ʊ] and [ɪ]) in our stimuli are shorter and more central (Table 1) than those in Darcy and Thomas (2019, p.7, Table 2). While the vowels in our stimuli can be seen as longer than typical schwas in English (Crystal and House, 1988), they might have been less pronounced than those in Darcy and Thomas (2019). In addition, the makeup of C1 obstruents in our stimuli differs from that in Darcy and Thomas (2019). Out of 24 words, we had 10 voiced and 14 voiceless C1s (see Appendix A), but Darcy and Thomas (2019) had 6 voiced C1s out of 30 words. These differences could have led the inserted vowels in our stimuli more likely to be perceived as part of consonant clusters rather than separate syllabic nuclei. This is particularly plausible when C1 is voiced, and the following consonant is a liquid, as liquids have well-defined formant structures, including higher formants such as F2 and F3, which resemble those of vowels. As a result, the formant structures between a voiced C1 and a liquid C2 may be attributed to the consonant cluster itself, rather than interpreted as a vowel between the consonants. This could partly explain why the proportion of vowel-inserted pseudowords accepted as words was much higher in our study than in Darcy and Thomas (2019). In the American talker condition, which is comparable to Darcy and Thomas's (2019) English listeners, the mean proportion of word response was 67% for [ʊ]-type and 50% for [ɪ]-type (see also Figure 1), much higher than those in Darcy and Thomas ([ʊ]-type: 14%, [ɪ]-type: 11.5%). We suspect that more conspicuous vowels in Darcy and Thomas (2019) could have arguably induced some sort of ceiling effects, increasing the likelihood of both [ʊ] and [ɪ] being correctly rejected, diminishing the difference between the two types of vowel insertion.

In contrast to the overall responses, the adaptation patterns for [ʊ]-type showed a significant talker identity effect. While reaction times to provide word responses to this type became faster over time across all talker conditions, the proportion of word responses increased only in the Korean talker condition. It warrants caution in interpreting these results, considering the accuracy-speed trade-off. That said, we tentatively suggest that the increase in the proportion of word responses reflects changes in more voluntary cognitive processes than reaction times, as the current experiment was designed to draw participants' attention to their word/nonword responses but not their reaction times. Thus, the results in Experiment 1 seem to indicate that the effect of talker information is more evident in the word response measure, which reflects more conscious decision-making mechanisms than reaction times, and that listeners not only differentiate native versus non-native talkers but also non-native talkers from different language backgrounds (Korean vs. Mexican).

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

That is, although the participants in Experiment 1 were not highly familiar with Korean-accented English, they were relatively more familiar with Spanish-accented English, which does not involve vowel insertion within onset clusters. This might have led to a mismatch between listeners' expectations and the incoming speech signals in the Mexican as well as American talker condition, preventing listeners from providing more word responses to [ʊ]-type even after repeated exposure to this type compared to the Korean talker condition. In contrast, in the Korean talker condition, listeners might have been less certain about whether a true mismatch existed, which made them more generous in their judgments and thus allowed for greater adaptation over time. This also explains the adaptation patterns shown for [ɪ] types.

On the other hand, the Mexican talker condition showed a larger decrease in reaction times to [ʊ]-type compared to the Korean talker condition in the second block. The results in the first block show that the proportion of word responses to [ʊ]-type is similar for Mexican talker and Korean talker conditions, while reaction times for word responses are much slower for the Mexican talker condition than the Korean talker condition. This suggests that participants in the Mexican talker condition initially needed more time to determine the [ʊ]-type stimuli as words during the first block of the experiment, but the reaction times decreased quickly in the second block. We suspect that this may also be due to the mismatch between the speech signal and the expectations stemming from the purported talker identity (McGowan, 2015). The speech signals were originally produced by a native speaker of American English without any hint of discernible foreign accents except the characteristics of Korean-accented English, an inserted vowel within onset clusters. When these signals were provided with a Mexican talker, the episodic memory traces appropriate for the talker's identity would have been activated, which did not match the incoming signals. This mismatch presumably led to longer reaction times for the listeners in the Mexican talker condition, but they quickly adjusted from the second block, as evidenced by the reduced reaction times (but note that this did not lead to increased word responses).

As a whole, the results in Experiment 1 expand the predictive coding account of speech perception, according to which the alignment between the expectations and the actual signals induces not only general benefits in processing degraded or foreign-accented speech but also greater online adaptation to it (McGowan, 2015; Sohoglu and Davis, 2016; Vaughn, 2019). We build on Vaughn (2019) and demonstrate that, even when talker identity is provided, it does not facilitate adaptation as effectively when the provided information about the talker mismatches the speech signals—the current results suggest that the adaptation is suppressed in the American and Mexican talker conditions (mismatch between talker and signals) compared to the Korean talker condition, where listeners do not have concrete expectations about the talker (neither match nor mismatch).

The participants in Experiment 1 were English speakers living in the US with limited exposure to Korean or Korean-accented English. In Experiment 2, we examine how prior exposure to Korean-accented English affects processing [ʊ]-insertion by recruiting English speakers living in Korea.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Experiment 2

Experiment 2 investigates the effect of linguistic experience on speech perception and adaptation by English speakers living in Korea.

Methods

Participants

A total of 43 English speakers living in Korea (13 males, 30 females, 0 other responses; age: mean = 26.77, SD = 5.82) participated in the study either in-person at a university in Seoul, Korea ($n = 27$) or online via Prolific ($n = 16$). They had lived in Korea for an average of 3.47 years (SD = 2.77), and a majority of them ($n = 31$) reported that Korean accounted for at least 10% of their daily language use. Five participants recruited in person were excluded because they reported having learned both Korean and English from birth. The remaining participants learned Korean as an additional language (mean age of acquisition = 18.44, SD = 6.67) and rated their familiarity with Korean-accented English as 5.82 (SD = 1.43) on average on a 1-7 Likert scale. They were assigned to either the Korean ($n = 22$) or the American ($n = 16$) talker condition. Mexican talker's condition was not tested in Experiment 2.

Materials and procedure

The stimuli and the procedure were the same as in Experiment 1.

Data analysis

Trials with excessively short or long reaction times (less than or greater than 2.5 SD from the mean per participant) were removed, with an exclusion rate of 3.1%. Accuracy rates for filler items were above 70% for all participants, so no participant was excluded based on this criterion.

Participants' responses (coded as 1 for word and 0 otherwise) were analyzed with a generalized mixed-effects regression model using the *glmer* function in R with a binomial family. Fixed effects included Stimuli type ([ʊ]-type, and [ɪ]-type, word), Talker (coded as −0.5 for Korean, and 0.5 for American), Order (Block 1, Block 2, and Block 3), and all possible interactions as fixed effects. Dummy coding was used for Stimuli type with [ʊ]-type as reference, and simple effect regression coding was used for Order with Block 1 as reference. The effects of C1 place of articulation (Labial, Alveolar, or Velar), C2 (r vs. l), and Platform (in-person vs. online) were excluded from the final model as they did not increase the model fit and did not yield statistically significant effects. The random effects included random intercepts for participant and item.

Reaction times with "word" responses measured from the offset of each stimulus were transformed according to the Box-Cox test ($(y^{lambda}-1)/lambda$) and were analyzed with a linear mixed-effects regression. This model had Stimuli type (word, [ʊ]-type, and [ɪ]-type), Talker (Korean, and American), Order (Block 1, Block 2, and Block 3), and the three-way interaction of these factors as fixed effects. Those factors

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

were coded in the same way as in the previous model. C1 place of articulation, C2, and Platform were removed from the model because they did not improve its fit or show any statistically significant effects. The random effects included a random slope for Order for participant, a random slope for Stimuli type for item, and random intercepts for both participant and item. Alpha was set to 0.05. See Appendix D and Appendix E for the full results of these models.

Results

Proportion of word responses

Figure 3 shows the proportion of word responses for each stimulus type and talker condition.

As in Experiment 1, participants responded as word less frequently for [ʊ]-type than for word type ($\beta = 4.90$, $CI = [3.83, 5.96]$, $p < 0.001$) but more frequently than [ɪ]-type ($\beta = -1.20$, $CI = [-1.63, -0.76]$, $p < 0.001$) across the two talker conditions and three block conditions.

The effect of Order 2 (Block 1 vs. Block 3) was statistically significant ($\beta = 0.58$, $CI = [0.10, 1.05]$, $p = 0.017$), and it interacted with Talker ($\beta = -1.08$, $CI = [-2.00, -0.16]$, $p = 0.021$). Post-hoc analyses showed that while the talker effect is not significant in Block 1 ($\beta = -0.31$, $CI = [-1.80, 1.17]$, $p = 0.681$), the increase in word responses for [ʊ]-type in Block 3 was statistically significant only for Korean talker ($\beta = 1.12$, $CI = [0.21, 2.02]$, $p = 0.006$) and not American talker ($\beta = -0.04$, $CI = [-0.97, 1.05]$, $p = 1.000$).

Reaction times

The reaction times (computed only for word responses) for each stimulus type and talker condition are presented in Figure 4.

The model with reaction times as the dependent variable revealed similar patterns as the proportion of word responses regarding Stimuli type: reaction times for [ʊ]-type were slower than word type ($\beta = -0.81$, $CI = [-1.22, -0.41]$, $p < 0.001$) and faster than [ɪ]-type ($\beta = 0.90$, $CI = [0.47, 1.33]$, $p < 0.001$) when decided as "word" across blocks and talker identities.

Also, both the effects of Order 1 (Block 1 vs. Block 2) ($\beta = -0.94$, $CI = [-1.47, -0.42]$, $p < 0.001$) and Order 2 (Block 1 vs. Block 3) ($\beta = -1.14$, $CI = [-1.67, -0.61]$, $p < 0.001$) were statistically significant, indicating that reaction times for [ʊ]-type decreased in Block 2 and Block 3 compared to Block 1 across talker conditions. The interaction between Stimuli 2 ([ʊ]-type vs. word type) and Order 2 (Block 1 vs. Block 3) ($\beta = 0.59$, $CI = [0.02, 1.17]$, $p = 0.044$) is also significant, with no significant decrease in reaction times for word type in Block 3 ($\beta = 0.40$, $CI = [-0.82, 1.63]$, $p = 0.520$) as opposed to [ʊ]-type.

Comparison between Experiment 1 and Experiment 2

In order to compare the responses in Experiment 1 and those in Experiment 2, additional analyses were conducted comparing results from the two experiments for Korean and American talker conditions.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

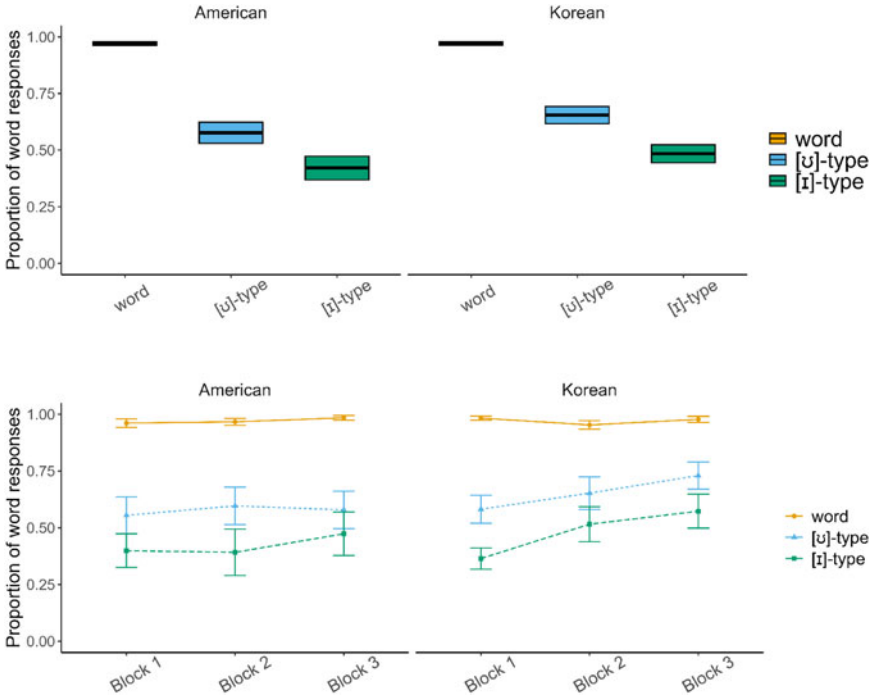


Figure 3. Proportion of word responses across blocks (top) and in each block (bottom) in Experiment 2.

A generalized mixed-effects regression model with a binomial family analyzed participants' responses to each trial (1 if responded as word, 0 if not). Reaction times for word responses were transformed based on the Box-Cox test ((y^{lambda}-1)/lambda) and analyzed with a linear mixed-effects regression model. In both models, fixed effects were Stimuli type (word, [ʊ]-type, and [ɪ]-type), Talker (coded as −0.5 for Korean and 0.5 for American), Order (Block 1, Block 2, and Block 3), testing Location (coded as −0.5 for United States in Experiment 1 and 0.5 for South Korea in Experiment 2) and their interactions. Stimuli type was dummy coded with [ʊ]-type as reference, and Order was coded using the simple effect regression coding scheme with Block 1 as reference. The random effects of both response and reaction time models included a random slope for Stimuli type for the item and random intercepts for both participant and item. In this section, we focus only on reporting effects related to the Location factor.

In the response model, a statistically significant interaction was observed between Stimuli 2 ([ʊ]-type vs. word type) and Location ($\beta = 0.83$, $CI = [0.05, 1.61]$, $p = 0.038$), due to a greater effect of Stimuli 2 for the South Korean participants ($\beta = 4.58$, $CI = [3.40, 5.76]$, $p < 0.001$) than the US participants ($\beta = 3.75$, $CI = [2.50, 5.00]$, $p < 0.001$). Other effects concerning Location did not reach statistical significance (all $p > 0.095$). In summary, the preference for word type over [ʊ]-type was greater for participants tested in South Korea and those in the US.

The reaction time model showed a statistically significant interaction of Order 1 (Block 1 vs. Block 2) and Location ($\beta = -0.43$, $CI = [-0.82, -0.04]$, $p = 0.030$) as

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

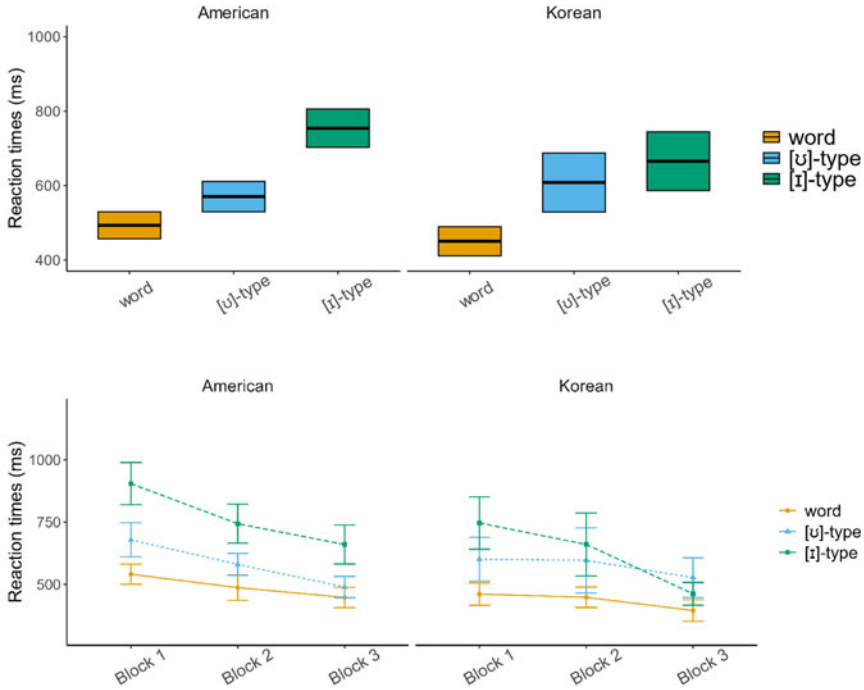


Figure 4. Reaction times for word responses across blocks (top) and in each block (bottom) in Experiment 2.

reaction times for [ʊ]-type decreased in Block 2 among the South Korean participants ($\beta = -0.54$, $CI = [-0.94, -0.14]$, $p = 0.002$) but not the US participants ($\beta = -0.11$, $CI = [-0.51, 0.29]$, $p = 0.970$). The interaction between Stimuli 1 ([ʊ]-type vs. [ɪ]-type) and Location was also statistically significant ($\beta = 0.34$, $CI = [0.10, 0.58]$, $p = 0.005$); the difference in reaction times for the two types was greater in the South Korean participants ($\beta = 0.60$, $CI = [0.22, 0.98]$, $p < 0.001$) than in the US participants ($\beta = 0.26$, $CI = [-0.11, 0.64]$, $p = 0.365$). Finally, the three-way interaction between Stimuli 1, Location, and Order 2 (Block 1 vs. Block 3) was significant ($\beta = -0.64$, $CI = [-1.23, -0.06]$, $p = 0.032$). Post-hoc analyses showed a statistically reliable interaction between Stimuli 1 and Location in Block 1 ($\beta = 0.58$, $CI = [0.18, 0.98]$, $p = 0.005$) but not in Block 3 ($\beta = -0.15$, $CI = [-0.54, 0.23]$, $p = 0.442$) due to a great amount of decrease in reaction times for [ɪ]-type in Block 3 by the South Korean participants ($\beta = -1.07$, $CI = [-1.54, -0.61]$, $p < 0.001$) compared to [ʊ]-type ($\beta = -0.67$, $CI = [-1.07, -0.28]$, $p < 0.001$). For the US participants, the decrease in reaction times in Block 3 was not significant for [ɪ]-type ($\beta = -0.42$, $CI = [-0.88, 0.04]$, $p = 0.10$) and comparable to South Korean participants for [ʊ]-type ($\beta = -0.66$, $CI = [-1.06, -0.26]$, $p < 0.001$). In other words, the South Korean participants showed faster reaction times for [ʊ]-type than [ɪ]-type at the beginning of the experiment compared to the US participants, but such group difference disappeared in the last block because they greatly adapted to [ɪ]-type over time.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

General discussion

The current study examined the effect of talker identity and prior linguistic exposure on speech perception, focusing on a feature commonly observed in Korean-accented English, namely the vowel insertion within onset clusters. In Experiment 1, where English speakers had minimal exposure to Korean-accented English, talker identity did not affect the overall responses but influenced the adaptation patterns, such that word responses for [ʊ]-type increased only for Korean talker. Experiment 2, which examined English speakers living in Korea, showed similar results—word responses for [ʊ]-type increased for Korean talker but not for American talker. On the other hand, when the results of Experiment 1 and Experiment 2 were compared, participants in Experiment 2 showed a greater advantage in reaction times for the [ʊ]-type stimuli over the [ɪ]-type control stimuli than those in Experiment 1. In addition, participants in Experiment 2 also showed earlier adaptation to [ʊ]-type in terms of reaction times than those in Experiment 1. In other words, English speakers with greater experience with Korean-accented English more readily recognized the words showing a feature of the accent than those with limited experience and adapted to the feature earlier during the experiment.

Adaptation patterns differ by talker identity

While no reliable effect of talker identity is found for overall responses or at the beginning of the experiment, the amount of adaptation differed by talker identity over the course of the experiment in both Experiment 1 and Experiment 2. The results from Experiment 1 indicate that listeners distinguish not just the native versus non-native status of the talker, but also among non-native talkers with different L1 backgrounds (i.e., Korean vs. Mexican). In other words, it is not the case that listeners adapt to any non-native talker in the same manner as opposed to the native talker, as predicted by the reverse linguistic stereotyping, according to which the non-native status of the talker affects how listeners perceive the speech (e.g., Kang and Rubin, 2009). Listeners instead make particular expectations about the talker based on their knowledge of the talker's background, which then affects the degree of the adaptation. A similar pattern was found in Experiment 2, in which participants were residing in South Korea and reported a higher familiarity with Korean-accented English than those in Experiment 1: word responses increased over time only in the Korean talker condition and not the American talker condition.

It is striking that there is no evidence for differential effects of talker identity on adaptation patterns between the two experiments. Although the South Korean participants are expected to have more robust representations of [ʊ]-inserted forms connected to the Korean talker than the US participants, the two groups of participants showed similar results, such that word responses to these forms increased only in the Korean talker condition. Despite limited exposure to Korean-accented English, the US participants at least were likely aware that neither the American nor the Mexican talker would produce [ʊ]-inserted forms, which may explain the absence of increased word responses to these forms produced by either

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

talker. Conversely, they were more flexible with the Korean talker, likely due to an expectation of greater pronunciation variations from a foreign speaker that they were not familiar with. This interpretation is also consistent with the outcome that the US participants in the Korean talker condition adapted to (showed an increase in word responses over the course of the experiment), not just the [ʊ]-inserted forms but also the [ɪ] inserted forms. It appears that the South Korean participants adopted a similar strategy, showing more flexibility to the Korean talker. Unlike the US participants, however, this flexibility may stem from their prior knowledge and expectations about Korean-accented English[6]. On the other hand, we observe the Location difference in reaction times, where the South Korean participants exhibited reduced reaction times to vowel insertion at an earlier time (i.e., in the second block) than the US participants, regardless of the talker identity. We return to this point in the next section.

In contrast to a long line of studies that report the effect of talker identity on speech perception, only a few studies (Melguy and Johnson, 2021; Vaughn, 2019) have examined the effect of talker identity on the *online* speech adaptation in a transcription task of foreign-accented speech. As far as we are concerned, the current study is the first to demonstrate that providing a matching versus a mismatching talker identity to the accented speech also affects adaptation in an auditory lexical decision task. These findings contribute to a more nuanced understanding of the relation between social information and speech adaptation.

Prior experience facilitates processing and adaptation to foreign-accented speech

When comparing the results in Experiment 1 (US) and Experiment 2 (South Korea), the reaction times in deciding [ʊ]-type words were faster than [ɪ]-type words, and this difference was greater for the South Korean participants than the US participants in the first block. This pattern is partially in line with what was found in Darcy and Thomas (2019), in which Korean-English bilinguals preferred [ʊ]-type to [ɪ]-type more than did English monolinguals in both proportion of word responses and reaction times. The current results further suggest that, in terms of reaction times, the perception of foreign-accented speech differs even among English speakers based on their familiarity with the accent.

In addition, the previous experience facilitated a decrease in reaction times in recognizing both [ʊ]-type and [ɪ]-type. The South Korean participants showed an earlier decrease in reaction times for [ʊ]-type: unlike the US participants, they responded to this type as words faster in the second block than in the first block. Also, they adapted to [ɪ]-type in the third block to a great extent, diminishing the initial preference they showed for [ʊ]-type. These results demonstrate an advantage of previous linguistic experience in speech adaptation, to both familiar and unfamiliar sound forms.[7] However, such an advantage did not manifest in the proportion of word responses.

Regarding participants' word responses, South Korean participants showed a greater preference for word type over [ʊ]-type. Unlike the difference between [ʊ]-type and [ɪ]-type, this contrast is not directly comparable to the results in Darcy and Thomas (2019) because they did not include the word type in their experiment.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

In order to identify when this pattern emerges during the experiment, we conducted post-hoc analyses that examined the interaction between Stimuli 2 and Location per block. The interaction is significant in block 3 ($\beta = 1.67$, $CI = [0.44, 2.90]$, $p = 0.008$) and not in block 1 and block 2, which means that the group difference is only evident in the final third of the experiment. This suggests that the Location difference is likely to be a result of the adaptation during the experiment, rather than a pattern that existed from the beginning.

Another important observation is that the Location difference for [ʊ]-type versus [ɪ]-type did not interact with talker identity. In other words, it was not the case that the difference between the US participants and the South Korean participants was greater for the Korean talker than the American talker. This means that although the South Korean participants had previous exposure to Korean-accented English that is likely associated with Korean talkers, the activation of the exemplars was not constrained to a particular talker identity during the experiment. This outcome corroborates the extensive literature that tested talker-independent generalization of speech adaptation (Bradlow and Bent, 2008; Xie et al., 2021; Xie and Myers, 2017). These studies show that adaptation to foreign-accented speech generalizes to different talkers speaking the same accent, especially when listeners are trained with multiple talkers, as their "long-term linguistic representations are updated in response to novel input" (Bradlow and Bent, 2008, p.14). While our study was not designed to explicitly test such generalizability, we speculate that the South Korean participants in our study would presumably have had exposure to Korean-accented English spoken by a large sample of Korean talkers compared to US participants, which led to more malleable categories among the South Korean participants for English words with consonant cluster onset. The richer set of exemplars could have enabled them to more readily recognize the [ʊ]-inserted forms at the outset of the experiment, regardless of the alleged talker identity. This suggests evidence for a broad generalization of adaptation. As far as we are aware, it has not been explicitly tested whether adaptation to novel speech patterns generalizes to talkers described as belonging to different races or even to native talkers, which we leave as a topic for future studies.

Taken together, the current results are broadly in line with the exemplar models of speech perception (e.g., Goldinger, 1998; Johnson, 1997), which posit that talker-specific details are stored in memory and used while perceiving speech sounds. Across two experiments, we show that listeners' perception of accented speech is influenced both by the previous experience with the accent and by the contextual plausibility of the talker-accent combination. Listeners with previous exposure to Korean-accented English recognize the stimuli modeled on the accent faster than the control stimuli, providing evidence for the perceptual benefit of experience. At the same time, a more plausible talker-accent pairing facilitates adaptation to the accent regardless of prior experience. These findings contribute to refining our understanding of how foreign-accented speech is perceived and processed, highlighting that listeners integrate socially grounded expectations based on talker identity to navigate various speech inputs more efficiently.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Acknowledgments. This research was funded by the Department of Linguistics at the University of Michigan to J.C. and the New Faculty Startup Fund at Seoul National University to H.K. We have no known conflict of interest to disclose. This study was not preregistered. The authors used ChatGPT (version 4o) to check grammar and spelling. All content was reviewed and edited by the authors, who take full responsibility for the final version.

Notes

1. The term 'adaptation' can possibly refer to both online and offline processes, but in the current study, we limit the use of the term to updating perceptual strategies online during the course of the experiment.
2. Although some participants self-reported their familiarity with Korean-accented English as high, we did not exclude them from data analysis since, in our view, their linguistic experiences are fundamentally different from the participants in Experiment 2, who have experience living in Korea. Additional analyses excluding those participants showed similar patterns to those reported in the Results section (see Supplementary materials for the results).
3. In accordance with a reviewer's suggestion, we also tested whether word acceptance differs by the degree to which each word is established as a loanword in Korean. This factor did not yield any statistically reliable results for both Experiment 1 and Experiment 2 (all $p > 0.157$). The data and codes for this analysis are available in the OSF repository.
4. This test finds the optimal lambda value that best transforms the data to approximate a normal distribution.
5. The mean reaction times for each stimuli type by other participants are 719.57 ms ([ʊ]-type), 769.44 ms ([ɪ]-type), and 614.32 ms (word type) among in-person participants, and 556.00 ms ([ʊ]-type), 580.00 ms ([ɪ]-type), and 452.72 ms (word type) among online participants.
6. This interpretation is supported by post-experiment verbal reports from participants who were assigned to the Korean talker condition (e.g., "I was aware of syllable separation as a strategy for spelling in Korean so when I heard common words with some difference it was mostly recognizable", "Reflecting on my own 'Konglish'").
7. As anonymous reviewers point out, the online adaptation to the [ɪ]-type as well as the [ʊ]-type by South Korean participants indicates that their adaptation was not solely based on the complete match between the stimuli and the talker identity, since [ɪ]-insertion is not a characteristic of Korean-accented English. Instead, the talker's Korean identity may have prompted participants to adopt a more lenient stance toward phonological variations in general during the experiment.

References

Adank, P., & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken-word comprehension. In J. Trouvain, & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (ICPhS 2007) (pp. 1925–1928). Dudweiler: Pirrot.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48.

Berent, I., Lennertz T., Jun J., Moreno M.A., & Smolensky P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, **105**, 5321–5325.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13–34). Amsterdam, Netherlands: John Benjamins Publishing.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.13). Retrieved from https://www.praat.org/.

Bouavichith, D. A., Calloway, I., Craft, J. T., Hildebrandt, T., Tobin, S. J., & Beddor, P. S. (2019). Perceptual influences of social and linguistic priming are bidirectional. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1039–1043). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211–252.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, **106**, 707–729.

Caffarra, S., & Martin, C. D. (2019). Not all errors are the same: ERP sensitivity to error typicality in foreign accented speech perception. *Cortex*, **116**, 308–320.

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, **98**, 73–101.

Clapp, W., Vaughn, C., & Sumner, M. (2023). The episodic encoding of talker voice attributes across diverse voices. *Journal of Memory and Language*, **128**, 104376.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, **116**(6), 3647–3658.

Crystal, T. H., & House, A. S. (1988). The duration of American-English vowels: an overview. *Journal of Phonetics*, **16**(3), 263–284.

Cuetos, F., Hallé, P. A., Domínguez, A., & Segui, J. (2011). Perception of prothetic /e/ in #sC utterances: Gating data. In W. S. Lee, & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 540–543). Hong Kong: City University of Hong Kong.

Darcy, I., & Thomas, T. (2019). When blue is a disyllabic word: perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism*, **22**(5), 1141–1159.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Timonium, MD: York Press.

Floccia, C., Goslin, J., & Girard, F. (2009). Regional and foreign accent processing in English. *Journal of Psycholinguistic Research*, **38**, 379–412.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, **105**(2), 251.

Hallé, P. A., Dominguez, A., Cuetos, F., & Segui, J. (2008). Phonological mediation in visual masked priming: Evidence from phonotactic repair. *Journal of Experimental Psychology: Human Perception and Performance*, **34**(1), 177.

Heo, Y., & Park, H. (2012). American learners' identification of Korean vowel categories. *Harvard Studies in Korean Linguistics*, **14**, 77–84.

Jakobson R., Fant C.G.M., & Halle M. (1952). *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates.* Acoustics Laboratory, M.I.T.

Johnson, K. (1991). Differential effects of speaker and vowel variability on fricative perception. *Language and Speech*, **34**(3), 265–279.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker Variability in Speech Processing*, 145–165.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, **28**, 441–456.

Kang, S. Y. (2012). The patterns of vowel insertion in Korean speakers' production of English C+/l/ and C+/r/ clusters. *Phonetics and Speech Sciences*, **4**(4), 3–17.

Kang, Y. (2011). Loanword phonology. In M. Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *Companion to Phonology* (pp. 2258–2281). Boston: Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0095

Kiparsky, K. (1973). How abstract is phonology? In O. Fujimura (Ed.), *Three Dimensions of Linguistic Theory* (pp.5–56). Tokyo: TEC.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Package "Emmeans". R Package Version 4.0-3. http://cran.r-project.org/package = emmeans

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: a free stimulus set of faces and norming data. *Behavior Research Methods*, **47**(4), 1122–1135.

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, **58**(4), 502–521.

Melguy, Y. V., & Johnson, K. (2021). General adaptation to accented English: Speech intelligibility unaffected by perceived source of non-native accent. *The Journal of the Acoustical Society of America*, **149**(4), 2602–2614.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, **20**(2), 139–154. https://doi.org/10.1017/S0272263198002022

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, **34**(2), 202–240.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, **60**(3), 355–376.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**(2), 309.

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods*, **162**, 8–13.

Pierrehumbert, J.B. (2016) Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, **2**, 33–52.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, **33**(4), 511–531.

Ryu, N. Y. (2018). Korean vowel identification by English and Mandarin listeners: Effects of L1-L2 vowel inventory size and acoustic relationship. *Toronto Working Papers in Linguistics*, 40.

Shin, D. J., & Iverson, P. (2011). Individual differences in vowel epenthesis among Korean learners of English. In W. S. Lee, & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1822–1825). Hong Kong: City University of Hong Kong.

Shin, D. J., & Iverson, P. (2014). Phonetic investigation of epenthetic vowels produced by Korean learners of English. *Phonetics and Speech Sciences*, **6**(4), 17–26.

Shoji, S., & Shoji, K. (2013). Vowel Epenthesis and Consonant Deletion in Japanese Loanwords from English. In J. Kingston, C. Moore-Cantwell, J. Pater, & R. Staubs. (Eds.), *Proceedings of the 2013 Annual Meetings on Phonology*. Washington, DC: Linguistic Society of America.

Sidaras, S.K., Alexander J.E., and Nygaard L.C.. (2008) Talker-specific and accent-general perceptual adaptation to Korean-accented English. *The Journal of the Acoustical Society of America* **123**(no. 5_Supplement), 3887.

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, **113**(12), E1747–E1756.

Stampe, D. (1979). *A Dissertation on Natural Phonology*. Garland Publishing.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, **4**, 1015.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, **20**(4), 580–591.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, **8**, 577.

Vaughn, C. R. (2019). Expectations about the source of a speaker's accent affect accent adaptation. *The Journal of the Acoustical Society of America*, **145**(5), 3218–3232.

Xie, X., Liu, L., & Jaeger, T. F. (2021). Cross-talker generalization in the perception of nonnative speech: A large-scale replication. *Journal of Experimental Psychology: General*, **150**(11), e22–e56.

Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, **97**, 30–46.

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

Yazawa, K., Konishi, T., Hanzawa, K., Short, G., & Kondo, M. (2015). Vowel epenthesis in Japanese speakers' L2 English. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix A. A full list of experimental stimuli and acoustic details of the inserted vowels

Word	[ʊ]-type	Duration (ms)	f0	F1	F2	[ɪ]-type	Duration (ms)	f0	F1	F2
blood	bulood	108.75	121	427	894	bilood	93.56	130	384	1349
blue	bulue	116.4	119	342	805	bilue	74.16	126	378	1366
bring	buring	86.77	109	393	980	biring	81.89	113	386	1417
broom	buroom	98.81	119	399	950	biroom	78.44	103	357	1529
club	culub	133.23	125	392	810	cilub	70.14	115	406	1142
clue	culue	93.21	152	400	1174	cilue	95.32	141	384	1251
cream	cuream	127.09	138	399	1074	ciream	85.15	139	375	1641
crew	curew	77.17	125	353	996	cirew	69.36	132	379	1409
dream	duream	77.83	121	367	848	diream	68.53	116	348	1650
drone	durone	96.67	111	405	1198	dirone	77.78	106	375	1571
glass	gulass	80.98	110	428	847	gilass	79.22	118	465	1441
glue	gulue	102.46	119	355	1061	gilue	81.81	119	386	1651
green	gureen	104.76	119	369	824	gireen	85.55	113	350	1706
group	guroup	95.48	116	392	792	giroup	74.13	115	348	1600
please	pulease	95.23	138	392	938	pilease	108.23	126	379	1016
ploy	puloy	95.87	152	402	771	piloy	79.16	110	405	905
price	purice	117.03	134	405	1037	pirice	78.31	136	414	1388
proof	puroof	123.59	101	371	980	piroof	70.4	106	391	1420
shrimp	shurimp	71.52	118	452	913	shirimp	81.06	120	428	1499
shroom	shuroom	81.11	135	405	1083	shiroom	87.78	122	364	1618
sleep	suleep	79.07	119	385	1078	sileep	91.52	120	407	1468
slow	sulow	113.42	134	405	1017	silow	58.71	129	410	1275
truck	turuck	93.81	159	451	891	tiruck	104.45	138	414	1508
truth	turuth	103.26	116	388	1070	tiruth	76.9	123	368	1678

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix B. Summary of results from the generalized mixed-effects model for word versus nonword responses in Experiment 1 (reference = [ʊ]-type).

Model equation: *response* ~ *Stimuli* * *Talker* * *Order* + *(1|participant)* + *(1+Stimuli|item)*

Predictor	Log-Odds	Standard Error	95% Conf. Interval	*p*
Intercept	**0.98**	**0.31**	**[0.37, 1.58]**	**0.002**
Stimuli 1 ([ʊ]-type vs. [ɪ]-type)	**−0.94**	**0.20**	**[−1.33, −0.55]**	**< .001**
Stimuli 2 ([ʊ]-type vs. word)	**3.53**	**0.34**	**[2.87, 4.19]**	**< .001**
Talker 1 (Korean vs. American)	0.09	0.51	[−0.90, 1.09]	0.852
Talker 2 (Korean vs. Mexican)	−0.38	0.52	[−1.39, 0.62]	0.455
Order 1 (Block 1 vs. Block 2)	**0.59**	**0.19**	**[0.22, 0.96]**	**0.002**
Order 2 (Block 1 vs. Block 3)	**0.58**	**0.19**	**[0.21, 0.95]**	**0.002**
Stimuli 1:Talker 1	−0.32	0.26	[−0.82, 0.19]	0.218
Stimuli 2:Talker 1	1.06	0.57	[−0.05, 2.17]	0.062
Stimuli 1:Talker 2	−0.46	0.27	[−0.99, 0.07]	0.091
Stimuli 2:Talker 2	0.14	0.42	[−0.69, 0.97]	0.741
Stimuli 1:Order 1	0.11	0.27	[−0.41, 0.63]	0.676
Stimuli 2:Order 1	−0.29	0.44	[−1.16, 0.58]	0.516
Stimuli 1:Order 2	**0.78**	**0.27**	**[0.25, 1.30]**	**0.004**
Stimuli 2:Order 2	0.03	0.52	[−0.99, 1.05]	0.954
Talker 1:Order 1	**−1.05**	**0.46**	**[−1.94, −0.15]**	**0.022**
Talker 2:Order 1	−0.56	0.46	[−1.47, 0.34]	0.222
Talker 1:Order 2	**−1.01**	**0.46**	**[−1.91, −0.10]**	**0.029**
Talker 2:Order 2	**−1.18**	**0.47**	**[−2.09, −0.26]**	**0.011**
Stimuli 1:Talker 1:Order 1	0.47	0.63	[−0.76, 1.71]	0.454
Stimuli 2:Talker 1:Order 1	1.50	1.19	[−0.84, 3.85]	0.208
Stimuli 1:Talker 2:Order 1	0.06	0.66	[−1.23, 1.35]	0.928
Stimuli 2:Talker 2:Order 1	0.90	1.05	[−1.15, 2.96]	0.390
Stimuli 1:Talker 1:Order 2	0.52	0.64	[−0.72, 1.77]	0.410
Stimuli 2:Talker 1:Order 2	**4.00**	**1.43**	**[1.20, 6.80]**	**0.005**
Stimuli 1:Talker 2:Order 2	1.14	0.67	[−0.16, 2.45]	0.087
Stimuli 2:Talker 2:Order 2	**2.42**	**1.03**	**[0.41, 4.44]**	**0.018**

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix C. Summary of results from the linear mixed-effects model for reaction times for word responses in Experiment 1 (reference = [ʊ]-type).

Model equation: *(rt^{lambda}-1)/lambda ~ Stimuli * Talker * Order * Platform + (1|participant) + (1+Stimuli|item)*

Predictor	*β*	Standard Error	95% Conf. Interval	*p*
Intercept	**8.96**	**0.12**	**[8.72, 9.19]**	**< .001**
Stimuli 1 ([ʊ]-type vs. [ɪ]-type)	0.16	0.09	[−0.01, 0.32]	0.068
Stimuli 2 ([ʊ]-type vs. word)	**−0.45**	**0.09**	**[−0.63, −0.28]**	**< .001**
Talker 1 (Korean vs. American)	0.09	0.20	[−0.30, 0.48]	0.655
Talker 2 (Korean vs. Mexican)	0.29	0.21	[−0.11, 0.70]	0.160
Order 1 (Block 1 vs. Block 2)	**−0.28**	**0.09**	**[−0.45, −0.12]**	**< .001**
Order 2 (Block 1 vs. Block 3)	**−0.47**	**0.09**	**[−0.64, −0.30]**	**< .001**
Platform	**−0.56**	**0.17**	**[−0.89, −0.24]**	**< .001**
Stimuli 1:Talker 1	−0.07	0.12	[−0.31, 0.17]	0.555
Stimuli 2:Talker 1	0.18	0.11	[−0.03, 0.39]	0.088
Stimuli 1:Talker 2	−0.16	0.13	[−0.42, 0.10]	0.222
Stimuli 2:Talker 2	0.03	0.11	[−0.19, 0.25]	0.805
Stimuli 1:Order 1	−0.04	0.13	[−0.31, 0.22]	0.743
Stimuli 2:Order 1	0.13	0.11	[−0.08, 0.35]	0.222
Stimuli 1:Order 2	0.08	0.13	[−0.18, 0.34]	0.561
Stimuli 2:Order 2	**0.33**	**0.11**	**[0.11, 0.55]**	**0.003**
Talker 1:Order 1	−0.29	0.20	[−0.69, 0.11]	0.153
Talker 2:Order 1	**−0.71**	**0.22**	**[−1.13, −0.29]**	**< .001**
Talker 1:Order 2	−0.22	0.2	[−0.61, 0.18]	0.282
Talker 2:Order 2	−0.04	0.22	[−0.47, 0.39]	0.853
Stimuli 1:Platform	0.09	0.11	[−0.12, 0.29]	0.410
Stimuli 2:Platform	**−0.20**	**0.09**	**[−0.38, −0.03]**	**0.022**
Talker 1:Platform	−0.10	0.40	[−0.87, 0.68]	0.810
Talker 2:Platform	0.17	0.41	[−0.64, 0.98]	0.679
Order 1:Platform	−0.23	0.17	[−0.57, 0.10]	0.172
Order 2:Platform	−0.32	0.17	[−0.66, 0.02]	0.061
Stimuli 1:Talker 1:Order 1	−0.07	0.31	[−0.67, 0.54]	0.833
Stimuli 2:Talker 1:Order 1	0.03	0.26	[−0.48, 0.54]	0.903
Stimuli 1:Talker 2:Order 1	0.25	0.34	[−0.41, 0.92]	0.453
Stimuli 2:Talker 2:Order 1	**0.57**	**0.27**	**[0.03, 1.10]**	**0.039**

(*Continued*)

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix C. (*Continued*)

Predictor	*β*	Standard Error	95% Conf. Interval	*p*
Stimuli 1:Talker 1:Order 2	0.18	0.31	[−0.42, 0.78]	0.558
Stimuli 2:Talker 1:Order 2	0.14	0.26	[−0.37, 0.65]	0.590
Stimuli 1:Talker 2:Order 2	−0.15	0.34	[−0.81, 0.51]	0.651
Stimuli 2:Talker 2:Order 2	0.03	0.28	[−0.52, 0.57]	0.926
Stimuli 1:Talker 1:Platform	0.12	0.24	[−0.36, 0.59]	0.635
Stimuli 2:Talker 1:Platform	**0.49**	**0.21**	**[0.08, 0.90]**	**0.020**
Stimuli 1:Talker 2:Platform	0.08	0.27	[−0.45, 0.60]	0.777
Stimuli 2:Talker 2:Platform	**0.54**	**0.22**	**[0.10, 0.97]**	**0.017**
Stimuli 1:Order 1:Platform	−0.36	0.27	[−0.88, 0.17]	0.181
Stimuli 2:Order 1:Platform	0.06	0.22	[−0.37, 0.49]	0.778
Stimuli 1:Order 2:Platform	−0.27	0.26	[−0.78, 0.25]	0.316
Stimuli 2:Order 2:Platform	0.10	0.22	[−0.33, 0.53]	0.641
Talker 1:Order 1:Platform	−0.24	0.40	[−1.04, 0.55]	0.547
Talker 2:Order 1:Platform	0.11	0.43	[−0.74, 0.95]	0.800
Talker 1:Order 2:Platform	0.52	0.41	[−0.28, 1.31]	0.202
Talker 2:Order 2:Platform	0.72	0.44	[−0.14, 1.58]	0.102
Stimuli 1:Talker 1:Order 1:Platform	1.06	0.62	[−0.15, 2.28]	0.087
Stimuli 2:Talker 1: Order 1:Platform	−0.07	0.52	[−1.09, 0.94]	0.886
Stimuli 1:Talker 2: Order 1:Platform	0.48	0.68	[−0.85, 1.81]	0.478
Stimuli 2:Talker 2: Order 1:Platform	−0.71	0.55	[−1.78, 0.37]	0.197
Stimuli 1:Talker 1: Order 2:Platform	0.53	0.61	[−0.67, 1.73]	0.384
Stimuli 2:Talker 1: Order 2:Platform	−0.43	0.52	[−1.45, 0.59]	0.411
Stimuli 1:Talker 2: Order 2:Platform	−0.29	0.67	[−1.61, 1.02]	0.662
Stimuli 2:Talker 2: Order 2:Platform	−0.66	0.56	[−1.75, 0.43]	0.236

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix D. Summary of results from the generalized mixed-effects model for word versus nonword response in Experiment 2 (reference = [ʊ]-type).

Model equation: *response* ~ *Stimuli* * *Talker* * *Order* + *(1|participant)* + *(1+Stimuli|item)*

Predictor	Log-Odds	Standard Error	95% Conf. Interval	*p*
Intercept	**0.85**	**0.4**	**[0.06, 1.64]**	**0.034**
Stimuli 1 ([ʊ]-type vs. [ɪ]-type)	**−1.20**	**0.22**	**[−1.63, −0.76]**	**< .001**
Stimuli 2 ([ʊ]-type vs. word)	**4.90**	**0.54**	**[3.83, 5.96]**	**< .001**
Talker (Korean vs. American)	−0.38	0.64	[−1.63, 0.87]	0.551
Order 1 (Block 1 vs. Block 2)	0.40	0.23	[−0.05, 0.86]	0.084
Order 2 (Block 1 vs. Block 3)	**0.58**	**0.24**	**[0.10, 1.05]**	**0.017**
Stimuli 1:Talker	−0.01	0.27	[−0.54, 0.52]	0.964
Stimuli 2:Talker	0.54	0.55	[−0.53, 1.61]	0.321
Stimuli 1:Order 1	0.03	0.33	[−0.62, 0.67]	0.929
Stimuli 2:Order 1	−0.58	0.62	[−1.78, 0.63]	0.352
Stimuli 1:Order 2	0.38	0.34	[−0.28, 1.04]	0.259
Stimuli 2:Order 2	0.05	0.74	[−1.40, 1.49]	0.949
Talker:Order 1	−0.21	0.46	[−1.12, 0.69]	0.643
Talker:Order 2	**−1.08**	**0.47**	**[−2.00, −0.16]**	**0.021**
Stimuli 1:Talker:Order 1	1.58	1.19	[−0.75, 3.90]	0.184
Stimuli 2:Talker:Order 1	0.42	0.66	[−0.86, 1.71]	0.519
Stimuli 1:Talker:Order 2	2.36	1.34	[−0.26, 4.98]	0.078
Stimuli 2:Talker:Order 2	0.85	0.40	[0.06, 1.64]	0.034

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix E. Summary of results from the linear mixed-effects model for reaction times for word responses in Experiment 2 (reference = [ʊ]-type).

Model equation: *(rt^{lambda}-1)/lambda ~ Stimuli * Talker * Order + (1+Order|participant) + (1+Stimuli|item)*

Predictor	*β*	Standard Error	95% Conf. Interval	*p*
Intercept	**13.36**	**0.31**	**[12.74, 13.97]**	**< .001**
Stimuli 1 ([ʊ]-type vs. [ɪ]-type)	**0.90**	**0.22**	**[0.47, 1.33]**	**< .001**
Stimuli 2 ([ʊ]-type vs. word)	**−0.81**	**0.21**	**[−1.22, −0.41]**	**< .001**
Talker (Korean vs. American)	0.41	0.52	[−0.61, 1.43]	0.428
Order 1 (Block 1 vs. Block 2)	**−0.94**	**0.27**	**[−1.47, −0.42]**	**< .001**
Order 2 (Block 1 vs. Block 3)	**−1.14**	**0.27**	**[−1.67, −0.61]**	**< .001**
Stimuli 1:Talker	0.38	0.28	[−0.17, 0.93]	0.179
Stimuli 2:Talker	0.26	0.24	[−0.20, 0.72]	0.273
Stimuli 1:Order 1	−0.17	0.36	[−0.87, 0.53]	0.633
Stimuli 2:Order 1	**0.59**	**0.29**	**[0.02, 1.17]**	**0.044**
Stimuli 1:Order 2	−0.58	0.35	[−1.27, 0.10]	0.095
Stimuli 2:Order 2	0.53	0.29	[−0.05, 1.10]	0.073
Talker:Order 1	0.09	0.53	[−0.95, 1.12]	0.867
Talker:Order 2	−0.47	0.54	[−1.53, 0.58]	0.380
Stimuli 1:Talker:Order 1	−0.13	0.71	[−1.52, 1.25]	0.851
Stimuli 2:Talker:Order 1	−0.57	0.58	[−1.70, 0.56]	0.320
Stimuli 1:Talker:Order 2	0.36	0.70	[−1.01, 1.72]	0.609
Stimuli 2:Talker:Order 2	0.23	0.58	[−0.90, 1.37]	0.685

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix F. Summary of results from the generalized mixed-effects model for word versus nonword responses comparing Experiment 1 and Experiment 2 (reference = [ʊ]-type).

Model equation: *response* ~ *Stimuli * Talker * Location * Order + (1|participant) + (1+Stimuli|item)*

Predictor	Log-Odds	Standard Error	95% Conf. Interval	*p*
Intercept	**0.99**	**0.31**	**[0.38, 1.59]**	**0.001**
Stimuli 1 ([ʊ]-type vs. [ɪ]-type)	**−1.02**	**0.20**	**[−1.41, −0.63]**	**< .001**
Stimuli 2 ([ʊ]-type vs. word)	**4.16**	**0.38**	**[3.42, 4.91]**	**< .001**
Talker (Korean vs. American)	−0.15	0.38	[−0.90, 0.61]	0.705
Location	−0.31	0.38	[−1.06, 0.44]	0.421
Order 1 (Block 1 vs. Block 2)	**0.49**	**0.16**	**[0.17, 0.80]**	**0.003**
Order 2 (Block 1 vs. Block 3)	**0.64**	**0.17**	**[0.32, 0.97]**	**< .001**
Stimuli 1:Talker	−0.16	0.19	[−0.53, 0.21]	0.390
Stimuli 2:Talker	**0.80**	**0.40**	**[0.02, 1.58]**	**0.045**
Stimuli 1:Location	−0.31	0.19	[−0.68, 0.05]	0.095
Stimuli 2:Location	**0.83**	**0.40**	**[0.05, 1.61]**	**0.038**
Talker:Location	−0.5	0.77	[−2.00, 1.01]	0.519
Stimuli 1:Order 1	0.13	0.23	[−0.32, 0.57]	0.577
Stimuli 2:Order 1	−0.46	0.42	[−1.29, 0.36]	0.272
Stimuli 1:Order 2	**0.48**	**0.23**	**[0.03, 0.94]**	**0.037**
Stimuli 2:Order 2	−0.10	0.50	[−1.09, 0.88]	0.836
Talker:Order 1	−0.69	0.33	[−1.33, −0.05]	0.034
Talker:Order 2	**−1.05**	**0.33**	**[−1.70, −0.40]**	**0.001**
Location:Order 1	−0.22	0.32	[−0.85, 0.42]	0.500
Location:Order 2	−0.27	0.33	[−0.91, 0.38]	0.421
Stimuli 1:Talker:Location	0.32	0.37	[−0.41, 1.05]	0.393
Stimuli 2:Talker:Location	−0.53	0.79	[−2.09, 1.02]	0.500
Stimuli 1:Talker:Order 1	−0.09	0.45	[−0.98, 0.80]	0.845
Stimuli 2:Talker:Order 1	1.55	0.86	[−0.13, 3.23]	0.071
Stimuli 1:Talker:Order 2	0.5	0.46	[−0.40, 1.39]	0.275
Stimuli 2:Talker:Order 2	**3.09**	**0.99**	**[1.15, 5.02]**	**0.002**
Stimuli 1:Location:Order 1	−0.13	0.45	[−1.02, 0.76]	0.774
Stimuli 2:Location:Order 1	−0.23	0.85	[−1.91, 1.44]	0.784
Stimuli 1:Location:Order 2	−0.06	0.46	[−0.96, 0.83]	0.889
Stimuli 2:Location:Order 2	−0.06	1.01	[−2.03, 1.92]	0.956

(*Continued*)

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix F. (*Continued*)

Predictor	Log-Odds	Standard Error	95% Conf. Interval	*p*
Talker:Location:Order 1	0.97	0.65	[−0.31, 2.25]	0.138
Talker:Location:Order 2	0.04	0.66	[−1.25, 1.32]	0.954
Stimuli 1:Talker:Location:Order 1	−1.48	0.91	[−3.26, 0.30]	0.103
Stimuli 2:Talker:Location:Order 1	−0.12	1.69	[−3.44, 3.19]	0.942
Stimuli 1:Talker:Location:Order 2	−0.19	0.91	[−1.97, 1.59]	0.836
Stimuli 2:Talker:Location:Order 2	−1.80	1.97	[−5.67, 2.07]	0.362

Appendix G. Summary of results from the linear mixed-effects model for reactions times for word responses comparing Experiment 1 and Experiment 2 (reference = [ʊ]-type).

Model equation: $(rt^{lambda}-1)/lambda \sim Stimuli * Talker * Location * Order + (1|participant) + (1+Stimuli|item)$

Predictor	*β*	Standard Error	95% Conf. Interval	*p*
Intercept	**10.10**	**0.16**	**[9.79, 10.41]**	**< .001**
Stimuli 1 ([ʊ]-type vs. [ɪ]-type)	**0.43**	**0.12**	**[0.20, 0.66]**	**< .001**
Stimuli 2 ([ʊ]-type vs. word)	**−0.52**	**0.12**	**[−0.75, −0.29]**	**< .001**
Talker (Korean vs. American)	0.18	0.23	[−0.26, 0.62]	0.423
Location	−0.44	0.23	[−0.88, 0.00]	0.052
Order 1 (Block 1 vs. Block 2)	**−0.32**	**0.10**	**[−0.52, −0.13]**	**0.001**
Order 2 (Block 1 vs. Block 3)	**−0.67**	**0.10**	**[−0.86, −0.47]**	**< .001**
Stimuli 1:Talker	0.05	0.12	[−0.19, 0.29]	0.688
Stimuli 2:Talker	0.18	0.10	[−0.02, 0.38]	0.086
Stimuli 1:Location	**0.34**	**0.12**	**[0.10, 0.58]**	**0.005**
Stimuli 2:Location	0.05	0.10	[−0.15, 0.25]	0.635
Talker:Location	0.16	0.45	[−0.73, 1.04]	0.731
Stimuli 1:Order 1	−0.14	0.15	[−0.44, 0.16]	0.354
Stimuli 2:Order 1	0.11	0.13	[−0.14, 0.36]	0.384
Stimuli 1:Order 2	−0.08	0.15	[−0.38, 0.22]	0.594
Stimuli 2:Order 2	**0.39**	**0.13**	**[0.14, 0.64]**	**0.002**
Talker:Order 1	−0.15	0.20	[−0.53, 0.24]	0.454
Talker:Order 2	−0.26	0.20	[−0.64, 0.12]	0.184
Location:Order 1	**−0.43**	**0.20**	**[−0.82, −0.04]**	**0.030**
Location:Order 2	−0.01	0.20	[−0.40, 0.37]	0.946

(*Continued*)

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press

Appendix G. (*Continued*)

Predictor	*β*	Standard Error	95% Conf. Interval	*p*
Stimuli 1:Talker:Location	0.36	0.24	[−0.11, 0.84]	0.132
Stimuli 2:Talker:Location	0.00	0.20	[−0.40, 0.40]	0.988
Stimuli 1:Talker:Order 1	−0.16	0.30	[−0.76, 0.44]	0.600
Stimuli 2:Talker:Order 1	−0.17	0.25	[−0.66, 0.32]	0.501
Stimuli 1:Talker:Order 2	0.14	0.30	[−0.44, 0.73]	0.630
Stimuli 2:Talker:Order 2	0.13	0.25	[−0.36, 0.62]	0.602
Stimuli 1:Location:Order 1	0.06	0.31	[−0.54, 0.66]	0.848
Stimuli 2:Location:Order 1	0.38	0.25	[−0.11, 0.87]	0.129
Stimuli 1:Location:Order 2	**−0.64**	**0.30**	**[−1.23, −0.06]**	**0.032**
Stimuli 2:Location:Order 2	−0.22	0.25	[−0.70, 0.27]	0.388
Talker:Location:Order 1	0.37	0.39	[−0.40, 1.14]	0.341
Talker:Location:Order 2	0.12	0.39	[−0.64, 0.89]	0.754
Stimuli 1:Talker:Location:Order 1	0.00	0.61	[−1.19, 1.20]	0.992
Stimuli 2:Talker:Location:Order 1	−0.32	0.50	[−1.30, 0.66]	0.520
Stimuli 1:Talker:Location:Order 2	−0.04	0.60	[−1.21, 1.14]	0.952
Stimuli 2:Talker:Location:Order 2	−0.10	0.50	[−1.07, 0.88]	0.842

Cite this article: Cho, J. & Kwon, H. (2026). The effect of talker identity and language experience in perceiving foreign-accented speech. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716425100416

https://doi.org/10.1017/S0142716425100416 Published online by Cambridge University Press