



Non-word repetition may reveal different errors in naive listeners and second language learners

Jeffrey J. Holliday* · Minkyoung Hong

Department of Korean Language and Literature, Korea University, Seoul, Korea

Abstract

The perceptual assimilation of a nonnative phonological contrast can change with linguistic experience, resulting in naïve listeners and novice second language (L2) learners potentially assimilating the members of a nonnative contrast to different native (L1) categories. While it has been shown that this sort of change can affect the discrimination of the nonnative contrast, it has not been tested whether such a change could have consequences for the production of the contrast. In this study, L1 speakers of Mandarin Chinese who were (1) naïve to Korean, (2) novice L2 learners, or (3) advanced L2 learners participated in a Korean non-word repetition task using word-initial sibilants. The initial CVs of their repetitions were then played to L1 Korean listeners who categorized the initial consonant. The naïve talkers were more likely to repeat an initial /s^ha/ as an affricate, whereas the L2 learners repeated it as a fricative, in line with how these listeners have been shown to assimilate Korean sibilants to Mandarin categories. This result suggests that errors in the production of new words presented auditorily to nonnative listeners may be driven by how they perceptually assimilate the nonnative sounds, emphasizing the need to better understand what drives changes in perceptual assimilation that accompany increased linguistic experience.

Keywords: perceptual assimilation, non-word repetition, sibilant fricatives, Korean, Mandarin Chinese

1. Introduction

The second language (L2) acquisition of a spoken language typically involves both the perception and production of L2 speech sounds. The theoretical and practical starting point of the development of these abilities in adult L2 learners is the non-native naïve listener: someone with ostensibly no exposure to the target language. Many studies have used perceptual assimilation and auditory discrimination tasks to investigate how naïve listeners

perceive non-native speech sounds. Best (1995) claimed that the way in which a naïve listener perceptually assimilates the members of a non-native phonological contrast to native (L1) categories will predict the listener's ability to auditorily discriminate between the non-native sounds. This claim has generally been supported by data from a wide range of segments and language pairings (e.g. Tyler et al., 2014).

Naïve listeners are fundamentally different from L2 learners in that the former have no phonological representations or knowledge of the target language, whereas the latter do. It is true

* holliday@korea.ac.kr, Corresponding author

Received 24 October 2019; Revised 30 January 2020; Accepted 30 January 2020

© Copyright 2020 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

that naïve listeners and L2 listeners from the same L1 background do not always differ in their perception of a non-native contrast (e.g. Wagner & Baker-Smemoe, 2013), and it is implicit in many study designs that the perception of naïve listeners represents the “starting point” of L2 acquisition (e.g. Sturman et al., 2016). While we do not dispute that point *per se*, we would like to suggest that naïve perception can differ from L2 perception in ways that have consequences for how we interpret the role of L1 influence in L2 phonological acquisition. Studies that explicitly compare the perception of these two types of listeners sometimes reveal differences that are ostensibly due to the additional linguistic experience of L2 learners that can be accumulated with even a very brief period of L2 exposure. To the extent that such differences are unpredictable from the naïve perception patterns themselves, the operationalization of naïve perception as the “starting point” of L2 acquisition becomes less useful.

In one such recent study of the perception of the Korean sibilant fricative contrast /s^h/-/s*/ by L1 speakers of Mandarin Chinese (hereafter, Mandarin), Holliday (2016) showed that L2 learners generally assimilated the initial consonant in both Korean /s^ha/ and /s*a/ to Mandarin /s/. Naïve listeners, on the other hand, were more likely than L2 learners to perceptually assimilate Korean /s^ha/ (but not /s*a/) to a Mandarin affricate category, such as /tʃh/ or /tʃh/, a result that was also found in Holliday (2014b). This difference in perceptual assimilation across listener groups in Holliday (2016) was reflected in their discrimination accuracy of the /s^ha/-/s*a/ contrast, with naïve listeners being more accurate than L2 learners in their discrimination of the contrast. Following the claims made in Best (1995) and Best & Tyler (2007), it was speculated that the naïve listeners were more accurate at discriminating between Korean /s^ha/-/s*a/ because they assimilated them to two different native categories, whereas the L2 learners more often assimilated them to a single category.

It has been suggested elsewhere that L1 Mandarin learners of Korean also struggle to produce Korean fricatives in a native-like way. Surveys of pronunciation errors in L1 Mandarin learners of Korean (Jeon, 2005; Li, 2015; Yao, 2007) have reported substitution errors in production, such as [s*] for word-initial /s^h/, although such reports were not supported by appropriate acoustic measurements. On the other hand, using the location and diffuseness of the peak in the frication spectrum, Kallay & Holliday (2012) found that L1 Mandarin speakers who were L2 learners of Korean produced Korean /s^h/ and /s*/ nearly identically. Before the vowel /a/, both sounds were produced acoustically similar to their Mandarin /s/ productions, whereas before the vowel /i/, both were palatalized, like Mandarin /e/. Thus, the L2 learners in that study produced Korean /s^h/ targets as unaspirated fricatives, either [s] or [e].

Taken together, the results of Holliday (2016) and Kallay & Holliday (2012) suggest that advanced L2 learners of Korean may produce Korean /s^ha/ as an unaspirated fricative [sa] because their L2 experience has somehow affected their perceptual assimilation (and hence articulatory target) of Korean /s^ha/ (Best & Tyler, 2007; Flege, 1995). In other words, Holliday (2016) showed that the perceptual assimilation and discrimination of Korean fricatives is substantially different between naïve listeners and L2 learners, and Kallay & Holliday (2012) showed that L2 learners produce Korean fricatives in a way

that is predicted by the perceptual targets laid out in Holliday (2016). An empirical gap that remains, and hence one goal of the current study, is to demonstrate that the production of Korean fricatives differs between L1 Mandarin speakers with and without L2 Korean experience.

Because naïve listeners cannot produce the target L2 using a traditional elicitation task, such as picture naming or wordlist reading, the current study used a non-word repetition paradigm that required no knowledge of the L2. Participants listened to non-word stimuli produced by a native Korean speaker and were asked to repeat the word, imitating its pronunciation as closely as possible. Non-word repetition has been used to assess language development in children (e.g. Edwards et al., 2004; Munson et al., 2005), including bilingual populations (Duncan & Paradis, 2016; Windsor et al., 2010). There has been substantial debate, however, over exactly what non-word repetition measures. In various previous studies, non-word repetition has been used to measure lexical access, speech production, motor planning, phonological processing, and phonological memory (Coady & Evans, 2008). The vast majority of this previous literature (summarized in Coady & Evans, 2008) is focused on children.

Non-word repetition may also be a suitable task for use with adult naïve listeners, and our motivations for using it in the current study are purely practical: non-word repetition may approximate classroom interactions in the earliest stages of L2 instruction. In novice L2 classrooms, learners are typically asked to listen to and repeat new words, and the articulatory targets of their productions would represent their L2 perceptual targets at that point in development (Flege, 1995).

Non-word repetition has been used with adult L2 learners or bilinguals in several previous studies, and we find it to be a suitable task for exploring the articulatory targets of both naïve listener and L2 learner populations. Cebrian (2007) and Zhang (2019) used a task in which bilingual listeners heard an /hVd/ non-word and were asked to repeat the word in an /hVb/ frame, thus forcing the listener to abstract away from the phonetic signal in the stimulus. Chakraborty et al. (2011) used 16 different non-words of different lengths, containing a range of phonemes, in a non-word repetition task to test the articulation of English-Bengali bilinguals. Thus, although we have not seen non-word repetition used with naïve listeners, and although there may be no consensus on the purpose of using non-word repetition in assessing child language development, it is nonetheless a task that approximates the earliest stages of L2 classroom exposure and is suitable for use with naïve listeners and L2 learners alike.

In the current study, the L1 Mandarin participants were a subset of those whose perception of Korean fricatives was reported in Holliday (2016), in which naïve listeners assimilated Korean /s^ha/ to a Mandarin affricate more frequently than L2 learners did. These participants include not only L1 Mandarin naïve talkers and novice L2 learners of Korean, but advanced L2 learners as well, which allows us to explore how the perception and production of Korean fricatives may change with long-term L2 experience. We have also included native Korean talkers, whose productions can be used to help contextualize the perception of non-native productions. For example, Holliday (2014a) showed that native listeners' identification accuracy of fellow native speakers' productions varied by vowel

context, with accuracy falling below 80% for fricatives followed by /u/ or /i/. Therefore, the misidentification of non-native talkers' productions in the current study may not always be due to some misperception on the part of the non-native talker who produced it, but to misperception on the part of the L1 Korean listener who perceived it.

The design of the current study is as follows. First, talkers were recruited from four populations: L1 Mandarin naïve talkers, L1 Mandarin novice L2 learners of Korean, L1 Mandarin advanced L2 learners of Korean, and native Korean talkers. These talkers then participated in a non-word repetition task in which they produced fricative- and affricate-initial disyllabic Korean non-words. The initial CVs of a subset of these productions were then used as stimuli in a perception experiment in which native Korean listeners identified and rated the goodness of the initial consonant.

2. Method

2.1. Talkers

As this study reports on the perception of speech elicited in a repetition task, the recording of the stimuli involved two stages: recording the original stimuli by a single original talker, and then playing these stimuli to four groups of participants. These participants were both "listeners", in that they listened to the target productions of the single original talker, but they were also "talkers", as their repetitions of these stimuli were recorded and then subsequently used in the construction of stimuli in the perception experiment reported in the current study. Thus, for the remainder of this paper, "original talker" will refer to the single original talker who produced the target stimuli; "talkers" will refer to the participants who both listened to and repeated these original stimuli; and "listeners" will refer to the listeners who identified and rated these repetitions.

The original stimuli were recorded by the original talker, a phonetically trained female native speaker of Seoul Korean. The speaker was aware that the stimuli would be played to listeners who would be asked to repeat them, and so every effort was made to produce the intended non-words as accurately and clearly as possible. A total of 54 non-words were recorded.

The talkers were recruited from four populations: naïve talkers (n=17), novice L2 learners (n=15), advanced L2 learners (n=17), and native Korean talkers (n=15). The naïve talkers were native Mandarin speakers with no experience learning Korean as an L2. These listeners were undergraduate students at Indiana University, and were recruited and tested in Bloomington, Indiana, USA. The novice L2 learners were L1 Mandarin speakers who were enrolled in a beginner-level full-time intensive Korean language course at Korea University. These listeners had completed three to five weeks of classes by the time of testing, and had arrived in Korea less than a week before those classes began. The advanced L2 learners were L1 Mandarin speakers who had completed at least one year of a full-time intensive Korean language course in Korea and were enrolled in undergraduate degree programs at Korea University at the time of testing, and had been living in Korea for one to five years. Lastly, the native Korean speakers all spoke a

non-Gyeongsang variety of Korean, and were undergraduate students at Korea University at the time of testing. More demographic information from the subset of these talkers whose productions were used to create the perception stimuli is provided in Table 1.

Table 1. Demographic information of the 48 talkers from whose productions stimuli were extracted for the perception experiment

Talker group	Age in years mean (SD)	Length of residence in Korea mean (SD)
Naïve	19.6 (1.0)	N/A
Novice L2	24.1 (3.0)	24 days (5 days)
Advanced L2	22.5 (1.6)	40.0 months (9.8 months)
L1 Korean	22.9 (2.4)	N/A

L2, second language; L1, native.

These talkers participated as listeners in the series of perception tasks reported in Holliday (2016). The recording of the stimuli used for the current experiment was the first task done in the experiment session, done before any of the other perception tasks, and not reported in Holliday (2016). The 54 non-word stimuli were presented auditorily in random order on a laptop computer running OpenSesame (ver. 2.8.1; Mathôt et al., 2012). In each trial, each stimulus was played twice, with a brief intervening pause, after which the talker was asked to repeat the word exactly as they heard it. The L2 learners and native Korean talkers were told that the stimuli were Korean words produced by a native speaker of Korean, but were words they had never heard before. This was said to ensure that the listeners did not think the words were English, or some other language, and would also not try to recognize them as real Korean words. The naïve talkers were not specifically told that the words were Korean, but just that they were from a language other than Chinese or English. They were not told that the stimuli were Korean in case they had preconceived ideas about how Korean should be pronounced. It was thus hoped that the talkers in each group would try to produce the words as closely as possible to how they were perceived.

2.2. Perception stimuli

12 talkers (8 female, 4 male) from each of the four talker groups described above were selected (total n=48), from whose productions CV stimuli were extracted. The balance between females and males was decided based on the number of talkers available in each group with useable productions (e.g. some talkers had skipped one or more words). The CV stimuli were extracted word-initially from non-words with CV.CV or CV.CVC syllabic structure, in which the second consonant was always a coronal obstruent. The word-initial CV consisted of one of the consonants /ʃ, s*, tɛ, tɛh, tɛ*/ followed by one of the vowels /a, i, u/. There were 2 non-words chosen for each CV combination, resulting in 30 target non-words (5 consonants×3 vowels×2 tokens) from which the CV stimuli were extracted for each talker. Each of these 30 target non-words from each of the 48 talkers had its intensity RMS normalized at 70 dB, the initial CV was extracted, and then the waveform of the CV was zeroed out over the final 35 ms of the vowel to reduce audible clipping.

This procedure resulted in 1,440 unique CV stimuli (30

CVs×48 talkers). These stimuli were divided into 6 lists, with each list containing 240 stimuli (15 CVs×16 talkers). The 16 talkers in each list consisted of 4 talkers from each of the 4 talker groups, and an individual talker's stimuli were always equally divided across 2 different lists. 4 of the lists contained only stimuli from female talkers, and 2 of the lists contained only stimuli from male talkers, ensuring that talker gender would vary across but not within listeners.

2.3. Listeners

24 native Korean listeners (14 females, 10 males) who had no experience with any variety of Chinese participated in the perception experiment. Their mean age was 27.3 years with a range of 20 to 48 years. They were living in Seoul at the time of the experiment but were from different regions of Korea. Although two of the listeners were from the Gyeongsang region, which carries a stereotype of not maintaining the phonological contrast between /sh/ and /s*/, studies of the production (Holliday, 2012; Lee & Jongman, 2016) and perception (Holliday, 2014a) of these fricatives by younger native Gyeongsang speakers revealed no significant differences with respect to native Seoul speakers.

All listeners had studied English in school. In addition, although a few reported studying an additional foreign language such as Japanese, German, or French, they reported their proficiency as low, having only minimal knowledge of the language or not remembering anything at all. Lastly, the listeners had limited to no instruction in phonetics or any other related fields such as teaching Korean as a foreign language. Listeners were paid their participation.

Four listeners were randomly assigned to each of the 6 stimuli lists, ensuring that each individual stimulus received 4 unique judgements, and because each talker's stimuli were always spread across 2 different lists, each talker's stimuli were presented to a total of 8 listeners.

2.4. Procedure

Listeners were tested in a quiet room at Korea University, seated in front of a Samsung NT900X-3B laptop computer running OpenSesame (ver. 3.2.4; Mathôt et al., 2012). Listeners were told that they would be hearing Korean sounds produced by native speakers of Mandarin. All instructions were presented on the screen in Korean, and stimuli were presented over high-quality headphones.

Each trial consisted of two parts: phoneme identification, and goodness rating. For each trial, listeners heard one stimulus twice and were asked to identify the Korean phoneme it sounded like the talker was trying to produce. Five buttons were presented on the screen, with each button labeled in Hangeul with one of the Korean fricative or affricate phonemes, and listeners pressed one of the buttons to make their selection. Then, listeners were instructed to rate the sound they had just heard in terms of how accurately it was pronounced with reference to the Korean sound they selected. The rating was done on an integer scale from 1 to 5, with 1 labeled as "very accurate" and 5 labeled "totally inaccurate". The response was made by clicking the mouse on the number on the scale.

2.5. Analysis

Throughout the discussion of the results, the term target will be used to refer to the consonant in the original non-word. There is no way to know what phonological target the talker had in mind when repeating the non-word, of course, and so in the discussion of accuracy we will always be referring to the original non-word that the talker was asked to repeat.

Identification accuracy was calculated using two different methods: overall segmental accuracy (the response was the same as the original target consonant), and manner accuracy (the response and the original target consonant were both fricatives or both affricates, but may have differed in terms of the phonation type). For example, if the target consonant was /tɕ*/ and the response was /tɕ/ it would be considered incorrect in terms of segmental accuracy (since the target and response are not the same), but correct in terms of manner accuracy (since the target and response are both affricates). Manner accuracy was calculated separately because the misperception and subsequent repetition of fricatives as affricates was one of the particular phenomena that this study aimed to investigate.

However, the primary result of interest is not simply how many stimuli were correctly perceived as the target, but rather what segment was perceived. For this reason, the bulk of the results section will focus on the interpretation of confusion matrices showing, for each talker group and vowel context, how frequently each target was perceived as each of the five consonants. Some error patterns turned out to be quite common across talker groups (e.g. /tɕ/ targets being perceived as aspirated /tɕ^h/), but we are specifically interested in error patterns that differ between naïve talkers and L2 learners.

Lastly, we also investigated the goodness ratings that listeners assigned to each production. There were three primary questions we aimed to answer with these data. First, are there differences across talker groups in terms of how good their productions were rated? Second, were the productions whose target consonant was correctly identified rated as better than the productions whose target was misidentified? For example, we will see that many /tɕ/ targets were perceived as /tɕ^h/, but was there any difference in the goodness ratings between correctly identified /tɕ^h/ targets and /tɕ/ targets that were misperceived as /tɕ^h/? Third, are there differences in mean goodness rating across response categories, regardless of the target consonant? It is possible that some phonological categories have looser perceptual criteria, and that, for example, producing "a good /s*/" is more easily achieved than "a good /s^h/". This idea is speculative, but these results could shed light on the question and suggest directions for future research. For ease of computation, the goodness rating scale of 1 ("very accurate") to 5 ("totally inaccurate") was transformed by reversing and shifting it, resulting in a scale of 0 ("totally inaccurate") to 4 ("very accurate"). All goodness ratings reported in this paper use the transformed scale.

All analyses were carried out in RStudio version 1.1.463 (RStudio Team, 2016) running R version 3.5.2 (R Core Team, 2018), and using the Tidyverse package version 1.2.1 (Wickham, 2017).

3. Results

3.1. Accuracy

Overall identification accuracy rates are presented in Table 2. As expected, L1 Korean talkers' productions were the most accurately identified, at 83.0% segmental accuracy, with naïve and novice L2 talkers' productions being the least accurately identified, at 59.8% and 58.2%, respectively. Advanced L2 talkers' productions fell in the middle, at 68.3%. Given that listeners had five responses from which to choose, these results are far above chance level, 20%. A look at manner accuracy rates shows that talkers' productions were rarely perceived as having incorrect manner (i.e. a fricative-for-affricate or affricate-for-fricative error), however, and that the vast majority of errors were due to incorrect phonation type.

Table 2. Segmental accuracy and manner accuracy rates by talker group

Talker group	Segmental accuracy (%)	Manner accuracy (%)
Naïve	59.8	93.8
Novice L2	58.2	96.7
Advanced L2	68.3	98.5
L1 Korean	83.0	99.6

L2, second language; L1, native.

The specific hypothesis that drove this study was that naïve talkers should be more likely to produce /s^h/ targets as an affricate, especially in the /a/ environment. To assess this hypothesis we further calculated manner accuracy just for the /s^h/ targets, according to vowel context, for each of the four groups. The L1 Korean talkers and both L2 groups had manner accuracy rates above 96% for every vowel context, suggesting that manner errors were rare. For the naïve talkers, manner accuracy rates for /a/, /i/, and /u/ environments were 46.9%, 90.6%, and 87.5%, respectively, indicating that although manner errors were slightly more frequent in naïve talkers, they were far more frequent specifically in the /a/ environment.

Table 3 shows the segmental accuracy rates broken down by target consonant and talker group. The most consistent trend across talker groups is that /tɛ/ and /s^h/ targets are on the whole more likely to be perceived incorrectly than /s*/, and especially /tɛ^h/ or /tɛ*/. Given the overall high rate of manner accuracy, and the fact that listeners could choose from only two fricative categories and three affricate categories, the segmental accuracy of /tɛ/ and /s^h/ are probably not directly comparable: among fricatives, chance level would be 50%, whereas among affricates it would be 33.3%. Thus, although the accuracy rates for these targets are both low, they should not necessarily be viewed as equally low.

Table 3. Segmental accuracy rates by target consonant and talker group, with overall means by target consonant

Target	Segmental accuracy (%)				
	Naïve	Novice L2	Advanced L2	L1 Korean	Overall
/tɛ/	35.4	34.0	55.2	70.5	48.8
/tɛ ^h /	83.3	83.7	90.6	92.0	87.4
/tɛ*/	78.5	76.0	85.8	97.2	84.4
/s ^h /	32.6	32.3	53.1	81.2	49.8
/s*/	69.1	64.9	56.6	64.0	66.1

L2, second language; L1, native.

The particularly low segmental accuracy rates for /tɛ/ and /s^h/ are

further broken down by vowel context in Table 4. An interesting trend appears, in which within talker group, vowel context does not seem to affect the segmental accuracy of /tɛ/ targets, but has a consistent effect on /s^h/ targets: naïve, novice L2, and advanced L2 talkers' /s^ha/ productions were less accurately perceived than their /s^hi/ and /s^hu/ productions, whereas for the L1 Korean talkers the trend is reversed, with their /s^ha/ productions being perceived more accurately than their /s^hi/ and /s^hu/ productions.

Table 4. Segmental accuracy rates for /tɛ/ and /s/ targets by talker group and vowel context

Target	Group	Segmental accuracy (%)		
		/a/	/i/	/u/
/tɛ/	Naïve	34.4	36.5	35.4
	Novice L2	36.5	29.2	36.5
	Advanced L2	57.3	50.0	58.3
	L1 Korean	68.8	69.8	72.9
/s ^h /	Naïve	16.7	43.8	37.5
	Novice L2	13.5	46.9	36.5
	Advanced L2	39.6	58.3	61.5
	L1 Korean	94.8	66.7	82.3

L2, second language; L1, native.

3.2. Error patterns

Tables 5, 6, and 7 show how frequently the repetitions of each target consonant were perceived as each consonant in the /a/, /i/, and /u/ vowel contexts, respectively, along with the transformed mean goodness ratings assigned by the listeners. Several trends can be observed.

First, we believe the high accuracy rates for the L1 Korean talkers' productions in the /a/ context demonstrate that the task is feasible. That is, when a native Korean speaker hears an unknown word that begins with a fricative or affricate followed by /a/, they are generally able to imitate it in a way that another listener would be able to correctly identify the original intended consonant. The accuracy rates for four of the five consonants were above 90%, and for the remaining consonant target, /tɛ/, it seems that listeners sometimes simply interpreted the phonetic aspiration in /tɛ/ as a cue to /tɛ^h/, which may be due to the lack of a reliable f0 cue in an isolated CV. It is also clear that /tɛ/ targets were perceived as /tɛ^h/ far more often than /tɛ^h/ targets were perceived as /tɛ/, indicating that the confusion between /tɛ/ and /tɛ^h/ was not symmetric. In any case, even though accuracy rates are not as high in the high vowel contexts or for the non-native talkers, the overall high accuracy rates for the L1 Korean talkers' productions in the /a/ context nevertheless demonstrate that the task is not inherently too difficult.

Second, manner errors are indeed rare. Repetitions of affricate targets were rarely perceived as fricatives, and repetitions of fricative targets were rarely perceived as affricates, with only two notable exceptions. One was the perception of /tɛa/ targets produced by non-native talkers as /s^h/, of which there was a total of 25 such responses. These responses represent 6.5% of non-native talker /tɛa/ target trials from 10 unique stimuli (3 naïve, 4 novice L2, and 3 advanced L2). Given that there were 24 unique /tɛa/ target stimuli per talker group, these /tɛa/ target stimuli perceived as /s^h/ do not seem to reflect a broad trend, but rather isolated cases. Also important to the current study is the fact that these cases were spread across both naïve and L2 talkers, and are thus not an example of a change in perception or production that accompanies L2 experience.

The other exception was the one that was predicted at the outset of the study: /s^ha/ targets produced by naïve talkers (and only naïve talkers) were frequently perceived as affricates, either /tɕ/ or /tɕ^h/. Combined, these affricate responses were more frequent (53.2%) than fricative responses (46.9%). This result was found to a much lesser degree in the /i/ and /u/ contexts: only 8.3% and 12.5% of /s^hi/ and /s^hu/ targets were perceived as /tɕ^h/. Crucially, however, this result was only ever found in the naïve talkers, and not for the

talkers that actually had experience with Korean.

Third, it can be seen that the two fricatives /s^h/ and /s*/ are quite confusable in all vowel contexts and by all talkers, with the exception of L1 Korean talkers in the /a/ context. Because manner errors are so rare, the vast majority of incorrect fricative trials are due to the target being perceived as the other fricative (see the segmental accuracy rates in Table 3). This result is unsurprising, and is in line with previous studies showing that both L1 listeners and

Table 5. Confusion matrix for each talker group in the /a/ vowel context. The first number in each cell represents the percentage of trials for that target given that response. Cells with responses less than 5% were removed for clarity. The second number, in parentheses, is the mean goodness rating

Naïve		Response					Novice L2		Response				
Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/	Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/		
/tɕ/	34.4 (2.82)	58.3 (2.46)		5.2 (1.20)		/tɕ/	36.5 (3.03)	39.6 (2.63)	9.4 (2.89)	14.6 (2.86)			
/tɕ ^h /	25.0 (2.38)	72.9 (2.73)				/tɕ ^h /	11.5 (2.82)	78.1 (2.87)		9.4 (3.00)			
/tɕ*/	13.5 (2.23)		86.5 (3.01)			/tɕ*/	7.3 (2.29)	5.2 (1.40)	83.3 (2.96)				
/s ^h /	18.8 (1.94)	34.4 (2.70)		16.7 (2.31)	30.2 (3.34)	/s ^h /				13.5 (2.46)	85.4 (3.12)		
/s*/				15.6 (2.27)	84.4 (3.30)	/s*/				20.8 (2.85)	75.0 (3.22)		
Advanced L2		Response					L1 Korean		Response				
Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/	Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/		
/tɕ/	57.3 (3.04)	35.4 (2.71)		6.3 (1.83)		/tɕ/	68.8 (3.17)	31.3 (2.97)					
/tɕ ^h /	13.5 (2.85)	85.4 (3.07)				/tɕ ^h /	9.4 (3.00)	90.6 (3.26)					
/tɕ*/	5.2 (2.20)	6.3 (2.50)	87.5 (3.20)			/tɕ*/			97.9 (3.35)				
/s ^h /				39.6 (3.37)	58.3 (2.98)	/s ^h /				94.8 (3.43)			
/s*/				25.0 (3.00)	75.0 (3.22)	/s*/					95.8 (3.51)		

L2, second language; L1, native.

Table 6. Confusion matrix for each talker group in the /i/ vowel context. The first number in each cell represents the percentage of trials for that target given that response. Cells with responses less than 5% were removed for clarity. The second number, in parentheses, is the mean goodness rating

Naïve		Response					Novice L2		Response				
Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/	Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/		
/tɕ/	36.5 (2.86)	57.3 (2.58)				/tɕ/	29.2 (2.89)	58.3 (2.48)	7.3 (2.57)	5.2 (2.20)			
/tɕ ^h /		93.8 (2.91)				/tɕ ^h /	5.2 (2.00)	89.6 (3.00)					
/tɕ*/	17.7 (2.24)		78.1 (3.11)			/tɕ*/	21.9 (2.48)	7.3 (0.71)	70.8 (3.24)				
/s ^h /		8.3 (2.00)		43.8 (3.12)	46.9 (3.16)	/s ^h /				46.9 (3.02)	49.0 (3.15)		
/s*/				46.9 (3.24)	53.1 (3.18)	/s*/				40.6 (2.67)	58.3 (3.14)		
Advanced L2		Response					L1 Korean		Response				
Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/	Target	/tɕ/	/tɕ ^h /	/tɕ*/	/s ^h /	/s*/		
/tɕ/	50.0 (2.85)	39.6 (2.21)	9.4 (2.89)			/tɕ/	69.8 (3.18)	20.8 (2.60)	8.3 (1.62)				
/tɕ ^h /	6.3 (2.17)	89.6 (2.99)				/tɕ ^h /	8.3 (1.75)	91.7 (3.25)					
/tɕ*/	10.4 (3.00)		84.4 (3.12)			/tɕ*/			96.9 (3.31)				
/s ^h /				58.3 (3.18)	39.6 (2.97)	/s ^h /				66.7 (3.34)	32.3 (2.81)		
/s*/				55.2 (3.08)	40.6 (2.69)	/s*/				40.6 (3.10)	58.3 (3.62)		

L2, second language; L1, native.

Table 7. Confusion matrix for each talker group in the /u/ vowel context. The first number in each cell represents the percentage of trials for that target given that response. Cells with responses less than 5% were removed for clarity. The second number, in parentheses, is the mean goodness rating

Naïve		Response					Novice L2		Response				
Target	/tɕ/	/tɕʰ/	/tɕ*/	/sʰ/	/s*/	Target	/tɕ/	/tɕʰ/	/tɕ*/	/sʰ/	/s*/		
/tɕ/	35.4 (2.44)	59.4 (2.49)				/tɕ/	36.5 (2.29)	52.1 (2.12)	10.4 (2.90)				
/tɕʰ/	15.6 (1.93)	83.3 (2.91)				/tɕʰ/	10.4 (2.00)	83.3 (2.74)	6.3 (2.33)				
/tɕ*/	22.9 (2.36)	5.2 (1.20)	70.8 (2.74)			/tɕ*/	18.8 (2.00)	7.3 (0.86)	74.0 (3.11)				
/sʰ/		12.5 (2.42)		37.5 (2.92)	50.0 (3.10)	/sʰ/				36.5 (3.26)	62.5 (3.22)		
/s*/				30.2 (2.79)	69.8 (3.03)	/s*/				38.5 (2.81)	61.5 (3.05)		
Advanced L2		Response					L1 Korean		Response				
Target	/tɕ/	/tɕʰ/	/tɕ*/	/sʰ/	/s*/	Target	/tɕ/	/tɕʰ/	/tɕ*/	/sʰ/	/s*/		
/tɕ/	58.3 (2.84)	24.0 (2.09)	17.7 (1.82)			/tɕ/	72.9 (2.93)	26.0 (2.00)					
/tɕʰ/		96.9 (2.84)				/tɕʰ/	6.3 (3.50)	93.8 (3.28)					
/tɕ*/	9.4 (1.33)	5.2 (1.20)	85.4 (2.74)			/tɕ*/			96.9 (3.38)				
/sʰ/				61.5 (3.15)	38.5 (2.95)	/sʰ/				82.3 (3.04)	15.6 (2.67)		
/s*/				45.8 (3.02)	54.2 (3.12)	/s*/				32.3 (2.55)	67.7 (3.28)		

L2, second language; L1, native.

(especially) L2 listeners have difficulty accurately identifying Korean /sʰ/ and /s*/ when followed by a high vowel (Holliday, 2014a). Table 6 shows the confusion matrix for the /i/ context, and that even L1 talkers' /sʰi/ and /s*i/ targets were identified at rates not far from chance level, a result indicating that either the talkers who produced the stimuli misperceived the intended targets in the original stimuli in the non-word repetition task, or the listeners in the current study misperceived the intended targets produced by the talkers. Regardless, the results demonstrate again that native listeners struggle to accurately identify Korean fricatives in high vowel contexts.

3.3. Goodness ratings

First, while there were marginal differences in overall mean goodness rating across talker groups, as shown in Table 8, these differences did not reach statistical significance [$F(3,44)=2.41, p=.08$]. This lack of a difference between the native and non-native talkers' productions could be because not only were the stimuli short, isolated CVs, but the listeners did not actually know what the original intended target was – the goodness rating was made with respect to the category chosen by the listener. Thus, a production could have been misidentified with respect to the intended target of the original non-word, but still sound like a good production of whatever it was perceived as. Another possible factor is that the listeners were told that the stimuli were produced by non-native speakers, whereas in reality they included some productions of native speakers as well. This could have biased listeners to not rate productions too highly overall, which would include the stimuli produced by the L1 Korean talkers. It could also have the opposite effect, in that L1 Korean talkers' productions could be considered very good for a non-native speaker. Given that the ratings could range from 0 to 4, with 4 being “excellent”,

mean ratings of 2.8 to 3.2 are relatively high.

Table 8. Mean goodness ratings assigned by listeners to the productions of each talker group

Talker group	Overall mean goodness rating	Mean rating of incorrect responses	Mean rating of correct responses
Naïve	2.81	2.61	2.94
Novice L2	2.85	2.65	2.99
Advanced L2	2.90	2.64	3.03
L1 Korean	3.18	2.62	3.29

L2, second language; L1, native.

With respect to the second question of whether the goodness ratings varied according to whether or not the production was correctly perceived as the target consonant in the original non-word, the two right columns in Table 8 show the mean ratings for incorrect and correct responses by talker group. It can be seen that in all four talker groups, productions whose intended consonant was correctly identified received higher goodness ratings than those that were incorrectly identified. These differences yielded statistically significant post-hoc paired *t*-tests [Naïve: $t(11)=2.97, p=.013$; Novice L2: $t(11)=3.09, p=.010$; Advanced L2: $t(11)=3.32, p=.007$; L1 Korean: $t(11)=6.69, p<.001$]. There exists a possible confound between talker group and target consonant, however, and so an additional post-hoc paired *t*-test was run, which confirmed that within each target consonant correctly identified productions were given higher goodness ratings than productions that were misidentified [$t(4)=3.48, p=.025$], regardless of which talker group the production came from.

Lastly, the mean goodness ratings assigned by response category, regardless of what the original intended target was, are shown in Table 9. Across productions from all talker groups, produc-

Table 9. Mean goodness ratings by response category and talker group

Response	Mean goodness rating by talker group					% of responses correct	Number of responses
	Naïve	Novice L2	Advanced L2	L1 Korean	Overall		
/tɕ/	2.45	2.53	2.80	3.02	2.72	68.1	825
/tɕʰ/	2.69	2.58	2.72	3.09	2.76	62.9	1,601
/tɕ*/	2.91	2.98	2.92	3.30	3.04	91.9	1,058
/sʰ/	2.80	2.90	3.08	3.16	3.01	56.9	1,009
/s*/	3.14	3.16	3.00	3.31	3.15	60.1	1,267

L2, second language; L1, native.

tions perceived as /s*/ are generally given the highest ratings, even though productions perceived as /s*/ are not any more likely to have been perceived as the original intended target (see the second column from the right). In addition, /s*/ was the second most frequent response, which suggests that listeners are overall less picky about what a good /s*/ is.

On the other hand, /tɕʰ/ responses were given, on average, lower goodness ratings, even though they were just about equally as likely to have been perceived as the original intended target. As Tables 5, 6, and 7 indicate, incorrect /tɕʰ/ responses were most frequently given in response to /tɕ/ targets. It is thus possible that when listeners hear an aspirated affricate it gets matched to their /tɕʰ/ category, but because the f0 is not as high it does not sound as good as true /tɕʰ/ productions produced with a higher f0.

4. Discussion

Non-word repetition can be imagined as what L2 learners do when presented with new words auditorily: they hear an unknown word in the L2, and then try to repeat it as accurately as they can. To the extent that their repetitions are perceived by native listeners as the intended string of phonemes in the original production, the L2 learner's perception and production of the intended phoneme string are at least satisfactory. On the other hand, to the extent that their repetitions are perceived by native listeners as something other than the intended string of phonemes in the original production, there is a problem somewhere along the chain: either in the perception or production of the L2 learner, or in the perception of the native listener. In this study, we examined three aspects of native listeners' perception of the non-word repetitions of native and non-native talkers: (1) how accurately they were perceived, (2) what types of errors were common in different vowel contexts and across talker groups, and (3) how good the listeners evaluated the repetitions to be.

When accuracy is defined as the listener perceiving the segment that was present in the original non-word, we found that initial consonants in L1 Korean talkers' productions were largely perceived accurately, especially when the following vowel was /a/. Productions from naïve talkers and L2 learners were, unsurprisingly, perceived less accurately, but there were differences across talker groups, consonant, and vowel context. Most notably, /tɕ/ and /sʰ/ targets were the most frequently misperceived, as /tɕ/ was often perceived as /tɕʰ/, and /sʰ/ was often perceived as /tɕʰ/ or /s*/. It was also found that manner errors – perceiving a fricative as an affricate or an affricate as a fricative – were very rare, with the main exception being the perception of /sʰa/ as an affricate. This specific manner

error was predicted by the perception results of Holliday (2016), in which naïve Mandarin listeners assimilated Korean /sʰa/ to an L1 aspirated affricate category, either /tɕʰa/ or /tɕʰa/.

This result constitutes support for the idea that the perception of naïve listeners can be quite different from that of L2 learners with only a few weeks of L2 experience, and that this difference has consequences for production as well. What is important to note is that the novice L2 learners' /sʰa/ productions were not accurately identified more frequently than those of naïve talkers: only the type of error changed. Instead of producing /sʰa/ sequences as [tea], [tɕʰa], or [s*a], like naïve talkers, the error produced by novice L2 learners for /sʰa/ targets was almost exclusively [s*a].

Ultimately, we believe this result highlights the need to understand why the perceptual assimilation of non-native segments might change after only a brief period of exposure to the language. If the ability to discriminate between L2 phones is driven by how they are perceptually assimilated, and if the perceptual assimilation of L2 phones by L2 learners cannot be confidently predicted from the perceptual assimilation of naïve listeners from the same L1 background, then what we need is a clearer understanding of how perceptual assimilation changes with linguistic experience.

References

- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). Baltimore, DE: York Press.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In Honor of James Emil Flege* (pp. 13-34). Amsterdam, The Netherlands: John Benjamins.
- Cebrian, J. (2007). Old sounds in new contrasts: L2 production of the English tense-lax vowel distinction. *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1637-1640). Saarbrücken, Germany.
- Chakraborty, R., Domsch, C., & Gonzales, M. D. (2011). Articulatory behaviors of nonnative speakers: Role of L2 proficiency and accent modification. *Perceptual and Motor Skills*, 113(1), 311-330.
- Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders*, 43(1), 1-40.
- Duncan, T. S., & Paradis, J. (2016). English language learners' nonword repetition performance: The influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal*

- of *Speech, Language, and Hearing Research*, 59(1), 39-48.
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421-436.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277) Baltimore, DE: York Press.
- Holliday, J. J. (2012). The acoustic realization of the Korean sibilant fricative contrast in Seoul and Daegu. *Phonetics and Speech Sciences*, 4(1), 67-74.
- Holliday, J. J. (2014a). The perception of Seoul Korean fricatives by listeners from five different native dialect and language groups. *Korean Linguistics*, 16(2), 91-108.
- Holliday, J. J. (2014b). The perceptual assimilation of Korean obstruents by native Mandarin listeners. *Journal of the Acoustical Society of America*, 135(3), 1585-1595.
- Holliday, J. J. (2016). Second language experience can hinder the discrimination of nonnative phonological contrasts. *Phonetica*, 73(1), 33-51.
- Jeon, W. (2005). *A study on Chinese students' mistakes in pronouncing Korean: Focusing on consonant* (Master's thesis). Sungkyunkwan University, Korea.
- Kallay, J., & Holliday, J. J. (2012, September). Using spectral measures to differentiate Mandarin and Korean sibilant fricatives. *Proceedings of the INTERSPEECH 2012* (pp. 118-121). Portland, OR.
- Lee, H., & Jongman, A. (2016). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association*, 42(2), 145-169.
- Li, H. (2015). Korean fricative /s, ss/ analysis' pronunciation based on Chinese beginners. *The Journal of Korean Language Education Research*, 2, 129-143.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314-324.
- Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research*, 48(1), 61-78.
- R Core Team (2018). R: A language and environment for statistical computing (version 3.5.2) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- RStudio Team (2016). RStudio: Integrated development for R (version 1.1.463) [Computer software]. Boston, MA: RStudio. Retrieved from <http://www.rstudio.com/>
- Sturman, H. W., Baker-Smemoe, W., Carreño, S., & Miller, B. B. (2016). Learning the Marshallese phonological system: The role of cross-language similarity on the perception and production of secondary articulations. *Language and Speech*, 59(4), 462-487.
- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71(1), 4-21.
- Wagner, K. O. C., & Baker-Smemoe, W. (2013). An investigation of the production of ejectives by native (L1) and second (L2) language speakers of Q'eqchi' Mayan. *Journal of Phonetics*, 41(6), 453-467.
- Wickham, H. (2019). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.3.0. <https://CRAN.R-project.org/package=tidyverse>
- Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, 19(4), 298-310.
- Yao, W. (2007). *Teaching pronunciation of the Korean language for the Chinese learners* (Master's thesis). Silla University, Korea.
- Zhang, J. (2019, August). Feature-specific advantages in L3 phonological acquisition. *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3740-3744). Melbourne, Australia.
- **Jeffrey J. Holliday**, corresponding author
Assistant Professor, Department of Korean Language and Literature
Korea University
145 Anam-ro Seongbuk-gu, Seoul 02841, Korea
Tel: +82-2-3290-1964
Email: holliday@korea.ac.kr
Areas of interest: L2 acquisition, phonetics, sociolinguistics
 - **Minkyong Hong**
PhD Student, Department of Korean Language and Literature
Korea University
145 Anam-ro Seongbuk-gu, Seoul 02841, Korea
Tel: +82-2-3290-1964
Email: gophang97@gmail.com
Areas of interest: L2 acquisition, phonetics