CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

The contribution of orthographic input, phonological skills, and rise time discrimination to the learning of non-native phonemic contrasts

Adeetee Bhide* (iD), Marta Ortega-Llebaria, Scott H. Fraundorf (iD) and Charles A. Perfetti

University of Pittsburgh
*Corresponding author. Email: arbhide89@gmail.com

Abstract

Although learning second language phonology is a difficult task, orthographic input may support the learning of difficult sound contrasts through a process known as orthographic facilitation. We extended this research by examining the effects of orthographic input together with individual differences in three different phonological learning processes, namely, the production of, perception of, and memorization of words containing three Marathi phonemic contrasts (i.e., [k-kʰ], [d̪-ɖ], and [t̪-ʈ]) by native English speakers. Moreover, because the [d̪-ɖ] and [t̪-ʈ] contrasts were particularly challenging in previous auditory training studies (e.g., Polka, 1991), we used cross-modal training in order to enhance learning by pairing auditory perception tasks with visual orthographic information, the amplification of relevant acoustic cues, and proprioceptive descriptions to the articulation of target phonemes. Results showed significant learning from the pre- to the posttest across tasks and contrasts, supporting the effectiveness of cross-modal training. Furthermore, incongruent orthographic input could inhibit perception, and orthographic input generally supported memory for word pronunciations. Moreover, individual differences regarding phonological skills and nonspeech auditory discrimination predicted participants' success in different phonological learning processes. These results provide a detailed picture of the complexity between different aspects of second language phonological learning and cross-modal training.

Keywords: dental/retroflex contrasts; L1 interference; non-native contrasts; orthographic facilitation; phonological awareness; rise time discrimination

Learning a second language (L2) that has phonemic contrasts different from the first language (L1) presents a perceptual challenge that can vary as a function of individual perceptual and phonological abilities and that can be mitigated by orthographic input that supports the representation of phonemic contrasts. Here, we extend research on these two factors—namely, orthographic input and two individual

CrossMark

difference measures (L1 phonological skills and rise time discrimination)—by examining their influence on three different aspects of phonological learning: English speakers' pronunciation of, auditory perception of, and learning of words containing three Marathi sound pairs. Although the effect of orthographic input on pronunciation and word learning has been studied, we are the first to examine the effect of orthographic input on non-native phonemic discrimination. Furthermore, although both L1 phonological skills and rise time discrimination have been shown to affect L2 acquisition, we are the first to examine them specifically in the context of the learning of non-native speech contrasts. The Marathi sound pairs studied are [k-kʰ], which contrast in voice onset time (VOT) duration, and the [d̪-ɖ] and [t̪-ʈ] sound pairs, which contrast a dental sound with its retroflex counterpart. The dental/retroflex sounds constitute a particularly interesting context in which to examine the effect of adding visual orthographic input to auditory training tasks because previous studies using auditory training for these sounds have had limited success with English-speaking adults (e.g., Polka, 1991; Tees & Werker, 1984; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1983).

Several lines of work suggest that orthographic support for phonology is valuable for building speech representations of less familiar speech segments. Theories of lexical organization stress strong links between orthography and phonology (Perfetti & Hart, 2002), and behavioral evidence suggests a benefit of orthography in L2 word learning (Chambré, Ehri, & Ness, 2017). However, the choice of a specific orthography is critical: although congruent orthographic input is often beneficial, incongruent orthographic input can impair learning (Hayes-Harb & Cheng, 2016). Here we compare non-native contrast learning in three orthographic conditions: (a) a novel, congruent orthography; (b) an incongruent, L1 transliteration; and (c) no orthography. We examine the effect of orthography on production, perception, and phonological memory. Because potential benefits of orthography can only be detected when learning occurs, we compared the orthographic conditions within the context of a comprehensive sound training strategy that incorporates articulation feedback and the amplification of auditory cues (Jamieson & Morosan, 1986; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002) to maximize learning.

The ease of learning L2 phonemic contrasts may also vary with perceptual and phonological abilities, especially L1 phonological knowledge and speech-relevant acoustic perception abilities. Specific candidate individual differences that may relate to learning L2 phoneme contrasts are the perception of rise time, a basic acoustic parameter, and English phonological awareness/decoding ability. Our study examines the influence of these two factors—orthographic support and perceptual/phonological abilities—on learning the pronunciation of, the perception of, and memorization of words containing Marathi sounds new to L1 English speakers.

Aspiration and place of articulation phonemic contrasts

We focus here on learning three sounds pairs that illustrate two contrasts that are absent in English: long versus short aspiration ([k-kʰ]) and dental versus retroflex

place of articulation ([d̪-ɖ] and [t̪-ʈ]). Among other languages, these contrasts occur in Hindi, which has been studied in previous research (e.g., Polka, 1991), and Marathi—two Indian languages that are phonologically very similar (Bhide & Perfetti, 2019). Correctly perceiving these contrasts is essential for both comprehension and production because these contrasts form many minimal pairs in Marathi: some words differ only in aspiration (e.g., <कान्दा> [kand̪a] "onion" and <खान्दा> [kʰand̪a] "shoulder"); others only in place of articulation (e.g., <ताक> [t̪ak] "buttermilk" and <टाक> [ʈak] "throw"). Furthermore, correct perception of contrasting phonemes speeds up spoken word recognition by reducing the number of competitors activated (Cutler, 2015). Finally, correct perception of these contrasts is important for writing because the different phonemes are represented with different graphs (e.g., [t̪] and [ʈ] are written as <त> and <ट>, respectively).

The dental/retroflex articulation contrast distinguishes the voiced ([d̪-ɖ]) and the voiceless ([t̪-ʈ]) stops. The *dental stops* [d̪] and [t̪] are articulated by the tongue touching the teeth. In contrast, the *retroflex stops* [ɖ] and [ʈ] are articulated by the slightly curled tongue touching the back of the roof of the mouth (Cibelli, 2015; Verma & Chawla, 2003; Werker, 1989; but see Hamann, 2003). Both of these differ from the English [d] and [t], in which the tip of the tongue touches the alveolar ridge.

The other sounds we examine, [k-kʰ], differ acoustically in their VOT, the time between the release of the stop closure in the velum and the onset of vocal fold vibration, which is perceived as aspiration (i.e., the aspiration in [kʰ] is perceived as longer than in [k]). English uses a two-way voicing contrast in which the absence of aspiration signals voiced stops and the presence of aspiration cues voiceless stops. However, in Hindi and Marathi, aspiration cues a three-way voicing contrast: voiced stops are defined by the absence of aspiration ([g]), as they are in English, but voiceless stops with shorter aspirations ([k]) additionally contrast with voiceless stops with longer aspirations ([kʰ]). The length of the aspiration in English voiceless stops falls in between the short and long voiceless stops in Hindi and Marathi.

We selected these two contrasts because they should differ in their difficulty for English speakers. English-speaking adults can typically detect the aspiration contrast (Aggarwal, 2012; Guion & Pederson, 2007), but they find the place-of-articulation contrast extremely difficult (Werker et al., 1981; Werker & Tees, 1983). One reason for the difference in difficulty is that aspiration contrasts primarily differ in terms of VOT duration, whereas the dental/retroflex place-of-articulation contrast is cued primarily by spectral differences (although dental consonants also have a longer VOT than do retroflex consonants; Hamann, 2003; Polka, 1991; Verma & Chawla, 2003). Specifically, retroflexes have a lowered third formant (F3) during the transition to the vowel (Guion & Pederson, 2007; Hamann, 2003).

This F3 cue in the dental/retroflex contrast may pose a challenge for native English speakers for two reasons. First, English does not have retroflex phonemes.[1] The role of L1 contrasts in L2 learning is well established and is explained in most models of phoneme L2 learning. For instance, according to the Unified Competition Model, people use L1 cues during L2 acquisition, and struggle to use cues not in the L1 (MacWhinney, 2005, 2012). Similarly, the Perceptual

Assimilation Model (Best, 1991, 1995; Best & Tyler, 2007) posits that non-native phonemes can be categorized as exemplars of L1 phonemes and that this categorization modulates the difficulty of L2 sound perception. For example, the aspiration contrast is relatively easy because the voiceless, unaspirated phoneme [k] may be categorized as the voiced L1 phoneme [g] whereas the voiceless, aspirated phoneme [kʰ] is categorized as a different L1 phoneme, the voiceless [k]. Thus, each L2 sound is perceived as a different L1 phoneme. However, discriminating place of articulation is very difficult because both dental and retroflexes are categorized as the same L1 alveolar phoneme.

Second, contrasts based on spectral information may be intrinsically more difficult. Burnham (1986) has proposed a general framework that posits that speech contrasts distinguished primarily on temporal dimensions, such as the aspiration difference, are *robust* contrasts that can easily be relearned in adulthood. However, the spectral-based contrasts, such as the dental/retroflex contrast, are not as robust, and more difficult to learn as an adult.

Perhaps for both of these reasons, it has been very difficult to teach English-speaking adults to discriminate dental/retroflex place-of-articulation contrasts. Tees and Werker (1984; see also Werker et al., 1981) had participants complete a category change discrimination paradigm with the [t̪-ʈ] distinction. After 300 trials of training, only 43% of participants could discriminate the phonemes, and only 20% could do so on a delayed posttest. Similarly, Polka (1991) found that participants' success in determining if two recorded phonemes are the same or different (an AX test) improved over the course of the experiment for the [t̪-ʈ] distinction, but not [ɖ-d]. In sum, learners generally find these phonemes difficult to discriminate even after extensive auditory training.

Orthographic input

Orthographic input can influence phonological representations in both the L1 (e.g., Bhide, Gadgil, Zelinsky, & Perfetti, 2014; Ehri & Wilce, 1980) and the L2 (e.g., Bassetti & Atkinson, 2015; Escudero, Hayes-Harb, & Mitterer, 2008; Jesry, 2005; Meng, 1998; Showalter & Hayes-Harb, 2013; Young-Scholten, 1997, 2002; Young-Scholten & Langer, 2015). For example, although L2 learners of Chinese generally do not omit vowels in speech, they do when those vowels are not represented in the orthography of pinyin, an alphabetic writing system that uses Roman letters to phonetically represent Chinese characters (Bassetti, 2006, 2007; Ye, Cui, & Lin, 1997).

In many cases, orthographic input benefits learning the phonological forms of new words—for both L1 and L2 learners—through a process called *orthographic facilitation*. Specifically, orthographic input may help learning because it establishes an association of speech to print and thus takes advantage of the bidirectional graph–phoneme connections (Chambré et al., 2017; Ehri, 2017; Jubenville, Sénéchal, & Malette, 2014; Miles, Ehri, & Lauterbach, 2016; Phillips, 2011). For example, Ricketts, Bishop, and Nation (2009) demonstrated that 8- to 9-year-old children better learned to associate nonwords with pictures if they were provided with both orthographic and oral representations of the nonwords, as opposed to just oral representations. Further, the orthographic knowledge of the participants

can mediate the degree of orthographic facilitation. For example, Zhang, Li, and Chen (2017) taught Chinese vocabulary words in three orthographic conditions: Chinese character, pinyin, and no orthography. They found that recall of the pronunciations was facilitated by both Chinese characters and pinyin for second-grade Chinese-speaking children, but only by the pinyin for English-speaking adults who were beginning learners of Chinese. This difference could reflect the fact that the Chinese-speaking children could effectively use the phonetic radicals to support their memory, but the adult beginning learners of Chinese did not have enough character knowledge to do so.

Although orthography can be beneficial, incongruent orthographic input can hinder learning (Hayes-Harb, Nicol, & Barker, 2010). Here, we define "incongruent" as different grapheme–phoneme correspondences in the L1 and the to-be-learned language. For example, Hayes-Harb and Cheng (2016) taught English-speaking participants the meanings of Chinese words using either pinyin or Zhuyin, a traditional character-derived phonetic orthography still used in Taiwan for children. The pinyin was either congruent with English spellings (e.g., hearing [nai] and seeing <nai>) or incongruent with English spellings (e.g., hearing [ɕiou] and seeing <xiu>, which is incongruent because an English speaker is more likely to associate the letter <x> with the sound [z] than with the sound [ɕ]). On congruent trials, both the pinyin and Zhuyin groups performed equally well. However, on incongruent trials, the Zhuyin group outperformed the pinyin group.

Orthography is also not beneficial when the two orthographic forms do not represent two contrasting phonemes. For example, Showalter and Hayes-Harb (2015) found that the letters <k> and <q> did not help native English speakers learn minimal pairs that differed by the Arabic velar–uvular contrast. The lack of orthographic effects could be either because the contrast is extremely difficult for L1 English learners or because the two letters chosen represented one English phoneme, [k] (e.g., the <k> in *rack* and the <q> in *Iraq* are pronounced as [k]). Jackson (2016) followed up on this study by using the same phonological contrast but different orthographic forms: <k/ḵ> and the novel letterlike form <k/ᛚ>. He found that the novel letterlike form (<ᛚ>) was more beneficial, suggesting that it was different enough to suggest to participants that they should expect a phonological difference. Similarly, Escudero, Simon, and Mulak (2014) taught Spanish speakers Dutch vocabulary either with Dutch orthography or with only audio. The orthography group outperformed the audio-only group when the orthography indicated a phonological difference. For instance, the orthographic difference between graphemes <i> and <u> indicates a phonological minimal pair in Dutch ([I] and [Y]) and also in Spanish ([i] and [u]). However, the audio-only group outperformed the orthography group when the orthography did not indicate a phonological difference. For example, the orthographic difference between the graphemes <u> and <uu> in Dutch corresponds to the phonological difference [Y] and [y] but would not suggest a phonological difference in Spanish, which does not use double graphemes to represent monopthongs.

Instruction can sometimes help participants better utilize orthographic input. For example, Jackson (2016) found that participants had difficulty utilizing the <k/ḵ> orthographic forms, perhaps because the diacritic did not suggest a phonological difference. However, when provided with explicit instruction, participants

were able to use the diacritic to learn the phonological difference. Although Jackson (2016) found that instruction helped participants better utilize minimally different orthographic cues in a word learning task, instruction is not necessarily beneficial in pronunciation tasks. For instance, Brown (2015) found that English speakers learning German continued to voice the final consonants of words that are spelled with a voiced consonant (e.g., <rad> "wheel" pronounced [rat]) even when given explicit instruction that German consonants are devoiced in the word-final position.

The beneficial effects of orthography may depend not only on orthographic forms but also on the properties of the phonological forms. Escudero (2015) demonstrated that English- and Spanish-speaking participants found orthography most beneficial when learning minimal pairs that differ by a moderately difficult phonemic contrast (e.g., [ɪ –y]), but not those that differ by an easy-to-discriminate phonemic contrast (e.g., [ɑ–y]) or by a very difficult to discriminate phonemic contrast (e.g., [a–ɑ]).

Here, we examine how orthography influences both phonological discrimination and phonological memory. We compare three orthographic conditions: no orthography, English transliterations, and Marathi orthography. Marathi has an alphasyllabic orthography whose graphs are known as *akshara*. All of the phonemes included in the present study are represented by different akshara in Marathi (e.g., [t̪] and [ʈ] are written as <त> and <ट>, respectively). Marathi orthography is of obvious relevance to learning L2 Marathi, and the effect of L1 (English) transliterations is also important because it allows us to examine how incongruent orthographic input influences learning (e.g., the alvolear <t> used to represent the dental [t̪]). This question is important because transliterations are frequently part of classroom instruction and also occur in the print environment. For example, when students learn Hindi as an L2 in the United States, they learn not only Hindi orthography but also L1 transliterations (e.g., for the word [t̪al] "rhythm," they learn both the Hindi orthographic representation <ताल> and the English transliteration <taal>; Bhatia, 2008). However, this transliteration fails to align with the L1 phoneme: English <t> is alveolar whereas the Hindi word [t̪al] uses a dental consonant. One might thus expect that the English transliteration causes interference with learning rather than facilitation by supporting an (incorrect) alveolar phonological representation. Furthermore, textbooks often use capitalization to suggest a phonological difference (e.g., <t> and <T> represent the dental and retroflex forms, respectively). However, capitalization may not suggest a phonological difference to students, as both letters represent the same English phoneme (i.e., [t]). Transliterations are also ubiquitous outside of the classroom environment; people are often exposed to English transliterations when sending text messages and viewing signs and billboards (e.g., see https://upload.wikimedia.org/wikipedia/en/5/54/Tashan_Poster.jpg and https://upload.wikimedia.org/wikipedia/en/8/83/Taal_film_poster.jpg). These transliterations are often inexact (e.g., using the letter <t> to represent both the dental and retroflex). This exposure to transliterations may shape the phonological representations of the many L2 learners of Hindi in India and also for learners elsewhere exposed to movies and media using transliterations.

Speech-related abilities in learning phonemic contrasts

Turning to the role of speech-related abilities, we detail below the theoretical rationale for the involvement of two abilities: perception of rise time and phonological awareness/decoding abilities.

Rise time discrimination

Rise time discrimination is a measure of sensitivity to the interval between the onset of an acoustic stimulus and its maximum amplitude; in other words, amplitude changes per time unit or *rise time*. Differences in rise time correlate with the speed in which articulators release their closure, a cue to major sound classes (Rosen, 1992). For example, stop sounds have a faster closure release than glides. As a result, the bilabials [b] and [w] contrast in rise time; [b], the stop, has a faster rise time than [w], the glide. Even beyond these specific contrasts, rise time discrimination may measure sensitivity to durational differences more generally, such as those relevant to distinguishing phonemes that differ in VOT (as in the present study). Rise time discrimination has predicted language learning in the L1 (Goswami, Fosker, Huss, Mead, & Szucs, 2011) and the L2 (Chung, Jarmulowicz, & Bidelman, 2017), as well as abilities related to phoneme perception, such as L1 phonological awareness and reading ability (Goswami, 2011).

Phonological awareness and decoding

Phonological awareness is explicit awareness of the units in oral language, and *decoding* is the ability to use grapheme–phoneme correspondences to pronounce novel words. Because there is a large correlation between L1 and L2 phonological awareness (see meta-analysis by Melby-Lervåg & Lervåg, 2011), English (L1) phonological skills may positively predict the ability of native English speakers to learn non-native (L2) phonemes.

Recent behavioral studies lend further support to this hypothesis. Wade-Woolley and Geva (2000) found that, in English–Hebrew bilinguals, English pseudoword decoding ability predicted sensitivity to a contrast that is productive in Hebrew but phonotactically constrained in English ([ts-s]). The Wade-Woolley and Geva study was conducted with bilingual children and with a contrast that is present in English (albeit only in certain word positions), so the results may not transfer to adults learning novel contrasts in a laboratory setting. Another study was conducted with adults in a laboratory setting: Gabay and Holt (2015) used an implicit learning paradigm to demonstrate that native English-speaking dyslexic adults were impaired at perceiving novel nonspeech sound categories, relative to control participants. Furthermore, both English phonological awareness and decoding positively predicted participants' abilities to perceive the novel sound categories. Although this study used nonspeech sound categories, Gabay and Holt argue that learning these sound categories is similar to learning L2 contrasts and that dyslexic participants' difficulty with acquiring the novel categories is associated with their impairment in procedural learning more generally. Nevertheless, two questions still remain, which we investigate in the present study. First, in adults, is there a positive relationship between phonological skills and perceiving speech categories, which requires participants to hear differences within their L1 phonemic categories, rather

than nonspeech sounds? Second, is the positive relationship between phonological skills and the learning of auditory categories discernible when using nondyslexic participants and an explicit learning paradigm in a laboratory setting?

Present study: Design, research questions, and hypotheses

To examine the contribution of orthography and individual differences in learning the pronunciation of, phonological discrimination of, and memory for words containing three Marathi contrasting sound pairs, the native English participants in the present study were taught three phonemic contrasts ([ʈ-t̪], [ɖ-d̪], and [k-kʰ]) under three orthographic conditions (no-orthography, English transliterations, and Marathi orthography), resulting in a 3 x 3 within-subject design. We also measured individual differences in auditory discrimination and phonological skills as a second factor influencing L2 phonological learning. (Due to data sparsity, we could not examine the interaction of these two factors.) We compared pretest and posttests results to evaluate the following four hypotheses:

Hypothesis 1: Given that congruent orthography facilitates L2 vocabulary learning (Chambré et al., 2017), including in the L2 (Zhang et al., 2017), via orthographic facilitation, the participants in the Marathi orthography group should outperform the other participants in remembering the phonemes that comprise words. Participants in the Marathi orthography group may also outperform the other participants in pronouncing and discriminating the phonemes, but to our knowledge, there are no studies that provide strong support for that hypothesis.

Hypothesis 2: We had competing hypotheses about how participants in the English orthography condition will perform on remembering the phonemes that comprise words. The English orthography may lead to poorer performance than the audio-only group (Hypothesis 2a) because the English orthography is incongruent with the phonological information (e.g., Escudero et al., 2014). Furthermore, some of the English graphs (e.g., <t-T>) may not indicate a phonological difference to the participants because both graphs represent the same English phoneme (in this case, [t]). However, we explicitly teach participants the L2 phonology corresponding to the English graphs, which may overcome the English graphs' shortcomings and allow orthographic facilitation of the L2 phoneme (Hypothesis 2b).

Hypothesis 3: Because incongruent orthography can elicit incorrect pronunciations among L2 learners despite instruction (Bassetti, 2006, 2007; Brown, 2015; Jesry, 2005; Ye et al., 1997), we hypothesize that the English orthography group will perform more poorly than the audio-only group in pronouncing the phonemes. They may also perform more poorly in discriminating the phonemes, but to our knowledge, there are no studies that provide strong support for that hypothesis.

Hypothesis 4: Given the evidence that rise time discrimination is a general measure of speakers' sensitivity to duration differences that strongly correlates with

phonological learning (e.g., Goswami et al., 2011), we hypothesize that rise time discrimination will predict phonological learning of one or more of the skills (pronunciation, perception, or memory).

Hypothesis 5: Given that L1 phonological awareness and decoding positively predicted people's perceptual abilities to learn novel sound categories in diverse populations (e.g., Gabay & Holt, 2015), and that there is evidence for a strong correlation between L1 and L2 phonological awareness (Melby-Lervåg & Lervåg, 2011) we hypothesize that L1 phonological skills will predict perception learning. L1 phonological skills may also predict pronunciation and memory for phonemes that comprise words, but to our knowledge, there are no studies that provide strong support for that hypothesis.

Method

Pilot work

Some prior research suggests that orthography is most beneficial when the to-be-learned phonemes are of an intermediate level of difficulty (Escudero, 2015). Thus, we wanted to test whether the place-of-articulation contrast is of an appropriate difficulty level to elicit orthographic effects—especially as prior studies (Burnham, 1986) have shown that learning the place-of-articulation contrast (e.g., [t̪-ʈ] in our study) is more difficult than learning duration differences (e.g., [k-kʰ] in our study). For this purpose, we first conducted a pilot study using natural utterances by native Marathi speakers. Participants ($n = 29$) completed AX tasks in which they heard phonemes by two different speakers and had to indicate if they were producing the same phoneme (e.g., both saying [k]) or different phonemes (e.g., one said [k], the other said [kʰ]). Although participants were able to differentiate productions of [k] from those of [kʰ], they were not able to differentiate the [ɖ-d] and [t̪-ʈ] contrasts. As a result of these pilot data, we modified the materials to make them easier to learn. We had participants compare utterances only within a speaker. Further, we manipulated stimuli by enlarging duration differences between dentals and retroflexes. Piloting on several additional participants showed that participants were able to learn under these conditions.

Participants

Seventy-seven undergraduate students at the University of Pittsburgh participated. We excluded 3 participants who did not finish all sessions and 5 participants whose procedures had experimenter error, leaving 69 participants (24 males, average age $= 19$ years, range $= 18$–26 years). Participants were recruited from an undergraduate subject pool and received course credit and a $10 bonus if they both came to all four appointments at their scheduled time and remained engaged during the tasks. All participants were native English speakers with normal or corrected-to-normal vision and hearing, and no diagnoses of language/reading disorders or speech impediments.

On our language history questionnaire (see Procedure, below), no participants reported any experience with any South Asian languages nor travel to South Asia.

Table 1. The graphs in the two orthography conditions

Phoneme	Marathi graph	English graph	Phoneme	Marathi graph	English graph
t̪	त	t	ʈ	ट	T
k	क	k	kʰ	ख	kʰ
d̪	द	d	ɖ	ड	D

Some participants reported experience with an L2, but none before the age of 12 years. The languages to which some participants reported exposure included Spanish, French, German, Russian, Hebrew, Korean, Tagalog, Polish, Slovakian, and Latin. All of these languages articulate their "d" and "t" consonants only at one location; they do not phonemically contrast two places of articulation. Thus, participants' language background should not help them with the present task. Korean and Latin have both [k] and [kʰ]. The other eight languages have a velar, voiceless, unaspirated stop consonant ([k]), but do not have an aspirated counterpart (Bolozky, 1997; Chung, 2007; Defior & Serrano, 2017; Grigorenko, Kornilov, & Rakhlin, 2017; Gussmann, 2007; Hanulíková & Hamann, 2010; Landerl, 2017; Llamzon, 1966; McCullagh, 2011; Walker, 1984). Therefore, two participants may have been able to better hear the aspiration contrast because they were exposed to Korean or Latin. However, because their language exposure should only help with one contrast, and that contrast was relatively easy for many participants, we elected to retain their data for analysis.

Overview

Participants learned three contrasts: [t̪-ʈ], [d̪-ɖ], and [k-kʰ]. They learned these contrasts in three orthography conditions: English, Marathi, and no-orthography. The pairing of contrasts with orthographic conditions and the order of the contrasts were counterbalanced across participants. For the Marathi orthography, we used the corresponding Devanagari akshara. (Devanagari is the name of the script used to write Hindi and Marathi.) For the English orthography, we used the letters from the Roman alphabet that are used to transliterate Devanagari akshara in textbooks (Bhatia, 2008; see Table 1).

Procedure

The experiment took place over 4 consecutive days. On the first day, participants completed a pretest, learning phase, and posttest for the first contrast (see Table 2). On the second day, they completed a delayed posttest on the first contrast and completed a pretest, learning phase, and posttest for the second contrast. On the third day, they completed a delayed posttest on the second contrast and completed a pretest, learning phase, and posttest for a third contrast. On the fourth day, participants completed a delayed posttest on the third contrast and several individual-difference measures. All tasks, which are described in detail below, were computerized[2] and were presented in the same order to all participants. The pre- and posttest included

Table 2. The tasks included in the pretest, learning, and posttest phases of the experiment

Pretest	Learning	Posttest
Repetition*	Phoneme PN learning**	Mispronunciation detection test****
Phoneme introduction*	Word PN learning***	Repetition*
Phoneme PN test**	Learning words (1st time through)****	Phoneme PN test**
Word PN test***	Mispronunciation detection learning****	Word PN test***
	Learning words (2nd time through)****	

Note: The same tasks were used for the immediate and delayed posttests. Related tasks are displayed using matching asterisks. PN = perception.

a production task (repetition) and two perception tasks (phoneme and word perception). In addition, the pretest included a phoneme introduction task to familiarize participants with the tested contrasts, and the posttest a mispronunciation detection test. In the learning phase, all tasks had immediate feedback after each trial. Participants performed the phoneme and word perception tasks, learned new words, and detected mispronunciations in these new learned words.

Pretest tasks
Repetition. In this task, participants had to reproduce the target phonemes. They heard each of the two phonemes in the contrast once. Then, they heard the first phoneme spoken by three different speakers, and they reproduced it once. The same procedure then followed for the second phoneme. Participants' reproductions were recorded.

Phoneme introduction. This task was a learning activity, but it was done before the phoneme perception pretest because it taught participants the phoneme–grapheme mappings they would need to do that task. In this task, participants heard both phonemes. Then, they heard the first phoneme and, if they were in the English or Marathi orthography conditions, saw its corresponding graph. The experimenter then taught them how to produce it in the following manner. For the [d̪-ɖ] and [t̪-ʈ] contrasts, participants were to produce the English alveolar phoneme, [d] or [t]. Then, they were to move their tongue either forward to touch their teeth or further back to produce the dentals or retroflexes, respectively. For the [k-kʰ] contrast, participants were asked to produce the English [k] with slight aspiration while holding their hand in front of their mouth to feel the air release. They were then taught to produce "no air" or "lots of air" to produce the unaspirated or aspirated forms, respectively. The experimenters worked with the participants until they produced the phonemes correctly at least once. This process was repeated for the next phoneme.

   After learning how to articulate the phonemes, participants heard both phonemes said by six speakers in a random order. In the orthography conditions, they also saw the corresponding graphs. During this task, participants were given reference cards (see Figure 1) to use during all tasks except for the repetition pretest.

Graph	Instruction to Participant	Description to Participant
t	Put tongue by teeth	Beginning of sound sharper
T	Put tongue in back of mouth	Beginning of sound rounder
द	Put tongue by teeth	Beginning of sound sharper
ड	Put tongue in back of mouth	Beginning of sound rounder
	Don't release air	Sound cuts off promptly
	Release air	Sound ends breathily

Figure 1. Examples of the reference cards participants could use during the experiment. Each reference card consisted of the graph (in the orthography conditions only), an instruction to participants as to how to produce the phoneme, and an informal description. The top reference card is the [t̪-ʈ] contrast in the English orthography condition, the middle reference card is the [d̪-ɖ] contrast in the Marathi orthography condition, and the bottom reference card is the [k-kʰ] contrast in the no-orthography condition.

Phoneme perception test. Participants heard the same person pronounce both phonemes in the contrast. Participants had to indicate which phoneme was the dental (for the [d̪-ɖ] or [t̪-ʈ] contrasts) or unaspirated (for the [k-kʰ] contrast). The instructions varied by learning condition. In the English and Marathi orthography conditions, participants were shown the graph of the target phoneme (i.e., the dental or unaspirated) and indicated which out of the two phonemes they heard was that target. In the no-orthography condition, during the instructions, participants heard one example of the target phoneme (i.e., the dental or aspirated) and one example of the other phoneme. To help participants remember which was the target phoneme, it was always listed first on their reference cards, and experimenters made sure participants were aware of that.

Phonemes from 6 different speakers were used, with each pair played 4 times, for a total of 24 presentations per phoneme. For each speaker, two of the pairs were in one order (e.g., dental followed by retroflex) and two were in the opposite order. The stimulus pairs were presented in a random order.

Word perception. This task was very similar to the phoneme perception task, but participants heard entire words that began with the phonemes of interest. Participants had to indicate which word began with the dental (for the [d̪-ɖ] and [t̪-ʈ] contrasts) or unaspirated (for the [k-kʰ] contrast) phoneme.

There were 20 stimulus pairs per phoneme contrast. Half began with one phoneme (e.g., [d̪]) when spoken correctly in Marathi; half began with the other phoneme (e.g., [ɖ]). The words were grouped by speaker to make the task easier and thus avoid floor effects. Participants received no feedback.

Learning-phase tasks
Phoneme and word perception learning. The learning phase was the same as the test above except that participants were given feedback after every response. For example, in the phoneme test, participants were shown *Correct* or *Incorrect* in the no-orthography condition. In the orthography conditions, they were shown

Figure 2. Example stimuli from the Learning Words task. Participants heard the word [d̪ar] 'door' and saw a picture of a door. Participants in the English and Marathi orthography conditions saw the corresponding text below the picture. Participants in the no-orthography condition only saw the picture.

Correct or *Incorrect* with reference to the graphs (e.g., *Correct! The first sound was d, the second sound was D*). In the word perception test, the feedback was *Correct* or *Incorrect* in the no-orthography condition and *Correct! The word began with a d sound, the second word began with a D sound* in the orthography conditions.

Learning words. Participants learned Marathi words that began with the phoneme of interest. They saw a picture that depicted the meaning of the word and heard the word pronounced. In the orthography conditions, they also saw the word spelled using English or Marathi graphs. The graph of interest was in red, and the rest of the word was in black (see Figure 2). Participants were instructed to repeat the word out loud to the best of their ability, while focusing on the first phoneme. There were 20 words for each phoneme contrast, half beginning with each phoneme. The participants cycled through each word three times in a random order. Each time, the word was pronounced by a different speaker. This task was done twice. None of the words in this task were used in the word perception task.

Mispronunciation detection learning. Participants saw the picture that depicted the meaning of the Marathi word they had just learned. They then heard two pronunciations of the word: one was correct, and one was incorrect in that the initial phoneme was replaced by the other phoneme in the contrast pair (e.g., if the word should begin with a [k], the mispronounced word began with a [kʰ]). Participants had to indicate which pronunciation was correct. The speaker during the

mispronunciation detection test was different from the three speakers who modeled the correct pronunciation during the learning phase. The participants went through each of the 20 words once; presentation order was random. In the orthography conditions, participants saw *Correct* or *Incorrect* and the graph representing the phoneme the word should begin with (e.g., *Correct! That word should begin with a **d** sound*). In the no-orthography condition, they saw *Correct/Incorrect! That word should begin with [audio recording of the phoneme]*.

Posttest tasks

The repetition, phoneme perception, and word perception tasks were identical to the pretests. For the mispronunciation detection test, the testing phase was the same as the learning phase except participants did not receive feedback.

Individual difference measures

On the fourth day, participants completed four individual difference measures: the Language History Questionnaire, a phonological awareness test, a decoding test, and a rise time discrimination test. All participants took the Language History Questionnaire. The other three individual difference measures were added halfway through data collection, so they were administered to only 39 participants. The following semester, the participants who did not complete those individual difference measures during their initial laboratory visit were invited to return to complete those tasks. This testing took 30 min, and participants were compensated $10. Seven participants elected to return, so those individual difference measures are available from 46 participants in total.[3]

Language history questionnaire. To probe language history, we administered the Language History Questionnaire (Tokowicz, Michael, & Kroll, 2004). We were most concerned about languages that participants were moderately proficient in and languages that participants heard from a young age because language exposure during the first couple of years of life can improve phonemic perception for the relevant contrasts (Tees & Werker, 1984). We used four questions in the Language History Questionnaire that required participants to report which languages they speak/read/write/understand fluently. Additional questions assessed which languages participants had been exposed to from a young age: "What languages were spoken in your home while you were a child and by whom?; Please list the language(s) your mother/father speak."

Phonological awareness test. In the phonological awareness test (PHAT; Olson, Wise, Conners, Rack, & Fulker, 1989; Perfetti & Hart, 2002), participants heard an English word (e.g., *middle*) and were asked to remove a phoneme to form another word (e.g., *middle* without the [d] is *mill*). They were then asked to add another phoneme in its place to make a new word (e.g., add [s] to make *missile*). Although orthography (mental spellings) can play a role in this task, the stimuli are chosen to reduce its influence (e.g., taking an orthographic approach to removing the [d] from *middle* results in the incorrect answer *mile*). The score is the number of correct answers.

Decoding test. The Real Word Test (Olson et al., 1989; Perfetti & Hart, 2002) was used to measure decoding skill. Participants were shown a list of pseudowords and marked which words are phonologically identical to real English words (e.g., *serkyouler* sounds like *circular*, but *dofter* does not sound like a real word). The score is a *d'* score.

Rise time discrimination test. Rise time discrimination threshold was measured using an adaptive staircase procedure (Huss, Verney, Fosker, Mead, & Goswami, 2011; Levitt, 1971). Participants judged which of three tones had a slower rise time (described to participants as "began more softly" and demonstrated via practice trials). Supplementary Audio Files 1 and 2 provide examples of tones with fast and slow rise times, respectively. The threshold at which participants could discriminate was the measure, with a lower threshold indicating better performance.

Materials

Materials were audio recordings from six native speakers of Marathi (3 male, 3 female). Each speaker produced each phoneme three times, and a heritage Marathi speaker selected the tokens spoken with the clearest diction. For each phoneme contrast, there were 40 pairs of a correctly pronounced word and a mispronounced word (e.g., a retroflex was replaced by a dental). Twenty pairs were used in the word perception task and 20 in the mispronunciation detection task. Each speaker produced each pair once. The audio recordings from all six speakers were not needed, so the best examples were selected by the heritage Marathi speaker.[4] For a few stimuli, there were not enough good recordings from the six speakers, so additional recordings were taken from another native speaker of Marathi and a heritage speaker of Marathi (both female). All of the recordings were cut and cleaned of background noise.

Piloting reviewed above showed that the [ḏ-ḍ] and [ṯ-ṭ] contrasts were particularly difficult. To make these place-of-articulation contrasts easier, VOT was manipulated using Praat (Boersma, 2002). To emphasize the fact that dentals have longer VOTs than do retroflexes (Hamann, 2003; Polka, 1991; Verma & Chawla, 2003), the VOT for all dentals was lengthened (see Supplementary Material Tables S.1 and S.2).

In addition, English speakers have difficulty distinguishing the [ḏ-ṯ] or [ḍ-ṭ] voicing contrasts (Polka, 1991). Because [ḏ] has shorter prevoicing than does [ḍ] (Verma & Chawla, 2003), and thus is more likely to be perceived as a "t," we also increased the prevoicing duration of the [ḏ] (see Supplementary Material Table S.2).

The duration values are in Tables 3–5, and labeled sound waveforms are shown in Supplementary Material Tables S.1–S.3. For the learning words task, we assessed whether the duration values significantly varied across conditions using unpaired *t* tests because this task had only correct pronunciations; there was no partner set with the incorrect pronunciation. For all other tasks, we used paired *t* tests because there were pairs of stimuli that only differed in one phoneme.

Table 3. The voice onset time (in seconds) of the (k-k^h) stimuli

	Unaspirated	Aspirated	*t* test
Phonemes	0.033 (0.005)	0.116 (0.027)	$t(5) = 6.652$, $p = .001^*$
Word PN learning	0.029 (0.010)	0.108 (0.024)	$t(19) = 13.788$, $p < .001^*$
Word PN test	0.030 (0.009)	0.094 (0.039)	$t(19) = 7.344$, $p < .001^*$
Learning words	0.023 (0.008)	0.095 (0.026)	$t(28) = 24.933$, $p < .001^*$
Mispronunciation detection	0.027 (0.010)	0.099 (0.021)	$t(19) = 12.214$, $p < .001^*$

Note: The durations are displayed as mean (standard deviation). $^*p < .05$.

As seen in Table 3, the VOTs of the [k^h] stimuli are reliably longer than the VOTs of the [k] stimuli, suggesting they were good examples of the phonemic contrast and can be discriminated using the VOT cue. Tables 4 and 5 show that the original [ʈ] and [ɖ] tokens typically have longer VOTs than do the [t̪] and [d̪] tokens, as we expected based on prior research, but the differences were not always reliable. By contrast, the manipulated [t̪] and [d̪] tokens do have reliably longer VOTs than their retroflex counterparts, so participants can use VOT to discriminate the manipulated phonemes. Unlike previous research (Verma & Chawla, 2003), in these stimuli, the [ɖ] and [d̪] tokens have equivalent prevoicing durations. However, this is not problematic because the goal of increasing the prevoicing duration was to help participants distinguish the voicing contrast (between "d" and "t" sounds) rather than the place-of-articulation contrast (between dentals and retroflexes).

The materials were validated by having a heritage Marathi speaker perform the tasks. She performed well on all the measures (all scores ≥87.5%; see Supplementary Material Table S.4), suggesting that the materials were sound. Furthermore, she reported not being able to detect the digital manipulations.

Results

Descriptive statistics

Performance on the phoneme contrast tasks is reported in Table 6. In general, the phoneme perception task was easier than the word perception task. On the [ɖ-d̪] phoneme perception pre-test and the [ɖ-d̪] and [t̪-ʈ] word perception pretests, average performance was close to chance (50%). On the other tasks, average performance was above chance. It is interesting to note that on some of the perception tasks, some participants were significantly below chance. It is unlikely that these participants were simply confused about the instructions because the perception learning task gave them feedback on every trial. Rather, we believe that these participants were able to discriminate the phonemes but were classifying them incorrectly.

Floor effects were not evident in any of the tasks. There may be ceiling effects for the [k-k^h] contrast on the forced-choice perception posttests; to account for this, we collapsed some categories to decrease the number of cells in the design, as discussed in greater detail below.

Table 4. The voice onset time (in seconds) of the (ṯ-t) stimuli

	Mean (standard deviation)			t test		
	Original dental	Manipulated dental	Retroflex	Original dental compared to manipulated dental	Original dental compared to retroflex	Manipulated dental compared to retroflex
Phonemes	0.014 (0.004)	0.023 (0.006)	0.011 (0.002)	t (5) = 5.116, p = .004*	t (5) = 1.697, p = .150	t (5) = 3.895, p = .011*
Word PN Learning	0.018 (0.008)	0.030 (0.012)	0.012 (0.003)	t (19) = 6.935, p < .001*	t (19) = 5.163, p < .001*	t (19) = 8.223, p < .001*
Word PN Test	0.016 (0.006)	0.031 (0.010)	0.014 (0.004)	t (19) = 10.777, p < .001*	t (19) = 2.550, p = .020*	t (19) = 9.566, p < .001*
Learning Words	0.016 (0.006)	0.031 (0.010)	0.014 (0.004)	t (29) = 10.777, p < .001*	t (28) = 2.550, p = .020*	t (28) = 9.566, p < .001*
Mispronunciation detection	0.018 (0.006)	0.022 (0.007)	0.017 (0.006)	t (19) = 2.241, p = .037*	t (19) = 1.159, p = .261	t (19) = 2.516, p = .021*

*p < .05.

Table 5. The voice onset time and prevoicing duration (in seconds) of the [d̪-ɖ] stimuli

		Mean (standard deviation)			t test		
		Original dental	Manipulated dental	Retroflex	Original dental compared to manipulated dental	Original dental compared to retroflex	Manipulated dental compared to retroflex
Prevoicing	Phonemes	0.130 (0.052)	0.166 (0.049)	0.176 (0.108)	$t(5) = 4.441$	$t(5) = 1.969$	$t(5) = 0.328$
					$p = .007^*$	$p = .106$	$p = .756$
	Word PN learning	0.143 (0.048)	0.164 (0.054)	0.138 (0.037)	$t(19) = 6.728$	$t(19) = 0.540$	$t(19) = 2.841$
					$p < .001^*$	$p = .596$	$p = .010^*$
	Word PN test	0.130 (0.061)	0.143 (0.066)	0.134 (0.035)	$t(19) = 2.595$	$t(19) = 0.357$	$t(19) = 0.741$
					$p = .018^*$	$p = .725$	$p = .468$
	Learning words	0.110 (0.038)	0.141 (0.044)	0.115 (0.031)	$t(29) = 15.235$	$t(28) = 0.471$	$t(28) = 2.781$
					$p < .001^*$	$p = .641$	$p = .010^*$
	Mispronunciation detection	0.117 (0.047)	0.145 (0.055)	.0114 (0.027)	$t(19) = 5.859$	$t(19) = 0.192$	$t(19) = 2.584$
					$p < .001^*$	$p = .850$	$p = .018^*$
Voice onset time	Phonemes	0.011 (0.003)	0.032 (0.007)	0.007 (0.004)	$t(5) = 6.314$	$t(5) = 1.808$	$t(5) = 6.322$
					$p = .001^*$	$p = .130$	$p = .001^*$
	Word PN learning	0.011 (0.003)	0.025 (0.007)	0.008 (0.004)	$t(19) = 8.896$	$t(19) = 2.728$	$t(19) = 11.678$
					$p < .001^*$	$p = .013^*$	$p < .001^*$
	Word PN test	0.011 (0.004)	0.021 (0.008)	0.008 (0.005)	$t(19) = 5.050$	$t(19) = 2.393$	$t(19) = 5.394$
					$p < .001^*$	$p = .027^*$	$p < .001^*$
	Learning words	0.013 (0.005)	0.028 (0.011)	0.008 (0.004)	$t(29) = 7.265$	$t(28) = 4.510$	$t(28) = 12.895$
					$p < .001^*$	$p < .001^*$	$p < .001^*$
	Mispronunciation detection	0.010 (0.003)	0.023 (0.005)	0.008 (0.003)	$t(19) = 8.122$	$t(19) = 1.833$	$t(19) = 10.868$
					$p < .001^*$	$p = .083$	$p < .001^*$

$^*p < .05$.

Table 6. Accuracy on the phoneme contrast tasks

P	Test	O	Repetition	Phoneme PN	Word PN	MD
[k-kʰ]	Pretest	English	65.9 (23.8)	86.6 (13.9)	77.7 (19.2)	
			50.0–100	45.8–100	40.0–100	
		Marathi	64.6 (27.5)	87.8 (12.9)	82.7 (15.3)	
			0–100	50.0–100	50.0–100	
		no-	67.4 (24.3)	85.1 (17.5)	77.6 (25.0)	
			50.0–100	41.7–100	20.0–100	
	Immediate posttest	English	81.8 (24.6)	95.1 (9.5)	90.0 (9.1)	79.5 (14.9)
			50.0–100	54.2–100	60.0–100	45.0–100
		Marathi	85.4 (23.2)	94.1 (10.6)	89.2 (14.6)	76.3 (14.5)
			50.0–100	62.5–100	35.0–100	50.0–100
		no-	73.9 (25.5)	90.8 (18.2)	86.7 (14.1)	66.7 (18.7)
			50.0–100	16.7–100	40.0–100	35.0–100
	Delayed posttest	English	72.7 (29.8)	95.3 (5.5)	91.1 (7.7)	69.0 (18.3)
			0–100	83.3–100	75.0–100	45.0–100
		Marathi	79.2 (25.2)	94.8 (8.4)	92.3 (10.2)	73.5 (18.2)
			50.0–100	70.8–100	60.0–100	35.0–100
		no-	73.9 (25.5)	93.8 (11.6)	91.5 (10.2)	65.5 (16.0)
			50.0–100	54.2–100	60.0–100	40.0–95.0
[ʈ-t̪]	Pretest	English	30.4 (39.1)	71.2 (21.9)	59.6 (24.8)	
			0–100	33.3–100	15.0–100	
		Marathi	52.2 (38.4)	75.4 (19.3)	58.7 (23.4)	
			0–100	41.7–100	15.0–95.0	
		no-	39.1 (42.5)	68.5 (22.8)	50.4 (22.9)	
			0–100	20.8–100	15.0–100	
	Immediate posttest	English	54.3 (45.0)	79.2 (17.2)	69.6 (21.3)	65.5 (11.1)
			0–100	54.2–100	30.0–100	50.0–90.0
		Marathi	65.2 (35.1)	79.5 (21.8)	70.2 (23.0)	72.1 (11.2)
			0–100	33.3–100	20.0–100	55.0–90.0
		no-	58.7 (35.8)	81.5 (16.0)	65.0 (19.9)	61.0 (16.1)
			0–100	45.8–100	35.0–100	40.0–85.0
	Delayed posttest	English	45.7 (42.4)	78.1 (22.6)	67.4 (22.7)	67.7 (15.1)
			0–100	12.5–100	25.0–100	50.0–90.0
		Marathi	56.5 (37.9)	79.3 (21.6)	67.8 (25.4)	59.2 (14.7)
			0–100	37.5–100	20.0–100	30.0–85.0
		no-	52.2 (38.4)	79.9 (18.8)	67.2 (21.7)	59.5 (12.1)
			0–100	25.0–100	20.0–100	40.0–85.0

(*Continued*)

Table 6. (*Continued*)

P	Test	O	Repetition	Phoneme PN	Word PN	MD
[d̪-ɖ]	Pretest	English	40.9 (33.2)	54.2 (25.2)	52.5 (18.8)	
			0–100	4.2–100	20.0–85.0	
		Marathi	40.9 (33.2)	61.6 (19.7)	50.5 (18.6)	
			0–100	29.2–100	15.0–85.0	
		no-	43.5 (40.7)	59.1 (25.6)	53.5 (17.2)	
			0–100	20.8–100	30.0–85.0	
	Immediate posttest	English	56.5 (34.7)	69.6 (19.9)	57.5 (16.4)	79.0 (19.2)
			0–100	33.3–100	30.0–90.0	50.0–95.0
		Marathi	47.7 (36.1)	75.8 (19.2)	66.8 (16.6)	72.3 (11.7)
			0–100	33.3–100	25.0–90.0	50.0–90.0
		no-	43.5 (40.7)	74.3 (16.0)	64.3 (15.6)	55.6 (14.0)
			0–100	45.8–95.8	35.0–90.0	35.0–85.0
	Delayed posttest	English	54.2 (38.8)	68.4 (24.9)	57.9 (18.2)	58.0 (12.0)
			0–100	8.3–100	25.0–95.0	45.0–75.0
		Marathi	50.0 (34.5)	79.0 (15.2)	65.7 (19.0)	63.6 (13.6)
			0–100	45.8–100	25.0–100	40.0–90.0
		no-	52.3 (39.3)	77.9 (15.5)	64.3 (16.4)	54.4 (18.8)
			0–100	45.8–100	35.0–95.0	30.0–90.0

Note: The top row shows mean (standard deviation) and the bottom row shows the range in percent accuracy. Not all participants were included in the mispronunciation detection column. Only the data of participants who scored over 70% on the immediate posttest of the phoneme perception and words tests were analyzed, and therefore only their scores are displayed. P = phoneme contrast. O = orthography. PN = perception. MD = mispronunciation detection.

The scores on the individual difference measures are as follows: phonological awareness task: $M = 19.52/38$, $SD = 6.77$, range = 12 – 38; decoding task (measured in d'): $M = 2.41$, $SD = 0.77$, range = −0.31 – 3.73; and rise time discrimination task (measured in ms threshold): $M = 90.40$, $SD = 68.61$, range = 26.57 – 260.83.

Analytic strategy

We analyzed the trial-level data using logit linear mixed effects models. All models were fit in R using the *lme4* package version 1.1–12 (Bates, Maechler, Bolker, & Walker, 2015). All significant and marginal effects are reported below. For the repetition and perception data, two models were used: one with all the participants ($n = 69$) and one with only the participants for whom individual difference measures were available ($n = 46$). The results from the two models were highly similar. For the sake of brevity, only the models with the individual difference measures are reported below; the models with all participants are reported in the Supplementary Material.

Table 7. Response types (in percentages) when participants were attempting to say the [ḍ-d̪] and [ṭ-t̪] contrasts

Phoneme	Correct	d	ḍ	d̠	t	t̪	ṭ	k	g	n	p	b	tʃ
d̠	33.0	20.7	25.1		1.5	4.9	13.8	0	0	0	0.5	0.5	0
ḍ	62.6	15.8		6.9	0	9.9	3.4	0.5	0.5	0.5	0	0	0
t̠	53.1	7.7	1.9	1.9	20.3	11.6		2.9	0.5	0	0	0	0
ṭ	47.8	13.0	13.5	1.0	15.0		5.8	1.9	1.4	0	0	0	0.5

Because the scores on the real word test and the PHAT were strongly and significantly correlated ($r = .53$, $p < .001$), we combined those measures into a composite phonological awareness/decoding (PAD) measure by converting scores on both tests into z scores and then averaging the z scores. Rise time discrimination was also z scored and was not correlated with either the real word test or the PHAT test, $r = .07$, $p = .65$ and $r = .02$, $p = .91$, respectively.

Repetition

Scoring

For the repetition task, three audio files were inaudible and six were not backed up properly, resulting in a loss of nine files. The contrasts were coded by heritage speakers of Hindi and/or Marathi. The interrater reliability was in the moderate to good range (Altman, 1991): the κ values for the accuracy rating for the [ḍ-d̪], [ṭ-t̪], and [k-kʰ] contrasts were .51, .52, and .67, respectively.

The [k-kʰ] contrast was coded as follows: when participants had minimal aspiration, it was coded as [k] (correct pronunciation of Marathi voiceless, unaspirated phoneme); when participants produced slight aspiration (i.e., a level consistent with English aspiration), it was coded as [kʰ]; when participants produced a lot of aspiration, it was coded as [kʰʰ] (correct pronunciation of Marathi voiceless, aspirated phoneme); and when participants produced prevoicing, it was coded as [g]. When participants were meant to produce the [k] phoneme, they were correct on 50.5% of trials. The [kʰʰ], [kʰ], and [g] errors were made on 31.1%, 12.6%, and 5.8% of trials, respectively. When participants were meant to say [kʰʰ], they were correct on 97.6% of trials; on 2.4% of trials, they mistakenly said [kʰ]. Because the [ḍ-d̪] and [ṭ-t̪] contrasts had a greater variety of errors, we present their accuracy rates and error types in Table 7.

Inferential statistics

We examined repetition accuracy as a function of the following predictor variables: phoneme contrast ([ḍ-d̪], [ṭ-t̪], and [k-kʰ]), the interaction between orthography (no-, English, and Marathi), and test (pretest, immediate posttest, delayed posttest),[5] and interactions between test and each of PAD and rise time discrimination (see Tables 8–9 and Supplementary Material Table S.5). Because of the same small

Table 8. Rise time discrimination ability predicts repetition and perception

Phoneme contrast	Test	Rise time median split	Repetition	Phoneme PN	Word PN
[k-kʰ]	Pretest	Low	60.9 (21.1)	86.1 (10.8)	74.3 (20.0)
			50–100	58.3–100	25.0–100
		High	76.1 (29.7)	86.2 (19.6)	82.2 (21.2)
			0–100	41.7–100	20.0–100
	Immediate posttest	Low	76.1 (25.5)	92.2 (11.0)	83.3 (17.0)
			50–100	62.5–100	35.0–100
		High	91.3 (19.4)	95.5 (11.0)	92.2 (10.6)
			50–100	54.2–100	60.0–100
	Delayed posttest	Low	76.1 (25.5)	93.3 (9.5)	90.7 (10.0)
			50–100	66.7–100	60.0–100
		High	80.4 (25.0)	96.2 (5.7)	94.3 (6.8)
			50–100	79.2–100	75.0–100
[t̪-t]	Pretest	Low	34.8 (35.1)	73.4 (18.6)	55.0 (21.5)
			0–100	45.8–100	15.0–100
		High	43.5 (43.4)	74.1 (25.6)	62.2 (22.9)
			0–100	20.8–100	20.0–100
	Immediate posttest	Low	45.7 (39.6)	79.3 (15.7)	64.1 (19.2)
			0–100	37.5–100	30.0–100
		High	67.4 (35.7)	83.3 (19.3)	75.0 (18.7)
			0–100	33.3–100	35.0–100
	Delayed posttest	Low	43.5 (37.9)	73.6 (22.7)	60.9 (21.2)
			0–100	12.5–100	25.0–95.0
		High	50.0 (39.9)	86.1 (18.3)	74.3 (23.9)
			0–100	37.5–100	20.0–100
[d̪-d]	Pretest	Low	26.1 (33.3)	65.4 (19.3)	54.6 (17.8)
			0–100	33.3–95.8	25.0–85.0
		High	50.0 (35.4)	55.8 (25.2)	56.7 (17.0)
			0–100	20.8–100	35.0–85.0
	Immediate posttest	Low	30.4 (36.1)	72.1 (18.1)	61.3 (16.7)
			0–100	33.3–100	35.0–90.0
		High	50.0 (34.5)	75.0 (17.9)	66.1 (16.9)
			0–100	37.5–100	30.0–90.0
	Delayed posttest	Low	40.9 (39.8)	71.7 (18.9)	60.9 (18.7)
			0–100	41.7–100	35.0–100
		High	50.0 (36.9)0–100	75.7 (21.0)8.3–100	67.0 (17.7)25.0–95.0

Note: $n = 46$. The top row shows average (standard deviation) and the bottom row shows the range in percent accuracy. This table shows a median split by rise time discrimination score. "Low" represents above the median because higher scores are associated with poorer performance. Although this table shows a median split to facilitate interpretation, the analyses were performed with continuous data. PN = perception.

Table 9. Phonological awareness and decoding ability predict repetition and perception

Phoneme contrast	Test	PAD median split	Repetition	Phoneme PN	Word PN
[k-kʰ]	Pretest	Low	65.2 (23.5)	88.4 (12.9)	77.0 (22.5)
			50–100	41.7–100	20.0–100
		High	71.7 (29.5)	83.9 (18.0)	79.6 (19.3)
			0–100	45.8–100	40.0–100
	Immediate posttest	Low	78.3 (25.3)	93.8 (9.6)	86.3 (17.1)
			50–100	62.5–100	35.0–100
		High	89.1 (21.1)	93.8 (12.5)	89.1 (12.1)
			50–100	54.2–100	60.0–100
	Delayed posttest	Low	71.7 (25.3)	94.0 (9.1)	92.0 (9.7)
			50–100	66.7–100	60.0–100
		High	84.8 (23.5)	95.5 (6.6)	93.0 (7.6)
			50–100	79.2–100	75.0–100
[t̪-t]	Pretest	Low	32.6 (35.7)	75.0 (20.6)	55.4 (22.3)
			0–100	37.5–100	20.0–100
		High	45.7 (42.4)	72.5 (24.0)	61.7 (22.3)
			0–100	20.8–100	15.0–100
	Immediate posttest	Low	50.0 (36.9)	78.6 (19.9)	69.8 (18.9)
			0–100	33.3–100	35.0–100
		High	63.0 (40.5)	84.1 (14.7)	69.3 (20.6)
			0–100	58.3–100	30.0–100
	Delayed posttest	Low	37.0 (37.6)	78.1 (19.1)	65.0 (24.2)
			0–100	41.7–100	25.0–95.0
		High	56.5 (37.9)	81.5 (23.7)	70.2 (22.7)
			0–100	12.5–100	20.0–100
[ɖ-d]	Pretest	Low	28.3 (33.1)	60.5 (21.8)	53.7 (15.1)
			0–100	20.8–100	35.0–85.0
		High	47.6 (37.0)	60.7 (24.2)	57.6 (19.4)
			0–100	29.2–100	25.0–85.0
	Immediate posttest	Low	41.3 (41.7)	71.2 (18.3)	62.0 (14.8)
			0–100	33.3–95.8	35.0–90.0
		High	38.6 (30.6)	75.9 (17.5)	65.4 (18.8)
			0–100	37.5–100	30.0–90.0
	Delayed posttest	Low	41.3 (38.9)	71.4 (17.1)	63.3 (17.2)
			0–100	41.7–100	40.0–100
		High	50.0 (37.8)0–100	76.1 (22.4)8.3–100	64.6 (19.7)25.0–95.0

Note: n = 46. The top row shows average (standard deviation) and the bottom row shows the range in percent accuracy. Although this table shows a median split to facilitate interpretation, the analyses were performed with continuous data. PN = perception.

number of trials (participants produced each phoneme only once at the three testing time points), not every variable of interest could be included in the analysis. Specifically, we excluded the order in which the phoneme was learned (which exploratory analysis indicated was not a significant predictor) as well as the three-way interaction among phoneme, orthography, and test.

We coded each predictor variable using orthogonal contrasts corresponding to our hypotheses of interest. For phoneme contrast, one contrast compared the two place-of-articulation contrasts ([ɖ-d̪] vs. [ʈ-t̪]), and one compared the place-of-articulation contrasts ([ɖ-d̪] and [ʈ-t̪]) to the aspiration contrast ([k-kʰ]). For test, one contrast compared the pretest to both posttests to assess the degree of learning within the task, and the other contrast compared the immediate posttest to the delayed posttest to assess forgetting over time. Finally, for orthography, we chose two contrasts to test the statistical reliability of the pattern we observed in the means: one contrast compared the lower performing English orthography to the no-orthography and Marathi conditions to determine whether English orthography is significantly detrimental, and the other contrast compared no-orthography to the superior Marathi condition to determine whether Marathi orthography was reliably beneficial.[6]

The model also included a random intercept for both subjects and items. A large portion of variance was explained by the random slope that allowed the effect of phoneme to vary by subjects; however, the model did not converge when this random slope was included, and excluding it did not change any of the conclusions presented below. The other subject and item random slopes were omitted because they did not explain a large portion of the variance.

The odds[7] of pronouncing the [k-kʰ] aspiration contrast correctly were 12.39 times higher than pronouncing the two place-of-articulation contrasts ([ɖ-d̪] and [ʈ-t̪]) correctly, $z = 2.27$, $p = .02$. The odds of correct pronunciation were 1.84 times higher on the posttests than the pretest, $z = 3.35$, $p = .001$. Both of these main effects were also seen in the analysis with all participants (see Supplementary Material Table S.5).

Both people with better rise time discrimination and people with better PAD performed better on the repetition task ($z = –3.41$, $p = .001$ and $z = 3.23$, $p = .001$, respectively). A 1 *SD* increase in both rise time discrimination and PAD was associated with a 1.51 and 1.57 times increase in the odds of correctly pronouncing the phonemes, respectively.

Perception analyses
Phoneme perception
The predictors were order (first, second, or third contrast learned), phoneme contrast ([ɖ-d̪], [ʈ-t̪], or [k-kʰ]), orthography (no-, English, or Marathi), and test (pretest, immediate posttest, or delayed posttest). For order, the first contrast learned was used as the baseline. The other three variables were coded the same as in the model of repetition above. Ideally, the model would have included an interaction between phoneme contrast, orthography, and test. However, it was not possible to include the three-way interaction with each combination of

Phoneme × Test because participants were nearly at ceiling on the [k-k^h] contrast at posttest, and logit models become unstable for proportions near 1 or 0. Therefore, the model included an interaction between orthography and test only. Interactions between test and both rise time discrimination and PAD were also included. The model also included all random effects explaining a large portion of the variance: (a) random intercept for subjects, (b) random intercept for the phoneme pair, (c) random intercept for the phoneme pair in a particular order, (d) random slope of phoneme contrast by subjects, and (e) random slope of test by subjects (see Supplementary Material Table S.6).

The odds of getting the answer correct were 5.70 times higher in the aspiration ([k-k^h]) condition than in the place-of-articulation ([ɖ-d̪] and [t̪-ʈ]) conditions, $z = 5.43$, $p < .001$, 2.21 times higher in the [t̪-ʈ] condition than in the [ɖ-d̪] condition, $z = 2.58$, $p = .01$, and 2.19 times higher on the posttests than on the pretest, $z = 7.97$, $p < .001$. Compared to the first phoneme they learned, participants performed marginally better on the second, $z = 1.76$, $p = .08$, and significantly better on the third, $z = 2.62$, $p = .01$, contrast learned. All of these effects were also seen in the model with all participants (see Supplementary Material Table S.6).

The main model (with individual differences) also yielded two marginal effects not present in the model with all participants. First, there was a main effect of orthography, with the no- and Marathi-orthography conditions outperforming the English orthography condition, $z = 1.78$, $p = .08$. Specifically, the English orthography condition lowered the odds of answering correctly by 1.38 times. Second, there was an interaction between test (pre/post contrast) and orthography (contrast comparing English to no- and Marathi-orthography), $z = 1.89$, $p = .06$. Although there was overall improvement from pretest to posttests, there was more improvement in the conditions with no orthography or Marathi orthography.

Individual differences significantly predicted performance on the task. There was a significant interaction between test (pre/post contrast) and PAD, $z = 2.49$, $p = .01$, and a marginal interaction between test (pre/post contrast) and rise time discrimination, $z = -1.82$, $p = .07$. Participants with better PAD and rise time discrimination improved more from the pretest to the posttests.

Word perception
This model had the same fixed effects structure as the model for phoneme perception. The model also included all random effects that explained a large portion of the variance: (a) random intercept for subjects, (b) random intercept for speaker, (c) random intercept for word itself, (d) random slope of phoneme contrast by subjects, and (e) random slope of test by subjects. Because nearly zero variance was explained by the random slope that allowed the change from immediate to delayed posttest to vary by subjects, that random slope was removed, and we retained only the random slope that allowed change from pretest to posttests to vary by subjects (see Supplementary Material Table S.7).

The odds of getting the answer correct were 4.60 times higher in the aspiration ([k-k^h]) condition than in the place-of-articulation ([ɖ-d̪] and [t̪-ʈ]) conditions, $z = 5.67$, $p < .001$, and 1.83 times higher on the posttests than on the pretest,

$z = 6.72$, $p < .001$. Participants performed significantly better on the second and third phoneme contrasts they learned than on the first contrast they learned, $z = 2.34$, $p = .02$ and $z = 3.63$, $p < .001$, respectively. There was a marginal interaction between orthography (contrast comparing English to no- and Marathi-orthography) and test (pre/posttest), $z = 1.89$, $p = .06$. Although there was overall improvement from the pretest to the posttests, there was more improvement in the conditions with no orthography or Marathi orthography. All of these effects were also seen in the model with all participants (see Supplementary Material Table S.7).

There was one effect in the main model that was not seen in the model with all participants: there was a marginal interaction between test (delayed/immediate contrast) and orthography (no-/Marathi contrast), $z = 1.81$, $p = .07$. Participants in the no-orthography condition improved from immediate to delayed posttest, but participants in the Marathi orthography condition did not. Individual differences predicted performance on the task: a 1 *SD* increase in rise time discrimination marginally improved the odds of getting a correct answer by 1.22 times, $z = -1.86$, $p = .06$.

Mispronunciation detection

Our primary interest in this task was how the type of orthography affected participants' abilities to remember the correct pronunciations for words. To do well on this task, participants needed to (a) recognize which phoneme the word began with during the learning phase, (b) discriminate the phonemes during the testing phase, and (c) remember which phoneme the word began with during the testing phase. Participants could not be expected to select the correct pronunciation of individual words during the test if they are unable to perceive the difference between the phonemes. Thus, to focus our analysis of the mispronunciation detection task on memory for the correct pronunciation, rather than phoneme discrimination, we restricted our analysis to participants who could successfully discriminate the phonemes, as evidenced by scoring above 70% on the immediate posttest for both the phoneme and the word perception tests. If participants met the criterion for one phoneme but not another, we included the phoneme for which they met the criterion. This provided [ɖ-d] data for 25 participants, [ʈ-t] data for 33 participants, and [k-kʰ] data for 64 participants. We did not test individual differences on this task because of the restrictions we placed on sampling, which resulted in smaller sample sizes and in participants who generally had good phonological skills and rise time discrimination.

Although the data were restricted to participants who scored over 70% correct in the immediate perception posttests, the ability to distinguish phonemes may still affect discrimination during the testing phase. Therefore, performance on the immediate posttest of the word perception task was included as a predictor. Because there were no significant differences in performance across the three phoneme contrasts when performance on the immediate word perception task was controlled, phoneme ([ɖ-d], [ʈ-t], or [k-kʰ]) was removed as a predictor. Thus, the predictors were order (first, second, or third contrast learned) and

the three-way interaction between orthography (no-, English, or Marathi), test (immediate posttest, or delayed posttest), and centered performance on the word perception immediate posttest. Order was coded as in the above models. For test, contrast coding was used to compare the two (immediate and delayed) posttests. Orthography was coded differently than in the previous analyses because the means supported Hypothesis 2b: that both orthographic conditions are better than the no-orthography condition. Specifically, for this analysis, one contrast compared no-orthography to the English and Marathi orthography conditions to determine the extent to which the presence of orthography is beneficial, and the other contrast compared English and Marathi orthography conditions to determine if the type of orthography differentially affects memory. The model also included random intercepts for subjects and items and a random slope of test by subjects (see Supplementary Material Table S.8).

Performance on the immediate word perception task was positively related to performance on the mispronunciation detection task, $z = 7.01$, $p < .001$; even among participants who could generally discriminate the phonemes, discrimination still predicted performance on the task. A marginal interaction between performance on the word perception task and the time of test, $z = 1.78$, $p = .08$, suggests that performance on the word perception task was more predictive of performance on the immediate mispronunciation detection test than on the delayed test.

The odds of a correct answer were 1.35 times lower in the no-orthography condition than in the two orthography conditions, $z = 3.81$, $p < .001$, and 1.33 times higher on the immediate posttest than on the delayed posttest, $z = 4.21$, $p < .001$. Both main effects were qualified by an interaction between the orthography (contrast comparing no orthography to English and Marathi orthography) and time of test, $z = 2.00$, $p = .05$. Participants in both orthography conditions outperformed participants in the no-orthography condition at delayed posttest, but the margin was smaller than at immediate posttest.

Discussion

In this study, we examined native English speakers' learning after cross-modal training of three non-native Marathi phonemic contrasts: two involving a place-of-articulation contrast and one involving aspiration with manipulated stimuli that amplified acoustic differences. These manipulations were included to improve performance and enable us to compare the orthographic conditions. We examined the effect of orthographic input on this learning process by comparing conditions with English orthography, Marathi orthography, or no orthography. Overall, participants' performance between pre- and posttest showed learning across phonemic contrasts and tasks, suggesting that cross-modal training with manipulated stimuli was beneficial. English transliterations were associated with marginally poorer perception, but the presence of orthography helped participants remember phonemes in new words (see Table 10). We also found two individual difference measures—rise time discrimination and phonological skills—both positively predicted people's abilities to perceive and produce the L2 sounds.

Table 10. Summary of key findings

	Repetition	Phoneme PN	Word PN	MD
K > T, D	*	*	*	n/a
T > D		*		n/a
Improve from pretest to posttests	*	*	*	n/a
No- & Marathi > English orthography		†		n/a
No- & Marathi improved more than did English orthography		†	†	n/a
Marathi & English > no-orthography	n/a	n/a	n/a	*
People with better phonological skills had higher accuracy	*			n/a
People with better phonological skills improved more		*		n/a
People with better rise time discrimination had higher accuracy	*		†	n/a
People with better rise time discrimination improved more		†		n/a

Note: "Improved more" indicates greater improvement from pretest to posttests. ">" indicates higher accuracy. PN = perception. MD = mispronunciation detection. *$p < .05$. †$p < .10$.

Learning Marathi contrasts

Manipulated place-of-articulation utterances

Many previous studies, including our pilot work, have attempted to teach participants dental/retroflex contrasts and failed to do so (e.g., Polka, 1991; Tees & Werker, 1984; Werker et al., 1981; Werker & Tees, 1983). The difficulty of these contrasts can be explained by the Unified Competition Model (MacWhinney, 2005, 2012), which posits that people struggle to use cues not in the L1 (such as the lowered F3). Furthermore, it aligns with Burnham's (1986) proposition that spectral-based contrasts are more difficult to learn as an adult.

Nevertheless, participants were able to learn the contrasts in our present study, perhaps because we used manipulated stimuli that increased the VOT of the dentals to make the VOT difference more salient to the native English participants. Thus, we created a cue that is present in the L1 (VOT) for our participants to use. We also changed the task from a spectral-based learning task to a duration-based learning task. Our results cohere with previous research findings showing that amplified acoustic differences constitute a useful tool for teaching non-native contrasts (e.g., Jamieson & Morosan, 1986). We also replicated past findings that the [t̪-ʈ] contrast is easier than the [d̪-ɖ] contrast (Polka, 1991), which demonstrates that the manipulated stimuli still elicited similar results to natural utterances.

A remaining issue is whether manipulation can be gradually faded out to help people learn the natural utterances. Answering this question requires one to carefully consider the acoustic differences between dental and retroflex consonants. Retroflexes have a lowered F3 during the transition to the vowel and a shorter

VOT, and our manipulation lengthened the dental VOT. Thus, the manipulated utterances may not help participants learn the formant differences, so training with manipulated utterances may not transfer to natural stimuli, in which formant differences are crucial for correctly perceiving the place-of-articulation contrast. However, the task-irrelevant perceptual learning (TIPL) framework (Kingston & Diehl, 1994; Vlahou, Protopapas, & Seitz, 2012) predicts that, even if participants are attending to the VOT differences, they will also learn the formant differences. According to the TIPL framework, learning occurs for task-irrelevant stimulus features if they are systematically paired with task targets. Thus, because "lowered F3" and "short VOT" are systematically paired, participants will learn to discriminate the formants even if they are attending to VOT differences during the task. Training on the manipulated utterances would likely help participants discriminate natural utterances, but experimental evidence of this conjecture is needed.

Aspiration contrast

This study also replicated prior findings (Aggarwal, 2012; Guion & Pederson, 2007) suggesting that the [k-kʰ] contrast is easier than place-of-articulation contrasts. However, we did not replicate prior research in that there were no ceiling effects at pretest (although ceiling effects were present at posttest). It is interesting to note that some participants continued to do poorly on this contrast even after training, so it was not universally easy. Prior studies have noted that participants may have done well on the [k-kʰ] contrast because they were perceiving it as the [g-k] contrast; this explanation is in line with the Perceptual Assimilation Model (Best, 1991, 1995; Best & Tyler, 2007), which posits that non-native phonemes can be categorized as exemplars of L1 phonemes. Although this helped participants discriminate the [k-kʰ] contrast in those studies, they would not be able to discriminate the three-way [g-k-kʰ] contrast that is present in Hindi and Marathi. One important difference between this study and previous studies is that we trained participants on articulation so that they knew that [k] was different than [g]. We also had participants produce the [k] and [kʰ] phonemes and found that participants said [g] only 5.8% of the time. Thus, it is likely that our participants had some understanding of the three-way voicing contrast that includes [k], [kʰ], and [g]. However, it is important to note that we did not have participants produce [g] and that perception does not necessarily follow production. Therefore, it is possible that our study may have better trained participants on the three-way voicing contrast than did previous studies, but more research is needed to confirm this.

Orthography

Our first hypothesis was that the presence of Marathi orthography enhances learning relative to no orthography in the mispronunciation detection task due to orthographic facilitation. Our results supported this hypothesis; the Marathi orthography condition significantly outperformed the no-orthography condition. We also hypothesized that the Marathi orthography may help participants with the repetition and perception tasks, although there were no previous studies that strongly supported this hypothesis. Contrary to this hypothesis, there were

no significant differences between the Marathi and no-orthography conditions for the repetition and perception tasks.

We had two alternative hypotheses for the effect of English orthography on the mispronunciation detection task: Hypothesis 2a: Participants would perform worst in the English orthography condition because the L1 transliteration induces interference and/or does not indicate a phonological difference, or Hypothesis 2b: Participants would perform better with English (as well as Marathi) orthography because participants were explicitly instructed on the grapheme–phoneme mappings, allowing them to successfully use orthography to bootstrap phonological learning. Our data supported Hypothesis 2b: in the mispronunciation detection task, participants performed better when they could access either orthography.

For the pronunciation task, we hypothesized that the English orthography condition would perform worse than the no-orthography condition because the incongruent orthography would induce interference. We thought that we may also see similar effects in the perception tasks, although there were not any previous studies to support this hypothesis (Hypothesis 3). To our surprise, we did not see any effects of orthography in the pronunciation task, but we did find that, in the perception tasks, participants performed marginally worse when the phonemes had been presented with English letters, compared with when they had been paired with Marathi graphs or presented auditorily only

It is interesting to note that the English orthography resulted in marginally poorer performance on the perception tasks, but in significantly better performance on the mispronunciation detection task. The reason for this task-dependent pattern may reflect the distinctive processes required by each task. The perception tasks required the representation of the target phoneme, the perception of each phoneme or word onset, and a comparison with the target. The mispronunciation detection task required a representation of each learned phoneme, the extraction of that phoneme from a word onset, associating that phoneme with a picture, and memory for that phoneme when presented with the same picture. Thus, the perception task required only discrimination of the phoneme pair, whereas the mispronunciation detection task required both discrimination and identification. Prior research suggests that the ability to detect acoustic differences and the ability to map those differences onto words with different meanings are distinct perception processes (Singh, 2018; Stager & Werker, 1997). A second potential explanation is that the tasks occur at different time points: the mispronunciation detection task came later in the experiment, after participants had time to learn new grapheme–phoneme mappings for the English graphs. A third possibility is that, in restricting our mispronunciation detection analysis to participants who could successfully discriminate the phonemes, we also restricted our analysis to participants who successfully learned the new grapheme–phoneme mappings for the English graphs.

Overall, our results suggest that using L1 transliterations can sometimes induce interference, which accords with Hayes-Harb and Cheng's (2016) finding that participants were impaired when the pinyin representation was incongruent with how an English speaker would expect a word to be spelled. However, if participants are explicitly taught the new grapheme–phoneme mappings and given time to practice, some participants can learn to use the L1 letter to cue a related phoneme in L2, which can be coded as a variant of its L1 value. This finding accords with

Zhang et al.'s (2017) finding that orthographic facilitation requires sufficient orthographic knowledge and with Jackson's (2016) finding that instruction is helpful when the graphs represent the same L1 phoneme (e.g., as explained, the <T> and <t> letters were used as transliterations for the Marathi [t̪] and [ʈ] phonemes; English speakers, however, relate <T> and <t> to the same English phoneme [t]).

Some prior research suggests that orthography is most beneficial for phoneme contrasts that are of an intermediate level of difficulty (Escudero, 2015). According to this perspective, all three contrasts likely fell in this zone of optimal difficulty because all three elicited orthographic effects (even though the aspiration contrast was easier to learn than the place-of-articulation contrasts). Orthographic effects likely would not have been seen had we used natural stimuli for the dental/retroflex contrasts because our pilot data suggests that natural stimuli did not permit participants to perceive the contrast between dental and retroflex sounds at all. Our results suggest that combining orthographic input with manipulated utterances facilitated the learning of the dental/retroflex contrast, and the TIPL framework (Vlahou et al., 2012) suggests that the gradually fading out the manipulation will help participants to transfer this newly learned contrast to natural stimuli.

Individual differences

We initially collected three individual-difference measures: rise time discrimination, phonological awareness, and decoding. However, we combined phonological awareness and decoding into one measure because they were strongly positively correlated in this sample. Rise time discrimination was not correlated with either measure in this sample. Research with English-speaking dyslexic adults and children as well as typically developing children has shown that rise time discrimination predicts phonological awareness and decoding (Goswami, 2011), so the lack of a significant correlation in this sample could reflect the fact that this sample consisted of skilled adult readers.

Rise time discrimination

Rise time discrimination ability was positively correlated with the ability to produce and discriminate the phonemic contrasts, in support of Hypothesis 4. This relation reflects phonemic learning rather than rise time per se because participants were being tasked with learning to discriminate stop consonants based on VOT variations rather than rise time variations. Thus, the rise time discrimination task may measure a sensitivity to durational differences more generally.

One caveat is that we manipulated the dental phonemes by increasing their VOT; it remains to be determined whether rise time discrimination predicts the ability to discriminate the dental/retroflex contrast given natural utterances. Nevertheless, the [k-kʰ] contrast used natural utterances, so rise time discrimination does predict ability to distinguish at least one unmanipulated contrast.

It is also important to note that we did not measure any other auditory perceptual abilities, such as discrimination of duration or frequency. A remaining question is whether rise time discrimination can predict the learning of non-native contrasts while controlling for other auditory perceptual abilities.

The finding that differences in the perception of nonspeech sounds can predict the L2 perception and production of consonantal minimal pairs supports and extends the results of Slevc and Miyake (2006). They studied Japanese–English bilinguals and likewise found that individual differences in L2 phonological perception and production could be predicted by differences in tonal perception and production. The present study generalized their results by using rise time discrimination as a predictor and studying English speakers learning Marathi.

Phonological awareness and decoding

In support of Hypothesis 5, we found a positive correlation between English PAD and performance on the repetition and perception tasks. Our results accord with Gabay and Holt's (2015) finding that auditory category learning is positively predicted by both L1 phonological awareness and decoding. Taken together, the results from the present study and those of Gabay and Holt suggest that L1 phonological awareness and decoding can predict the learning of both speech and nonspeech categories, using both implicit and explicit learning paradigms, and within both dyslexic and typically developing populations. These results suggest that a general phonological ability, which may be related to procedural learning, contributes to phonological awareness in both the L1 and the L2.

Conclusion

As hypothesized, L2 orthography benefited learning of a non-native phonemic contrast in at least one task. For L1 orthography, we had competing hypotheses: L1 orthography may hinder learning (because of incongruent orthographic–phonological information) or may facilitate it (because we explicitly taught participants the orthography–phonology mappings, which may allow for orthographic facilitation). Our results provided partial support for both hypotheses: although L1 transliterations can induce interference, participants can overcome this interference with explicit instruction. Moreover, the study suggests that both rise time discrimination and L1 phonological skills positively predict L2 phonemic production and perception, thus identifying sources of individual differences in L2 phonological learning. Although we studied these individual differences in the context of learning Marathi phonemes, general phonological and auditory skills should theoretically predict the learning of L2 phonology more broadly. Overall, this study demonstrates that both individual differences and the presence of appropriate orthographic support affect L2 phonological learning.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0142716419000511

Notes

1. English occasionally uses a retroflex rhotic (Hamann, 2003).

2. Note that for the repetition task, the instructions were given via a computer but experimenters recorded and later scored the responses. For the phoneme introduction task, instructions were given via a computer but experimenters gave the participants feedback and coached them on their productions.

3. Phonological awareness and auditory perception should not greatly change within a college student over the course of several months, so we do not think that it is problematic that some of the data were collected later.

4. For each pair in word perception we used recordings from two speakers, one for the learning and one for the test. For mispronunciation detection, we used correct pronunciations from three speakers during the learning words task. We used the correct pronunciation/mispronunciation pair from one other speaker for the mispronunciation detection test.

5. When we say that the model included an interaction, the model also included lower level interactions and main effects. For example, if the model includes a three-way interaction, it also includes all three two-way interactions and all three main effects.

6. We had two hypotheses: Hypothesis 2a: the English orthography is worse than Marathi and no-orthography, and Hypothesis 2b: both the Marathi and English orthographies are better than no-orthography. The means supported Hypothesis 2a, so we chose this set of contrasts.

7. The model parameter estimates are in log odds. The odds are obtained by back-transforming the parameter estimates from the model.

References

Aggarwal, A. (2012). Acquisition of Hindi contrasts by English speakers: An optimality theoretic account. *Acta Linguistica Asiatica*, **1**, 9–24. doi: 10.4312/ala.1.2.9-24

Altman, D. (1991). *Practical statistics for medical research*. London: Chapman and Hall.

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy*, **9**, 95–114.

Bassetti, B. (2007). Effects of Hanyu pinyin on pronunciation in learners of Chinese as a foreign language. In A. Guder, X. Jian, & Y. Wan (Eds.), *The cognition, learning and teaching of Chinese characters*. Beijing, China: Beijing Language and Culture University Press.

Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics*, **36**, 67–91.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. doi: 10.18637/jss.v067.i01

Best, C. T. (1991). *The emergence of native-language phonological influences in infants: A perceptual assimilation model*. Haskins Laboratories Status Report on Speech Research, SR-107/108. New Haven, CT: Haskins Laboratories.

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. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: Benjamins.

Bhatia, T. K. (2008). *Colloquial Hindi: The complete course for beginners* (2nd ed.). New York: Routledge.

Bhide, A., Gadgil, S., Zelinsky, C. M., & Perfetti, C. A. (2014). Does reading in an alphasyllabary affect phonemic awareness? Inherent schwa effects in Marathi-English bilinguals. *Writing Systems Research*, **6**, 73–93. doi: 10.1080/17586801.2013.855619

Bhide, A., & Perfetti, C. (2019). Marathi. In R. M. Joshi & C. McBride (Eds.), *Handbook of literacy in akshara orthographies*. Cham, Switzerland: Springer.

Boersma, P. P. G. (2002). Praat: A system for doing phonetics by computer. *Glot International*, **5**, 341–345.

Bolozky, S. (1997). Israeli Hebrew phonology. In A. S. Kaye & P. T. Daniels (Eds.), *Phonologies of Asia and Africa* (pp. 287–311). Winona Lake, IN: Eisenbrauns.

Brown, K. E. (2015). *The influence of explicit instruction on failure to acquire a phonological rule due to orthographic input: The case of native English speakers learning German* (Unpublished master's thesis, University of Utah).

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, **7**, 207–239.

Chambré, S. J., Ehri, L. C., & Ness, M. (2017). Orthographic facilitation of first graders' vocabulary learning: Does directing attention to print enhance the effect? *Reading and Writing*, **30**, 1137–1156. doi: 10.1007/s11145-016-9715-z

Chung, I. (2007). *Ecology of PF: A study of Korean phonology and morphology in a derivational approach* (Unpublished doctoral dissertation, University of Connecticut).

Chung, W.-L., Jarmulowicz, L., & Bidelman, G. M. (2017). *Amplitude envelope onset, native prosodic and phonological awareness, and nonnative word learning.* Paper presented at the 24th Annual Conference for the Society for the Scientific Study of Reading, Halifax, Nova Scotia, Canada.

Cibelli, E. S. (2015). *Aspects of articulatory and perceptual learning in novel phoneme acquisition* (Unpublished doctoral dissertation, University of California–Berkeley).

Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, **36**, 115–128. doi: 10.1017/S0142716414000459

Defior, S., & Serrano, F. (2017). Learning to read Spanish. In L. Verhoeven & C. A. Perfetti (Eds.), *Reading acquisition across languages and writing systems* (pp. 270–298). Cambridge: Cambridge University Press.

Ehri, L. C. (2017). *Orthographic mapping explains facilitation of vocabulary learning: Theory and evidence.* Paper presented at the 24th Annual Conference for the Society for the Scientific Study of Reading, Halifax, Nova Scotia, Canada.

Ehri, L. C., & Wilce, L. S. (1980). The influence of orthography on readers' conceptualization of the phonemic structure of words. *Applied Psycholinguistics*, **1**, 371–385.

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, **36**, 7–22. doi: 10.1017/S014271641400040X

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, **36**, 345–360. doi: 10.1016/j.wocn.2007.11.002

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism*, **17**, 384–395. doi: 10.1017/S1366728913000436

Gabay, Y., & Holt, L. L. (2015). Incidental learning of sound categories is impaired in developmental dyslexia. *Cortex*, **73**, 131–143. doi: 10.1016/j.cortex.2015.08.008

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, **15**, 3–10.

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szucs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: The Ba-Wa distinction in developmental dyslexia. *Developmental Science*, **14**, 34–43.

Grigorenko, E. L., Kornilov, S. A., & Rakhlin, N. (2017). Learning to read Russian. In L. Verhoeven & C. A. Perfetti (Eds.), *Reading acquisition across languages and writing systems* (pp. 393–415). Cambridge: Cambridge University Press.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 57–77). Philadelphia, PA: Benjamins.

Gussmann, E. (2007). *The phonology of Polish.* New York: Oxford University Press.

Hamann, S. (2003). *The phonetics and phonology of retroflexes* (Unpublished doctoral dissertation, Utrecht University),

Hanulíková, A., & Hamann, S. (2010). Slovak. *Journal of the International Phonetic Association*, **40**, 373–378. doi: 10.1017/S0025100310000162

Hayes-Harb, R., & Cheng, H. W. (2016). The influence of the pinyin and Zhuyin writing systems on the acquisition of Mandarin word forms by native English speakers. *Frontiers in Psychology*, **7**, 1–13. doi: 10.3389/fpsyg.2016.00785

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, **53**, 367–381. doi: 10.1177/0023830910371460

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time percepion and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, **47**, 674–689.

Jackson, J. L. (2016). *The effects of novel orthographic input and phonetic instruction in second language phonological acquisition* (Unpublished master's thesis, University of Utah).

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, **40**, 205–215. doi: 10.3758/BF03211500

Jesry, M. M. (2005). Theoretically-based practical recommendations for improving EFL / ESL students' pronunciation. *Journal of King Saud University*, **18**, 1–33.

Jubenville, K., Sénéchal, M., & Malette, M. (2014). The moderating effect of orthographic consistency on oral vocabulary learning in monolingual and bilingual children. *Journal of Experimental Child Psychology*, **126**, 245–263. doi: 10.1016/j.jecp.2014.05.002

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, **70**, 419–454.

Landerl, K. (2017). Learning to read German. In L. Verhoeven & C. A. Perfetti (Eds.), *Reading acquisition across languages and writing systems* (pp. 299–322). Cambridge: Cambridge University Press.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustics Society of America*, **49**, 467–477. doi: 10.1121/1.1912375

Llamzon, T. A. (1966). Tagalog phonology. *Anthropological Linguistics*, **8**, 30–39.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). New York: Oxford University Press.

Macwhinney, B. (2012). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 211–227). New York: Routledge.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience*, **2**, 89–108. doi: 10.3758/CABN.2.2.89

McCullagh, M. (2011). The sounds of Latin: Phonology. *A Companion to the Latin Language*, 81–91. doi: 10.1002/9781444343397.ch6

Melby-Lervåg, M., & Lervåg, A (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, **34**, 114–135.

Meng, Z. (1998). Duiwai hanyu yuyin jiaoxue zhong shiyong "hanyu pinyin fang" an'de jige wenti [Some issues related to using Hanyu pinyin in the teaching of Chinese as a foreign language]. *Yuyin Yanjiu Yu Duiwai Hanyu Jiaoxue* (pp. 322–329). Beijing: Beijing Yuyan Wenhua Daxue Chubanshe.

Miles, K. P., Ehri, L. C., & Lauterbach, M. D. (2016). Mnemonic value of orthography for vocabulary learning in monolinguals and language minority English-speaking college students. *Journal of College Reading and Learning*, **46**, 99–112.

Olson, R., Wise, B., Conners, F., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities*, **22**, 339–348. doi: 10.1177/002221948902200604

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam: Benjamins.

Phillips, A. M. (2011). *Orthographic representations affect adult second language word learning* (Unpublished master's thesis, University of Pittsburgh).

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, **89**, 2961–2977. doi: 10.1121/1.400734

Ricketts, J., Bishop, D. V. M., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *Quarterly Journal of Experimental Psychology*, **62**, 1948–1966. doi: 10.1080/17470210802696104

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London*, **336**, 367–373. doi: 10.1098/rstb.1992.0070

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, **29**, 185–200. doi: 10.1177/0267658313480154

Showalter, C. E., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, **36**, 23–42. doi: 10.1017/S0142716414000411

Singh, L. (2018). Bilingual infants demonstrate advantages in learning words in a third language. *Child Development*, **89**, 397–413. doi: 10.1111/cdev.12852

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency. *Psychological Science*, **17**, 675–681. doi: 10.1111/j.1467-9280.2006.01765.x

Stager, C. L., & Werker, J. F. (1997). Infants listen to more phonetic details in speech perception tasks than in word-learning tasks. *Nature*, **388**, 381–382.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, **38**, 579–590. doi: 10.1037/h0080868

Tokowicz, N., Michael, E., & Kroll, J. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, **7**, 255–272.

Verma, R., & Chawla, P. (2003). Comparative analysis of Hindi retroflex and dental CV syllables and their synthesis. *Proceedings of the Workshop on Spoken Language Processing*. Mumbai, India: ISCA.

Vlahou, E. L., Protopapas, A., & Seitz, A. R. (2012). Implicit training of nonnative speech stimuli. *Journal of Experimental Psychology: General*, **141**, 363–381. doi: 10.1037/a0025014

Wade-Woolley, L., & Geva, E. (2000). Processing novel phonemic contrasts in the acquisition of L2 word reading. *Scientific Studies of Reading*, **4**, 295–311.

Walker, D. C. (1984). *The pronunciation of Canadian French*. Ottawa, Canada: Ottawa Press.

Werker, J. F. (1989). Becoming a native listener. *American Scientist*, **77**, 54–59. doi: 10.2307/27855552

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, **52**, 349–355.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, **37**, 278–286. doi: 10.1037/h0080725

Ye, J., Cui, L., & Lin, X. (1997). *Waiguo xuesheng hanyu yuyin xuexi duice [Chinese phonetics for foreign students]*. Beijing, China: Yuwen Chubanshe.

Young-Scholten, M. (1997). Second-language syllable simplification: Deviant development or deviant input. In J. Leather & A. James (Eds.), *New Sounds 97: Proceedings of the Third Symposium on the Acquisition of Second-Language Speech* (pp. 351–360). Klagenfurt, Austria: University of Klagenfurt.

Young-Scholten, M. (2002). Orthographic input in L2 phonological development. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 263–279). Trier, Germany: Wissenschaftlicher Verlag Trier.

Young-Scholten, M., & Langer, M. (2015). The role of orthographic input in second language German: Evidence from naturalistic adult learners ' production. *Applied Psycholinguistics*, **36**, 93–114. doi: 10.1017/S0142716414000447

Zhang, J., Li, H., & Chen, Y. (2017). *Orthographic facilitation in oral vocabulary learning in native and nonnative Chinese speakers*. Paper presented at the 24th Annual Conference for the Society for the Scientific Study Reading, Halifax, Nova Scotia, Canada.