Research Article

Phonetica 2020;77:350–393
DOI: 10.1159/000505298

Received: May 22, 2018
Accepted: August 29, 2019
Published online: January 31, 2020

Effects of Ultrasound Familiarization on Production and Perception of Nonnative Contrasts

Kevin D. Roon[a, b] Jaekoo Kang[a, b] D.H. Whalen[a–c]

[a] Program in Speech-Language-Hearing Sciences, CUNY Graduate Center, New York, NY, USA; [b] Haskins Laboratories, New Haven, CT, USA; [c] Department of Linguistics, Yale University, New Haven, CT, USA

Abstract

Background/Aims: We investigated the efficacy of ultrasound imaging of the tongue as a tool for familiarizing naïve learners with the production of a class of nonnative speech sounds: palatalized Russian consonants. ***Methods:*** Two learner groups were familiarized, one with ultrasound and one with audio only. Learners performed pre- and postfamiliarization production and discrimination tasks. ***Results:*** Ratings of productions of word-final palatalized consonants by learners from both groups improved after familiarization, as did discrimination of the palatalization contrast word-finally. There were no significant differences in the improvement between groups in either task. All learners were able to generalize to novel contexts in production and discrimination. The presence of palatalization interfered with discrimination of word-initial manner, and ultrasound learners were more successful in overcoming that interference. ***Conclusion:*** Ultrasound familiarization resulted in improvements in production and discrimination comparable to audio only. Ultrasound familiarization additionally helped learners overcome difficulties in manner discrimination introduced by palatalization. When familiarizing learners with a novel, nonnative class of sounds, a small set of stimuli in different contexts may be more beneficial than using a larger set in one context. Although untrained production can disrupt discrimination training, we found that production familiarization was not disruptive to discrimination or production.

1 Introduction

One of the many challenges facing the learner of a nonnative language is producing sounds that do not occur in the learner's native language (see, e.g., Couper, 2003; Derwing and Munro, 2005, and references therein). Many studies

karger@karger.com
www.karger.com/pho

Kevin D. Roon
Program in Speech-Language-Hearing Sciences
CUNY Graduate Center
365 Fifth Avenue, New York, NY 10016 (USA)
E-Mail kroon@gc.cuny.edu

have shown that providing explicit articulatory instruction on how to produce nonnative speech sounds can improve learners' productions, compared to learners relying solely on reproducing acoustic input. This has been shown to be possible using nothing more than detailed verbal instruction of the required articulation (e.g., Catford and Pisoni, 1970). Visually detailed explanations of varying types and complexity have also been found to help, for example, by showing static images of the articulatory targets (e.g., Saito, 2013) and animated models of the vocal tract producing the sounds (e.g., Wang et al., 2014). While these various types of articulation-based instructions have been shown to help nonnative productions, one problem that they share is that they do not give learners any means of evaluating how well or whether they themselves are implementing the articulations that they have been instructed to make. The main purpose of this study was to examine how familiarizing naïve learners with explicit real-time articulatory feedback would impact their production and perception of nonnative sounds.

1.1 Articulatory Feedback for L2 Speech Sounds

Many studies in the domain of speech-language therapy and communication disorders have used various technologies including electropalatography (EPG), electromyography, and ultrasound to provide patients with different types of articulatory feedback, resulting in improvement in misarticulations (see surveys provided by, e.g., Cleland et al., 2015; Hitchcock & Byun, 2015). Compared with these clinical applications, relatively few studies have investigated the efficacy of such biofeedback in the acquisition of nonnative speech sounds. Bliss et al. (2018) provide an extensive review, but we summarize key studies and findings here.

EPG involves fitting a custom false palate containing electrodes that can measure lingual contact (Hardcastle, 1972). Gibbon et al. (1991) report improvement in 2 Japanese speakers' production of the /ɹ/-/l/ contrast in English after training with EPG. Schmidt and Beamer (1998) report using EPG successfully to improve the pronunciation of certain contrasting sounds in English to adult speakers of Thai. Schmidt (2012) reports similar results for 2 speakers of Korean on similar English contrasts. Hacking et al. (2017) used EPG to train English-speaking learners of Russian to produce palatalized consonants, resulting in improvements to certain acoustic properties associated with palatalization but not in significant improvements in identification of these sounds by native Russian listeners. While these results are encouraging, the use of EPG as a feedback tool faces substantial hurdles in that the false palates must be custom-made for each participant. In addition, EPG is only useful for studying the production of speech sounds that involve lingual contact with the palate.

Katz and Mehta (2015) successfully used electromagnetic articulography (EMA, Hoole & Zierdt, 2010) to provide real-time feedback of participants' tongue-tip position in producing a voiced, coronal, palatal stop. When using EMA, small sensors are glued to the speaker's articulators, and the speaker sits in a magnetic field. Their participants were able to adjust their productions in response to an on-screen articulatory target corresponding to a possible but unattested speech sound, indicating that they were able to interpret and use the EMA feedback data usefully. However, this novel speech sound is unattested in any language, so strictly speaking it is not an instance of learning an L2 speech sound. Suemitsu et al. (2015) report success in using an EMA-based feedback system to train Japanese learners of English to produce the vowel /æ/. This system created learner-specific targets for /æ/ for 3 lingual EMA sensors based on that speaker's productions of Japanese /a/, /i/, and /ɯ/. Those targets were then shown on a computer display along with the real-time positions of the EMA sensors on the learner's tongue while the learner tried to produce English words containing /æ/. Suemitsu et al. (2015) found that the acoustics of the /æ/ produced by learners using this system were closer to native-speaker productions than those of learners who got acoustic training only. However, EMA is rather impractical for wider applications: participants must have sensors glued to their articulators and speak with wires coming out of their mouths, and EMA systems are relatively expensive, require highly trained staff to operate, and need custom software for visualizing real-time sensor movement relative to an articulatory target.

Wilson and Gick (2006) and Wilson (2014) discuss the research and pedagogical benefits of using ultrasound feedback in the acquisition of nonnative speech sounds (see, e.g., Stone, 2005, for an overview of using ultrasound in speech research in general), but the number of studies that have used ultrasound feedback for L2 training is relatively small. Gick et al. (2008) report pilot data showing that ultrasound feedback was useful in teaching Japanese learners of English better production of the English /ɹ/-/l/ contrast, while Tateishi and Winters (2013) found some improvement for productions of English onset /l/ by Japanese learners who received ultrasound biofeedback, but not for /ɹ/. Antolík et al. (2013) report some success in using ultrasound feedback as a training tool in teaching native Japanese speakers to differentiate French /u/ versus /y/. King and Ferragne (2017) report pilot data showing that French learners of the English light-dark allophonic contrast for /l/ improved their productions of the coda dark /ɫ/ when presented with ultrasound videos of a native speaker making the sound. In summary, these 4 studies indicate that using ultrasound feedback for L2 training has yielded encouraging results, but more work is needed to establish whether ultrasound is an effective tool for this purpose.

Ultrasound has several advantages over the other imaging technologies discussed above. The technology is safe and noninvasive. The cost of ultrasound machines adequate for these feedback purposes continues to decrease. The procedure for visualizing tongue movements for these feedback purposes is simple and requires minimal training on the part of the instructor and even less for the learner. Cleland et al. (2013) showed that ultrasound movies of the tongue are intuitively interpretable to naïve participants, based on the fact that they were able to classify ultrasound videos of a speaker producing segments of their native language. No custom software is needed to see the feedback. Ultrasound was therefore used in this study to familiarize learners with native articulation of L2 speech sounds.

1.2 Relationship between L2 Perception and Production

It is reassuring that when L2 learners receive training on discrimination, their discrimination improves (e.g., Strange & Dittman, 1984; Jamieson & Morosan, 1986; Logan et al., 1991; Bradlow et al., 1997; Wang et al., 1999), and when they receive training on sound production, their production improves (e.g., Catford & Pisoni, 1970; Macdonald et al., 1994; Derwing et al., 1998; Couper, 2003; Derwing & Rossiter, 2003; Saito, 2013). A question of long-standing theoretical debate and practical pedagogical interest is how speech perception and production are linked, both for the native language (Liberman & Whalen, 2000; see Galantucci et al., 2006; Lotto et al., 2009, for surveys, discussion, and further references) as well as for nonnative languages (e.g., Best, 1995; Flege, 1995, 1999; Baese-Berk, 2010).

The results from studies that have investigated how perception and production interact for nonnative languages have been mixed. Bradlow et al. (1997) showed that high-variability perceptual training of the English /l/-/ɹ/ contrast improved nonnative productions by Japanese learners. Wang et al. (2003) found that perceptual training of Mandarin tones improved the productions of English-speaking learners, while Bent (2005) found no correlation between production and perception of Mandarin tones by naïve English learners. However, Catford and Pisoni (1970) showed that training English speakers by verbally explaining what articulation was required to make a set of nonnative speech sounds – even without any biofeedback – not only resulted in better productions of those sounds, but also in better performance in a forced-choice identification task. However, Goto (1971) found that for Japanese learners of English, more proficiency in production was not related to a leaner’s ability to discriminate the English /l/-/ɹ/ contrast. Schmidt (2012) provided extensive training to 2 Korean speakers on producing contrasting English sounds, and found that both learners improved on identifying contrasts they had been trained on, but not on untrained contrasts. Kartushina and Frauenfelder (2014) found no correlation between how well Spanish learners of French discriminated and produced a nonnative vowel contrast. Kartushina et al. (2015) found that production training of French speakers on Danish vowels using a real-time display of the learners’ formants resulted not only in improved production, but also in improved discrimination of the vowels, although within-individual changes in production were not correlated with changes in perception. Hazan et al. (2005) found that audiovisual perceptual training resulted not only in improved perception of the English /v/-/b/-/p/ distinction by Japanese learners, but also in improved production of the English /ɹ/-/l/ contrast. Although King and Ferragne (2017) and Tateishi and Winters (2013) found some improvements in production of English /l/ by French and Japanese learners, respectively, who received ultrasound biofeedback, they found no improvements in discrimination or identification, respectively, for those learners.

There is also a set of studies that show, rather surprisingly, that production can be deleterious to perception, both in L2 (Baese-Berk, 2010; Baese-Berk & Samuel, 2015) as well as L1 (Leach & Samuel, 2007). For example, Baese-Berk and Samuel (2015) trained 2 groups of Castilian Spanish speakers on the nonnative /s/-/ʃ/ contrast using an ABX task with feedback. One group of learners had to speak the X stimulus on every trial in training, and one group did not. The group that was required to produce the stimulus did worse on the ABX task after training than the group that did not. Given the wide variety of findings in the above studies, another goal of the present study was to investigate further what effects explicit production familiarization would have on perception, as measured by discrimination.

1.3 Generalization

Most if not all studies that have trained learners on production of nonnative sounds have focused on single sounds (e.g., Japanese learners of English /ɹ/, Saito & Lyster, 2012) or contrast pairs (e.g., Japanese learners of English /ɹ/-/l/, Hazan et al., 2005; French learners of the Danish vowel contrasts /e/ vs. /ɛ/ and /y/ vs. /ø/, Kartushina et al., 2015; Korean learners of English /i/ vs. /ɪ/, Lee and Lyster, 2016). Perceiving and producing nonnative, individual segments and contrasts certainly presents a challenge for certain L2 learning. Consequently, the predominant theories of L2 perception and production – the Speech Learning Model (Flege, 2007, inter alia) and the Perceptual Assimilation Model (PAM, Best, 1995; Best & Tyler, 2007) – have been largely concerned with how L1 and L2 sound systems interact to account for how well a particular L2 sound or pair of sounds will be perceived and/or produced based on influences of L1.

In addition to individual sounds or pairs of sounds, an L2 can also make use of classes of sounds that are not used in a learner's L1. There is far less known about how well L2 learners are able to generalize what they learn to novel contexts. One recent study that investigated this question was from Pajak and Levy (2014), who found that L2 listeners whose L1 makes use of a durational difference only in vowels showed an enhanced ability to discriminate durational contrasts in L2 consonants, compared to L2 learners whose L1 does not make use of contrastive length distinctions at all. Pajak and Levy (2014) conclude that L1 learners make higher-order generalizations over phonetic dimensions, which increases their sensitivity to a novel L2 context that makes use of the same phonetic dimension.

Hacking et al. (2017) explored a different type of generalization, namely, whether L2 learners trained on a secondary articulation for one consonant could generalize to another consonant. Specifically, they used EPG to train learners on producing secondary palatalization in Russian (details on palatalization are presented in section 1.4 below). One group of learners was trained on /s/ vs. /s^j/, while another group was trained on /t/ vs. /t^j/. As mentioned in section 1.1 above, Hacking et al. (2017) found that learners improved in the acoustic production of palatalized consonants as measured by changes in the second formant transitions of the adjacent vowel, but these relevant phonetic changes after training did not result in improved identification of palatalization by native Russian listeners. In terms of generalization, Hacking et al. (2017) found no differences in production based on which pair a learner was trained on. Given the modest improvements shown by the learners in that study, the conclusions that can be drawn about generalization are limited: learners were able to generalize whatever they learned in training that resulted in improvement to consonants they were trained on to untrained consonants, to more or less the same incremental degree.

The present study was designed to test further the ability of L2 learners to generalize familiarization with the production of this class of palatalized consonants to novel environments. It expands upon the design used by Hacking et al. (2017) by familiarizing learners on >1 pair of palatalized/nonpalatalized consonants and investigates the ability of learners to generalize in both production and discrimination.

Table 1. Russian palatalized/nonpalatalized contrasts (in bold) in minimal and near-minimal pairs

Stops	Lower lip	Tongue tip
Word-initial	[**b**u.ˈdʲitʲ] – [**b**ʲu.ˈdʐ̩et] “to wake” – “budget” будить – бюджет [**p**ot] – [**p**ʲotr] “sweat” – “Peter” пот – Пётр	[ˈ**d**a.ʐ̩e] – [ˈ**d**ʲa.dʲa] “even” – “uncle” даже – дядя [**t**as] – [**t**ʲaʂ] “washbasin” – “rod” таз – тяж
Word-final	[rʲep] – [sʲtʲe**p**ʲ] “turnip (gen. pl.)” – “steppe” реп – степь	[bɨ**t**] – [bɨ**t**ʲ] “existence” – “to be” быт – быть
Liquids	Rhotic	Lateral
Word-initial	[**r**ump] – [ˈ**r**ʲum.ka] “point” – “cordial glass” румб – рюмка	[**l**uk] – [**l**ʲuk] “onion” – “hatch” лук – люк
Word-final	[tar] – [t͡sa**r**ʲ] “packaging (gen. pl.)” – “tsar” тар – царь	[mo**l**] – [mo**l**ʲ] “pier” – “moth” мол – моль

1.4 Test Case: Russian Palatalization

The nonnative class of sounds used in this study is consonant palatalization in Russian, produced and discriminated by native English speakers with no knowledge of Russian. Russian systematically contrasts palatalized versus nonpalatalized consonants across primary oral articulator, manner (stops, fricatives, nasals, and liquids), voicing, and word/syllable position, in both stressed and unstressed syllables (Jones & Ward, 1969; Halle, 1971; Padgett, 2001; Kochetov, 2002; Timberlake, 2004). While there is debate as to whether the contrast in Russian is properly characterized as palatalized versus velarized rather than palatalized versus “plain” (see, e.g., Evans-Romaine, 1998; Padgett, 2001; Proctor, 2011, for discussion and references), this detail is not material for the present study, which concerns only the production of palatalization. Examples of stops and liquids are shown in Table 1. Palatalization is usually characterized as a secondary articulation achieved by raising the tongue dorsum toward the palate (as in the production of the glide /j/) concurrently with the production of the primary articulation(s) required for the consonant (Avanesov, 1974; Ladefoged & Maddieson, 1996; Kochetov, 2002, chapter 3). We refer to palatalized consonants in general as “Cʲ” and their nonpalatalized counterparts as “C.” There is ample experimental evidence that the palatalization contrast is very salient to Russian speakers (Diehm, 1998; Kavitskaya, 2006; Babel & Johnson, 2007; Kochetov & Smith, 2009; Bolaños, 2017). In fact, Kavitskaya (2006) showed that to Russians, palatalization is no less salient perceptually than voicing or place of articulation. This is hardly surprising given its prevalence in the sound system of Russian. Palatalized sounds are commonly referred to in Russian as "soft" sounds, while their nonpalatalized counterparts are referred to as "hard" sounds.

This contrast was chosen for the present study for several reasons. From a methodological point of view, the main articulator used in the secondary palatalization gesture is the tongue dorsum, which is optimal for imaging with ultrasound. From a pedagogical point of view, the pervasiveness of palatalization in Russian phonology means that any reasonable competency in producing Russian speech must involve making this contrast. In addition, palatalization is not contrastive in English, although there are a few consonant-/j/ sequences that contrast with the singleton consonant. These contrasts are limited to word-medial nasals, as in "canon" versus "canyon" (/ˈkænən/ vs. /ˈkænjən/), and word-initially after singleton labial or velar consonants: for example, "booty" versus "beauty" (/ˈbuti/ vs. /ˈbjuti/), "coup" versus "cue" (/ku/ vs. /kju/). While comparable to some degree with true palatalized consonants, these are produced as consonant-glide sequences in (American) English, not as palatalized singletons (Diehm, 1998). These Cj sequences are possible only in certain word/syllable positions: for example, such consonant-/j/ clusters are completely unattested word-finally in English. It is therefore not surprising that the Russian palatalization contrast is often challenging for native speakers of American English to master (Diehm, 1998). Bolaños (2017) used a repetition task to show that the production of the palatalization contrasts by English speakers who were completely unfamiliar with Russian were systematically different from productions of native Russian speakers. Hacking et al. (2016, 2017) have found that this difficulty persists even with advanced L2 speakers of Russian. Hacking (2011) found that native Russian listeners were unable to correctly identify as palatalized (as opposed to not palatalized) word-final /p^j, s^j, r^j, t^j/ produced by advanced English-speaking learners and were only able to correctly identify word-final /l^j, n^j/ 28 and 19% of the time, respectively. Hacking et al. (2016) found acoustic analyses of the productions of the palatalization contrast in word-final consonants by 6 very proficient English-speaking students of Russian were significantly less palatalized than those of native Russian speakers.

One goal of the present study was to investigate what effects, if any, production familiarization with ultrasound would have on perception (more specifically, discrimination) of the L2 contrast. Several studies have investigated how well English listeners discriminate the Russian palatalized/nonpalatalized contrast (summarized in Table 2) and have found that while English listeners do not perform as well as Russian listeners, they do seem able to discriminate the contrast reasonably well, at least when the contrast is presented word-initially or intervocalically. Using a forced-choice identification task of nonsense CV/C^jV syllables, Diehm (1998) found that Russian listeners did numerically, but not statistically significantly, better than English-speaking Russian learners who had at least 2 years of Russian language study. English learners misidentified CV as C^jV 8.0% of the time, compared to Russian listeners who did this 0.4% of the time. English learners misidentified C^jV as CV 8.6% of the time, whereas Russians never made this mistake. Babel and Johnson (2007) used a speeded AX task to test the ability of English listeners who had no exposure to Russian to discriminate nonsense CV from C^jV syllables. The English listeners' discrimination was not statistically different from the Russian listeners, either in terms

Table 2. Summary of studies testing English perception of Russian palatalization contrasts

Study	Task	Consonants	Contexts
Babel and Johnson (2007)	Speeded AX	b, v, m, d, l, r	Word/utterance-initial
Bolaños (2017)	AXB	t, p	Syllable-initial and -final, always intervocalic
Diehm (1998)	Forced-choice ID	b, v, m, d, z, l, r	85% word/utterance-initial, 15% word/utterance-final
Kochetov and Smith (2009)	AX	l, r	Word-medial intervocalic
Kulikov (2011)	AX	p, b, f, v, m, t, d, s, z, n, r, l	Word-initial and word-final
Rice (2015)	ABX	p, b, f, v, m, t, d, s, z, n, l, r	Word-initial, -medial, and -final (2 vowel contexts)

of accuracy or speed. Both groups effectively discriminated over 95% of the time, and response times for both groups were not statistically different, and the quality of the vowel (/i, u, a/) had no effect on accuracy or on response times of the CV/C^{j}V discrimination. Kochetov and Smith (2009) reported no significant difference between English and Russian listeners in distinguishing word-medial /r/ from /r^j/ and /l/ from /l^j/ (in a /taXap/ template), though the English listeners were numerically worse at the discrimination than the Russian listeners for both pairs. Bolaños (2017) found that English speakers were reasonably good at discriminating palatalized from nonpalatalized Russian word-initial consonants, but did worse (though still above 85% correct) at discriminating the contrast word-finally. This word-final condition, however, was not utterance-final, so a following vowel was present to provide acoustic information following the palatalized consonant. We would expect discrimination to be worse in utterance-final position, where this information would not be present.

The most extensive investigation of the discrimination of the Russian palatalization contrast by naïve English learners is provided by Rice (2015), who used an ABX design to test how well English listeners could discriminate stimuli that differed only in palatalization. The contrast was tested for a wide range of Russian consonants (Table 2) in word-initial, word-medial, and word-final position. In addition, a novel feature of the Rice (2015) study was that she used multiple talkers in the ABX task. Each of the 3 stimuli in a given ABX triad were produced by 3 different talkers, with the A and B stimuli produced by 2 talkers of the same gender and the X stimulus produced by a talker of the opposite gender. It has been shown that high-variability perceptual training, especially with stimuli produced by multiple talkers, leads to better discrimination of the general categorical differences of the target contrast, rather than lower-level, talker-specific acoustic detail (Lively et al., 1993; Bradlow, 2008). An issue common to the previously discussed studies is that all of their stimuli were produced by one talker. Results from the Rice (2015) study are therefore more likely to represent the listeners' ability to discriminate the relevant categorical differences rather than individual speaker idiosyncrasies. Discrimination was much better in the prevocalic positions than in word-final position (a result also found by Kulikov, 2011); there was a slight advantage overall for discriminating the contrast for labial consonants, and discrimination of liquids (/r, l/) was distinctly disadvantaged. There were no other consistent patterns of discrimination of palatalization based on manner or primary oral articulator, with discrimination depending on the combination of the individual consonant and word position.

The PAM (Best, 1995; Best et al., 2001) is the most relevant model of L2 speech sound perception for making predictions concerning the ability of naïve listeners to discriminate nonnative contrasts. The central concept behind PAM is that a listener's ability to discriminate 2 contrasting, nonnative sounds will be a function of how well or whether each of those 2 L2 sounds map to L1 categories. The easiest discrimination is predicted when the 2 contrasting L2 sounds are perceived by the listener as belonging to 2 contrasting sound categories in the listener's L1, even if they are not prototypical. If the 2 sounds are both perceived as belonging to the same L1 category but differ materially in how prototypical each sound is for that category (say, with one being prototypical and one not), then the listener will discriminate the sounds well, but not as well as if they clearly corresponded to 2 different L1 categories. The worst discrimination is predicted when both sounds are classified as belonging to the same L1 category with no relevant differentiation of the goodness of fit in that category being perceived by the listener. An inherent challenge in testing the predictions of PAM is determining how a given L2 sound is perceived with respect to L1 categories. As pointed out by Rice (2015), that determination is often made by the researcher based on some set of phonetic – often acoustic – properties (e.g., Escudero et al., 2012; Fabra & Romero, 2012; Mokari & Werner, 2017, among a great many others), though Strange (2007) has shown these acoustic properties do not reflect exactly the perception of listeners.

For the perception of palatalized consonants, it might be possible for English speakers to identify a separate segment roughly equivalent to /j/. Rice (2015) ran an experiment in which she explored whether English listeners perceived palatalized Russian consonants as "containing a 'y' sound" along with the primary consonant. Overall, the glide percept was much less frequent word-finally than word-initially, but that the glide was perceived much more frequently for word-final /p^j/ than for /t^j, f^j, s^j, r^j, l^j/. She found that the discrimination of the palatalization contrast was better for consonants where listeners were able to identify a "y" sound, suggesting that Russian palatalized consonants were likely perceived as nonprototypical exemplars of the English Cj, but only when that palatalization was perceptible as a glide.

The prior results, therefore, indicate that palatalization is difficult to perceive in final position and for all manners of articulation. Because the palatalized/nonpalatalized contrast is cross-classifying with manner and primary articulator in Russian across word positions, it provides a useful case for testing how well learners are able to generalize the relevant aspects of palatalization to novel contexts. By familiarizing learners with subsets of stimuli of the same manner, it is possible to examine whether and in what circumstances learners would be able to generalize the contrast to new contexts, both in production and in discrimination.

Lastly, another reason that palatalization makes for an especially interesting case study is that the effects of palatalization in the discrimination of L2 sounds are not limited to listeners being able to tell whether 2 otherwise similar consonants differ in terms of palatalization. Smith and Kochetov (2009) report that the presence of palatalization can also affect discrimination of other aspects of nonnative sounds. Specifically, they found that Korean, Taiwanese Mandarin, Japanese, and Cantonese listeners (all bilingual with English) performed worse at discriminating the Russian palatalized pair /r^j/-/l^j/ than their nonpalatalized counterparts /r/-/l/. However, this effect of the presence of palatalization was not just limited to L2 discrimination. Smith and Kochetov (2009) found that the presence of palatalization also resulted in worse discrimination of /r^j/-/l^j/ compared to /r/-/l/ by native Russian listeners (though still better than the L2 listeners). So, while it is clear that the presence of palatalization can interfere with the discrimination of nonnative L2 contrasts, the results from Smith and Kochetov (2009) suggest that the presence of palatalization could also interfere with the discrimination of L2 contrasts that are shared with L1.

The design of the discrimination task in the present study is similar in many ways to the task used by Rice (2015) so that the baseline discrimination could be compared to the results from that study. We also used this task to investigate whether the presence of palatalization had an effect on learners' ability to discriminate along another phonetic dimension that is used in the learners' L1, namely, manner (details in section 2.3.1 below). This discrimination task was also used to investigate the effect of production familiarization on perception.

1.5 Summary of Goals of the Study

The primary goal of the present study was to assess whether access to real-time ultrasound video imaging of the tongue during production familiarization of a nonnative contrast would be more beneficial than familiarization without ultrasound imaging (acoustic only). This goal had 2 parts. The first part involved testing the hypothesis that the production of this contrast should improve after familiarization, and more so for learners with access to real-time ultrasound imaging than for those without. The second part was to test competing hypotheses concerning the impact of production familiarization on discrimination. Studies have shown that production of nonnative speech can have deleterious effects on perceptual training (Baese-Berk, 2010; Baese-Berk & Samuel, 2015). If, as these studies seem to suggest, production is inherently detrimental to perception, then we might expect that learners should do worse on a discrimination task after production familiarization. On the other hand, if production familiarization enhances the relevant properties that need to be discriminated, then learners should improve after production familiarization. As a separate question, we investigated whether there was a difference in effect on discrimination depending on whether the learner had access to ultrasound imaging during familiarization.

Another goal of the present study was to investigate whether and to what degree production familiarization would be generalizable by learners to new environments. This goal also had 2 parts. The first was to familiarize learners only with palatalization within one manner of articulation (stops or fricatives),

Table 3. Tasks in the present study, in chronological order

Task				Learners	Stimuli	Approximate duration
1	(pre)	AX discrimination		18	144 pairs	20 min
2	(pre)	Repetition		18	48 utterances	15 min
3		Familiarization			8 pairs	25 min
		Familiarization stimuli				
			Ultrasound	Audio only		
		Stops	6	6		
		Fricatives	6	–		
4	(post)	Repetition		18	48 utterances	15 min
5	(post)	AX discrimination		18	144 pairs	20 min

and then see whether they would be able to generalize that familiarization to other manners, in both production and discrimination. The second part was to determine whether the presence of palatalization would impede another type of discrimination (namely, manner), and if so, whether learners would be able to generalize what they learned in familiarization about palatalization to improve in that discrimination.

In order to assess the effects of ultrasound familiarization on production and discrimination, we collected baseline data from naïve English-speaking learners on their production and discrimination of Russian palatalization across 6 consonant pairs in 2 word positions. These baseline production data provide the most comprehensive view to-date of how well completely naïve learners produce secondary palatalization.

2 Materials and Methods

Learners performed 5 tasks in this experiment in the order shown in Table 3: (1) a prefamiliarization AX discrimination task, (2) a prefamiliarization repetition task, (3) familiarization, (4) a postfamiliarization repetition task, (5) a postfamiliarization AX discrimination task. Each participant was assigned to 1 of 3 groups in the familiarization task, which determined what type of familiarization they received and what type of stimuli were used in the familiarization. Details of each task are provided in the procedure section (2.3) below.

All experiments were conducted in the Speech Production, Acoustics, and Perception Lab at the City University of New York (CUNY) Graduate Center. Procedures were approved by the CUNY Human Research Protection Program.

2.1 Participants

There were 18 naïve learners (9 male, 9 female, ranging in age from 19 to 44 years). All of the learners were native speakers of American English, had never studied any Slavic language or Irish Gaelic (which also has contrastive palatalization; Ní Chasaide, 1999), and had never spoken any such language at any point in their lives. Stimuli were produced by 2

Table 4. Consonant stimuli for both tasks

Manner	Lower lip		Tongue tip	
Stop	/pam/–/p^{j}am/ /map/–/map^{j}/	**пам – пям** **мап – мапь**	/tam/–/t^{j}am/ /mat/–/mat^{j}/	**там – тям** **мат – мать**
Fricative	/fam/–/f^{j}am/ /maf/–/maf^{j}/	**фам – фям** **маф – мафь**	/sam/–/s^{j}am/ /mas/–/mas^{j}/	**сам – сям** **мас – мась**
Liquid				
Lateral			/lam/–l^{j}am/ /mal/–mal^{j}/	**лам – лям** **мал – маль**
Trill			/ram/–/r^{j}am/ /mar/–/mar^{j}/	**рам – рям** **мар – марь**

Real Russian words (adverbs or nominative-case nouns/adjectives) are single-underlined. Double-underlined words are phonological words in Russian, but not in nominative case. The stimuli in Russian orthography are shown in bold to the right of each pair.

speakers; both were native speakers of Moscow Russian (a 32-year-old male and 28-year-old female) and were naïve to the purpose of the experiment. In addition, we had Russian speakers evaluate the productions of the learners (see section 3.1 for details). Previous studies that have used native speakers to rate the productions of learners have varied greatly in the number of raters they used (e.g., Wang et al., 2003, used 5; Tateishi and Winters, 2013, used 3; Hacking et al., 2017, used 3; King and Ferragne, 2017, used 15). Schmid and Hopp (2014) recommend having 10–20 raters to determine the “foreign accentedness” of learner productions, but their recommendation is for determining “global foreign accent,” that is, overall impressions of accent in utterances without focusing on any one particular phonetic aspect of the utterances. Since our study was focused specifically on evaluating how well the learners produced palatalized consonants, we had 8 Russian-speaking raters (ranging in age from 30 to 49, 6 women). Seven of the 8 raters were native speakers of Russian (none of whom was a speaker who produced the stimuli). The eighth rater was the first author, who is a trained phonetician and proficient in Russian. Although he is not a native speaker, inclusion of this rater’s ratings did not materially change correlation among the raters (see section 3.1 for details).

All participants provided informed consent and received payment for their time. All participants self-reported that they had no speech, language, or hearing disorders.

2.2 Stimuli

The stimuli for all of the tasks in the experiment consisted of monosyllabic C_1VC_2 syllables, listed in Table 4. All target consonants occurred both word-initially (C_1) and word-finally (C_2). Since word position is manipulated in this experiment, voiceless obstruents were chosen because word-final voiced obstruents devoice in Russian (Halle, 1971) and could therefore not be used. The vowel (V) was always /a/. The vowel was kept constant to control for any coarticulatory effects of the vowel on the palatalization gesture. The vowel /a/ was chosen so that the lingual articulation of the vowel would be maximally different from the palatalization gesture. For each consonant/position/vowel combination, one stimulus had a palatalized consonant, and the other was not palatalized. The nontarget consonant in each stimulus was the bilabial nasal stop (/m/). The stimuli were a combination of real words (10/24) and nonwords (14/24) in Russian, because a complete design is not possible using only words or only nonwords. The stimuli that are words are underlined in Table 4. Single-underlined words are real Russian words, either nouns or adjectives inflected for nominative case (which is what the carrier phrase required), or adverbs. Double-underlined words are real words in cases other than nominative. Regardless of whether the stimulus is a word, the palatalized/nonpalatalized variants of the consonants are all attested in both word-initial and -final positions. Since the learners did not know Russian, all stimuli were effectively nonwords to them.

Each Russian speaker produced all of the target stimuli in the carrier phrase "a ɛtə ____" ("*and this is a* _____") presented on a computer screen in Russian orthography. Each stimulus appeared 5 times, in randomized order. Video images of the lingual articulation along the midsagittal plane were recorded using an Ultrasonix SonixTouch ultrasound machine (BK Ultrasound, www.bkultrasound.com). The display of the ultrasound machine was streamed to a PC and captured at 59.9402 Hz with an Osprey 260e video capture card using a lossless codec (Magic YUV). Concurrent audio was recorded by a Sennheiser shotgun microphone attached to the same video capture card, sampled at 44.1 kHz.

2.2.1 Stimuli for the AX and Repetition Tasks

One production of each stimulus (Table 4) was chosen per speaker, that is, there was one token of each stimulus from the male speaker and one from the female speaker. Tokens containing creaky phonation or unusual intonation were excluded as possible stimuli. For the stop and fricative stimuli, only tokens where the lingual contour was clearly visible in the ultrasound videos were used. The audio of each utterance was excised from the video, and the resultant sound file was scaled to have an average intensity of 70 dB SPL using Praat (Boersma & Weenink, 2018). The same stimuli were used for both the prefamiliarization and postfamiliarization AX discrimination and repetition tasks.

2.2.2 Stimuli for Familiarization Task

The familiarization task required only the stimuli from Table 4 that included stops and fricatives (see section 2.3.3 for details on the procedure). Two sets of stimuli were created, one that was audio-visual and one that was audio-only. For the audio-visual stimuli, video stimulus pairs were extracted and concatenated together so that the nonpalatalized variant of a particular consonant was shown, followed by the palatalized variant, for example, /tam/ followed by /t^{j}am/, both spoken by the same speaker. The combination of 2 consonants, 2 word positions, and 2 speakers resulted in 8 familiarization pairs for each group. The videos were digitally manipulated so that the first and last frames of the 2 stimuli in each pair were held still so that learners could orient themselves to the tongue contour in the video before each stimulus in the pair played. Within each stimulus pair, the initial frame of the nonpalatalized stimulus was held still for 0.83 s (corresponding to 50 video frames), followed by the video of the nonpalatalized stimulus, then the last frame of the nonpalatalized stimulus was held still for 0.83 s, then the initial frame of the palatalized stimulus was held still for 0.83 s, followed by the video of the palatalized stimulus, then the last frame of the palatalized stimulus was held still for 0.83 s. The portions of the video where a single frame was held still were accompanied by silence, while the videos of the productions were accompanied by the corresponding audio from the ultrasound recording. The audio-only stimuli were created by extracting the audio track from the concatenated pairs described above, including the intervals of silence.

2.3 Procedure

The 5 tasks relied on 3 procedures, described here. The AX discrimination and the repetition tasks were performed both before and after familiarization (Table 3), during which learners sat in a sound-attenuated booth wearing over-ear headphones, with a PCB Piezotronics 1/2" free-field, prepolarized condenser microphone in front of them, slightly superior to their line of sight to a computer monitor.

Table 5. Example of the AX design for one consonant (/t/)

	A^1	X^2	A^1	X^2	A^2	X^1	A^2	X^1
1	tam	tam	mat	mat	tam	tam	mat	mat
2	t^{j}am	t^{j}am	mat^j	mat^j	t^{j}am	t^{j}am	mat^j	mat^j
3	tam	t^{j}am	mat	mat^j	tam	t^{j}am	mat	mat^j
4	t^{j}am	tam	mat^j	mat	t^{j}am	tam	mat^j	mat
5	tam	sam	mat	mas	tam	sam	mas	mas
6	t^{j}am	s^{j}am	mat^j	mas^j	t^{j}am	s^{j}am	mat^j	mas^j

^{1} Produced by speaker 1. ^{2} Produced by speaker 2.

2.3.1 AX Discrimination

An AX discrimination task was used to determine to what degree English speakers could discriminate the categorial difference of the palatalized versus nonpalatalized consonant contrast in Russian, both before and after familiarization, that is, tasks 1 and 5 in Table 3 (see, e.g., Beddor & Gottfried, 1995; Davidson & Shaw, 2012, for discussion of the benefits of and issues with various tasks that can be used for testing discrimination). On each trial learners heard 2 stimuli, after which they had to indicate whether they were the same or different. One known issue with the AX task for short stimuli like those used in this experiment is that it can lead people to rely on fine phonetic detail rather than categorical information (Pisoni, 1973). To address this issue, the A and X stimuli were produced by different speakers (e.g., Gottfried et al., 1985), one male and one female, with the intent of forcing the listeners to de-prioritize speaker-specific phonetic detail. It has also been argued by Gerrits and Schouten (2004) that learners may tend to behave conservatively in an AX task and only respond "different" when they are confident of the difference. To address this issue, we had an unbalanced number of same and different trials, and learners were told in the instructions that many of the stimuli were different.

There were 6 consonants tested (Table 4) in a fully crossed design, with Table 5 showing all of the trials for /t/ as an example. All combinations of 6 consonants, 2 palatalization possibilities (present or absent), 2 word positions, and 2 speaker orders yielded 48 trials on which the correct answer was "same" (corresponding to the examples in rows 1 and 2 of Table 5). Pairs in which the only difference from the above was palatalization added another 48 trials on which the correct answer was "different": 24 trials where A produced by speaker 1 was palatalized and X produced by speaker 2 was not, and 24 trials where X was palatalized and A was not (corresponding to the examples in rows 3 and 4 of Table 5). Eight trials were also added for each consonant where the mismatching consonant differed in manner but shared the same articulator and obstruence/sonorance (i.e., /p/-/f/, /t/-/s/, /r/-/l/, corresponding to the examples in rows 5 and 6 of Table 5). There were 96 "different" trials, for a total of 144 trials for each learner in each of the 2 sessions (pre- and post-familiarization).

The experiment was run using DMDX (version 5.1.4.0, Forster and Forster, 2003). The time between the end of the A stimulus and the beginning of the X stimulus (the interstimulus interval) was 500 ms. There was a 500-ms pause after the key press before the presentation of the next trial. The instructions to the learners were: "On each trial you will hear 2 different speakers say one word each in a foreign language. Your task is to indicate whether you think the 2 speakers said the same word or different words. Press the RIGHT SHIFT key if you think they said the SAME word. Press the LEFT SHIFT key if you think they said DIFFERENT words. There will be many trials when they say different words." The left and right shift keys had labels above them that said "SAME" and "DIFFERENT," respectively. Learners could use 1 or 2 hands to respond, as we were not collecting response times.

Learners had 2 practice trials with the experimenter to ensure that they understood the task. Learners received no feedback during the task as to whether a given answer was correct. Trials were pseudorandomized for each session. One native speaker of Russian, who was also one of the raters for the repetition task, performed the AX discrimination task. This Russian speaker made 1 mistake out of 144 trials, demonstrating that the contrasts in the stimuli used were readily discriminable by a native speaker in this task.

2.3.2 Repetition

Learners were instructed to repeat what they heard over the headphones as closely as they could. Learners heard 4 full sets of the stimuli in Table 4 presented binaurally, 2 sets produced by the male speaker, the other 2 by the female speaker. The stimuli were not blocked by speaker and were pseudorandomized, also using DMDX. Audio of the learners' productions was recorded via a PCB Piezotronics signal conditioner which fed into one channel of a stereo recording using PowerLab signal acquisition hardware and LabChart software (AD Instruments, www.adinstruments.com), sampled at 40 kHz. The output from the computer that was playing the audio stimuli to the learner's headphones was split so that the stimuli were also recorded on the other channel of the stereo recording so that the Russian stimulus that the learner was trying to repeat could be readily identified from the audio recording. Each participant performed the repetition task twice (the rows labeled 2 and 4 in Table 3), producing all of the stimuli in each repetition task.

2.3.3 Familiarization

There were 3 familiarization groups in this task ("familiarization stimuli" groups shown under task 3 in Table 3): learners who were familiarized with the palatalization contrast with ultrasound videos (and concurrent audio) using the stimuli shown in Table 4 with fricative consonants only ("ultrasound fricatives"), those familiarized with ultrasound using stop consonants only ("ultrasound stops"), and those familiarized with audio stimuli only using stop consonants only ("audio stops"). This allowed us to investigate the effect of the mode of familiarization (audio-only vs. ultrasound) by comparing results from the ultrasound stops and audio stops groups, and the effect of familiarization stimulus type (stops vs. fricatives) by comparing results from the ultrasound stops and ultrasound fricatives groups. Within a given manner of articulation (stops or fricatives), each group was familiarized with 8 such pairs: 2 primary oral articulators (lower lip, tongue tip), 2 word positions (initial, final), produced by 2 speakers (male and female). Familiarization stimuli from 2 Russian speakers were used because in perception studies, training on multiple talkers has been shown to lead to better discrimination of the relevant categorical differences rather than lower-level, talker-specific acoustic detail (Lively et al., 1993; Bradlow, 2008). Learners in the 2 ultrasound familiarization groups wore headphones and sat in front of 2 screens, one showing the familiarization stimuli and the other showing the real-time ultrasound video display (Fig. 1).

The ultrasound machine was the same one used to record the familiarization stimuli. The experimenter instructed the learner on how to hold the probe and place it correctly under the chin, and explained what ultrasound shows. The learners were given an articulatory explanation of palatalization, with examples of nonpalatalized and palatalized sounds in text, still images, and ultrasound videos of the 2 native speakers. The familiarization videos were shown to the learners on a laptop using PsychoPy2 (Peirce, 2007). Learners were told to watch the video and wait until they saw and heard both of the stimuli in each pair. They then turned to the monitor of the ultrasound machine, which showed their lingual articulation in real time, and were told to try to match what they heard and what they saw as best as they could. Learners practiced with the experimenter in the room. Once the instructions and ultrasound were clear to the learner and the practice was finished, the experimenter explained that learner's productions were not being recorded, and that the experimenter would leave the room so that the learner would not be self-conscious about making mistakes. The familiarization was self-paced, with no corrective feedback provided during familiarization, similar in this regard to the

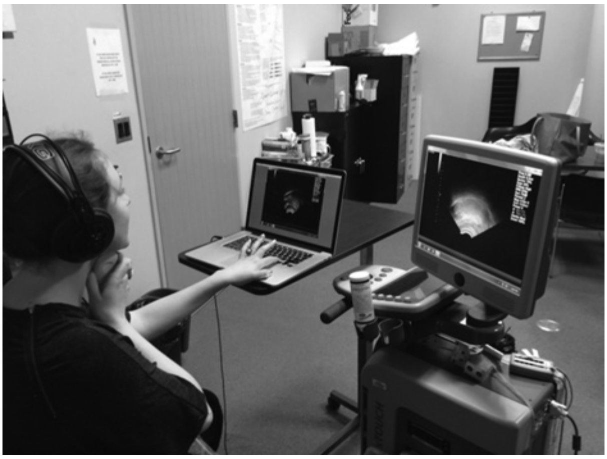

Fig. 1. Setup of the familiarization using ultrasound.

design of the experiment of Hacking et al. (2017). Learners could replay each of the 8 pairs up to 10 times. The familiarization task was relatively short, with about 10 min of instruction and the self-administered portion of the task taking about 15 min. Learners in the audio-only familiarization group received the same articulatory explanation as the ultrasound group. The instructions and familiarization routine for the audio-only group were as close as possible to the ultrasound group, except that learners in this group did not see ultrasound videos of the native speaker productions and did not see their own articulations with ultrasound imaging.

2.3.4 Ratings

In order to assess any effects of familiarization on production, we had raters judge the accentedness of the pre- and postfamiliarization audio productions of the learners in the 2 repetition tasks. The recordings of just the palatalized productions were labeled by hand using Praat and extracted into individual sound files. Each rater sat at a computer wearing binaural headphones. The experimenter explained to each rater that what they were going to hear were utterances made by learners who were trying to learn the palatalization contrast, and that each utterance they heard would be of a learner trying to produce a palatalized sound. They were instructed that their task was to try to score how well the learner produced that palatalized sound, using a Likert scale with the ratings shown in Table 6, and to ignore as much as possible any other aspects of the utterance (e.g., the vowel). The target stimulus that the learner was attempting to produce was shown at the beginning of each trial in Russian orthography, which always unambiguously indicates palatalization. No indication was given to the raters as to whether a particular utterance was a pre- or postfamiliarization production. Each rater heard all 1,728 palatalized stimuli produced by each of the 18 learners, yielding a total of 13,824 ratings. The ratings were split into 2 sessions, each containing one half of each learner's productions and each lasting just over 1 h. Both sessions were completed by a given rater on the same day, except for one rater who rated the sets 12 days apart. Presentation of the stimuli to the raters was controlled by PsychoPy2.

Interrater reliability was assessed by calculating a two-way intraclass correlation coefficient of agreement (Shrout & Fleiss, 1979; McGowan et al., 1990) with the *irr* package (Gamer et al., 2019) for R (R Development Core Team, 2018). The calculated intraclass correlation coefficient value of 0.522 ($F[1,518, 116] = 12.9$, $p = 2.44 \times 10^{-42}$) indicates "moderate" agreement among the raters according to Koo and Li (2016). The inclusion of the nonnative Russian speaker in the ratings did not have a material impact on the intraclass cor-

Table 6. Likert scale used by raters for rating learners' productions of palatalization

Rating	Instruction
Native-like	The best rating, meaning it sounds like a Russian said it
Good	Does not sound perfect but is recognizably soft
OK	Sounds more soft than hard
Poor	You can barely tell that it is soft, but it is not obviously hard
Not soft: hard	They made the right sound but the hard version
Other problem	They made the wrong sound, or it was not understandable

relation coefficient (which was 0.532 for the 7 raters excluding the nonnative speaker), so all ratings from 8 raters were used. To account for the variance across raters, rater was included in all statistical models as a random variable in the next sections.

3 Results

The results from the production task are presented first, then the discrimination.

3.1 Repetition

3.1.1 Baseline Repetition

Before investigating the effects of familiarization, it is informative to establish a baseline of how well learners produced the palatalized consonants before any familiarization. Figure 2 uses beanplots (Kampstra, 2008) to show the distribution of ratings of the prefamiliarization productions by consonant and word position, where the width of the beanplot at a particular rating reflects the number of responses for that category. Two aspects of the data that are shown in Figure 2 warrant discussion before further analyses are presented. The first is that learners' productions of word-final palatalization were rated much worse than word-initial palatalization. A cumulative-link mixed-effect model (CLMM) using the *ordinal* package (Christensen, 2019) for R with rating as the predicted value (excluding ratings of "other," see below), rater and learner as random effects, and word position as a fixed effect predictor showed that the effects of word-final productions were rated significantly worse than word-initial productions ($z = -39.44$, $p < 2 \times 10^{-16}$). One notable difference in ratings that was dependent on word position was that while all consonants had some word-final productions that were rated as "not palatalized" (i.e., "hard", 1,580 of 6,912 ratings, 22.9%), ratings of "not palatalized" were exceptionally rare for word-initial productions (58 of 6,912 ratings, 0.8%). Given the large differences in ratings based on word position, all subsequent analyses are presented separately within word position. This separation reduces the number of interactions between fixed-effect predictors included in the analyses, which simplifies both the presentation and interpretation of the results (Table 6).

DOI: 10.1159/000505298

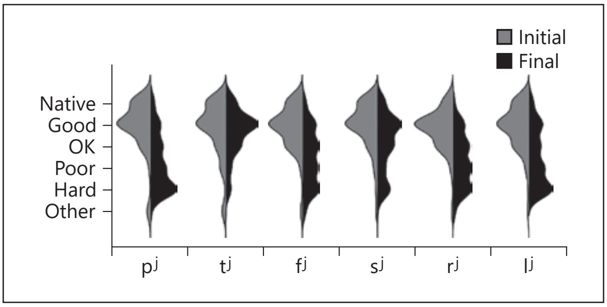

Fig. 2. Ratings of learners' palatalized consonants before familiarization by segment and word position.

The other aspect of the data is that "other" ratings were rare (351 of 13,824 ratings, 2.5%). A recurrent error of this category (which occurred both before- and after familiarization), was the learner producing [m^{j}aC] instead of [maC^j] (although some raters treated this as a "hard" rating, since the target C was not palatalized). Other examples included mistakes in place, for example, producing [mat^j] instead of [map^j], or producing a completely different consonant, for example, producing [maθ] instead of [maf^j]. Productions with a rating of "other" were excluded from all further analyses, since there was a very small number of them, and they did not involve palatalization per se.

Two CLMMs (one for each word position) were fit to the prefamiliarization rating data in order to assess whether the ratings of the productions differed significantly by consonant. The CLMMs included consonant as a fixed factor and learner and rater as random effects. The consonant /p^j/ was arbitrarily chosen as the reference level. To answer the question of which consonants were different from which others within word position, the CLMMs were assessed using estimated marginal means (Searle et al., 1980) using the *emmeans* package (Lenth et al., 2019) for R. The results of the CLMMs are shown in Table 7.

The 6 palatalized consonants (/p^j, t^j, f^j, s^j, r^j, l^j/) yielded 15 post hoc combinations to be tested, so the Bonferroni-corrected α of 0.0033 was used; significantly contrasting pairs are shown with an asterisk in Table 7. Word-initially, the most common rating for all consonants was "good," and there were few significant differences. The only significant differences were that /r^j, l^j/ were rated worse than /p^j/, and /r^j/ was rated worse than /s^j/. Word-finally, however, there were far more significant differences. /t^j/ was rated significantly better than all other consonants, /s^j/ was rated significantly better than /p^j, f^j, r^j, l^j/, and /p^j/ was rated significantly worse than /f^j/ (though not worse than /r^j, l^j/). The results of the word-final ratings can be summarized as /t^j/ >> /s^j/ >> /p^j, f^j, r^j, l^j/.

3.1.2 Effects of Familiarization on Repetition

The ratings of learners' productions of all palatalized consonants, by familiarization group, before and after familiarization are shown in Figure 3. Ratings of word-initial productions are shown in Figure 3a and ratings of word-final productions in Figure 3b.

Table 7. Results of pairwise comparisons from 2 cumulative link mixed-effects models for ratings of learners' productions of word-initial and word-final palatalized consonants, showing differences between specific consonant pairs

Word position	Contrast	Estimate	SE	*z* ratio	*p* value
Word-initial					
	$p^j \sim t^j$	–0.295	0.122	–2.422	0.1487
	$p^j \sim f^j$	–0.428	0.118	–3.618	0.0040
	$p^j \sim s^j$	–0.182	0.121	–1.508	0.6588
	$p^j \sim r^j$	–0.721	0.118	–6.088	<0.0001*
	$p^j \sim l^j$	–0.533	0.119	–4.474	0.0001*
	$t^j \sim f^j$	–0.132	0.120	–1.102	0.8806
	$t^j \sim s^j$	0.113	0.123	0.922	0.9411
	$t^j \sim r^j$	–0.425	0.120	–3.546	0.0052
	$t^j \sim l^j$	–0.237	0.121	–1.965	0.3624
	$f^j \sim s^j$	0.245	0.119	2.061	0.3079
	$f^j \sim r^j$	–0.293	0.116	–2.529	0.1158
	$f^j \sim l^j$	–0.105	0.117	–0.897	0.9473
	$s^j \sim r^j$	–0.539	0.119	–4.521	0.0001*
	$s^j \sim l^j$	–0.350	0.120	–2.923	0.0405
	$r^j \sim l^j$	0.188	0.116	1.617	0.5872
Word-final					
	$p^j \sim t^j$	2.2744	0.117	19.450	<0.0001*
	$p^j \sim f^j$	0.5721	0.113	5.042	<0.0001*
	$p^j \sim s^j$	1.4761	0.115	12.821	<0.0001*
	$p^j \sim r^j$	0.2739	0.113	2.425	0.1474
	$p^j \sim l^j$	0.3397	0.115	2.962	0.0362
	$t^j \sim f^j$	–1.7023	0.112	–15.170	<0.0001*
	$t^j \sim s^j$	–0.7982	0.109	–7.296	<0.0001*
	$t^j \sim r^j$	–2.0004	0.113	–17.662	<0.0001*
	$t^j \sim l^j$	–1.9346	0.114	–16.901	<0.0001*
	$f^j \sim s^j$	0.9040	0.111	8.134	<0.0001*
	$f^j \sim r^j$	–0.2982	0.110	–2.706	0.0741
	$f^j \sim l^j$	–0.2324	0.112	–2.076	0.3001
	$s^j \sim r^j$	–1.2022	0.112	–10.770	<0.0001*
	$s^j \sim l^j$	–1.1364	0.113	–10.047	<0.0001*
	$r^j \sim l^j$	0.0658	0.112	0.589	0.9918

An asterisk indicates a significant predictor (Bonferroni corrected $\alpha = 0.0033$).

Two CLMMs were fit to the data in order to assess the effects of the different familiarization groups on the ratings of productions, one for word-initial productions and another for word-final ones. Each model had rating as the predicted value, learner and rater as random effects, and included the fixed effects of session (pre- or postfamiliarization), familiarization group, and the interaction between the two. In each model, prefamiliarization productions by the audio-only familiarization group served as the reference level. Significant effects with

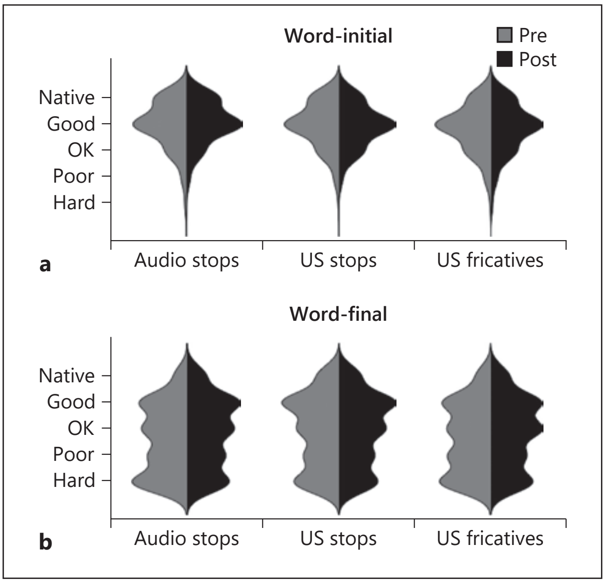

Fig. 3. Ratings of learners' productions of all palatalized consonants, by familiarization group, pre- and postfamiliarization production, word-initially (**a**) and word-finally (**b**).

a positive estimate therefore indicate improvement compared to this baseline case. The results of the CLMMs are shown in Table 8.

For word-initial productions, the model shows that there was no significant difference in the ratings across the prefamiliarization groups. There was also no significant improvement from pre- to postfamiliarization assessment. Postfamiliarization productions of the learners familiarized with stop stimuli using ultrasound feedback were rated significantly worse than postfamiliarization productions of the learners familiarized with stop stimuli using audio-only feedback. For word-final productions, the model shows that again there was no significant difference in the ratings across the prefamiliarization groups. There was a significant improvement in ratings of postfamiliarization productions compared to ratings of prefamiliarization productions and no significant interaction with group. That is, ratings improved overall from pre- to postfamiliarization assessment with no significant differences in that improvement based on familiarization group.

3.1.3 Generalization in Repetition

The results presented in the previous section include the ratings of all of the productions of all of the learners, regardless of what type of stimuli (stops or fricatives) a given learner was familiarized with. Recall that the stimuli used to familiarize the learners differed based on group. Some learners were familiarized using only stop stimuli and others with only fricatives (Table 3, 4). A given learner had been familiarized with 2 of the 6 consonants they produced. The other 4 consonants were not part of their familiarization. We can examine how well learners generalized how to produce palatalization by looking at the ratings of each of the consonants. However, adding consonant to the CLMMs presented

Table 8. Results of the 2 cumulative link mixed-effects models for ratings of learners' productions of palatalized consonants (one for word-initial productions and one for word-final), before and after familiarization ("session" pre, post) by familiarization group

Word position	Coefficients	Estimate	SE	*z*	Pr (> |*z*|)
Word-initial					
	Session: post	0.20736	0.12873	1.611	0.1072
	Group: US stops	–0.03424	0.36330	–0.094	0.9249
	Group: US fricatives	0.42884	0.36325	–1.181	0.2378
	Post × US stops	–0.36489	0.18153	–2.010	0.0444*
	Post × US fricatives	–0.17123	0.18157	–0.943	0.3457
Word-final					
	Session: post	0.38058	0.13878	2.742	0.0061*
	Group: US stops	0.44238	0.42406	1.043	0.2969
	Group: US fricatives	0.06387	0.42395	0.151	0.8803
	Post × US stops	–0.32160	0.19651	–1.637	0.1017
	Post × US fricatives	–0.09895	0.19629	–0.504	0.6142

US, ultrasound. An asterisk indicates a significant predictor ($\alpha = 0.05$).

in Table 8 would introduce a number of comparisons that would be nearly impossible to interpret, and would likely not have sufficient statistical power.

In order to get a sense of how well learners generalized the familiarization they received, we calculated changes in ratings within consonant. The categorical ratings were converted to integers, with "hard" being 1 and "native-like" being 5 ("other problem" ratings were excluded, as above). The mean rating within learner and rater for each consonant/word position/session combination was calculated, yielding the mean rating of 4 productions (2 repetitions of the stimulus spoken by the male Russian speaker and 2 of the female, minus any productions with ratings of "other"). The mean of the 4 prefamiliarization productions was then subtracted from the mean of the 4 postfamiliarization productions so that positive values indicate improvement and negative values indicate that productions got worse. This change in ratings assumes an equal distance between category values, but the validity of this assumption cannot be known and is probably inaccurate. Therefore, this measure is used to present a qualitative analysis of the data, so it is not appropriate to run statistical analyses on these changes in ratings. The results of the change in ratings are shown in Figure 4. Each bar represents the mean of the change in mean ratings across 96 ratings, and the error bar represents one standard error of the mean of the means. For present purposes, we interpret any mean change where the error bar does not encompass zero as indicating a meaningful change.

The changes in ratings for the word-initial productions within consonant were all small, indicating that the changes seen in Figure 3 were not driven by large differences across consonants. Figure 4a indicates that the ratings for the audio stop group improved for the consonants that were stimuli in their familiarization (/p^{j}/ and /t^{j}/), as well as those for /s^{j}/ and /l^{j}/, and that ratings for

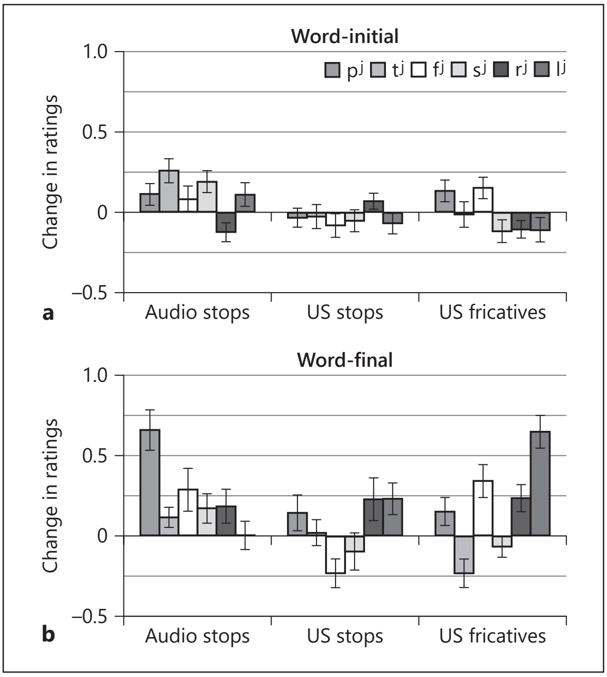

Fig. 4. Changes in ratings from pre- to postfamiliarization assessment by consonant within familiarization group. **a** Word-initial. **b** Word-final.

/rʲ/ got slightly worse. The ultrasound stop group showed a slight improvement in ratings for word-initial /rʲ/, but no changes for any other consonants. The ratings for the ultrasound fricative group improved for /pʲ/ as well as /tʲ/, but got slightly worse for /sʲ, rʲ, lʲ/.

The changes in ratings for word-final productions were greater and showed more variation across groups and consonants. The changes in ratings for the audio stop group show that ratings improved not only for the 2 consonants they had been familiarized with, but also for 3 of the 4 consonants with which they had not been familiarized. The ratings for /lʲ/ did not change. Ratings for the ultrasound stop group improved for /pʲ/ as well as for /rʲ, lʲ/ but got worse for /fʲ/. Ratings for the ultrasound fricative group improved for /fʲ/ (but not for /sʲ/) as well as for /pʲ, rʲ, lʲ/, and they got worse for /tʲ/. It is also worth noting that the change in ratings for /pʲ/ for the audio stop group was relatively large, as was the change for /lʲ/ for the ultrasound fricative group. Ratings for /lʲ/ also improved for the ultrasound stop group.

3.1.4 Discussion

The baseline ratings from the pre-familiarization repetition task examined how well completely naïve English-speaking learners were able to produce Russian palatalization with 6 different consonants across 2 word positions, when imitating recordings of native speakers. The primary finding of the baseline ratings was that there was a significant difference between word-initial and word-final consonants, with word-initial productions being significantly better than word-final ones. This finding is very much in line with the results reported by Hacking (2011). The productions of word-initial palatalized consonants by the learners in the present study were predominantly rated "good," rarely as "native-like," and almost never as "not palatalized" (Fig. 2). These ratings indicate that while learners may not have produced word-initial palatalization like a native speaker, they almost always produced some approximation of palatalization that made the utterance recognizably different from a nonpalatalized consonant. This is consistent with the results from Diehm (1998) and Bolaños (2017) showing that English speakers tend to produce onset C^{j}V sequences like CjV sequences, which while being close to producing a palatalized consonant, is not what Russians produce. Unlike word-initial productions, word-final productions were frequently rated as "not palatalized," indicating that the learners often either did not perceive the palatalization in the utterance they were repeating (see section 3.2 below) or did not know how to produce even an approximation of the palatalization word-finally. Ignoring productions that were rated "not palatalized," word-final productions were still rated worse than the word-initial productions. This suggests that learners were not sure how to produce word-final palatalization, even when they perceived that what they were repeating was not a nonpalatalized consonant. This difference based on word position is potentially due to the fact that while English does have word-initial CjV sequences, which are to some degree comparable to Russian C^{j}V, English lacks word-final VCj (and VC^j) sequences.

There were some differences in prefamiliarization ratings based on consonant, which were themselves dependent on word position. Word-initial /p^j/ was one of the 2 consonants that was rated significantly higher than any others: specifically, it was rated higher than the palatalized liquids /r^j, l^j/. In stark contrast, word-final /p^j/ was rated significantly worse than /t^j, s^j, f^j/, but was rated no differently from /r^j, l^j/. Hacking (2011) suggests that the successful production of word-final palatalized consonants may be related to sonority, with more sonorous consonants being easier to produce. This explanation is not sufficient for the present results, since /t^j/ (the other stop in the data) was rated no better or worse than any other consonant word-initially, but word-final /t^j/ was rated better than all other word-final consonants. One possibility is that the L2 learners may have perceived word-final /p^j/ as /p/ and simply produced a nonpalatalized /p/, but perceived the difference between word-final /t^j/ and /t/, and differentiated them in their productions. This explanation is plausible, since Kochetov (2004) found that even native Russian listeners identified word-final /p^j/ as /p/ (as produced by native Russian talkers) significantly more frequently than they identified as word-final /t^j/ as /t/. Another possibility is that if (some) learners did perceive a difference between word-final /p^j/ and /p/ as well as word-final /t^j/ and /t/, their strategies for replicating /p^j/ may have sounded less like palatalization than the strategies for replicating /t^j/. For example, word-final /p^j/ and /t^j/ are both characterized by longer releases than their nonpalatalized counterparts (Kochetov, 2002), so some learners may have produced them as /p^h/ and /t^h/. An extended release of word-final /t^h/ would result in some period of turbulent airflow above the tongue following the release, which may have been rated as plausibly being attributable to palatalization by the raters. No such turbulent airflow above the tongue would be expected after the release of a word-final /pʰ/, so the consonant would possibly be more likely to be rated as not palatalized. This latter explanation is also consistent with the fact that word-final /tʲ/ must have been produced with some acoustic properties that caused raters to not just rarely rate it as nonpalatalized ("hard") /t/, but in fact most often rate it as "good."

The other consonant that was rated significantly differently from other consonants was /sʲ/, which was the other of the 2 word-initial consonants (the other being /pʲ/) rated better than any other (in this case, better than /rʲ/). Word-final /sʲ/ was rated significantly better than all consonants other than /tʲ/, which was rated significantly better than /sʲ/. According to Kochetov (2017), the primary acoustic differences between word-initial /sʲ/ and /s/ were the formant transitions (F1 and F2) from the fricative to the following vowel, but durational differences were not significant. The acoustic differences for word-final /sʲ/ and /s/ in Russian were comparable to the word-initial differences. These acoustic differences must have been sufficiently salient to the learners that they were able to both perceive differences between word-final /sʲ/ and /s/, and reliably approximate some aspect(s) of them in production. Assuming that a salient aspect of word-final /sʲ/ was the transition of F1 and F2 from the /a/ into the fricative, and that what the learners produced was some approximation of that, it is unclear why they were perceived and produced better for /sʲ/ than for /pʲ, fʲ, rʲ, lʲ/.

The results from the present study are seemingly at odds with those obtained by Hacking (2011), especially for word-final productions. Hacking (2011) reports that in a forced-choice task native Russian listeners were unable to correctly identify as palatalized word-final /pʲ, sʲ, rʲ, tʲ/ produced by native-English L2 Russian speakers, but did so more successfully for word-final /lʲ, nʲ/ (though still below 30% accuracy). In the present study, word-final /sʲ, tʲ/ were rated as reasonably good, and word-final /lʲ/ was rated as poor. It is important to point out that the task for the Russian listeners in the Hacking (2011) study was to identify whether each production was palatalized, whereas the raters in the present study heard only palatalized productions and were asked to rate how good they were. It is not known what particular acoustic aspects of the productions the raters in the present study attuned to in order to determine their ratings, but they did not have to decide whether the target consonants they heard were supposed to be palatalized. It is possible that the same raters would have had difficulty making that determination given the productions from the same learners in a task similar to that used by Hacking (2011). Nevertheless, to the degree that the studies can be compared, the results from the present study do not support the notion that goodness of palatalization word-finally correlates positively with sonority. In the present study, sonority showed the opposite effect on ratings within coronals: the obstruents /tʲ, sʲ/ were rated relatively highly, while the sonorants /rʲ, lʲ/ were rated very poorly. However, the labial obstruents did not pattern with the coronal obstruents but rather with the coronal sonorants. If sonority is a factor influencing how well learners produce palatalization, then it seems to depend on primary oral articulator. Regardless of the differences in the studies, the results from both Hacking (2011) and Hacking et al. (2016) show that the difficulties in producing word-final palatalized consonants experienced by the present learners – who were completely untrained in palatalization before the study – persist even after years of studying Russian. Further investigation will hopefully provide a better understanding of the source of these difficulties.

There were no significant differences in the prefamiliarization ratings between the familiarization groups; no differences were predicted since learners were randomly assigned to groups. As far as the effects of familiarization on the change in ratings of all productions from pre- to postfamiliarization assessment are concerned, there were no significant improvements in ratings for word-initial palatalized consonants. This may be due to the fact that word-initial ratings started off comparatively high (especially when compared to word-final ratings), so there was less room for improvement with the word-initial productions. The ratings of the word-initial productions from the group that was familiarized with ultrasound using stop stimuli was significantly worse after familiarization than the postfamiliarization productions from the group that was familiarized with audio-only using stop stimuli. However, a CLMM with the same random effects as above fit to only the data from the ultrasound stop familiarization group showed that there was no significant difference in ratings for this group from pre- to postfamiliarization production (estimate = –0.1586, z = –1.402, Pr[>|z|] = 0.161, with prefamiliarization production as the reference level). Ratings of word-final productions improved significantly from pre- to postfamiliarization assessment, with no significant differences based on familiarization group.

The additional details provided by the changes in ratings by consonant indicated that the overall significant improvement for the word-final productions was attributable to a variety of changes in ratings based on familiarization group and consonant. Ratings of productions of learners from all 3 familiarization groups improved for at least 1 of the 2 consonants with which they had been familiarized, though only the audio stop group improved on both. Learners from all groups also showed improvement in production for consonants with which they had not been familiarized. As mentioned above, the change in ratings for productions of consonants that started out relatively well rated were small, as there was little room for improvement. It is therefore particularly interesting to look at word-final /p^j/ and /r^j/. These 2 consonants (along with /l^j/) were rated significantly worse than the other consonants before familiarization, and ratings of the productions of these 2 consonants improved for all 3 groups regardless of whether that consonant was used in familiarization, indicating that both familiarization techniques were beneficial and that all groups were able to generalize what they had learned in familiarization to novel consonants. That said, there were differences across the groups. The audio stop group showed the most consistent improvement in ratings for word-final productions, with all consonants except /l^j/ showing an improvement in ratings. The groups familiarized using ultrasound showed more mixed results, improving on one familiarized consonant but not both. Both ultrasound groups also receiving slightly worse ratings for one consonant each, though in both cases this was for a consonant that they had not been familiarized with (/f^j/ for the stop group and /t^j/ for the fricative group). The case of word-final /l^j/ is interesting. As noted by Hacking (2011), word-final /l^j/ is particularly difficult for English-speaking learners of Russian to produce. Word-final /l^j/ was rated as very poor before familiarization in the present study, and it was not used in the stimuli for any of the familiarization groups. Only the 2 ultrasound groups showed improvement in ratings for this consonant after familiarization, suggesting that the ultrasound imaging may have been especially helpful for this very difficult case.

3.2 AX Discrimination

Each participant performed the AX discrimination task 2 times (tasks 1 and 5 in Table 3), with 144 trials each time for a total of 288 trials. The 18 learners thus yielded 5,184 same/different judgments. Any trial on which the response time was >3 s was discarded, assuming the learner was inattentive on that trial, resulting in 221 trials being discarded (4.3% of the data). Performance on the AX discrimination tasks was measured using *d*' (Macmillan and Creelman, 2004), calculated using the correction method for zero values of Hautus (1995). Higher *d*' scores indicate better performance on the discrimination task. Unlike percentage correct, *d*' addresses the response bias of each learner by taking into account correct and incorrect responses on both same and different trials. A *d*' value cannot therefore be calculated for each trial, but rather is calculated across sets of trials that must include trials on which the learner was supposed to respond "same" and trials on which the learner was supposed to respond "different." The calculation of a single *d*' value therefore requires an absolute minimum of 4 trials, but is more reliable with more trials. The specific sets of trials from which *d*' values were calculated are detailed in the following subsections. *d*' values were calculated separately for prefamiliarization responses and for postfamiliarization responses. Recall that the AX discrimination task was designed not only to establish how well learners discriminate the contrast between palatalized and nonpalatalized consonants before familiarization, but also whether the presence of palatalization affects learners' ability to discriminate the manner contrast between stops and fricatives, a contrast that exists in the learners' L1. The AX discrimination analyses and results were split into 2 sets, one in which we examined how well learners discriminated pairs differing only in palatalization, and another in which we examined how well learners discriminated pairs differing only in manner. We present baseline discrimination of these contrasts first, and then examine the effects of familiarization on changes in discrimination for both the palatalization and the manner contrasts.

3.2.1 Baseline Discrimination

The number of trials required for the calculation of *d*' precluded calculating *d*' for individual consonant pairs in the present data set, so *d*' for palatalization discrimination was calculated within all consonants sharing manner. Each learner's performance in discriminating the palatalization contrast was assessed by calculating within-learner *d*' values for all of the trials from the prefamiliarization AX task that were the same (corresponding to the examples in rows 1 and 2 of Table 5) plus those trials on which the stimuli mismatched on palatalization (e.g., /tam/~/t^{j}am/, /mat^j/~/mat/, corresponding to the examples in rows 3 and 4 of Table 5) within word position and manner.

Figure 5 shows boxplots of the distributions of the *d*' values across all participants by manner within word position. A linear mixed-effects model (LME) was created using the *lme4* package (Bates et al., 2015) for R with *d*' as the pre-

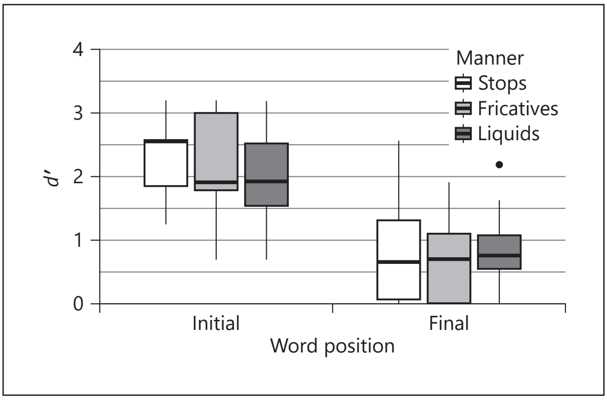

Fig. 5. Prefamiliarization AX discrimination (measured by *d*') of the palatalization contrast (e.g., /p^{j}am/ vs. /pam/, /map^j/ vs. /map/), by manner within word position for all familiarization groups.

Table 9. Results of the linear mixed-effects model for *d*' for prefamiliarization discrimination of the palatalization contrast

Fixed effects	Estimate	SE	*t*
Intercept	*2.36687*	*0.15994*	*14.799*
WP: final	–1.56797	0.18177	–8.626*
Manner: fricatives	–0.17505	0.18177	–0.963
Manner: liquids	–0.46117	0.18177	–2.537*
WP final: fricatives	0.06899	0.25706	0.268
WP final: liquids	0.46041	0.25706	1.791

WP, word position. An asterisk indicates a significant predictor (|*t*| >2).

dicted variable, participant as a random effect, and with word position, manner, and the interaction between word position and manner as fixed effects. The results of the model are shown in Table 9. The intercept represents the *d*' for word-initial stops (chosen arbitrarily). Significant effects were determined as those having |*t*| >2 (Gelman and Hill, 2007, p. 42). Results of the model show that the effect of word position was significant, with discrimination of palatalization better word-initially than word-finally, and that discrimination of palatalization of word-initial liquids was worse than word-initial stops. No other effects were significant.

In order to assess the discrimination of manner contrasts in the presence or absence of palatalization, we calculated within-learner *d*' based on the "different" trials on which the A and X differed only in manner, and their corresponding "same" trials. That is, discrimination of manner contrasts when both A and X were not palatalized but differed in manner was assessed by calculating *d*' based on the trials corresponding to the examples in row 5 of Table 5 (e.g., /tam/~/sam/) plus the trials corresponding to the examples in row 1 of Table 5 (e.g., /tam/~/tam/). Discrimination of manner contrasts when both A and X were palatalized but differed in manner was assessed by calculating *d*' based on the trials corresponding to the examples in row 6 of Table 5 (e.g., /t^{j}am/~/s^{j}am/)

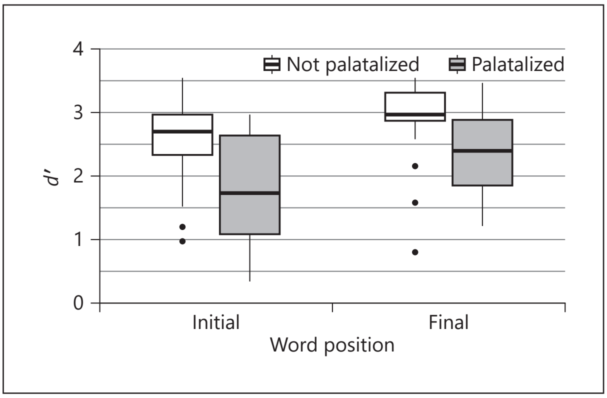

Fig. 6. Prefamiliarization AX discrimination (measured by *d*') of manner contrasts within word position for all familiarization groups. "Not palatalized" indicates discrimination of manner when neither A nor X was palatalized (e.g., /tam/ vs. /sam/). "Palatalized" indicates discrimination of manner when both A and X were palatalized (e.g., /t^{j}am/ vs. /s^{j}am/).

Table 10. Results of the linear mixed-effects model for *d*' for prefamiliarization discrimination of the manner contrasts

Coefficients	Estimate	SE	*t*
Intercept	*1.8443*	*0.1729*	*10.669*
WP final	0.5412	0.1669	3.242*
Nonpalatalized	0.6905	0.1669	4.137*
WP final: nonpalatalized	–0.2344	0.2361	–0.993

WP, word position. An asterisk indicates a significant predictor (|*t*| >2).

plus the trials corresponding to the examples in row 2 of Table 5 (e.g., /t^{j}am/~ /t^{j}am/). All *d*' values were calculated separately within word position.

Figure 6 shows the *d*' values for manner discrimination across all participants within word position, grouped based on whether the stimuli in the pair were palatalized. An LME model was created with *d*' as the dependent variable, participant as a random effect, and with word position, pair palatalization, and the interaction between word position and pair palatalization as fixed effects. The results of the model are shown in Table 10. The intercept represents the *d*' for word-initial palatalized pairs, which had the lowest *d*' values. Results of the model show that the effect of word position was significant, showing that discrimination of manner for palatalized pairs was better word-finally than word-initially. The effect of palatalization was also significant, showing that discrimination of manner for nonpalatalized pairs was better than for palatalized pairs word-initially. The interaction between word position and palatalization was not significant.

3.2.2 Effects of Familiarization on Discrimination

Since there was a significant effect of word position on the discrimination of both the palatalization and manner contrasts, the analyses of the effects of

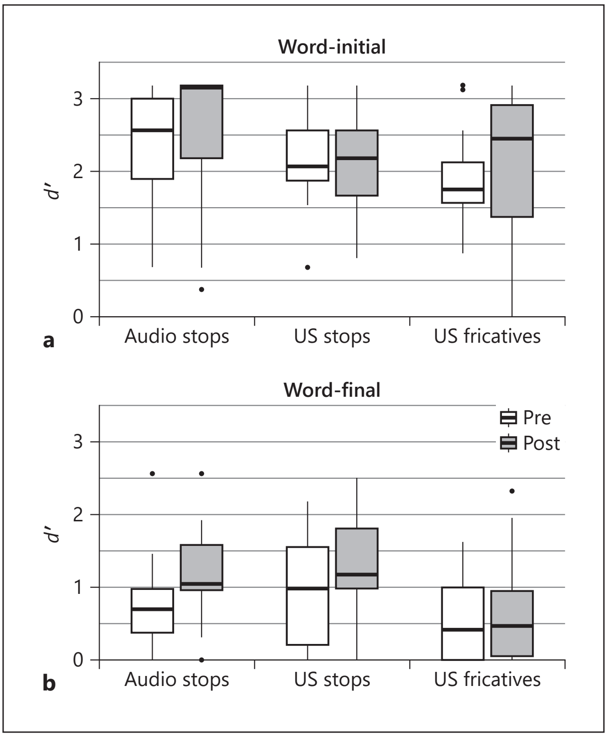

Fig. 7. Change in AX discrimination (measured by *d'*) of the palatalization contrast pre- and postfamiliarization assessment, within familiarization group, word-initially (**a**) and word-finally (**b**).

familiarization on discrimination were also conducted separately within word position, as with the analyses for the effects of familiarization on ratings in repetition above (an LME identical to the one presented in Table 10 but including only postfamiliarization *d'* values confirmed that the word position effect remained after familiarization as well, $|t| = -5.436$). The *d'* values for discriminating the palatalization contrast before and after familiarization within familiarization group are shown in Figure 7, with Figure 7a showing the *d'* values for word-initial discrimination and word-final shown in Figure 7b.

Two LME models were fit to the AX discrimination data to determine the significance of familiarization within familiarization group, one model for word-initial discrimination and one for word-final. Each model had *d'* as the predicted value, and fixed effects of session (pre- or postfamiliarization), familiarization group, and the interaction between the two. Learner was included as a random effect as were random slopes for session by learner. The intercept was the *d'* for the audio stop group before familiarization. The results of the models are presented in Table 11.

As expected (and as with the repetition task reported above in section 3.1.2), there were no significant differences based on familiarization group before familiarization in either word position. The model for word-initial discrimination indicates that, numerically, *d'* improved from pre- to postfamiliarization production, but this difference was not significant. However, for word-final dis-

Table 11. Results of the linear mixed-effects models for effects of familiarization on the discrimination of the palatalization contrast within familiarization group, one model for word-initial and another for word-final

Word position	Coefficients	Estimate	Standard error	*t*
Word-initial	Intercept	*2.37794*	*0.21879*	*10.868*
	Session: post	0.16507	0.23772	0.694
	Group: US stops	–0.21191	0.30942	–0.685
	Group: US fricatives	–0.45752	0.30942	–1.479
	Post × US stops	–0.13000	0.33619	–0.387
	Post × US fricatives	–0.05838	0.33619	–0.174
Word-final	Intercept	*0.7464*	*0.2212*	*3.374*
	Session: post	0.4633	0.2155	2.150*
	Group: US stops	0.2172	0.3128	0.694
	Group: US fricatives	–0.1664	0.3128	–0.532
	Post × US stops	–0.1579	0.3048	–0.518
	Post × US fricatives	–0.3710	0.3048	–1.217

Post, after familiarization; US, ultrasound.

crimination, *d*' did improve significantly from pre- to postfamiliarization production. The interactions between session and familiarization group were not significant in either word position.

3.2.3 Generalization in Discrimination

In this section we explore 2 different ways in which familiarization with palatalization may have been generalized in discrimination by the learners to environments that were not part of their familiarization. The first was to examine how the *d*' values compared before and after familiarization in the discrimination of palatalization, looking at discrimination of palatalization for manners that were not part of their familiarization. The second was to examine how the presence of palatalization affected the discrimination of manner differences, since this was not the purpose of the familiarization, and to see whether and how the different familiarization types affected the discrimination of manner. Given the way *d*' is calculated, these analyses resulted in only 4 *d*' values per manner per learner (2 before and 2 after familiarization) for the within-manner discrimination of palatalization and 2 *d*' values per manner per learner (1 before and 1 after familiarization) for the discrimination of manner, which is insufficient data for statistical modeling. Therefore, as in the generalization section for repetition (3.1.3), qualitative analyses only were used to explore these questions of generalization.

Similar to the changes in ratings presented in the generalization analyses in section 3.1.3, we calculated the change in *d*' values from pre- to postfamiliarization assessment. The value Δ*d*' was calculated by taking the mean of the prefamiliarization *d*' values for the 2 consonants that shared manner, and subtracting it from the mean of the postfamiliarization *d*' values for the same 2 consonants,

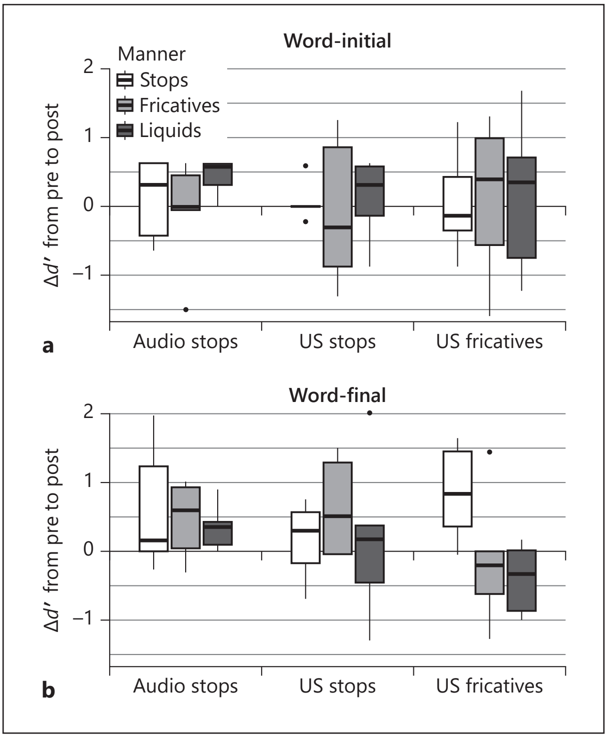

Fig. 8. Change in AX discrimination (measured by Δ*d*′) of the palatalization contrast before and after familiarization by manner, within familiarization group, word-initially (**a**) and word-finally (**b**).

for each learner. Figure 8 shows the mean Δ*d*′ values within each manner for each familiarization group, with the word-initial palatalization contrasts shown in Figure 8a and the word-final palatalization contrasts shown in Figure 8b (which provide by-manner detail for the differences in the grouped boxplots of Fig. 7a, b, respectively). Word-initially, the audio stops group and ultrasound fricatives group both improved in discriminating the manner that was used in their familiarization, while the ultrasound stops group showed no change. All 3 familiarization groups showed improvements in discrimination of the palatalization contrast in liquids, which were not part of the familiarization for any group. Word-finally, the audio stops group and ultrasound stops group both improved in discriminating stops, while the ultrasound fricatives group showed slightly worse discrimination of the palatalization contrast in word-final fricatives. However, each of the 3 groups had the largest improvement in discrimination for consonants of manners that were not part of their familiarization, resulting in the overall improvement in the discrimination of the word-final palatalization contrast found in the previous section.

As far as discriminating the manner contrast is concerned, we saw in section 3.2.1 that the discrimination of manner when neither consonant was palatalized and within word-final pairs regardless of palatalization was very good. Therefore we did not examine the discrimination of manner in these contexts further. However, the discrimination of manner was significantly hampered

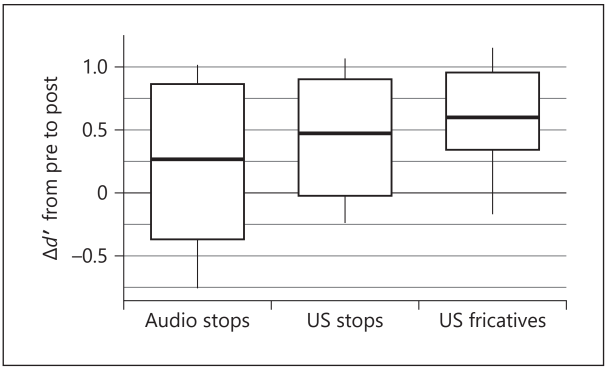

Fig. 9. Change in AX discrimination (measured by Δ*d*′) of the word-initial manner contrasts before and after familiarization for palatalized pairs, by familiarization group.

when both stimuli were palatalized and word-initial, so the effect of familiarization on manner discrimination in this context was analyzed. Figure 9 shows that discrimination improved for all 3 familiarization groups, and that the increases were numerically greater for the 2 ultrasound groups than for the audio group.

3.2.4 Discussion

The present results show that prefamiliarization discrimination of the Russian palatalization contrast by naïve English listeners is good when that contrast is presented in word-initial, prevocalic position. This is in line with a consistent set of findings in other studies (Diehm, 1998; Babel & Johnson, 2007; Kulikov, 2011; Rice, 2015; Bolaños, 2017). The discrimination of this nonnative contrast in this position was roughly comparable to the discrimination of word-initial manner for nonpalatalized pairs (compare Fig. 5, 6), which does occur in English. The present results also show that discrimination of the palatalization contrast is significantly worse word-/utterance-finally than word-/utterance-initially. This is also consistent with results found by other researchers for both nonnative listeners (Kochetov, 2004; Kulikov, 2011; Rice, 2015) as well as for Russian listeners (Kochetov, 2004). These effects of word position are not surprising. In general, acoustic information is more salient word-initially (see, e.g., Wright, 2004, for a summary), and the information for palatalization is no exception. The specific case of discriminating Russian palatalization in coda position may be additionally challenging since the palatalization gesture has been shown to have lesser magnitude and to be timed with the primary gesture differently in coda than in onset (Kochetov, 2002), and may therefore result in lessened acoustic information indicating its presence.

In addition to these language-independent acoustic considerations, the influence of the native English sound categories may also have been a factor in the learners' ability to discriminate the palatalization contrast. While English does not contrast word-initial consonants based on palatalization, it does contrast consonant-/j/ sequences with single labial consonants, for example., *pure* /pjuɹ/ vs. *poor* /puɹ/ and *food* /fud/ vs. *feud* /fjud/. While these contrasts are not equivalent to the Russian palatalization contrast, the palatalization contrast may be similar enough acoustically to these English contrasts to be perceived easily as different (cf. Flege, 1986; Best, 1995). However, word-final consonant-/j/ sequences do not exist in English, so in this word position this (potentially) similar category is not available. In terms of discriminating palatalization for liquids, our results are consistent to some degree with the finding of Rice (2015) in that they were discriminated worse than stops and fricatives; however, in our results this held only word-initially. The difficulty in the discrimination of word-initial /r/-/r^{j}/ was likely to be due both to the fact that English does not have a trilled /r/, and that it does not contrast word-initial /ɹ/-/ɹj/. The difficulty with discriminating /l/-/l^{j}/ may be due in part to the fact that English /l/ is sometimes realized as [l] and sometimes as [ɫ]. This velarized allophone is comparable to the Russian nonpalatalized /l/, but this allophony is predictable in English based on word position (Sproat and Fujimora, 1993). Differentiation among different variants of /l/ within a word position may therefore be additionally challenging for English speakers. In summary, word-final palatalization contrasts may have been more difficult because they were both in a less favorable position acoustically, and because the potentially helpful English category distinctions that exist word-initially do not exist word-finally.

As far as the effects of familiarization are concerned, familiarization with audio with or without ultrasound visualization resulted in improved discrimination of the palatalization contrast word-finally but not word-initially. Lack of improvement in the discrimination of word-initial palatalization is most likely due in large part to the fact that discrimination of word-initial palatalization was reasonably good before familiarization, so there was not much room for improvement. Although all familiarization groups improved in word-final discrimination, discrimination of this contrast word-finally was still significantly worse than word-initially. There was no significant difference in the improvement shown based on familiarization group. While the results shown in Figure 7b suggest that it may be more beneficial to use stop stimuli in familiarization, Figure 8b shows that the improvements for the stop groups (audio and ultrasound) were in fact driven more by improvements in discriminating fricatives than by stops. Conversely, the ultrasound fricative group did not improve much numerically in word-final discrimination (Fig. 7b), but Figure 8b shows that their discrimination of stops improved more than any other group/manner combination. The consistent theme for all 3 groups is that improvements in discrimination were not due to improvements in the type of stimuli they were exposed to in familiarization. All learners demonstrated that the improvements in discrimination involved some degree of generalization to other environments. It is unclear why this was the case but is an interesting question to explore in further study.

The results from the discrimination of manner show that the challenges presented by the Russian palatalization contrast to the learner were not limited to learning the palatalization contrast itself: the presence of palatalization also affected the learner's ability to discriminate manner. Before familiarization, learners discriminated manner well when neither stimulus was palatalized, which is not surprising given the contrasts between /p/-/f/, /t/-/s/, and /ɹ/-/l/ (although not /r/-/l/) all exist in English. However, our results show that manner discrimination was not as good word-initially as word-finally, even for nonpalatalized pairs. This is the opposite of what would be expected based on the generalization that acoustic information is more salient word-initially than word-finally. However, even under good listening conditions, Miller and Nicely (1955, p. 342, their Table VI) found that native English listeners misidentified manner more often than voicing or nasality in an identification task of English singleton consonants before /a/. The present results are therefore consistent with this earlier finding.

A novel finding of the present study is that when both the A and X stimuli were palatalized, the presence of palatalization negatively impacted learners' ability to discriminate manner in general in the prefamiliarization task. Learners' performance on the manner discrimination was worse compared to when neither was palatalized, with the effect of word-initial discrimination (e.g., /t^{j}am/ vs. /s^{j}am/) being worse than word-final (e.g., /mat^j/ vs. /mas^j/). Therefore, learners had the most difficulty discriminating manner contrasts word-initially when both stimuli were palatalized. Further examination of the changes in discrimination of this most challenging environment suggests that learners who were familiarized with palatalization using ultrasound (regardless of whether they were familiarized with stop or fricative stimuli) improved in discriminating this contrast better after familiarization more than the audio group did. One possible explanation for this difference could be that the articulation required for palatalization results in multiple acoustic consequences that vary depending on the consonant, environment, and interaction between the two. For stops, formant transitions before postvocalic and after prevocalic palatalized stops have a lower first formant (F1) and higher second formant (F2) than their nonpalatalized counterparts (Halle, 1971; Bolla, 1981; Kochetov, 2002). The spectral properties of the release burst for stops also differ based on palatalization (Halle, 1971; Bolla, 1981; Iskarous and Kavitskaya, 2018). In addition, nonpalatalized voiceless stops in Russian have short-lag voice onset time (Ringen and Kulikov, 2012), while palatalized voiceless stops have a prolonged period of aperiodic energy concentrated in higher frequencies before the onset of phonation associated with the vowel (Kochetov, 2002). For fricatives, there are differences in F1 and F2 that are similar to those found with stops. While there are spectral differences due to palatalization for labial voiceless fricatives (/f/ vs. /f^j/), the differences for the coronal voiceless fricatives (/s/ vs. /s^j/) are minimal (Bolla, 1981; Iskarous and Kavitskaya, 2018). As for the liquids, the palatalized trilled /r^j/ has fewer vibratory contacts of the tongue tip with the palate than its nonpalatalized counterpart /r/ (Iskarous and Kavitskaya, 2010), and the formant structures of the laterals /l/ vs. /l^j/ are distinct (Bolla, 1981; Iskarous and Kavitskaya, 2018). When the focus of familiarization was on the acoustical differences between C and C^j pairs, the learner may grasp that there are many things to which they need to attune but may not have a cohesive idea of what those are indicating. This more fractured attention to various acoustic properties might be sufficiently distracting that they might not attune to the relevant differentiating properties of manner. For example, if presented with the pair /t^{j}am/-/s^{j}am/, the /t^{j}am/ stimulus will start with a relatively short release burst followed by a periodic of frication that is much longer than the release burst. If the learner is trying to attend to multiple potential acoustic aspects of palatalization, he or she could miss the relatively short release burst of an initial /t^j/ and perceive it as /s^j/ due to the extended frication associated with /t^j/.

However, secondary palatalization can be characterized relatively straightforwardly articulatorily as an approximation of the tongue body toward the palate concurrently with the primary oral articulations (Ladefoged and Maddieson, 1996, pp. 363–365). Even though the nature of the articulation associated with palatalization was explained to all leaners regardless of familiarization group, only the learners who were familiarized with ultrasound imaging saw the productions of native speakers as well as their own lingual articulation. If this articulatory familiarization via ultrasound imaging guided learners toward a more encompassing goal of detecting the articulatory movement corresponding to palatalization based on whatever acoustic evidence was available, then they may have been less distracted by any one particular acoustic goal in isolation and therefore less likely to miss other acoustic manifestations of manner.

4 General Discussion and Conclusion

In the present study, naïve learners of the Russian palatalization contrast performed an AX discrimination task and a repetition task. The learners were then familiarized with the palatalization contrast, with one group of learners having access to real-time ultrasound imaging of the vocal tract during familiarization, and another group having access to audio materials only. After familiarization, the learners again completed the repetition and discrimination tasks. The ratings of the productions of the learners in the prefamiliarization repetition task by Russian speakers were significantly worse word-finally than word-initially. The results of the discrimination task were similar, in that learners were significantly worse at perceiving the contrast between palatalized and nonpalatalized pairs word-finally compared to word-initially. The relatively poor ratings of word-final productions of palatalized consonants were therefore likely due at least in part to learners often not perceiving the palatalization contrast in this word position. For word-initial consonants, there was no significant improvement from pre- to postfamiliarization assessment in the ratings of the productions of palatalized consonants, or in the discrimination of the palatalization contrast for word-initial consonants. This was most likely due to there being little room for improvement in both tasks, since the prefamiliarization performance in both tasks was reasonably good. There were, however, significant improvements for word-final consonants from pre- to postfamiliarization task in both production and perception of palatalization.

Our first goal was to assess whether access to ultrasound imaging during familiarization would be more effective than familiarization with audio stimuli only, looking at performance in both perception and discrimination. The results showed that the ratings of word-final productions by all learners improved after familiarization, but that there was no significant difference between learners who had access to ultrasound imaging and those who did not. Nevertheless, there were some differences in the details of the ratings across the groups that suggest that the ultrasound imaging was uniquely helpful in some ways. Specifically, production of word-final /l^j/ is known to be particularly challenging for English learners of Russian (Hacking, 2011), and the learners in this study were no exception. Baseline productions of word-final /l^j/ were rated poor. Ratings of the productions of word-final /l^j/ improved only for learners who had access to ultrasound in familiarization, suggesting that this additional articulatory information may have been especially useful in a particularly challenging specific case.

As with production, the discrimination of the word-final palatalization contrast improved for all learners after familiarization, but again there was no significant difference between the familiarization groups. It is also worth noting that the results from the present study were not what one might expect given the results from Baese-Berk (2010) and Baese-Berk and Samuel (2015). The primary task in their experiments was perceptual training, aimed at the discrimination of nonnative L2 contrasts. They found that discrimination training was disrupted by having learners produce nonnative speech sounds during training, to the point where learners who produced these sounds were unable to reliably discriminate the target contrast, while learners who did not have to produce the sounds were. In contrast, the primary task for all of the naïve learners in the present study was production familiarization (with or without ultrasound) of the nonnative contrast of Russian palatalization. Even with this focus on production, performance of learners on the discrimination of the contrast they were producing did improve from pre- to postfamiliarization assessment. The results from the present study do not, however, inherently contradict the findings of Baese-Berk (2010) and Baese-Berk and Samuel (2015). A key difference between the 2 experimental tasks is that it is possible to have learners focus on perception without any requirement to produce speech overtly, but the reverse is not true. Given a sufficiently difficult nonnative contrast, it may well be that preventing learners from focusing all of their attention exclusively on discrimination by adding the inherent demands of production to the discrimination task makes the learner's task difficult enough that they are unable to discriminate the contrast. However, it is virtually impossible to focus on production without any involvement of perception. Indeed, in the Baese-Berk and Samuel (2015) experiments, the participants heard their own, presumably inaccurate, productions during the course of their discrimination task. In order to produce the sounds with which they are being familiarized, learners must of course first perceive them. Any improvement in production as a result of familiarization must be contingent in no small part to successful perception, even if that perception is not native-like. It is not surprising then that improvement in discrimination was found in the present study, as this improvement likely reflects this requisite – although likely imperfect – perceptual attuning to the nonnative contrast. There were also other material differences between the studies, including the tasks, experimental paradigms, and specific contrasts used, each of which complicates direct comparison between the studies. The differences between these studies underscore the complexity of the interactions between speech perception and production, especially in L2, and highlight the importance of considering these myriad factors when comparing results.

Another goal of the present experiment was to investigate whether learners would be able to generalize what they had learned in production familiarization in new environments, again in both production and discrimination. There was good evidence in the present results that learners were able to generalize reasonably well. The stimuli used in the familiarization for each group were restricted to consonants of the same manner, either only stops or only fricatives.

An analysis of the changes in ratings of the individual consonants showed that the ratings of learners' productions for all familiarization groups included improvements in ratings for consonants that were not part of their familiarization. Learners did generalize in this way, and this was the case regardless of whether the learners had access to ultrasound imaging during familiarization. A similar investigation of effects of familiarization on discrimination showed even stronger evidence of generalization than in production. The improvements in discrimination from pre- to postfamiliarization production were driven for all groups by improved discrimination of the palatalization contrast for consonants having a manner that was not part of their familiarization stimuli. Further evidence of the learners' ability to generalize came from the improvement in discrimination of the manner contrast in word-initial pairs where both consonants were palatalized (e.g., /t^{j}am/~/s^{j}am/). Prefamiliarization discrimination in such pairs showed that the presence of palatalization had a negative impact on discriminating manner contrasts, but in the otherwise easier word-initial position. Learners familiarized using ultrasound improved in discriminating word-initial manner contrasts within palatalized consonants, while the learners familiarized with audio stimuli showed the least improvement in this discrimination.

Although the effects of using ultrasound imaging of the vocal tract during production familiarization that we found were not significantly different from those who were familiarized with audio stimuli only, the results highlight the usefulness of ultrasound in familiarizing learners with this type of articulatory information. Recall from section 2.3.3 that the audio familiarization group did receive a detailed explanation of the articulation involved in palatalization, including short ultrasound videos of native speakers producing palatalized and nonpalatalized pairs. Therefore, even this control group did have a brief exposure to visual articulatory information from ultrasound. This visual information may have helped them, even without seeing further videos of more consonants from native speakers or their own articulations. The duration of the familiarization was very short compared with other studies in which learners have been trained on novel nonnative segments. Familiarization lasted 15 min or less in the present study, whereas training times have been much longer in other studies, for example, 1 h per segment in the study by Kartushina et al. (2015) and 4 h in the study by Saito (2013). It is reasonable to expect that additional improvements could be achieved with longer familiarization sessions, which would be the most straightforward extension of the existing experimental design.

Perhaps the most relevant study with which the present results can be compared is Hacking et al. (2017), in which EPG was used to train learners to produce Russian palatalized consonants. The protocol used by Hacking et al. (2017) involved 8 relatively short weekly training sessions of about 15 min each, during which the learners received no corrective feedback. The learners in the study of Hacking et al. (2017) did show improvement in certain acoustic measures (F2 transitions into the palatalized consonant), though these improvements did not translate into improved identification of these learners' utterances as palatalized by Russian listeners. The duration of the training in the present study was roughly the same as a single session from Hacking et al. (2017), and like that study, the learners in the present study did not receive any corrective feedback during familiarization. The most striking difference between the present results from those of Hacking et al. (2017) was that ratings of the learners in the present experiment did improve after familiarization. While it may be the case that ultrasound is a more appropriate tool than EPG for visualizing the relevant lingual articulation that is involved in palatalization (as mentioned in section 1.4), ultrasound imaging alone cannot account for the fact that ratings improved in the present study but identification did not in Hacking et al. (2017), since the improvements held across familiarization groups in the present study. Another material difference between the studies is that the participants in Hacking et al. (2017) were reasonably advanced students of Russian, while the learners in the present study were completely naïve. It may have been a more difficult task to get experienced students from Hacking et al. (2017) to change ingrained habits of producing palatalized consonants, than to instruct naïve learners who had no such ingrained habits. It could also be that Hacking et al. (2017) did not have enough Russian speakers evaluate the learners' productions for the identification results to reach significance. Schmid and Hopp (2014) recommend at least 10 raters for determining "foreign accentedness." While the task used by Hacking et al. (2017) was much more targeted than generic accent rating (i.e., identification), they employed only 3 listeners and found moderate but nonsignificant changes. Including more raters may have yielded different results. As we point out in section 3.1.4, another important difference between the studies is that the present study required the Russian-speaking listeners to focus specifically on evaluating the goodness of the palatalization; they did not have to determine whether the productions themselves were palatalized. The results from the 2 studies are therefore not directly comparable, since it is not known what the rate of identification of the productions from the present data would have been. Another material difference between the studies is that the learners in the study of Hacking et al. (2017) were trained on word- and sentence-final /p/-/p^{j}/, /t/-/t^{j}/, /s/-/s^{j}/, /n/-/n^{j}/, /l/-/l^{j}/, and /r/-/r^{j}/, whereas the present familiarization stimuli contained only 2 consonants that shared manner per learner, but in both word-initial and word-final position (and from 2 talkers). Even taking the results from Hacking et al. (2017) at face value, and inasmuch as the identification results are comparable with the present ratings, another possibility that could explain the difference across the 2 studies is that for teaching a class of sounds like palatalization (as opposed to a single novel L2 speech sound), it may be more effective to expose learners to a smaller set of stimuli in different contexts (prosodic positions and talkers) than to a larger set in one context. This possibility is supported by the observation that the learners in the present study showed a reasonably good ability to generalize what they had been exposed to in familiarization to novel environments in both production and discrimination. If this assessment is accurate, then there could be important ramifications from a pedagogical standpoint in terms of training nonnative classes of speech sound.

Lastly, the learners did not receive any correctional feedback during familiarization, which has shown to improve the efficacy of training learners to discriminate (Lee & Lyster, 2016) and produce (Saito & Lyster, 2012) nonnative speech sounds (though see Maas et al., 2008, for more detailed discussion of the relative benefits of corrective feedback). Despite the challenges of the task, learners in the present study who were familiarized using ultrasound showed gains in production that were comparable to the audio-only group. In addition, familiarization with ultrasound resulted in improved discrimination of manner contrasts in the presence of palatalization, whereas familiarization with audio stimuli only did not. The fact that the improvements in the present study were found without any corrective feedback is of practical pedagogical value, since the type of familiarization used does not require the active involvement of an instructor. The potential pedagogical benefits of this approach notwithstanding, it would also be useful to investigate the efficacy of the familiarization with ultrasound in conjunction with corrective feedback.

In summary, familiarizing naïve learners with the production of a class of nonnative speech sounds by providing real-time ultrasound imaging of the vocal tract was shown to result in some improvement to those learners' ability to both produce and discriminate those sounds. This improvement was comparable to the improvement attained by another group of learners who were familiarized without ultrasound imaging, but the improvements shown by the 2 groups differed in where and to what degree the 2 methods were effective, and neither method was detrimental to production or discrimination. We conclude that ultrasound can be a useful complement to other methods of providing training in the acquisition of nonnative speech sounds, and further study of its efficacy in articulatory training is warranted and promising.

Acknowledgment

We are grateful to Ann R. Bradlow for valuable discussion, as well as to the audiences at the conferences Formal Approaches to Slavic Linguistics 26 (in Urbana-Champaign, IL, USA) and Ultrafest VIII (in Potsdam, Germany), and to the members of the Language, Learning, and Cognition Lab at Royal Holloway University of London for valuable feedback on earlier presentations of these data. We thank 3 anonymous reviewers and editor Alexei Kochetov for further suggestions.

Disclosure Statement

The authors have no conflicts of interest to declare.

Funding Sources

The authors gratefully acknowledge support from NIH Grant DC-002717 to Haskins Laboratories and the City University of New York.

Author Contributions

K.D.R. designed and conducted the experiments in the present study and was primary author of the manuscript. J.K. was involved in most of the programming aspects of the present study, including the software that was used to control the familiarization experiments, and several of the quantitative analyses. D.H.W. conceived of the fundamental idea of using ultrasound imaging to aid in the production of unfamiliar foreign-language sounds. He oversaw all aspects of the experimental design, the quantification of the data, the analyses of the results, and the writing of the manuscript.

References

Antolík, T. K., Pillot-Loiseau, C., & Kamiyama, T. (2013). Comparing the effect of pronunciation and ultrasound trainings to pronunciation training only for the improvement of the production of the French /y/-/u/ contrast by four Japanese learners of French. Presented at *Ultrafest VI Workshop*, Edinburgh.

Avanesov, R. I. (1974). *Russkaja literaturnaja i dialektnaja fonetika* [Russian literary and dialectal phonetics]. Moscow, Russia: Prosveshchenije.

Babel, M., & Johnson, K. (2007). Cross-linguistic differences in the perception of palatalization. In J. Trouvain & W. J. Barry (Eds.), Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI) (pp. 749–752). Saarbrücken, Germany: University of Saarland.

Baese-Berk, M.M. (2010). An examination of the relationship between speech perception and production (PhD. thesis), Northwestern University, IL.

Baese-Berk, M.M., & Samuel, A.G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language, 89*, 23–36.

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

Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research (pp. 207–232). Timonium, MD: York Press.

Bent, T. (2005). *Perception and production of non-native prosodic categories* (PhD. thesis), Northwestern University, IL.

Best, C.T. (1995). A direct-realist perspective on cross-language perception. In W. Strange (Ed.), Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research (pp. 167–200). Timonium, MD: York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustic Society of America, 109*(2), pp. 775–794.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn, & M. J. Munro (Eds.), Language experience in second language speech learning: In honor of James Emil Flege (pp. 13–34). Amsterdam, the Netherlands: Benjamins.

Bliss, H., Abel, J., & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction. Journal of Secondondary Language Pronunciation, 4(1), pp. 127–151.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* [computer program], 6.0.21. Retrieved from http://www.praat.org

Bolaños, L. (2017). Perception and production of timing in non-native speech: Russian palatalization (PhD thesis), Yale University, CT.

Bolla, K. (1981). *A conspectus of Russian speech sounds*. Budapest, Hungary: Akadémiai Kiado.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. G. Hansen Edwards, & M. L. Zampini (Eds), *Phonology and second language acquisition* (pp. 287–308). Amsterdam, the Netherlands: Benjamins.

Bradlow, A. R, Pisoni, D. B, Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/. IV. Some effects of perceptual learning on speech production. *Journal of the Acoustic Society of America, 101*(4), 2299–2310.

Catford, J. C., & Pisoni, D.B. (1970). Auditory vs. articulatory training in exotic sounds. *Modern Language Journal, 54*, 477–481.

Christensen, R. H. B. (2019). *Ordinal – Regression models for Ordinal* [computer program], R package version 2019.4-25. Retrieved from: http://www.cran.r-project.org/package=ordinal/

Cleland, J., McCron, C., & Scobbie, J. M. (2013). Tongue reading: Comparing the interpretation of visual information from inside the mouth, from electropalatographic and ultrasound displays of speech sounds. *Clinical Linguistics and Phonetics, 27*(4), 299–311.

Cleland, J., Scobbie, J. M., & Wrench, A.A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics and Phonetics, 29*(8-10):575–597.

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect, 18*, 53–70.

Davidson, L., & Shaw, J.A. (2012). Sources of illusion in consonant cluster perception. Journal of Phonetics, 40(2), 234–248.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379–397.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393–410.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*, 1–17.

Diehm, E. E. (1998). Gestures and linguistic function in learning Russian: Production and perception studies of Russian palatalized consonants (PhD thesis), Ohio State University, OH.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics, 40*(2), 280–288.

Evans-Romaine, D. K. (1998). *Palatalization and coarticulation in Russian* (PhD thesis), University of Michigan, MI.

Fabra, L..R., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics, 40*(3), 491–508.

Flege, J. E. (1986). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders*. Norwood, NJ: Ablex Publishing.

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research. Baltimore, MD: York Press.

Flege, J. E. (1999). The relation between L2 production and perception. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), Proceedings of the XIVth International Congress of Phonetics Sciences (pp. 1273–1276). San Francisco, CA.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 353–381). Berlin, Germany: Mouton de Gruyter.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers, 35*(1), 116–124. https://doi.org/10.3758/BF03195503.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review, 13*(3), 361–377. https://doi.org/10.3758/BF03193857.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). irr: *Various coefficients of interrater reliability and agreement*. R package version 0.84.1 [computer program]. Retrieved from https://CRAN.R-project.org/package=irr

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge, UK: Cambridge University Press.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics, 66*(3), 363–376. https://doi.org/10.3758/BF03194885

Gibbon, F. E., Hardcastle, W. J., & Suzuki, H. (1991). An electropalatographic study of the /r/, /l/ distinction for Japanese learners of English. *Computer-Assisted Language Learning, 4*(3), 153–171. https://doi.org/10.1080/0958822910040304.

Gick, B., Bernhardt, B. M., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 309–322). Amsterdam, the Netherlands: Benjamins.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia, 9*(3), 317–323. https://doi.org/10.1016/0028-3932(71)90027-3

Gottfried, T. L., Jenkins, J. J., & Strange, W. (1985). Categorial discrimination of vowels produced in syllable context and in isolation. *Bulletin of the Psychonomic Society, 23*(2), 101–104. https://doi.org/10.3758/BF03329794

Hacking, J. F. (2011). The production of palatalized and unpalatalized consonants in Russian by American learners. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), Achievements and perspectives in the acquisition of second language speech: New Sounds 2011 (pp. 93–101). Frankfurt am Main, Germany: Peter Lang.

Hacking, J. F., Smith, B. L., & Johnson, E. M. (2017). Utilizing electropalatography to train palatalized versus unpalatalized consonant productions by native speakers of American English learning Russian. *Journal of Second Language Pronunciation, 3*(1), 9–33. https://doi.org/10.1075/jslp.3.1.01hac

Hacking, J. F., Smith, B. L., Nissen, S. L., & Allen, H. (2016). Russian palatalized and unpalatalized coda consonants: An electropalatographic and acoustic analysis of native speaker and L2 learner productions. *Journal of Phonetics, 54*, 98–108. https://doi.org/10.1016/j.wocn.2015.09.007

Halle, M. (1971). *The sound pattern of Russian*. The Hague, the Netherlands: Mouton. https://doi.org/10.1515/9783110869453

Hardcastle, W. J. (1972). The use of electropalatography in phonetic research. *Phonetica, 25*(4), 197–215. https://doi.org/10.1159/000259382

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments and Computers, 27*(1), 46–51. https://doi.org/10.3758/BF03203619

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication, 47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Hitchcock, E. R., & Byun, T. M. (2015). Enhancing generalisation in biofeedback intervention using the challenge point framework: A case study. *Clinical Linguistics and Phonetics, 29*(1), 59–75. https://doi.org/10.3109/02699206.2014.956232

Hoole, P., & Zierdt, A. (2010). Five-dimensional articulography. In B. Maasen & P. H. H. M. van Lieshout (Eds.), Speech motor control: New developments in basic and applied research (pp. 331–349). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199235797.003.0020

Iskarous, K., & Kavitskaya, D. (2010). The interaction between contrast, prosody, and coarticulation in structuring phonetic variability. *Journal of Phonetics, 38*(4), 625–639. https://doi.org/10.1016/j.wocn.2010.09.004

Iskarous, K., & Kavitskaya, D. (2018). Sound change and the structure of synchronic variability: Phonetic and phonological factors in Slavic palatalization. *Language, 94*(1), 43–83. https://doi.org/10.1353/lan.2018.0001

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /delta/-/θ/ contrast by francophones. *Perception and Psychophysics, 40*(4), 205–215. https://doi.org/10.3758/BF03211500

Jones, D., & Ward, D. (1969). *The phonetics of Russian*. Cambridge, UK: Cambridge University Press.

Kampstra, P. (2008). Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software. Code Snippets, 28*, 1–9.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology, 5*, 1246. https://doi.org/10.3389/fpsyg.2014.01246

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138*(2), 817–832. https://doi.org/10.1121/1.4926561

Katz, W. F., & Mehta, S. (2015). Visual feedback of tongue movement for novel speech sound learning. *Frontiers in Human Neuroscience, 9*, 612. https://doi.org/10.3389/fnhum.2015.00612

Kavitskaya, D. (2006). Perceptual salience and palatalization in Russian. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory Phonology 8* (pp. 589–610). Berlin, Germany: Mouton de Gruyter.

King, H., & Ferragne, E. (2017). The effect of ultrasound and video feedback on the production and perception of English liquids by French learners. Presented at Phonetics & Phonology in Europe 2017, Cologne, Germany.

Kochetov, A. (2002). Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization. New York, NY: Routledge.

Kochetov, A. (2004). Perception of place and secondary articulation contrasts in different syllable positions: Language-particular and language-independent asymmetries. *Language and Speech, 47*(Pt 4), 351–382. https://doi.org/10.1177/00238309040470040201

Kochetov, A. (2017). Acoustics of Russian voiceless sibilant fricatives. *Journal of the International Phonetic Association, 47*(3), 321–348. https://doi.org/10.1017/S0025100317000019

Kochetov, A., & Smith, J. (2009). Cross-language perception of Russian plain/palatalized laterals and rhotics. *The Journal of the Acoustical Society of America, 126*(4), 2313. https://doi.org/10.1121/1.3249537

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kulikov, V. (2011). Features, cues, and syllable structure in the acquisition of Russian palatalization by L2 American learners. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), Achievements and perspectives in SLA of speech: New sounds 2010 (pp. 193–204). Frankfurt am Main, Germany: Peter Lang.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Malden, MA: Blackwell Publishing.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition, 38*(1), 35–64. https://doi.org/10.1017/S0272263115000194

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). R package 'emmeans': Estimated marginal means, aka least-squares means [computer program]. Retrieved from https://github.com/rvlenth/emmeans

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5), 187–196. https://doi.org/10.1016/S1364-6613(00)01471-6

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3 Pt 1), 1242–1255. https://doi.org/10.1121/1.408177

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874–886. https://doi.org/10.1121/1.1894649

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences, 13*(3), 110–114. https://doi.org/10.1016/j.tics.2008.11.008

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning, 44*(1), 75–100. https://doi.org/10.1111/j.1467-1770.1994.tb01449.x

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Hove, UK: Psychology Press. https://doi.org/10.4324/9781410611147

McGowan, R. S., Smith, C. L., Browman, C. P., & Kay, B. A. (1990). Methods for least-squares parameter identification for articulatory movement and the program PARFIT. Haskins Laboratories Status Report on Speech Research, 101/102, 220–230.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338–352. https://doi.org/10.1121/1.1907526

Mokari, P. G., & Werner, S. (2017). Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners' production and perception of Standard Southern British English vowels. *Lingua, 185*, 81–95. https://doi.org/10.1016/j.lingua.2016.07.008

Ní Chasaide, A. (1999). Irish. In T. I. P. Association (Ed.), *Handbook of the International Phonetic Association* (pp. 111–122). Cambridge, UK: Cambridge University Press.

Padgett, J. (2001). Contrast dispersion and Russian palatalization. In E. V. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 187–218). San Diego, CA: Academic Press.

Pajak, B., & Levy, R. (2014). The role of abstraction in non-native speech perception. *Journal of Phonetics, 46*, 147–160. https://doi.org/10.1016/j.wocn.2014.07.001

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics, 13*(2), 253–260. https://doi.org/10.3758/BF03214136

Proctor, M. (2011). Towards a gestural characterization of liquids: Evidence from Spanish and Russian. *Laboratory Phonology, 2*(2), 451–485. https://doi.org/10.1515/labphon.2011.017

R Development Core Team (2018). *R: A language and environment for statistical computing* [computer program], version 3.2.4. Retrieved from http://www.R-project.org

Rice, H. R. (2015). Perceptual acquisition of secondary palatalization in L2: Strengthening the bonds between identificaiton and discrimination through multi-sequence category mapping (PhD thesis), University of Indiana, IN.

Ringen, C., & Kulikov, V. (2012). Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics, 20*(2), 269–286. https://doi.org/10.1353/jsl.2012.0012

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development. *Studies in Second Language Acquisition, 35*(1), 1–29. https://doi.org/10.1017/S0272263112000666

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning, 62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Schmid, M. S., & Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing, 31*(3), 367–388. https://doi.org/10.1177/0265532214526175

Schmidt, A. M. (2012). Effects of EPG treatment for English consonant contrasts on L2 perception and production. *Clinical Linguistics and Phonetics, 26*(11-12), 909–925. https://doi.org/10.3109/02699206.2012.718036

Schmidt, A. M., & Beamer, J. (1998). Electropalatography treatment for training Thai speakers of English. *Clinical Linguistics and Phonetics, 12*(5), 389–403. https://doi.org/10.1080/02699209808985233

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means. *The American Statistician, 34*, 216–221.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Smith J, Kochetov A (2009): Categorization of non-native liquid contrasts by Cantonese, Japanese, Korean, and Mandarin listeners. *Toronto Working Papers in Linguistics, 34*, 1–15.

Sproat, R., & Fujimora, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21*(3), 291–311. https://doi.org/10.1016/S0095-4470(19)31340-3

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics, 19*(6-7), 455–501. https://doi.org/10.1080/02699200500113558

Strange, W. (2007). Cross-language phonetic similarity of vowels. Theoretical and methodological issues. In O.-S. Bohn & M. J. Munro (Eds.), Language experience in second language speech learning: In honor of James E. Flege (pp. 35–55). Amsterdam, the Netherlands: Benjamins. https://doi.org/10.1075/lllt.17.08str

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*(2), 131–145. https://doi.org/10.3758/BF03202673

Suemitsu, A., Dang, J., Ito, T., & Tiede, M. (2015). A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning. *The Journal of the Acoustical Society of America, 138*(4), EL382–EL387. https://doi.org/10.1121/1.4931827

Tateishi, M., & Winters, S. (2013). Does ultrasound training lead to improved perception of a nonnative sound contrast? Evidence from Japanese learners of English. In Luo S (Ed.), Proceedings of the 2013 Annual Conference of the Canadian Linguistic Association (pp. 1–15). Victoria, BC.

Timberlake, A. (2004). *A reference grammar of Russian*. Cambridge, UK: Cambridge University Press.

Wang, X., Hueber, T., & Badin, P. (2014). On the use of an articulatory talking head for second language pronunciation training: the case of Chinese learners of French. In S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (Eds.), Proceedings of the 10th International Seminar on Speech Production (ISSP) (pp. 449–452). Cologne, Germany.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America, 113*(2), 1033–1043. https://doi.org/10.1121/1.1531176

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649–3658. https://doi.org/10.1121/1.428217

Wilson, I., & Gick, B. (2006). Ultrasound technology and second language acquisition research. Presented at 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006): *The Banff Conference*, Banff, AB.

Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology, 6*(6), 285–289. https://doi.org/10.1250/ast.35.285

Wright, R. A. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 34–57). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511486401.002