

Articulatory and perceptual cues to non-native phoneme perception: Cross-modal training for early learners

Second Language Research

1–31

© The Author(s) 2020

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0267658320921217

journals.sagepub.com/home/slr**Emily Cibelli** 

Northwestern University, USA

Abstract

Non-native phoneme perception can be challenging for adult learners. This article explores two routes to strengthening early representations of non-native targets: perceptual training, which focuses on auditory discrimination of novel contrasts, and articulatory training, which highlights the articulatory gestures of non-native categories. Of particular interest is whether cross-modal transfer from production to perception is beneficial to improving discrimination. A longitudinal experiment integrating both training types found that articulatory training did not improve discrimination once perceptual learning had taken place. However, a follow-up experiment found an equivalent benefit for perceptual and articulatory training when each was presented as the only learning style to separate groups of learners. These findings suggest that articulatory learning can ‘cross over’ to assist acquisition in the perceptual domain, and may play a key role for second language (L2) learners struggling with both perception and production of novel phoneme categories.

Keywords

articulatory training, cross-modal transfer, English, Hindi, non-native perception, speech perception, stop contrasts

1 Introduction

Perception of novel phonemes in a second language (L2) is a well-known challenge for adult learners. Many studies have documented challenges in discrimination of non-native categories in listeners with no significant exposure to the language (Best and Avery,

Corresponding author:

Emily Cibelli, Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208, USA.

Email: emily.cibelli@gmail.com

2007; Best et al., 2001, 2009; Golestani and Zatorre, 2004; Lim and Holt, 2011; Pruitt et al., 2006; Song et al., 2008). However, experienced second language speakers have also been shown to have difficulties (Akahane-Yamada et al., 1996; Bradlow et al., 1997, 1999; Diaz et al., 2008; Flege et al., 1997; Hattori and Iverson, 2009; Lai, 2009). Theoretical accounts of this challenge (e.g. Best et al., 2001; Flege, 1995; Kuhl et al., 2008) focus on potential interference from the native language (L1); when there is a mismatch between the inventory of the L1 and L2, native language phoneme categories may bias perception towards them, inhibiting recognition of non-native categories as distinct. Therefore, a central focus in acquisition research concerns intervention strategies that train learners to overcome their L1 biases. A crucial component to improved discrimination is the need for learners to recognize that a novel category (or two contrasting categories) exist, and are distinct from native phonemes. During the process of second language acquisition, learners receive support from the lexicon and contextual information in conversation as evidence for novel categories and contrasts that may otherwise be difficult for a learner to identify (Hayes-Harb, 2007). However, listeners (both novice and experienced learners) in the lab are often tested on syllables or words outside of the context of a discourse; in these cases, researchers must find other ways to cue listeners to the existence of non-native contrasts.

Feedback on performance in identification or discrimination tasks can be used to help learners recognize novel categories that sound similar to L1 categories, and to discriminate non-native contrasts (e.g. Goudbeek et al., 2008; McCandliss et al., 2002) The benefits of feedback has also been found for learners acquiring new tonal contrasts (e.g. Wang et al., 1999, 2003; Wang, 2013). Directing learners' attention to particular components in the speech signal can also help; Pederson and Guion-Anderson (2010) found that learners' discrimination of non-native consonants improved when told to attend to consonants during training, but did not when their attention was instead focused on vowels.

However, explicit training is not uncontroversial; a series of studies (Gulian et al., 2007; Lim and Holt, 2011; Seitz and Watanabe, 2003) have argued that implicit tasks leads to more robust category learning. Vlahou et al. (2011) argue that explicit feedback runs the risk of learners forming incorrect hypotheses about new categories, inhibiting learning of the true target category. (However, generalization to new exemplars was not found in their implicit paradigm, suggesting an advantage for explicit training methods which employ high-variability stimuli; see, e.g. Lively et al., 1993; Sadakata and McQueen, 2013, 2014). Crucially, implicit learning studies pair exemplars of a target category with a correlated cue (whether acoustic, lexical, or visual), ensuring that exposure is not the only source of information. Taken together, these studies suggest that information about a category should be unambiguous if it is explicitly stated to learners; otherwise, correlated cues may be safer so as to avoid incorrect hypotheses by learners.

The manipulation of stimuli acoustics can also be used to cue non-native categories. One approach, adaptive fading (Terrace, 1963), presents exaggerated or well-separated tokens of a contrast pair at the onset of training, and reduces the distance between them on subsequent exposures. This allows learners to begin to learn novel categories from distinct examples, and then to bootstrap to cases where the contrast is more acoustically ambiguous (Jamieson and Morosan, 1986; McCandliss et al., 2002; Pruitt, 1995). Escudero et al.

(2011) argued that cue enhancement in adaptive fading bears some similarity to first language acquisition, where infant-directed speech may emphasize acoustic cues critical to L1 category contrasts.

The above approaches are perceptual strategies most commonly employed to assess and enhance perceptual skills in learners. But another line of research has attempted to leverage the two halves of phoneme acquisition – perception and production – in an integrated approach. Speakers receive both auditory and somatosensory feedback when they produce speech (Lametti et al., 2012; Tremblay et al., 2003), and so learners may draw on representations in both the acoustic and articulatory domains as they are exposed to non-native categories.

Theoretical accounts of second-language phoneme acquisition ascribe different roles to the interaction of perceptual and articulatory representations. The Perceptual Assimilation Model (Best and Avery, 2007; Best et al., 2001) is grounded in a tradition that emphasizes the importance of motor or articulatory representations as the basis for perception (Best, 1995; Fowler, 1986, 1989, 2008). It proposes a range of difficulty for non-native category discrimination by adult learners, depending on how the L2 sounds assimilate to L1 categories. A corollary of the model, the Articulatory Organ Hypothesis (Best et al., 2009; Goldstein and Fowler, 2003; Studdert-Kennedy and Goldstein, 2003), proposes that assimilation is especially acute for categories sharing place of articulation, emphasizing the importance of articulators as a driver for perceptual similarity. The Speech Learning Model (SLM; Baker et al., 2002; Flege, 1995; Guion et al., 2000) also proposes a schema to evaluate variable difficulty of non-native contrasts. However, phonetic or acoustic distance, rather than articulator distance, is the primary factor explaining difficulty in this model. Even so, SLM argues that a link between perceptual and articulatory representations is a natural consequence of category invariance as experienced learners develop more abstract phonemic representations of a category (Flege et al., 1997).

A third account, the Native Language Magnet (NLM) theory (Iverson et al., 2003; Kuhl, 2000; Kuhl et al., 2008) describes a mechanism for first language acquisition that accounts for some of the difficulties that adult learners have in separating L2 phonemes from L1 categories. In this account, developing L1 categories begin to warp the perception of all incoming speech sounds towards the center of these categories. More experience with the L1 will strengthen these categories and the attractor effect - with the consequence that later in life, L2 sounds will also be attracted to these L1 categories. While NLM is centered around acoustic representations, articulatory representations are proposed to play a role in challenging listening conditions, such as noisy signals or unfamiliar speech sounds, suggesting their potential utility in discrimination for L2 learners.

Much of the experimental work on cross-modal transfer in L2 phoneme acquisition (defined here as training in the perceptual domain for the benefit of production, and vice-versa) has focused on transfers from perception to production, perhaps reflecting an ongoing interest in helping L2 learners reduce their accent when producing their second language. These studies find support for facilitatory transfer from perception to production, although the effects vary both in rate of change and in substantial individual differences (Akahane-Yamada et al., 1996; Bradlow et al., 1999; Pimsleur, 1963; Wang

et al., 2003; Wilson et al., 2014). Furthermore, improvements in the two domains within a single individual do not always develop at the same rate (Baese-Berk, 2010; Bradlow et al., 1997; Sheldon and Strange, 1982; Zampini, 1998).

Fewer studies have trained articulatory targets with the aim of improving perceptual categorization, although production is also sometimes included as part of a joint perceptual-articulatory training paradigm (e.g. Wang, 2013). (In the present article, the approach of explicit instruction on articulatory targets is referred to as articulatory training, but has also been referred to as form-focused instruction, e.g. Saito and Lyster, 2012.) A seminal study by Catford and Pisoni (1970) took an instructional approach to teaching non-native sounds from multiple languages to native English learners, with lessons describing place of articulation, manner of articulation, and airstream mechanisms in detail. This was compared to training of a second group who received auditory training. Participants in the articulatory training group outperformed those in auditory training on both production and perception. More modest improvement was reported by Lacabex et al. (2008), who trained Spanish learners on the full-vowel/schwa contrast in English. After three months of training, perceptual and articulatory training had comparable effects on discrimination.

Visual information can also be used to provide articulatory training. Hazan et al. (2005) trained native Japanese speakers acquiring English consonant categories in this way; they found that for visually-salient contrasts (/b/-/p/-/v/), audiovisual training outperformed audio-only training. Acoustic information can also be presented visually in acoustic analysis software to give learners real-time feedback about contrasts without visually-observable articulators; this approach was used by Kartushina et al. (2015) to improve production and perception of non-native vowels by French learners of Danish.

While the small set of production-to-perception studies has generally found beneficial effects, results are not uniformly positive. Schneiderman et al. (1988) found a complex pattern of results from a training series designed around a semester of French instruction. They argued that initial improvement after perceptual training could actually be disrupted by articulatory training. (The complex interplay of production and perception in learners has been found in short-term laboratory contexts as well; see Baese-Berk, 2010; Levy and Law, 2010.) Because approaches to articulatory-based training and their outcomes have varied, more data is needed to clarify the effects of cross-modal transfer in these training paradigms.

The current study presents two experiments with perceptual and articulatory training designed for learners at the very beginning of the acquisition process (i.e. with no prior exposure to the language). The structure of articulatory training was inspired by the approach of Catford and Pisoni (1970), and asks whether very explicit awareness of gestural targets during speech production provides an unambiguous basis for category learning in both the perceptual and articulatory domains. In other words, given the challenges associated with perceiving novel contrasts in second language acquisition, can these contrasts be reinforced by giving learners explicit information about the different articulators and gestures required to produce them? Perceptual training was designed with both performance feedback and adaptive fading components. Of particular interest in the current study is (a) the relative and additive contributions of perceptual

and articulatory training, and (b) how learning rates are influenced by the differing ways that novel categories relate to categories in the native language.

1 The current study

This study consists of two experiments designed to examine the efficacy of training in two domains – perceptual and articulatory – on novel phoneme discrimination. Perceptual training is an intuitive approach to improving perception, as learning and performance take place in the same domain. The hypothesis motivating this study is that cross-modal information (transfer from production to perception) can also stimulate learning, by making learners explicitly aware of the existence of a novel category or contrast (particularly in cases where it may otherwise assimilate to an L1 category). By providing physical articulatory landmarks, learners will have a more conscious and concrete anchor on which to build a novel category. This hypothesis is motivated by past studies showing improvement in perceptual discrimination after articulatory training (Catford and Pisoni, 1970; Hazan et al., 2005), as well as theoretical accounts of non-native phoneme perception that place articulatory information as central to (Best et al., 2001, 2009) or facilitatory for (Flege et al., 1997; Kuhl et al., 2008) category representations. It was predicted that articulatory training would confer an additional benefit beyond perceptual training, boosting participants' abilities to discriminate novel contrasts.

In Experiment 1, a multi-session learning paradigm was designed to sequentially teach learners perceptual, and then articulatory, cues to novel contrasts. In self-paced experimental sessions, native English speakers were trained on a series of coronal stop contrasts that differ from their L1 in both place and voicing features. To preview the results, learners' discrimination improved after perceptual training, but did not change further after articulatory training.

The design of Experiment 1 leaves open the question of whether articulatory training was ineffective, or whether it simply was not powerful enough to further improve on the successful perceptual training manipulation. Experiment 2 disambiguates these possibilities in a between-participants design by comparing sets of learners who received short forms of either perceptual or articulatory training.

2 Mapping Hindi to English

Both experiments focus on the Hindi series of coronal stop consonants, a common set of contrasts in the study of second language phoneme acquisition. Hindi has a place of articulation contrast between dental and retroflex stops that has been shown to be a particular challenge for native English speakers (Golestani, 2014; Pederson and Guion-Anderson, 2010; Pruitt et al., 2006; Tees and Werker, 1984; Vlahou et al., 2011), as they tend to perceive both categories as an English alveolar stop.

Hindi also has a four-way voicing distinction in its stop series, contrasting (voiceless) unaspirated, (voiceless) aspirated, voiced, and breathy voiced stops. While the voicing contrasts tend to be more responsive to training, they still present a challenge for English-speaking adult learners (Tees and Werker, 1984). The relationship of these four

categories to the two-way voice-onset-time (VOT) contrast in English is somewhat complex, but assimilation to the L1 VOT categories often shows the following patterns. The Hindi aspirated and unaspirated stops are good matches to the English voiceless (/t/) and voiced (/d/) stops, respectively; the former has long-lag positive VOT or aspiration after the stop burst, and the latter has short-lag positive VOT, with minimal aspiration. Because this maps onto an English category contrast, discrimination of these two Hindi categories is typically good even prior to training. The Hindi voiced stop maps onto the voiced allophone of English /d/, with pre-voicing during the stop-closure (typically found in English unstressed vowel-medial positions, but with some variation by speaker and context; Lisker and Abramson, 1964; Davidson, 2016). As a result, the Hindi voiced and unaspirated stops are often perceived as allophones of the same English category and are difficult for learners to discriminate. The Hindi breathy voiced category has properties of both voiced and voiceless English stops (with both pre-voicing and long-lag positive VOT, but with voicing throughout). As a result, this category could assimilate to either (or neither) English VOT category.

Because these four voicing categories map in variable ways to the English system, some are easier for native English-speaking listeners to discriminate than others. Coupling the place contrast with a voicing contrast in a pair of novel sounds can provide additional evidence to listeners to help them discriminate a contrast (the number of contrastive features often – but not always – decreases the confusability of two consonants; see Bailey and Hahn, 2005). As a result, this series of stop consonants is a rich set of categories with variable difficulty for native English learners, providing an ideal context to test the hypotheses of the current study.

II Experiment I

Experiment 1 was designed to integrate well-established perceptual training paradigms (performance feedback and adaptive fading) with precise instruction on articulatory targets. This design assesses whether articulatory training provides a benefit to learners beyond perceptual training alone. The experiment was designed as a multi-session study, with discrimination tests before and after several perceptual training sessions, and a final test after articulatory training.

I Methods

Participants. Native English speakers were recruited to participate in the experiment via flyers posted on the University of California's Berkeley campus. Prior to enrollment, individuals who responded to recruitment materials were excluded from enrollment in the study if they reported any significant exposure to Hindi in the home, classroom, or in their family or community, whether or not they considered themselves to be fluent speakers. Respondents were also excluded if they had exposure to another language with a dental-retroflex contrast or a four-way VOT contrast. Of the 29 participants who were ultimately enrolled in the study, eight were excluded from analysis (three for experimenter error,¹ and five for failing to complete all eight sessions). Twenty-one participants were included in the final analysis (15 female; mean age: 22.6 years, SD 9.4).

Table 1. The eight Hindi coronal stop consonants.

Consonant	Voicing	Positive VOT	Negative VOT	Place
t̪	unaspirated	short-lag	none	dental
t̪ ^h	aspirated	long-lag	none	dental
d̪	voiced	short-lag	pre-voicing	dental
d̪ ^h	breathy	long lag (breathy)	pre-voicing	dental
ʈ	unaspirated	short-lag	none	retroflex
ʈ ^h	aspirated	long-lag	none	retroflex
ɖ	voiced	short-lag	pre-voicing	retroflex
ɖ ^h	breathy	long lag (breathy)	pre-voicing	retroflex

Note. VOT = voice-onset-time

Participants were paid \$10 per hour. As incentive to complete the full study, a bonus of \$20 was awarded to any participant who completed all eight sessions.

Participants' language experience was assessed in a pre-screening questionnaire. Eighteen of 21 participants reported some familiarity with or exposure to a second language; 10 reported exposure to a third language. Participants rated their abilities in speaking, writing, reading, and understanding in each language on a scale from 1 to 4 (where 1 = not at all proficient, and 4 = fluently proficient). The average proficiency score across all skills for a second language was 2.49 (SD 0.79); for an L3, the average score was 2.43 (SD 0.71).

Stimuli. A female native speaker of Hindi recorded consonant–vowel (CV) and vowel–consonant–vowel (VCV) syllables with one of eight consonants (see Table 1) and the vowels /a/, /i/, or /u/. Two series of stimuli were recorded: a 'careful' series, in which the speaker was instructed to speak clearly with emphasis on the contrast between consonants, and a 'natural' series, where tokens were spoken without particular emphasis. Ten tokens of each style (careful or natural), syllable (CVC or CV), vowel (/a, /i/, /u/), and consonant (the eight in Table 1) combinations were recorded, for a total of 960 tokens. From these, 384 (four of each combination) were selected to use in the experiment, based on the most reliably rated tokens in an eight-alternative forced choice identification task with two additional native Hindi speakers.

Syllables were recorded in blocks; an unintended result of this was that the speaker used contrastive pitch to distinguish some CVC syllable types (e.g. /uɖ^hu/ with low-high pitch vs. /uɖu/ with high-low pitch). Experimenter instructions were not sufficient to eliminate this during the recording session; as a result, all stimuli were pitch-flattened in Praat (Boersma and Weenink, 2019) to the F0 mean across the stimulus set. This removed F0 correlates of voicing that are known to cue breathy stops (Hombert et al., 1979; Schiefer, 1986); however, it was necessary in order to avoid pitch contours as an additional cue to the identity of a category.

Experiment structure. Data was collected in eight sessions as part of a larger study to test perceptual and articulatory learning (results on pronunciation learning in this population

Table 2. Structure of Experiment I.

Session	Perception task	Perception feedback	Production task	Production feedback	Stimuli
Pre-test	AX discrimination	–	Repetition	–	CV natural
Perceptual training 1	AX discrimination	Accuracy feedback	–	–	VCV careful
Perceptual training 2	AX discrimination	Accuracy feedback	–	–	VCV natural
Perceptual training 3	AX discrimination	Accuracy feedback	–	–	CV careful
Perceptual training 4	AX discrimination	Accuracy feedback	–	–	CV natural
Post-test	AX discrimination	–	Repetition	–	CV natural
Articulatory training	–	–	Repetition	Visual cues	CV natural
Re-test	AX discrimination	–	Repetition	–	CV natural

are reported in Cibelli, 2020). Sessions and tasks are summarized in Table 2. There is evidence that sleep can assist the development of novel phoneme categories (Earle and Myers, 2013, 2015; Fenn et al., 2003); to reflect this, there was always at least one night's break after a training session before completing a test session. The median number of days to complete the eight sessions was 16 (range: 7–29 days). Custom scripts in OpenSesame were used to present experiment materials (Mathôt et al., 2012), with auditory stimuli presented over headphones. A serial response button box was used to collect accuracy and reaction time data (Psychology Software Tools, Inc.).

Perceptual training consisted of four sessions. In each, participants completed an AX discrimination task and received trial-level accuracy feedback. These sessions were designed as an adaptive fading paradigm (Jamieson and Morosan, 1986; McCandliss et al., 2002; Protopapas and Calhoun, 2000; Terrace, 1963). To progress from well-separated stimuli to more perceptually-challenging pairs of stimuli, the first session presented VCV stimuli spoken in a careful style. The second used VCV tokens spoken in a more 'natural' style, the third used CV careful tokens, and the fourth CV natural tokens. VCV tokens were taken to be 'easier' for listeners because they provide more acoustic information to disambiguate non-native stimuli. The formant transitions out of the first vowel cue the place of articulation of the consonant, and the contrast between the first vowel voicing and the closure may assist in cuing consonant voicing.

Test sessions consisted of two tasks: AX discrimination without feedback, and a repetition task. In both, participants heard CV natural tokens. The no-feedback discrimination task was used to measure participants' discrimination performance at baseline (pre-test), after perceptual training (post-test), and after all training (re-test).

The articulatory training session gave participants explicit information about the gestures necessary to produce the target categories; as a result, participants also received explicit information about the number of categories they were trying to learn (which they

may not have been aware of if they failed to discriminate some novel categories). Training was presented as a self-paced lesson; example training slides are presented in Figure 1. Training began with place of articulation, with information about tongue placement for dental and retroflex consonants and how they differ from alveolar stops. Participants were taught to read sagittal sections, which were paired with color cues to place of articulation (red for retroflex, green for dental: Figures 1A and 1B) throughout the session.

Training then introduced the concept of voicing, starting with the voiceless unaspirated/voiceless aspirated contrast familiar to learners as the English /t-/d/ contrast. Participants were taught about the ‘puff of air’ in aspiration, and its absence in unaspirated consonants, by holding their hands in front of their face while hyperarticulating English ‘t’ and ‘d’. These were paired with visual cues for the presence and absence of aspiration (a puffing cloud and an X: see Figures 1C and 1D).

Pre-voicing was introduced next; participants learned to identify the presence or absence of voicing by holding their fingers on their throat while humming, with a corresponding visual cue. When participants felt comfortable producing pre-voicing in voiced stops, they combined pre-voicing and aspiration to produce the breathy stop. Participants practiced all combinations of voicing and place features (Figure 1E). The end of the lesson contained a repetition task, with visual cues to the target category (Figure 1F).

2 Results


Modeling approach. To assess performance, d' was calculated from the discrimination data in each test session, to capture sensitivity to contrasts while accounting for bias towards ‘different’ responses (for details on calculating d' , see Macmillan and Creelman, 2004). Prior to calculation, trials with outlier reaction times (RTs) – defined as responses less than 100 ms, or values greater than 3 standard deviations from an individual participant’s mean reaction time – were removed. A d' value was calculated for each combination of participant ($n = 21$), test session (3), and contrast type (3: place contrast, voicing contrast, or place + voicing contrast), resulting in 189 unique values for analysis.

The d' data from ‘different’ trials was analyzed with a linear mixed-effects regression model. The model included two reverse Helmert-coded fixed effects for session, comparing (1) post-test to pre-test, and (2) re-test to the two previous test sessions. The model structure also included contrast-coded fixed effects for place of articulation ($-0.5 =$ dental, $0.5 =$ retroflex), number of contrasting features ($-0.5 =$ one, $0.5 =$ two), and the interaction these and each session predictor. Mean L2 and L3 experience (on a 4-point scale), as well as the number of days it took to complete all eight sessions, were centered and included as control variables.

The model was fit in R (R Core Team, 2018) using the lme4 (Bates et al., 2015b) and RePsychLing (Bates et al., 2015a) packages. By-participant random intercepts were included, as well as the maximal random slopes justified by the data, following the recommendations in Bates et al. (2015a). The final random slope structure included de-correlated random slopes for contrast type, number of features, and both session

(a)


Here is a picture of the inside of your mouth when you say a dental "t".



Notice how the tip of the tongue touches the upper teeth, indicated in red. To orient you, a blue arrow is pointing to your nose.

(b)


Here is a picture of your tongue during a retroflex "t". Notice the placement and how the tongue tip curls back.



Press any key to continue.

(c)

That puff of air is a voicing feature of "t". We'll use it in this experiment to remind you about that little puff of air.




Whenever you see it, you'll make a sound with voicing like "t".

Press any key to continue.


(d)

One of the ways that "t" is different from "d" is that "d" does not have that little puff of air.

When you see this picture for "t", think "puff of air".







When you see this picture for "d", think "no puff of air".



Press any key to continue.

(e)


Let's review the place and voicing cues. Try saying each of these:

Dental "t"		"tah" "tee" "too"
Retroflex "t"		"tah" "tee" "too"
Dental "d"		"dah" "dee" "doo"
Retroflex "d"		"dah" "dee" "doo"

Press any key to continue.

(f)

Listen carefully, then repeat.



When you have repeated the syllable, press any key to continue.

Figure 1. Example training slides from articulatory training.

Notes. Figures 1A and 1B show training of the dental/retroflex place contrast by introducing participants to major articulatory landmarks using sagittal sections. Color cues remind learners of dental (green) and retroflex (red) place of articulation. Figure 1C introduces the concept of aspiration, and a picture to associate with the concept. Figure 1D compares the aspirated 't' to the unaspirated 't' (English orthography). Figure 1E asks participants to practice combining place and voicing with visual cues. Figure 1F demonstrates a repetition trial for a syllable with the target consonant /t^hu/, with visual cues.

Table 3. Fixed effects of the d-prime (d') model, Experiment 1.

	Estimate	Standard error	t	χ^2	p (χ^2)
Contrast type	1.417	0.080	17.73	57.99	< 0.001
Number of features	1.313	0.042	31.19	80.70	< 0.001
Session (pre-test vs. post-test)	0.383	0.090	4.26	13.04	< 0.001
Session (pre-/post-test vs. re-test)	0.155	0.103	1.51	2.17	0.141
L2 proficiency	0.123	0.084	1.46	2.49	0.154
L3 proficiency	0.004	0.083	0.05	0.00	0.961
Days to completion	-0.103	0.080	-1.28	1.59	0.208
Contrast * session (pre vs. post)	-0.980	0.086	-1.13	1.28	0.259
Features * session (pre vs. post)	0.959	0.097	0.98	0.96	0.327
Contrast * session (pre/post vs. re)	0.044	0.074	0.60	0.36	0.549
Features * session (pre/post vs. re)	-0.124	0.084	-1.47	2.19	0.144

predictors. The model was re-fit after excluding extreme residuals (> 2.5 SD; 97.9% of data retained after exclusion; Baayen, 2008). Nested model comparisons were used to assess the significance of fixed effects.

Findings. Participants as a group achieved 67.7% accuracy in the pre-test, 80.9% accuracy at post-test, and 79.9% accuracy at re-test. The fixed effects of the d' model are reported in Table 3; performance by session and contrast type is summarized in Figure 2. There was a significant effect of contrast type ($\beta = 1.417$, $\chi^2(1) = 57.99$, $p < 0.001$); discrimination was better for trials that contrasted in voicing than place features. The effect of number of contrasting features was also significant ($\beta = 1.313$, $\chi^2(1) = 80.70$, $p < 0.001$); discrimination was higher when a pair contrasted in both place and voicing features. L2 proficiency, L3 proficiency, and the number of days to complete the study did not have a significant effect on discrimination (all $\chi^2(1) < 2$, $p > 0.10$).

In test sessions, errors on ‘different trials’ were made on 22.3% of voicing trials, and 45.2% of place trials. Figure 3 shows the breakdown of errors in voicing trials during all test sessions. As expected, contrasts that map closely to the English VOT distinction (i.e. Hindi aspirated and voiceless stops) were well-discriminated. Targets with short-lag positive VOT – voiceless and voiced trials – were most often confused with one another; this pattern is also expected, as these categories can both be perceived as allophones of the English voiceless category. Breathy tokens were confused with both voiced and aspirated tokens, but the latter (a match in positive VOT) was more common. This indicates that listeners were especially influenced by positive VOT when they made errors, but that both aspiration and voicing features caused some discrimination errors.

Turning to the by-session predictors, there was a significant main effect comparing pre-test and post-test ($\beta = 0.383$, $\chi^2(1) = 13.08$, $p < 0.001$), with discrimination higher at the post-test (after perceptual training). The effect at re-test did not reach significance ($\beta = 0.156$, $\chi^2(1) = 2.17$, $p = 0.141$), indicating that discrimination did not improve or decline after articulatory training.² None of the interactions of session and contrast type or number of features reached significance (all $\chi^2(1) < 3$, $p > 0.10$).

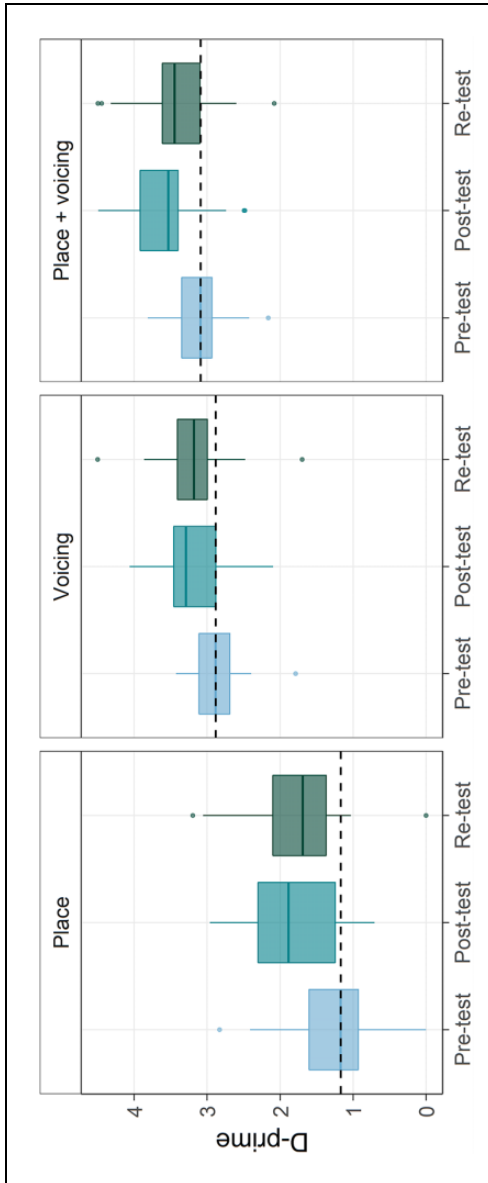


Figure 2. Boxplot of discrimination performance by session and contrast type, Experiment 1. Note. The dotted horizontal line shows the median d' for each contrast type at pre-test.

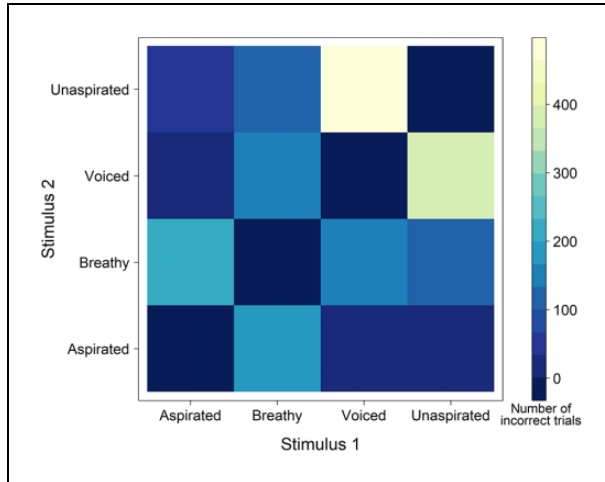


Figure 3. Confusion matrix of incorrect voicing trials (cases where participants incorrectly responded ‘same’) during test sessions, Experiment 1.

Note. Unaspirated and voiced consonants were most often confused for one another.

Reaction time analysis. To further assess performance in Experiment 1, a model of log-transformed reaction time (RT) was constructed with data from 15,131 correct ‘different’ trials in the pre-test, post-test, and re-test. The model structure was identical to the d' model, with the exception of an extra predictor for trial count (centered around 0) to account for changes to RTs over the duration of each session. The full model table is reported in Appendix 1. RTs were faster at post-test ($\beta = -0.413$, $\chi^2(1) = 15.61$, $p < 0.001$) and re-test ($\beta = -0.279$, $\chi^2(1) = 15.79$, $p < 0.001$) than previous sessions. As in the d' model, RTs were faster for voicing trials ($\beta = -0.169$, $\chi^2(1) = 35.85$, $p < 0.001$) and trials where two features contrasted ($\beta = -0.170$, $\chi^2(1) = 50.40$, $p < 0.001$).

While the RT model is largely consistent with the d' model, the increased number of data points available when the data is not aggregated provides additional sensitivity to detect trial-level changes in performance. It was possible to observe that correct judgments were being made more quickly in the re-test session, suggesting that there was improvement to performance after articulatory training.

3 Experiment 1B: Control study

Introduction. The adaptive fading component of Experiment 1 required that multiple sessions of perceptual training be run. As a result, participants received more exposure to the target categories during this portion of training than during articulatory training. There is some evidence that repeated exposure may improve perception of non-contrastive phones; for example, Pegg and Werker (1997) found that discrimination of allophones of the same phoneme improved with exposure. Therefore, it is important to ensure that exposure alone was not responsible for improvement after perceptual training. In Experiment 1B, a new set of nine participants completed four sessions of the

Table 4. Structure of Experiment 1B.

Session	Perception task	Perception feedback	Production task	Production feedback	Stimuli
Session 1	AX discrimination	–	Repetition	–	CV natural
Session 2	AX discrimination	–	Repetition	–	CV natural
Session 3	AX discrimination	–	–	–	CV natural
Session 4	AX discrimination	–	Repetition	–	CV natural

discrimination task, with no feedback given on performance. The same recruitment criteria used in Experiment 1 applied, and the language experience of this set of participants was similar to the first study (average L2 experience: 2.57, SD 0.87; average L3 experience: 2.33, SD 0.98). All sessions used CV natural stimuli, that is, there was no adaptive fading manipulation. In addition, there was no performance feedback given during these tasks. The first, second, and fourth sessions did include the production repetition task, to make these sessions comparable to the pre-test, post-test, and re-test of the main study and control the amount of exposure to the target stimuli. A summary of the study format is given in Table 4.

Results. Across the four sessions, participants responded correctly to 70.9%, 68.7%, 69.7% and 68.8% of trials. Calculation of d' and model selection procedure followed the procedures described for Experiment 1; the full model table is reported in Appendix 2. The main effects of contrast type ($\beta = 1.731$, $\chi^2(1) = 32.92$, $p < 0.001$) and two-feature contrast ($\beta = 1.474$, $\chi^2(1) = 35.29$, $p < 0.001$) were significant; trials with a voicing contrast or with two contrasting features were more accurately perceived than place trials or one-feature contrasts.

None of the main effects of session reached significance (all $\chi^2(1) < 2$, $p > 0.10$). However, the interaction of contrast type and session was significant for session 2 ($\beta = 0.210$, $\chi^2(1) = 4.92$, $p = 0.028$) and session 3 ($\beta = 0.399$, $\chi^2(1) = 9.17$, $p = 0.003$). Separate follow-up models on place and voicing trials were used to determine the reliability and direction of the effect. In both cases, the discrimination of place trials was poorer in later sessions (session 2: $\beta = -0.350$, $\chi^2(1) = 5.55$, $p = 0.018$; session 3: $\beta = -0.416$, $\chi^2(1) = 7.93$, $p = 0.005$). The effect was not significant for follow-up models of voicing trials (all $\chi^2(1) < 2$, $p > 0.10$). The significant interaction of number of features and session 1 vs. 2 ($\beta = 0.272$, $\chi^2(1) = 5.99$, $p = 0.014$) and sessions 1/2 vs. 3 ($\beta = 0.302$, $\chi^2(1) = 9.16$, $p = 0.003$) also failed to reach significance in follow-up models (all $\chi^2(1) < 1$, $p > 0.10$).

The findings of the control study indicate that repeated exposure, in the absence of other training, is not sufficient to improve discrimination of the target contrasts. Discrimination of voicing trials and two-feature trials did not change. While this by itself could indicate an under-powered study, detection of the place contrast reliably declined³ in the middle two sessions, suggesting that the pattern is more likely attributable to the ineffectiveness of exposure alone. These results clarify that the effects found in the post-test in Experiment 1 reflect the impact of accuracy feedback and adaptive fading.

4 Discussion

Experiment 1 assessed discrimination performance of non-native categories before and after two types of training: perceptual training with adaptive fading and performance feedback, and articulatory training with explicit information about articulatory targets. The study found improvement in discrimination after perceptual training. There was no additional change in discrimination after articulatory training, although reaction times suggested improvements in the speed at which participants could correctly discriminate categories. A control study (Experiment 1B) confirmed that exposure to the Hindi stimuli alone was not sufficient to improve discrimination, indicating that the adaptive fading and feedback manipulations were effective.

Contra Catford and Pisoni (1970), articulatory training did not improve participants' ability to discriminate novel contrasts, although – as reaction times sped up and performance did not decrease at the re-test – it is possible that articulatory training helped to maintain what was learned during earlier study sessions. However, the experiment design was not identical in the two cases. Critically, participants in the current study received articulatory training only after perceptual training, meaning that any detectable effects would have been additive on the benefit gained during those training sessions. Articulatory training was also much shorter than perceptual training (one vs. four sessions). Given this, it is difficult to know whether articulatory training was ineffective altogether, or whether its contribution would be stronger under other conditions. To clarify this, Experiment 2 was designed to test several training conditions in a between-participants design. In each, participants received one of four training types designed to address different potential limitations of the Experiment 1 study design.

III Experiment 2

I Introduction

Experiment 1 lacked a detectable effect of articulatory training on discrimination ability. Experiment 2 was designed to clarify whether this represents a true null effect or reflects the sequential, within-participants design of the experiment. Three hypotheses were entertained to address possible aspects of the study design that could have resulted in a null effect:

- The 'local ceiling' hypothesis: There may be an upper limit or diminishing returns on the amount of improvement demonstrable in the scope of a laboratory study. Wright et al. (2015) found that discrimination of non-native categories improved after 60 trials of feedback, but that increasing training to 120 or 240 trials had no additional benefit. Similarly, Iverson and Evans (2009) observed a flattening in performance of English vowel identification by native Spanish speakers after three of five total training sessions. If a similar process applied in Experiment 1, learners may have reached their ceiling by the post-test, prior to articulatory training, leaving no room for additional improvement within the structure of that experiment.
- The 'inflexible instructor' hypothesis: Participants may show individual variation in their learning styles. Using a computer program for training does not provide

an opportunity for the teacher to adapt to the needs of the learner. By contrast, Catford and Pisoni (1970) allowed for in-person interaction with a trained phonetician during learning, which may have provided crucial flexibility at a more individualized level.

- The exposure hypothesis: Perceptual training was longer than articulatory training in Experiment 1 (four sessions vs. one session). The duration of exposure certainly plays a role in acquisition at long time scales (Jia et al., 2006); therefore, participants may have shown an effect of articulatory training if it had been extended. In particular, a longer repetition task – where articulatory cues are reinforced in production, and perceptual cues are also provided – may assist learners with unstable category representations.

Crucially, these hypotheses are not mutually exclusive; multiple factors may have contributed to the pattern of results in Experiment 1.

Design. To address these hypotheses, a short version of Experiment 1 was designed, with participants randomly assigned to one of four conditions in a between-participants design. The study consisted of three phases: pre-test, training, and post-test, with only the training phase varying by condition. For recruitment reasons, all phases took place on the same day. The pre-test and post-test phases were identical to the test sessions from Experiment 1. The four training conditions were as follows:

1. Condition 2A: Basic articulatory training. This condition was identical to the articulatory training session from Experiment 1. This condition directly tests the local ceiling hypothesis, by providing only articulatory training and not the preceding perceptual training sessions to learners. By contrast, improvement in Conditions 2B and 2C below, combined with a lack of improvement in 2A, would argue against the local ceiling hypothesis.
2. Condition 2B: Long articulatory training. This condition replicated the training slides from articulatory training in Experiment 1. However, the length of the repetition task at the end of training was increased, so that participants completed four times as many repetition trials (270 total) in order to reinforce the link between perceptual and articulatory cues. This condition tests the exposure hypothesis.
3. Condition 2C: Interactive articulatory training. This condition used the same articulatory training session from Experiment 1, but the experimenter, a trained phonetician, was present and sat next to the participant while they completed the training. Every participant in this condition opted to ask clarifying questions about the training to the experimenter, and the experimenter also offered feedback and corrections during training. This condition tests the inflexible instructor hypothesis.
4. Condition 2D: Perceptual training. Conditions 2A–2C are not directly comparable to Experiment 1, where participants received multiple days of perceptual training. Condition 2D controls for this by running the first session of discrimination training from Experiment 1 (with CVC careful stimuli). This allows for a comparison of the efficacy of perceptual and articulatory training at roughly equal lengths.

Predictions. Performance in the four conditions allow for direct tests of the hypotheses outlined above. If performance improves in condition 2A, when participants received only the standard articulatory training, it suggests that the local ceiling inhibited performance in Experiment 1. Improvement in condition 2B suggests that the exposure during articulatory training in Experiment 1 was too short. Improvement in condition 2C at post-test suggests that the ‘inflexible instructor’ may have hampered performance in Experiment 1. Finally, differences between condition 2D and the other training conditions would indicate a difference in the efficacy of perceptual and articulatory training.

2 Methods

Sixty participants were recruited and randomly assigned to one of the four conditions (2A–2D). Pre-screening for language background followed the procedures in Experiment 1. In addition, due to observed patterns of educational backgrounds in interested participants in Experiment 2, potential participants were excluded if they had any linguistics background beyond an introductory course, in order to control for knowledge of phonetics. Fifty-two participants reported some exposure to a second language, with a mean proficiency score of 2.68 on a 4-point scale (SD 0.87). Twenty-three reported experience with a third language (mean proficiency: 2.28, SD 0.91).

The structure of Experiment 2 consisted of three phases, all completed in a single two-hour session. The pre-test and post-test were identical to the test sessions in Experiment 1, and were the same across conditions. The training stage varied by conditions; 15 participants received each of the training types described in Section III.1. Optional five-minute breaks were provided between each phase of the session. All test sessions presented CV natural tokens.

Modeling approach. Prior to modeling, trials with extreme reaction times – shorter than 100 ms, or greater than 3 SD from individual means – were excluded; 97.2% of the data was retained. The dependent variable d' was calculated as described in Section II.2, generating 360 values for analysis (15 participants * 2 sessions * 3 contrast types). A linear-mixed effects regression model was constructed, with fixed effects for session (−0.5 = pre-test, 0.5 = post-test), contrast type (−0.5 = place, 0.5 = voicing), number of contrasting features (−0.5 = one, 0.5 = two), and mean L2 and mean L3 proficiency (both centered). The model also included three predictors contrasting the training conditions: training style (0.5 = perceptual, −0.167 = others), experimenter intervention (−0.5 = basic articulatory training, 0.5 = interactive articulatory training, 0 = others), and study length (−0.5 = basic articulatory training, 0.5 = long articulatory training, 0 = others).

The model included two-way interactions between session and all non-proficiency predictors, all condition predictors and contrast type, and all condition predictors and number of features, and three-way interactions of session, all condition predictors, and either contrast type or number of features. The maximum random effects structure supported by the data included correlated participant slopes for session, contrast type, and number of features.

Table 5. Pre-test and post-test accuracy by condition, Experiment 2 (percentages).

Condition	Pre-test accuracy	Post-test accuracy
2A: Basic articulatory training	66.6	72.2
2B: Long articulatory training	66.8	68.5
2C: Interactive articulatory training	65.7	74.9
2D: Perceptual training	66.5	77.1

Table 6. Fixed effects for the d-prime (d') model, Experiment 2.

	Estimate	Standard error	t	χ^2	$p(\chi^2)$
Session	0.215	0.081	2.66	6.67	0.01
Contrast type	1.519	0.049	30.88	168.11	< 0.001
Number of features	1.279	0.034	37.39	194.09	< 0.001
Training style	0.017	0.199	0.08	0.01	0.932
Experimenter intervention	-0.089	0.189	-0.47	0.22	0.639
Study length	0.13	0.185	0.70	0.49	0.483
L2 proficiency	-0.014	0.039	-0.35	0.12	0.733
L3 proficiency	0.007	0.04	0.18	0.03	0.860
Session * contrast type	-0.087	0.043	-2.03	4.06	0.044
Session * number of features	-0.012	0.048	-0.24	0.06	0.809
Session * training style	-0.267	0.281	-0.95	0.90	0.344
Session * experimenter intervention	0.029	0.265	0.11	0.01	0.913
Session * study length	0.233	0.265	0.88	0.77	0.380
Contrast type * training style	-0.368	0.171	-2.15	4.46	0.035
Number of features * training style	0.065	0.118	0.55	0.30	0.586
Contrast type * study length	0.19	0.161	1.18	1.35	0.245
Number of features * study length	0.065	0.112	0.57	0.33	0.567
Contrast type * experimenter intervention	-0.145	0.16	-0.91	0.81	0.367
Number of features * experimenter Intervention	-0.019	0.112	-0.17	0.03	0.868
Session * contrast type * training type	-0.397	0.149	-2.67	6.99	0.008
Session * number of features * training type	-0.097	0.168	-0.58	0.33	0.566
Session * contrast type * study length	-0.233	0.139	-1.67	2.77	0.096
Session * number of features * study length	-0.332	0.157	-2.12	4.43	0.035
Session * contrast type * experimenter intervention	0.278	0.139	1.99	3.92	0.048
Session * number of features * experimenter intervention	0.300	0.157	1.91	3.61	0.058

Results. Accuracy data for the four conditions in Experiment 2 are reported in Table 5, and the fixed effects of the d' model are reported in Table 6. There was a significant effect of session ($\beta = 0.215$, $\chi^2(1) = 6.67$, $p = 0.010$), indicating that performance improved at the post-test compared to the pre-test. There was a significant effect of the two predictors describing stimulus properties: contrast type ($\beta = 1.519$, $\chi^2(1) = 168.11$, $p < 0.001$) and number of features ($\beta = 1.279$, $\chi^2(1) = 194.09$, $p < 0.001$); voicing

contrasts and two-feature contrasts were more accurately discriminated than place or one-feature contrasts, respectively. Session and contrast type significantly interacted ($\beta = -0.087$, $\chi^2(1) = 4.06$, $p = 0.044$); follow-up models of place and voicing trials separately indicated that improvement from pre-test to post-test was larger for place contrast trials (mean d' improvement: 0.262, $\beta = 0.266$, $\chi^2(1) = 8.47$, $p = 0.004$) than for voicing contrast trials (mean d' improvement: 0.197, $\beta = 0.193$, $\chi^2(1) = 4.91$, $p = 0.027$). The interaction of session and number of features was not significant ($\beta = -0.012$, $\chi^2(1) = 0.060$, $p = 0.809$).

None of the condition main effects – training style ($\beta = 0.017$, $\chi^2(1) = 0.01$, $p = 0.932$), experimenter intervention ($\beta = -0.089$, $\chi^2(1) = 0.22$, $p = 0.639$), and study length ($\beta = 0.130$, $\chi^2(1) = 0.49$, $p = 0.483$) – reached significance. Crucially, all two-way interactions of session and condition also failed to reach significance (all $\chi^2(1) < 1$, $p > 0.10$), indicating that improvement from pre-test to post-test did not strictly depend on training style.

In the two-way interactions between condition and stimulus properties, only the study type by contrast type interaction reached significance ($\beta = -0.087$, $\chi^2(1) = 4.46$, $p = 0.035$). Follow-up models split by training type (perceptual training vs. all production conditions) showed that the distance between discrimination of voicing and place trials was larger in articulatory training conditions (mean d' difference: 1.57, $\beta = 1.579$, $\chi^2(1) = 125.60$, $p < 0.001$) than in the perceptual training condition (mean difference: 1.34, $\beta = 1.349$, $\chi^2(1) = 41.96$, $p < 0.001$).

However, there were several three-way interactions which suggested differences between conditions that were conditional on stimulus properties; see Figure 4 for a comparison. Turning first to the distinction between perceptual training and all articulatory training sessions: the interaction of session, contrast type, and training style was significant ($\beta = -3.97$, $\chi^2(1) = 6.99$, $p = 0.008$). Follow-up models found a significant contrast type by session interaction only in the perceptual training model ($\beta = -0.300$, $\chi^2(1) = 13.71$, $p < 0.001$). In this condition, there was only improvement on place trials from pre-test to post-test (Figure 4D, left panel; mean d' difference: 0.202), not on voicing trials (Figure 4D, middle panel; mean difference: 0.016). Conversely, in the three articulatory training conditions (Figures 4A–4C), there was improvement in both place trials (mean difference: 0.379) and voicing trials (mean difference: 0.255), resulting in a lack of a session by contrast type interaction ($\beta = -0.221$, $\chi^2(1) = 0.18$, $p = 0.674$) in the follow-up model.

Session and number of features significantly interacted with study length ($\beta = -0.331$, $\chi^2(1) = 4.43$, $p = 0.035$). Follow-up models determined that this was driven by the long articulatory training condition (Figure 4B; $\beta = -0.163$, $\chi^2(1) = 3.23$, $p = 0.072$), where the d' score for one-feature contrasts improved from pre-test to post-test (difference: 0.44) more than two-feature contrasts (difference: 0.30). The marginal effect of session, contrast type, and study length ($\beta = -0.233$, $\chi^2(1) = 2.77$, $p = 0.096$) was not reliable in follow-up models.

The three-way interaction of experimenter intervention, session, and contrast type was significant ($\beta = 0.278$, $\chi^2(1) = 3.92$, $p = 0.047$). Follow-up models comparing the basic articulatory condition to the interactive condition found that the effect was driven by the latter, with a marginal session by contrast type interaction ($\beta = 0.128$, $\chi^2(1) =$

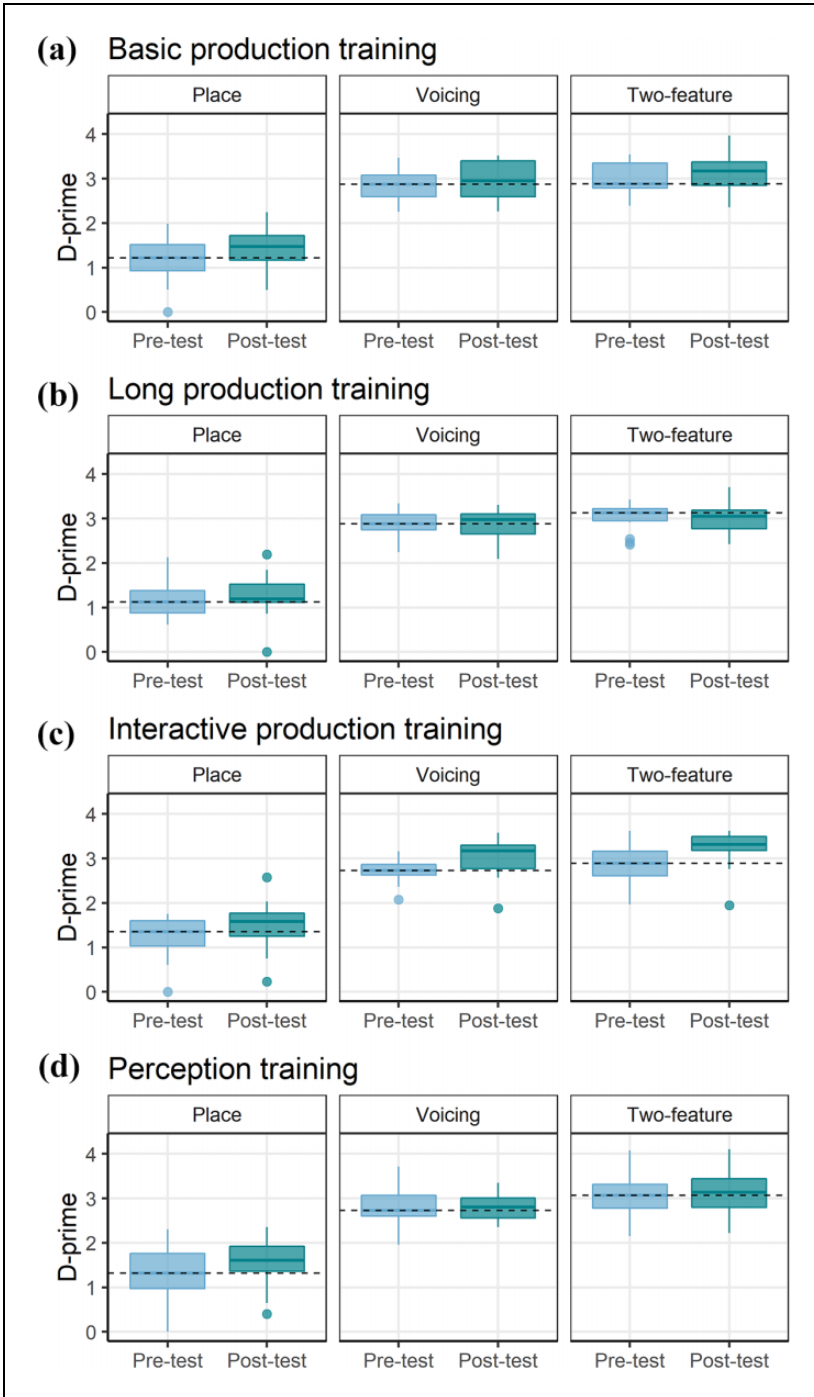


Figure 4. By-contrast boxplots of d' performance in the pre-test and post-test, Experiment 2.

Note. Each panel shows one of the four between-participant conditions.

3.51, $p = 0.061$). In the interactive condition (Figure 4C), the improvement from pre-test to post-test was larger for place trials (mean d' difference: 0.295) than for voicing trials (mean difference: 0.154). A marginal interaction of number of features, session, and experimenter intervention ($\beta = 0.300$, $\chi^2(1) = 3.61$, $p = 0.058$), coupled with follow-up models, found that improvement in the interactive condition was also larger for two-feature trials ($\beta = 0.160$, $\chi^2(1) = 4.33$, $p = 0.038$, mean difference 0.351) than for one-feature trials (mean difference: 0.227). L2 and L3 proficiency failed to reach significance ($\chi^2(1) < 1$, $p > 0.10$).

Reaction time model. As in Experiment 1, a reaction time model of 25,601 'different' trials with correct responses was constructed, with the same structure as the d' model but including a predictor for trial count (for full model table, see Appendix 3). There were significant effects of session ($\beta = -0.246$, $\chi^2(1) = 43.44$, $p < 0.001$), contrast type ($\beta = -0.202$, $\chi^2(1) = 51.83$, $p < 0.001$), and number of features ($\beta = -0.168$, $\chi^2(1) = 56.63$, $p < 0.001$); consistent with the direction of the d' model effects, post-test, voicing, and two-feature trials were all faster. However, some study type differences emerged in reaction times that did not in d' . The effect of training style was significant ($\beta = -0.325$, $\chi^2(1) = 6.99$, $p = 0.008$); participants in the perceptual training condition were faster to identify contrasts correctly than those in articulatory training conditions. This effect interacted with session ($\beta = -0.345$, $\chi^2(1) = 9.81$, $p = 0.002$); participants in perceptual training condition also showed a larger RT reduction at post-test.

A significant interaction of training style and number of features ($\beta = 0.147$, $\chi^2(1) = 6.01$, $p = 0.014$) revealed that the speed advantage for two-feature contrasts was driven by participants in the articulatory training conditions; the marginal interaction of experimenter intervention and number of features ($\beta = 0.108$, $\chi^2(1) = 3.64$, $p = 0.056$) revealed that this two-feature advantage was larger for participants in basic training than those in the interactive condition. Furthermore, the disparity between place and voicing trials was larger in the basic condition than the interactive condition ($\beta = 0.191$, $\chi^2(1) = 6.43$, $p = 0.011$).

3 Discussion

Experiment 2 provides evidence that participants can use both perceptual training and articulatory training to improve their discrimination of non-native categories. Even though the training was relatively short, participants showed improvement from pre-test to post-test. None of the two-way session by condition interactions were significant, indicating that participants did not differ in the overall degree of improvement based on the condition they were in.

Three-way interactions of condition, stimulus features, and session revealed that both place and voicing trials were improved by articulatory training conditions, while improvement in perceptual training was limited to place trials. This suggests that articulatory training helped learners with a more diverse set of cues. Place trials also improved more than voicing trials in the interactive condition, and one-feature trials more than two-feature trials in the long condition. As place trials, and to an extent all trials that contrasted in a single feature, were more difficult for learners at pre-test, these

patterns indicate that the more in-depth articulatory training conditions were helpful for especially difficult contrasts.

The reaction time model did reveal an advantage for learners in the perceptual training condition. This may reflect an advantage for perceptual training in the speed that it takes learners to reach a judgment about two stimuli. However, perceptual training in Experiment 2 was unique in using the same task (AX discrimination) during both training and testing, so it is also possible that participants in this condition were faster due to greater familiarity with the task.

Turning to the hypotheses outlined in Section III.1, Experiment 2 provides evidence against the null hypothesis that articulatory training is ineffective for improving discrimination of non-native contrasts. The results of Experiment 2 provide evidence for the local ceiling hypothesis; when participants started at baseline and received articulatory training first, their discrimination ability matched those of participants who received perceptual training. Evidence for the exposure hypothesis and the inflexible instructor hypotheses is weaker, as there were not strong differences between the articulatory training conditions in Experiment 2. However, the benefit of long articulatory training for one-feature trials, and interactive training for place trials – the more difficult contrasts – may reflect that length and quality do matter, even in an early and short training intervention.

IV General discussion

This study was designed to examine the efficacy of explicit articulatory training to improve perceptual discrimination of challenging non-native contrasts by novice listeners. Experiment 1 integrated perceptual and articulatory training into a single multi-session paradigm; evidence was found for improvement after perceptual training, but no added benefit for articulatory training was detected. Experiment 2 separated perceptual training and articulatory training into separate conditions for direct comparison. In that experiment, improvement in perceptual and articulatory training conditions were comparable, albeit with a speed advantage for perceptual training. Taken together, these studies suggest that articulatory training may be most effective for rapid perceptual acquisition when introduced at the beginning of learning, rather than as a supplement for learners with some prior exposure to the phoneme inventory of the second language. A study similar in structure to Experiment 1, but which presented articulatory training prior to perceptual training, could confirm this. This paradigm could also illuminate whether perceptual learning can have an additive beneficial effect on top of articulatory training, or whether a ceiling would be reached regardless of which training type preceded the other. In other words, are these training types truly qualitatively different, or is order of presentation the primary driver of these effects for novice learners?

Unlike Catford and Pisoni (1970), there was not a clear advantage for articulatory training over perceptual training in the present study. Given the ease of implementing perceptual training in the lab, and because of the speed advantage demonstrated in Experiment 2, it may be preferable in research contexts where the primary goal is to test the limits of non-native phoneme perception. However, for L2 learners, there is reason to believe that there are substantial benefits to targeted articulatory training. The acquisition

of categories in an L2 is integrated, not isolated to one domain, as learners typically must both comprehend and produce the L2. Therefore, training paradigms that build more integrated category representations may help learners improve both skills more quickly. A study of production targets using the paradigm reported here (Cibelli, 2020) found significant improvements in the production of non-native sounds after a single session of articulatory training. Given this, it may be a useful part of classroom approaches to L2 acquisition, for both production (Lord, 2005; Olson, 2014) and perception.

Experiment 2 does not provide unambiguous support for one training method over another. However, there was some evidence that discrimination showed greater improvement for challenging contrasts (place trials, one-feature trials) in the interactive and long articulatory training conditions, compared to basic articulatory training. This suggests that the quality of training can have an impact, particularly on the types of targets that learners are most likely to struggle with. The varied training conditions also highlight the possibility of individual variation. In the interactive session, some participants may have responded well to the teaching style of the experimenter, while others may not have. And in the long training sessions, the quality of one's own productions had the potential to reinforce – or derail – the development of perceptual categories. It is also possible that conducting the post-tests in Experiment 2 on a separate day, allowing for sleep consolidation (Earle and Myers, 2013, 2015; Fenn et al., 2003), may have shown more separation between training types.

1 Theoretical implications

While results from naive learners in Experiment 2 are encouraging for the role of articulatory training, the current study does not provide support for a strong version of theories that assume articulatory targets underlying perception, such as the Perceptual Assimilation Model (Best et al., 2001, 2009) and Direct Realist implementations of Motor Theory (Fowler, 2008; Galantucci et al., 2006). At the very least, articulatory representations do not have a distinct advantage as a pathway to perceptual discrimination for these learners.

However, the findings are consistent with accounts that argue for a facilitatory role for cross-modal information. Several accounts of speech perception argue that support from the motor system can be recruited to assist in adverse or challenging perceptual conditions. The extended Native Language Magnet Theory (Kuhl et al., 2008) emphasizes a link between perception and production as reinforcement during first language acquisition, as infants map from caregiver input to their own early productions. Recent neuroimaging data from this lab (Kuhl et al., 2014) suggests that motor pathways are most active during perception when target percepts are unfamiliar, while auditory pathways take precedence with well-known, familiar sounds. Work on the neural pathways of speech perception in adults provides an additional line of support for this account, arguing that motor pathways are more strongly recruited in noisy conditions, when the identity of a percept is less certain (Davis and Johnsruide, 2007; Hickok et al., 2011). The perception of challenging or unfamiliar L2 contrasts, arguably another 'adverse' listening condition, maps neatly onto this account. Even if the perceptual system does not depend on an articulatory representation for these sounds, it is

reasonable to expect that when available, these representations can be recruited to support a weak or fledgling perceptual category.

2 Future directions

While the differences between the stop consonant paradigms of English and Hindi provide a range of novel features to test, it stands to reason that other L1–L2 system mismatches may provide different challenges to the present account. In a well-designed articulatory learning paradigm, the specifics of the gestures being taught would need to be tailored to each novel phoneme in light of the L1 of the learning group. Therefore, the literature would benefit from additional studies which test the limits of articulatory-focused training on perceptual development. In addition, means of presenting articulatory information with real-time feedback have tremendous potential. For example, novel contrasts with visually-apparent features such as lip rounding may benefit from audiovisual instruction (Faytak, 2016; Hazan et al., 2005; Matsui, 1995), while contrasts marked by tongue position differences may benefit from ultrasound feedback (Tsui, 2012; Wilson, 2014).

Finally, the present study focused on novice learners at the earliest stages of acquisition. The results of Experiment 1 suggest that some experience with perceptual targets, however brief, may suppress the utility of cross-modal information. However, that does not exclude the possibility that articulatory training could be helpful at other stages of acquisition. Perhaps there is a U-shaped function of utility; production targets may be very useful to learners with no prior percepts, less helpful when some perceptual experience has been gained, and then beneficial again to experienced learners with stable percepts who can use the additional information to fine-tune their perception of challenging categories. Because non-native phonemes can remain challenging even at higher levels of L2 proficiency, studies with very experienced learners may reveal an alternative mechanism of perceptual learning in this population.

Acknowledgements

Dorothy Dao, Amanda Geib, Charlotte Hoerber, Aaliyah Ichino, Anna Lundborg, and Jocelyn Takahashi assisted with data collection and processing. The author thanks Susanne Gahl, Keith Johnson, and Robert Knight for comments on an early draft of this article.


Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by NSF Grant DGE-1106400 and funding from the Phi Beta Kappa Northern California Association.

ORCID iD

Emily Cibelli  <https://orcid.org/0000-0003-3137-4817>

Notes

1. In these cases, the wrong experiment script was run during the post-test session, impacting the ability to compare these three participants to those who had completed the study design as intended.
2. To ensure that the Helmert-style grouping did not obscure differences in comparing the re-test to previous sessions, two follow-up models comparing it to each previous session were run, using the same model selection procedure described above. These models confirmed that there was a significant difference between pre-test and re-test performance ($\beta = 0.353$, $\chi^2(1) = 6.58$, $p = 0.010$) but not between post-test and re-test ($\beta = -0.052$, $\chi^2(1) = 0.33$, $p = 0.563$). In other words, the numerical difference between post-test and re-test d-prime (d') values did not reflect a return to baseline performance at re-test.
3. This decline was not predicted, and is somewhat difficult to explain. One possibility is that the place contrast is difficult enough for listeners that, in the absence of experimental evidence cueing the contrast, listeners became even *more* confident in the assimilation of dental and retroflex tokens to a single coronal category.

References

- Akahane-Yamada R, Tohkura Y, Bradlow AR, and Pisoni DB (1996) Does training in speech perception modify speech production? *Proceedings of the Fourth International Conference on Spoken Language Processing*. Philadelphia, PA: IEEE, pp. 606–09.
- Baayen RH (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baese-Berk MM (2010) An examination of the relationship between speech perception and production. Unpublished PhD thesis, Northwestern University, Evanston, IL, USA.
- Baker W, Trofimovich P, Mack M, and Flege JE (2002) The effect of perceived phonetic similarity on non-native sound learning by children and adults. In: Skarabela B, Fish S, and Do AHJ (eds) *Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla, pp. 36–47.
- Bailey TM and Hahn U (2005) Phoneme similarity and confusability. *Journal of Memory and Language* 52: 339–62.
- Bates D, Kliegl R, Vasishth S, and Baayen H (2015a) *Parsimonious mixed models*. arXiv preprint 1506.04967.
- Bates D, Mächler M, Bolker B, and Walker S (2015b) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.
- Best CT (1995) A direct realist perspective on cross-language speech perception. In: Strange W (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. York: Timonium, MD, pp. 167–200.
- Best CT and Avery M (2007) Nonnative and second-language speech perception: Commonalities and complementarities. In: Flege JE, Bohn OS, and Munro MJ (eds) *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, pp. 13–34.
- Best CT, McRoberts GW, and Goodell E (2001) Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109: 775–94.

- Best CT, Goldstein L, Tyler MD, and Nam H (2009) Articulating the Perceptual Assimilation Model (PAM): Perceptual assimilation in relation to articulatory organs and their constriction gestures. *The Journal of the Acoustical Society of America* 125: 2758–58.
- Boersma P and Weenink D (2019) *Praat: Doing phonetics by computer: Version 6.0.50* [computer program]. Available at: <http://www.praat.org/> (accessed April 2020).
- Bradlow AR, Akahane-Yamada R, Pisoni DB, and Tohkura Y (1999) Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics* 61: 977–85.
- Bradlow AR, Pisoni DB, Akahane-Yamada R, and Tohkura T (1997) Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299–2310.
- Catford JC and Pisoni DB (1970) Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal* 54: 477–81.
- Cibelli EC (2020) Training non-native consonant production with perceptual and articulatory cues. *Phonetica* 77: 1–28.
- Davidson L (2016) Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54: 35–50.
- Davis MH and Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing Research* 229: 132–47.
- Diaz B, Baus C, Escera C, Costa A, and Sebastián-Gallés N (2008) Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning. *Proceedings of the National Academy of Sciences* 105: 16083–88.
- Earle FS and Myers EB (2013) Building phonetic categories: an argument for the role of sleep. *Frontiers in Psychology* 5: 1192.
- Earle FS and Myers EB (2015) Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance* 41: 1680–95.
- Escudero P, Benders T, and Wanrooij K (2011) Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America* 130: EL206–12.
- Faytak M (2016) Articulatory habit versus adaptive flexibility in L2 phone learning. *The Journal of the Acoustical Society of America* 140: 3340–3340.
- Fenn KM, Nusbaum HC, and Margoliash D (2003) Consolidation during sleep of perceptual learning of spoken language. *Nature* 425: 614–16.
- Flege JE (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (ed) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore, MD: York Press, 233–77.
- Flege JE, Bohn OS, and Jang S (1997) Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25: 437–70.
- Fowler CA (1986) An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14: 3–28.
- Fowler CA (1989) Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology* 1: 145–60.
- Fowler C (2008) The FLMP STMPed. *Psychonomic Bulletin and Review* 15: 458–62.
- Galantucci B, Fowler C, and Turvey M (2006) The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review* 13: 361–77.
- Goldstein L and Fowler CA (2003) Articulatory phonology: A phonology for public language use. In: Schiller NO and Meyer AS (eds) *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin: Mouton de Gruyter, pp. 159–207.

- Golestani N (2014) Brain structural correlates of individual differences at low-to-high levels of the language processing hierarchy: A review of new approaches to imaging research. *International Journal of Bilingualism* 18: 6–34.
- Golestani N and Zatorre RJ (2004) Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage* 21: 494–506.
- Goudbeek M, Cutler A, and Smits R (2008) Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication* 50: 109–25.
- Guion S, Flege J, Akahane-Yamad AR, and Pruitt J (2000) An investigation of current models of second language speech perception: The case of Japanese adults perception of English consonants. *The Journal of the Acoustical Society of America* 107: 2711–24.
- Gulian M, Escudero P, and Boersma P (2007) Supervision hampers distributional learning of vowel contrasts. In: *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: University of Saarbrücken, pp. 1893–96.
- Hattori K and Iverson P (2009) English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America* 125: 469–79.
- Hayes-Harb R (2007) Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research* 23: 65–94.
- Hazan V, Sennema A, Iba M, and Faulkner A (2005) Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47: 360–78.
- Hickok G, Houde J, and Rong F (2011) Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69: 407–22.
- Hombert JM, Ohala JJ, and Ewan WG (1979) Phonetic explanations for the development of tones. *Language* 55: 37–58.
- Iverson P and Evans BG (2009) Learning English vowels with different first-language vowel systems ii: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America* 126: 866–77.
- Iverson P, Kuhl PK, Akahane-Yamada R, et al. (2003) A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–57.
- Jamieson DG and Morosan DE (1986) Training non-native speech contrasts in adults: Acquisition of the English /delta/-/theta/ contrast by Francophones. *Perception and Psychophysics* 40: 205–15.
- Jia G, Strange W, Wu Y, Collado J, and Guan Q (2006) Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America* 119: 1118–30.
- Kartushina N, Hervais-Adelman A, Frauenfelder UH, and Golestani N (2015) The effect of phonetic articulatory training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America* 138: 817–32.
- Kuhl PK (2000) A new view of language acquisition. *Proceedings of the National Academy of Sciences* 97: 11850–57.
- Kuhl PK, Conboy B, and Coffey-Corina S (2008) Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 979–1000.
- Kuhl PK, Ramirez RR, Bosseler A, Lin JFL, and Imada T (2014) Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences*: 11238–45.
- Lacabex EG, García Lecumberri M, and Cooke M (2008) Identification of the contrast full vowel–schwa: Training effects and generalization to a new perceptual context. *Ilha do Desterro* 55: 173–96.

- Lai YH (2009) Asymmetry in Mandarin affricate perception by learners of Mandarin Chinese. *Language and Cognitive Processes* 24: 1265–85.
- Lametti DR, Nasir SM, and Ostry DJ (2012) Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience* 32: 9351–58.
- Levy ES and Law FF (2010). Production of French vowels by American–English learners of French: Language experience, consonantal context, and the perception–production relationship. *The Journal of the Acoustical Society of America* 128: 1290–1305.
- Lim SJ and Holt LL (2011) Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science* 35): 1390–1405.
- Lisker L and Abramson AS (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422.
- Lively SE, Logan JS, and Pisoni DB (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: 1242–55.
- Lord G (2005) (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania* 88: 557–67.
- Macmillan NA and Creelman CD (2004) *Detection theory: A user's guide*. Hove: Psychology Press.
- Mathôt S, Schreij D, and Theeuwes J (2012) OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavioral Research Methods* 44: 314–24.
- Matsui JK (1995) The use of audio-visual aids in the language laboratory to teach lip-rounding. *Language Laboratory* 32: 169–76.
- McCandliss BD, Fiez JA, Protopapas A, Conway M, and McClelland JL (2002) Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience* 2: 89–108.
- Olson DJ (2014) Phonetics and technology in the classroom: a practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania* 97: 47–68.
- Pederson E and Guion-Anderson S (2010) Orienting attention during phonetic training facilitates learning. *Journal of the Acoustical Society of America* 127: EL54–59.
- Pegg JE and Werker JF (1997) Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America* 102: 3742–53.
- Pimsleur P (1963) Discrimination training in the teaching of French pronunciation. *The Modern Language Journal* 47: 199–203.
- Protopapas A and Calhoun B (2000) Adaptive phonetic training for second language learners. In: Delcloque P (ed) *Proceedings of the 2nd International Workshop on Integrating Speech Technology in Language Learning*. Dundee: University of Abertay, pp. 31–38.
- Pruitt JSJ (1995) The perception of Hindi dental and retroflex stop consonants by native speakers of Japanese and American English. Unpublished PhD thesis, University of South Florida, Tampa, FL, USA.
- Pruitt JSJ, Jenkins JJ, and Strange W (2006) Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America* 119: 1684–96.
- R Core Team (2018) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: <https://www.R-project.org> (accessed April 2020).
- Sadakata M and McQueen JM (2013) High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America* 134: 1324–35.

- Sadakata M and McQueen JM (2014) Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology* 5: 1318.
- Saito K and 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: 595–633.
- Schiefer L (1986) F0 in the production and perception of breathy stops: Evidence from Hindi. *Phonetica* 43: 43–69.
- Schneiderman E, Bourdages J, and Champagne C (1988) Second-language accent: The relationship between discrimination and perception in acquisition. *Language Learning* 38: 1–19.
- Seitz AR and Watanabe T (2003) Psychophysics: Is subliminal learning really passive? *Nature* 422: 36.
- Sheldon A and Strange W (1982) The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3: 243–61.
- Song JH, Skoe E, Wong PCM, and Kraus N (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience* 20: 1892–1902.
- Studdert-Kennedy M and Goldstein L (2003). Launching language: The gestural origin of discrete infinity. *Studies in the Evolution of Language* 3: 235–54.
- Tees RC and Werker JF (1984) Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38: 579–90.
- Terrace HS (1963) Discrimination learning with and without 'errors'. *Journal of the Experimental Analysis of Behavior* 6: 1–27.
- Tremblay S, Shiller DM, and Ostry DJ (2003). Somatosensory basis of speech production. *Nature* 423: 866–69.
- Tsui HML (2012) Ultrasound speech training for Japanese adults learning English as a second language. Unpublished master's thesis, University of British Columbia, Vancouver, BC, Canada.
- Vlahou E, Protopapas A, and Seitz A (2011) Implicit training of nonnative speech stimuli. *Journal of Experimental Psychology: General* 141: 1–19.
- Wang X (2013) Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal* 97: 144–60.
- Wang Y, Jongman A, and Sereno JA (2003) Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America* 113: 1033–43.
- Wang Y, Spence MM, Jongman A, and Sereno JA (1999) Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America* 106: 3649–58.
- Wilson C, Davidson L, and Martin S (2014) Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language* 77: 1–24.
- Wilson I (2014) Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology* 35: 285–89.
- Wright BA, LeBlanc EK, Conderman JS, and Coburn CS (2015) Contributions of practice with feedback and testing without feedback to learning of a non-native phonetic contrast. *The Journal of the Acoustical Society of America* 137: 2384–84.
- Zampini ML (1998). The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education* 3: 85–100.

Appendix 1. Fixed effects of the Experiment I reaction time model.

	Estimate	Standard error	t	χ^2	p (χ^2)
Trial count	-0.026	0.008	-3.37	11.35	< 0.001
Contrast type	-0.169	0.028	-6.00	35.85	< 0.001
Number of features	-0.171	0.024	-7.11	50.40	< 0.001
Session (pre-test vs. post-test)	-0.413	0.069	-5.98	15.61	< 0.001
Session (pre-/post-test vs. re-test)	-0.279	0.046	-6.00	15.69	< 0.001
L2 proficiency	-0.069	0.081	-0.85	0.70	0.402
L3 proficiency	-0.020	0.091	-0.22	DNC	DNC
Days to completion	-0.040	0.090	-0.45	0.20	0.657
Contrast * session (pre-test vs. post-test)	-0.002	0.072	-0.03	0.00	0.976
Contrast * session (pre-/post-test vs. re-test)	-0.008	0.061	-0.13	0.02	0.898
Features * session (pre-test vs. post-test)	-0.032	0.057	-0.56	0.31	0.577
Features * session (pre-/post-test vs. re-test)	0.048	0.049	0.97	0.94	0.332

Note. The model included decorrelated by-participant random slopes for both session predictors. DNC indicates a χ^2 model comparison in which the model with the held-out predictor did not converge.

Appendix 2. Fixed effects for the Experiment 1B (control study) d-prime (d') model.

	Estimate	Standard error	t	χ^2	p (χ^2)
Contrast type	1.731	0.094	18.36	32.92	< 0.001
Number of features	1.474	0.070	21.07	35.29	< 0.001
Session 1 vs. 2	-0.172	0.134	-1.29	1.52	0.218
Session 1/2 vs. 3	-0.08	0.088	-0.92	0.79	0.374
Session 1/2/3 vs. 4	0.056	0.089	0.63	0.39	0.534
L2 proficiency	-0.129	0.106	-1.22	1.38	0.240
L3 proficiency	0.138	0.106	1.30	1.55	0.212
Contrast * session (1 vs. 2)	0.210	0.093	2.25	4.82	0.028
Features * session (1 vs. 2)	0.272	0.108	2.52	5.99	0.014
Contrast * session (1/2 vs. 3)	0.399	0.088	4.56	17.86	< 0.001
Features * session (1/2 vs. 3)	0.302	0.096	3.15	9.16	0.003
Contrast * session (1/2/3 vs. 4)	0.143	0.080	1.79	3.13	0.077
Features * session (1/2/3 vs. 4)	0.058	0.089	0.65	0.42	0.516

Note. The model included decorrelated by-participant random slopes for contrast type, number of features, and all three session predictors.

Appendix 3. Fixed effects of the Experiment 2 reaction time model.

	Estimate	Standard error	t	χ^2	p (χ^2)
Trial count	-0.064	0.009	-7.51	39.82	< 0.001
Session	-0.246	0.031	-7.92	43.44	< 0.001
Contrast type	-0.202	0.023	-8.86	51.83	< 0.001
Number of features	-0.168	0.018	-9.61	56.63	< 0.001
Training style	-0.325	0.118	-2.74	6.99	0.008
Experimenter intervention	0.035	0.114	0.31	0.10	0.755
Study length	0.105	0.109	0.96	0.89	0.344
L2 proficiency	-0.064	0.033	-1.92	3.37	0.067
L3 proficiency	-0.016	0.035	-0.46	0.20	0.652
Session * contrast type	-0.04	0.032	-1.27	1.62	0.204
Session * number of features	-0.018	0.024	-0.75	0.56	0.454
Session * training style	-0.345	0.105	-3.29	9.81	0.002
Session * experimenter intervention	0.056	0.099	0.56	0.32	0.575
Session * study length	0.069	0.100	0.70	0.47	0.491
Contrast type * training style	0.112	0.076	1.46	2.09	0.148
Number of features * training style	0.147	0.058	2.52	6.01	0.014
Contrast type * study length	-0.100	0.075	-1.33	1.72	0.190
Number of features * study length	-0.012	0.057	-0.20	0.04	0.841
Contrast type * experimenter intervention	0.191	0.073	2.60	6.43	0.011
Number of features * experimenter intervention	0.108	0.056	1.93	3.64	0.056
Session * contrast type * train. type	0.024	0.105	0.23	0.05	0.822
Session * number of features * train. type	-0.161	0.081	-2.00	3.98	0.046
Session * contrast type * study length	0.041	0.108	0.38	0.14	0.709
Session * number of features * study length	0.091	0.082	1.11	1.24	0.266
Session * contrast type * experimenter intervention	0.034	0.102	0.33	0.11	0.741
Session * number of features * experimenter intervention	-0.089	0.078	-1.15	1.31	0.252

Note. The model included correlated by-participant random slopes for trial count, session, contrast type, number of features, and session*contrast type.