HHS Public Access

Author manuscript

J Phon. Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

J Phon. 2020 January ; 78: . doi:10.1016/j.wocn.2019.100940.

Training a non-native vowel contrast with a distributional learning paradigm results in improved perception and production

Heather Kabakoff,

Department of Communicative Sciences & Disorders, New York University, New York, NY

Gretchen Go,

Department of Communicative Sciences & Disorders, New York University, New York, NY

Susannah V. Levi

Department of Communicative Sciences & Disorders, New York University, New York, NY

Abstract

Previous distributional learning research suggests that adults can improve perception of a non-native contrast more efficiently when exposed to a bimodal than a unimodal distribution. Studies have also suggested that perceptual learning can transfer to production. The current study tested whether the addition of visual images to reinforce the contrast and active learning with feedback would result in lcearning in both conditions and would transfer to gains in production. Native English-speaking adults heard stimuli from a bimodal or unimodal /o/-/œ/ continuum. No group differences were found on a discrimination task, possibly suggesting that the supports eliminated previously documented group differences. On an identification task, listeners in the bimodal group showed better performance than the unimodal group on the endpoint stimuli. Production results indicated that both groups showed increased Euclidean distance between the target vowels after training, suggesting that perceptual training improved production skills in both conditions. Contrary to expectations, degree of perception and production learning were not correlated. Together, these results suggest that a bimodal distribution may aid learning, but that adding images to reinforce the contrast and active learning to the training paradigm could mitigate disadvantages found previously for participants exposed to a unimodal distribution.

Introduction

Learning to speak a second language with native-like proficiency can be challenging, especially for adults. Part of the challenge is learning to perceive and produce speech sounds that are not in the first language (L1). Research on the perception of native and non-native contrasts shows a shift in perception across the first year of life. Infants begin life able to discriminate acoustic variants within and across speech sound categories in all languages (Trehub, 1973, 1976; Werker & Tees, 1984). Within a year, they lose the ability to perceive non-native phoneme contrasts (Polka & Werker, 1994; Werker & Tees, 1984), while

Corresponding Author, heather.kabakoff@nyu.edu.

simultaneously learning to ignore some within-category acoustic-phonetic variability for native phonemes (Kuhl et al., 1992). These findings led researchers to explore possible underlying mechanisms that support this process of perceptual attunement.

Distributional learning as a mechanism for learning non-native contrasts

One possible mechanism is that perceptual attunement stems from sensitivity to statistical patterns in the linguistic input. Maye et al. (2002) tested whether sensitivity to distributions in the input would affect perception of non-native sound contrasts. English-learning infants were presented with tokens drawn from an eight-step acoustic continuum of voice onset time (VOT), ranging from pre-voicing to short-lag VOT. Importantly, this contrast does not exist in English. Infants were either presented with tokens drawn mostly from near the endpoints of the continuum (stimuli 2 and 7), which resulted in a bimodal distribution of input, or with tokens drawn primarily from the center of the continuum (stimuli 4 and 5), which resulted in a unimodal distribution of sounds in the input. Following training, Maye et al. (2002) tested whether infants could discriminate stimuli 3 and 6, which were presented to both groups an equal number of times. Infants in the bimodal condition successfully discriminated these stimuli, whereas infants in the unimodal condition did not. Maye et al. (2002) argued that exposure to different distributions resulted in a change in perception, where infants in the bimodal condition inferred two categories, but those in the unimodal group did not. Maye et al. (2002) concluded that infants learn speech sound categories with sensitivity to the distributions of the sounds to which they are exposed.

Researchers have also sought to examine whether similar paradigms could be used to train non-native sound contrasts in adult listeners. The acquisition of non-native speech sound categories is different for adults compared to infants, as infants are tuning perception to their native language, whereas adults learning a non-native contrast already have a phonological system in place for their native language. Abundant research on the perception of second language (L2) sound categories shows that the relationship between the L1 and L2 sounds affects perception of the L2 sounds (Best, 1991; Escudero, 2005; Flege, 2003; Kuhl & Iverson, 1995). Despite the difference between learning a novel contrast for infants versus adults, a few studies have provided some evidence for the benefits of using a distributional learning paradigm to train L2 speech sounds in adults.

Similar to the study with infants, Maye and Gerken (2000) presented English-speaking adults with synthesized stimuli along an eight-step VOT continuum ranging from pre-voiced to short-lag for an alveolar contrast. As in the infant study, training stimuli were either drawn from a bimodal or unimodal distribution. To ensure that adults were attending to the training task, they were asked to check a box following presentation of each stimulus, but were otherwise listening passively to the stimuli. Using a discrimination task with only the endpoint stimuli (1 and 8), which were presented the same number of times in both the unimodal and bimodal conditions, the authors found that adults in the bimodal group discriminated these stimuli significantly better than those in the unimodal group. Similar results were found for a velar stop contrast in Maye and Gerken (2001). The authors interpreted these findings as evidence that adults can learn a non-native contrast through exposure to a bimodal distribution of a speech sound contrast.

Hayes-Harb (2007) tested whether a distributional learning paradigm or a lexical contrast would lead to better discrimination by adults. Participants were exposed to sounds along a pre-voiced to short-lag VOT eight-step continuum in one of six conditions: unimodal with no images; bimodal with no images; stimuli 2 and 7 with no images ("two-seven"); stimuli 2 and 7 with one image ("no contrast"); stimuli 2 and 7 with two images ("contrast"); no training at all ("control"). The contrast condition was included to determine whether the addition of pictures used to reinforce the contrast (termed "lexical support" in the original study) would facilitate learning. Similar to the previous studies, the adults listened passively as the stimuli were presented during the exposure phase. As found in previous studies, participants in the bimodal group performed better than those in the unimodal group. In addition, participants in the contrast group (with two images who only heard stimuli 2 and 7) performed significantly better on the discrimination task than those in either the no contrast group (who had one image) or the two-seven group (no images), showing that the addition of contrasting pictures improves discrimination of a novel sound contrast. Two other important findings emerged. First, participants in the contrast group outperformed those in the bimodal group, suggesting that the addition of images to support a contrast leads to better discrimination than exposure to a distribution alone. Second, the no training control group showed relatively good discrimination despite a lack of training. Analyses showed that this group did not significantly differ from either the bimodal group or the contrast group, but did perform significantly better than the unimodal group, suggesting that the unimodal condition actually serves to suppress the contrast. In her dissertation, Hayes (2003) tested four groups of participants where half in both the unimodal and bimodal conditions were shown one image for all 8 stimuli, and half were shown one image for stimuli 1–4 and a different image for stimuli 5–8. To ensure that participants were attending to the task, they were asked to check a box after each trial, but otherwise were passive during the listening portion. The bimodal condition with two images led to significantly better discrimination than a unimodal condition with one image, suggesting some benefit of two images, as assigning sounds to one of two images could help draw attention to the contrast. Critical to the current study, Hayes (2003) made no explicit comparison between the bimodal condition with two images and the unimodal condition with two images. The data indicate a numerical difference (29.6% accuracy for the bimodal with two images and 17.4% for unimodal with two images), but another comparison of more extreme differences (29.6% accuracy for the bimodal with two images and 15.9% for bimodal with one image) yielded a non-significant finding with a two-tailed test. Thus, it remains unclear whether the addition of contrasting images could mitigate the suppression of the contrast, as was found in Hayes-Harb (2007), where the unimodal group with one image performed worse than the bimodal group with one image.

Baese-Berk (2010) also replicated and extended the findings from the previous studies of distributional learning with adults. Also using a negative to short-lag VOT continuum, Baese-Berk (2010) presented listeners with only the extreme combination of conditions: either a bimodal distribution with two images or a unimodal distribution with a single image. As with the other studies using adult participants, the exposure task involved implicit learning, where listeners passively listened to each stimulus and then indicated that they completed a trial by pressing a key to advance to the next trial. An important modification of

Baese-Berk (2010) was that training spanned two days. As expected, the bimodal group with two images demonstrated incremental learning across the two days on a discrimination task, whereas the unimodal group with one image did not.

Whereas the above studies all tested the perception of a temporal VOT contrast, Escudero et al. (2011) and Wanrooij et al. (2015) used a distributional learning paradigm to test the perception of vowels. In Escudero et al. (2011), native Spanish speakers were trained on the Dutch /ɑ/-/aː/ contrast. Rather than comparing unimodal versus bimodal groups, they compared two different bimodal groups, one where the continuum was enhanced (F1 range: 600–885 Hz, F2 range: 1000–1430 Hz) and one where it was compressed (F1 range: 700–795 Hz, F2 range: 1115–1330 Hz). They found a trend where participants in the enhanced group performed better than those in the compressed group. In a follow-up study, Wanrooij et al. (2015) created two conditions where various measures of dispersion (e.g., range, standard deviation) were matched between the bimodal and unimodal distributions. In contrast to previous work, they found no differences between the two groups. They concluded that previous studies of distributional learning may have found differences in learning not due to the number of peaks (two versus one), but instead due to differences in the dispersion of the stimuli that were presented. Importantly for the current study, these researchers found no effect of input distribution.

Finally, the aforementioned studies on distributional learning utilized a passive learning paradigm (i.e., implicit/reflexive/unsupervised) to demonstrate whether speech categorization can occur implicitly. As this line of research has evolved into a line of research attempting to optimize the training paradigm for learning a non-native contrast, it is important to consider how training approaches that incorporate an active learning paradigm (i.e., explicit/reflective/supervised) have yielded promising results. A recent distributional learning study examined how active learning with feedback could cause listeners to shift which acoustic cues they used to identify a stop-voicing contrast in the native language of the listeners (Harmon et al., 2019). Participants were presented with either a unimodal or bimodal distribution of VOT and also with support of a secondary cue to voicing in stops (fundamental frequency). Importantly for the current study, no effect of input distribution was found when feedback was provided. Instead, participants whose feedback reinforced the alternate voicing cue (fundamental frequency) changed their perception to depend on the secondary cue. In a similar study, Goudbeek et al. (2008) trained Spanish and English listeners on non-native Dutch vowel contrasts differing in either duration (/ʏ/~/ø/) or formant frequency (/ʏ/~/ø/). They found that listeners more easily learned the contrast with the dimension most relevant in their language (formant frequency in Spanish; duration in English), and that accuracy feedback increased learning in both conditions. This suggests that accuracy feedback can improve learning of contrasts whether familiar or novel acoustic cues are trained. The incorporation of accuracy feedback may actually lead to such strong learning effects that previously documented differences found between passive learning paradigms are no longer present when feedback is provided. Training Japanese listeners on the /ɹ/~/l/ contrast, McCandliss et al. (2002) found the expected advantage for listeners exposed to an adaptive condition (i.e., two sounds start out maximally different and become more similar over the course of training) compared with a fixed condition. However, this group difference was no longer present when participants were provided accuracy feedback.

This finding indicates that accuracy feedback leads to more robust learning than when listeners are provided adjusted input in a passive learning paradigm.

Taken together, previous work suggests that adults exposed passively to a bimodal distribution can learn a non-native contrast. However, these previous studies also suggest that a variety of other factors may contribute to, or even eliminate, changes in perception following training. In particular, these factors include the addition of images to support one versus two labels and the incorporation of active learning with feedback. Furthermore, possible differences in the mechanisms involved in learning consonant categories versus vowel categories may also impact individuals' abilities to learn novel speech categories. Thus, a goal of the current study is to explore whether the combination of these factors could enhance learning of a non-native contrast in the unimodal condition or whether the benefits of a bimodal distribution would still remain.

Relationship between perception and production

When learning a second language, individuals must not only learn to perceive the sound contrasts in a second language, they must also learn to produce these same contrasts. Studies of spontaneous imitation within a speaker's native language suggest that the auditory input can change a speaker's productions. For example in shadowing studies where participants repeat words or sentences heard over headphones, shadowed productions are more similar perceptually and acoustically to the productions from the target speaker than are baseline productions (Goldinger, 1998; Mitterer & Ernestus, 2008; Shockley et al., 2004). In terms of explicit perceptual training, several studies from different disciplines have explored whether perceptual training can transfer to improvement in production.

Bradlow et al. (1997) investigated how perceptual training impacted both perception and production of /r/-/l/ by Japanese learners of English. During training, participants listened to five different speakers producing minimal pairs containing the target phonemes, identified whether they heard an /r/ or an /l/, and then were provided with feedback. Across training, listeners' perception of this non-native contrast improved. In addition, even though production was not directly trained, words produced after training were rated as better productions than those produced prior to training. This suggests that learning to perceive the contrast between /r/ and /l/ transferred to improved production of the same contrast. However, even though the learners improved in both perception and production overall, individual perception and production abilities were not found to be correlated, highlighting the presence of individual variation. In a follow-up study, Bradlow et al. (1999) found that these improvements in perception and production were maintained three months later. These findings suggest that there was a reliable and lasting transfer of learning from perception to production.

Baese-Berk (2010) also examined both perception and production of non-native contrasts in a perceptual training task. As mentioned above, participants completed training with a distributional learning paradigm to learn the non-native pre-voiced to short-lag VOT contrast. In addition to perceptual measures, participants also completed a production task in which they were asked to repeat the endpoint stimuli from the acoustic continuum. A three-way interaction was found between training group (bimodal with two images vs. unimodal

with one image), day (1 vs 2) and endpoint token (1 vs 8). Numerically, participants in the bimodal group showed a larger change in the VOT difference (2.3 ms on day 1 versus 3.8 ms on day 2), although post-hoc comparisons were not conducted. An additional analysis in the bimodal group examined the relationship between perceptual learning and production to determine whether individual differences in perception predict performance on the production task. The regression model with perception (discrimination) performance resulted in a significantly better fit than one without, suggesting that perception and production are linked such that those who reach the greatest level of accuracy on the perception task are more likely to also show the greatest distinction between the two target sounds in production.

In addition to second language learning, researchers have also explored whether perceptual training can improve speech production in children with speech sound disorders. Rvachew (1994) explored the benefits of perceptual training on the production of /ʃ/ for preschoolers who exhibited difficulty producing this sound. The preschool-aged children were either given perception training on the word *shoe* (trained on /ʃ/), the words *shoe* and *moo* (trained on /ʃ/), or the words *cat and Pete* (not trained on /ʃ/). Children were instructed to identify whether the stimulus was produced correctly or with an error and were provided with immediate feedback. Following six weeks of perceptual training, a post-training production test was administered. The groups that received perceptual training for /ʃ/ showed greater gains in production than the control group that did not receive training for /ʃ/. This suggests that perceptual training transfers to gains in production for children with speech sound disorders. Similar results were found in Jamieson and Rvachew (1992) and Rvachew et al. (2004).

Individual differences in perceptual learning

While some previous studies of the perception-production link have examined whether individual differences in perceptual skills predict which participants will improve in production, few studies have examined which participants are likely to improve in the perceptual domain. Several researchers have suggested that a participant's performance on perceptual tasks may be related to working memory (Kong & Edwards, 2016; Manis et al., 1997; McBride-Chang, 1996). Kong and Edwards (2016) found that inhibition and task shifting ability were related to individual categorization patterns (degree of categoricity) and hypothesized that working memory plays a role in learning new categories. In studies investigating speech perception in young children, the relationship between performance on perceptual tasks and working memory has been described as bidirectional, where the process of perceptual attunement builds working memory, which in turn, leads to greater perception of speech sound contrasts (Manis et al., 1997; McBride-Chang, 1996). For these reasons, we administered several tests of working memory to ensure that there are no differences in working memory skills between the groups.

Current study

The current study extends the research on distributional learning by modifying the procedure in two ways in order to optimize learning and allow us to test whether participants in the unimodal condition can learn a non-native contrast. First, as in previous studies including

Hayes (2003), the current procedure involves the use of different images to support category learning. One difference in our design is that participants in both the unimodal and bimodal group were provided with two images. We used two images in both the bimodal and unimodal conditions to directly assess whether those in a bimodal condition would continue to outperform those in a unimodal condition given this additional support. Second, we also included accuracy feedback to add an active component to the learning task. Previous studies of distributional learning with adults have involved passive learning tasks during which participants check a box or press a button after hearing each stimulus. As intended, this paradigm closely mirrors the passive way in which infants are exposed to sounds in their native language, but it may be that some form of active engagement with each stimulus is beneficial to adults. Given the previous discussion of learning benefits found in the presence of accuracy feedback in various training paradigms, we were interested in whether active learning would lead to enhanced learning for a non-native contrast. Additionally, we were interested in whether previously found differences between unimodal and bimodal groups would be affected by the incorporation of accuracy feedback, as found with adjusted versus fixed conditions in McCandliss et al. (2002) and in the cue-reweighting in Harmon et al. (2019).

The majority of studies of distributional learning have used a temporal VOT contrast. Like Escudero et al. (2011) and Gulian et al. (2007), the current study examines perceptual learning of a vowel contrast. Previous research has found differences in how individuals perceive vowels versus consonants (Fry et al., 1962), where both identification and discrimination performance appears to be more categorical for consonants than for vowels. Thus, perceptual training of a vowel contrast could lead to different learning patterns than for a consonantal contrast, as in most previous studies using a distributional learning paradigm. In addition, in American English, most dialectal differences are manifested as differences in vowels (Labov et al., 2006), thus adult listeners' perception may be more flexible in learning to perceive a vowel contrast, which may manifest as longer learning trajectories for non-native vowel contrasts than for consonant contrasts.

When selecting a vowel contrast, we considered which non-native vowel contrast would be most difficult for speakers of American English to learn. Several studies have demonstrated that the front-back contrast between rounded vowels in French (Levy, 2009a, 2009b; Levy & Strange, 2008) and German (Polka, 1995; Strange et al., 2009) is difficult for American English listeners to perceive. In an alveolar context, the discrimination error rate was 27% for the /y/-/u/ contrast and 38% for the /œ/-/o/ contrast (Levy, 2009b). In the current study, we use the French /œ/-/o/ contrast in an alveolar context because of its high confusability and because production of /d/ in the /dyt/ context is often affricated, thus providing prevocalic acoustic cues.

Finally, in addition to these changes in study design, we also tested working memory for the participants to ensure that there were no group differences. The training paradigm in the current study included two images and accuracy feedback while participants were exposed to either a unimodal or bimodal distribution of an /œ/-/o/ contrast. In addition to these two primary supports, the training spanned two days to provide multiple training blocks and to allow for the inclusion of the working memory tests. Though not intended as an additional

support, we acknowledge that the ability to consolidate information overnight can facilitate learning new categories (Earle et al., 2017; McGregor, 2014). We are interested in whether the addition of these supports will allow listeners in both conditions to learn to perceive and produce a novel contrast. Results supporting this hypothesis would contradict previous research demonstrating a unimodal disadvantage (i.e., bimodal advantage) characterized by suppression of the perception of a novel category (Hayes-Harb, 2007). We ask the following four questions:

1. Do the participants in the unimodal condition learn to perceive the non-native vowel contrast differently than those in the bimodal condition based on a perceptual discrimination task?
2. Do the participants in the unimodal condition learn to perceive the non-native vowel contrast differently than those in the bimodal condition based on a perceptual identification task?
3. Does perceptual training transfer to gains in production of the same non-native vowel contrast?
4. Is perceptual skill (discrimination/identification) associated with production abilities after training?

Experimental/Materials and methods

Participants

Thirty-two adults ages 18–30 participated in the study (6 male, 26 female). All were native speakers of American English, passed a hearing screening at 500, 1000, 2000, and 4000 Hz at 25 dB HL, and had no history of a speech or language disorder. All participants were compensated for their time. Participants reported having spoken or studied the following languages: Spanish (n = 23), Italian (n = 2), Hindi (n = 2), Gujarati (n = 1), Polish (n = 1), and American Sign Language (n = 1). Half were randomly assigned to the bimodal (13 females; 3 males) and half to the unimodal condition (13 females; 3 males). Within these groups, half were assigned to stimulus order A and half to stimulus order B (see procedure for description). The age breakdown of each condition is included in Table 1. Two additional participants completed the experiment but were not included in data analysis. Exclusion of these participants is described in the statistical analysis section.

Stimuli

A 29-year-old male native speaker of Parisian French who had been living in the United States for five years recorded a set of French nonwords. The speaker was recorded in a sound-attenuated booth using a head-mounted Shure 10-A unidirectional (cardioid) condenser microphone with a flat frequency response from 40 to 20,000 Hz. Productions were digitized into 16-bit stereo recordings via a Fostex FR-2LE field recorder at 44.1 kHz and transferred via Compact Flash card to a computer. The speaker produced five repetitions of nine French nonwords in a carrier phrase as they appeared in random order on a computer screen. The carrier phrase, "*J'ai dit* _____ *à des amis*" ("I said ___ to my friends"), was adapted from Levy (2009a). The nonwords consisted of each of the vowels (/œ, o, i, y, e, ɛ,

a, ɔ, u/) in the /radVt/ context, and were presented following French orthographic conventions.

The stimuli recordings were downsampled to 11025 Hz and amplitude normalized in Praat (Boersma & Weenink, 2019). Vowels were spliced out of the /radVt/ context and included the release burst of the /d/ up through the final visible peak in the acoustic waveform before the /t/. Duration measures were taken from the waveform and confirmed with the spectrogram. A Praat script was used to extract the first three formant frequencies (F1, F2, F3) from the midpoint of the vowel.

The productions of /œ/ and /o/ that were most similar in F1 and duration while showing the greatest differences in F2 values were selected as the base vowels for synthesis. To create an eight-step continuum, stimuli were synthesized with linear predictive coding in Matlab (MathWorks Inc., 2000). A window size of 256 samples was used (corresponding with the sampling rate) with a hop size of 128 samples and an LPC order of 12 (corresponding to the order of the synthesis filter). The /œ/ sample was used as the source, providing the excitation portion of the output signal, and the /o/ sample was used as the destination, providing the spectral filter to be applied to the source signal. The mix percentage was increased from 0% (the source signal synthesized with its own spectral filter) up to 100% (the source signal synthesized with the destination's spectral filter completely) in 2% increments. Filter coefficients describing the spectral envelope of each input signal were computed using Levinson-Durbin recursive autocorrelation. To obtain interim stimuli, the source and destination filter coefficients were interpolated based on the mix percentage. This resulted in a resynthesis filter representing the interpolation of the input signals' spectral filters. For each sample of the frame being processed, the excitation signal obtained from the source signal was multiplied by an interpolated gain factor determined from the mix percentage, yielding a scaled excitation value to which the synthesis filter was applied. This process resulted in synthesized output signals which have frame-accurate applications of the spectral filter described by the LPC order and mix percentage.

To select eight steps from this 51 step continuum, the original F2 values of /œ/ (1467 Hz) and /o/ (1045 Hz) were converted to a Bark scale using the f2bark function in the 'hqmisc' package (Quené, 2014) in R (RStudio Team, 2017). These converted Bark values were used as the ideal endpoints to select the Bark values that were equally spaced along the perceptual scale of F2. The continuum used in the current study comprised the 8 steps that were the closest matches to these ideal 8 steps from the set of 51 synthesized steps. Although the actual values did not match the ideal values exactly, no selected step differed by more than 0.03 Bark from the ideal value. The remaining vowels (/i/, /y/, /e/, / ɛ/, /a/, /ɔ/, and /u/) were each selected based on similarity in duration to the selected /œ/ and /o/ tokens and the centrality of their formant values among the five repetitions. Formant and duration information about all selected stimuli are presented in Table 2.

Procedure

Participants attended two one-hour sessions on consecutive days and completed the tasks presented in Table 3. Each task will be described in more detail below. All experimental tasks except the hearing screening were presented with E-Prime 2.0 software (Psychology

Software Tools) on a laptop. Stimuli were presented at a comfortable listening level with Sennheiser HD-280 headphones in the training, discrimination, and identification tasks. For the repetition task, recordings were made using a Sennheiser HMD 280-XQ-2 combination headphones and microphone with a Fostex FR-2LE recorder. For all tasks, participants sat in a sound-attenuated booth. To indicate responses, the screen was labeled with the possible responses and participants either used keys 1–8 on the presentation laptop or an Empirisoft DirectIN response box with eight buttons.

Training Task

Participants completed a self-paced training task in which they heard stimuli from the 8-step continuum and were asked to identify the sound by selecting one of two images: an orange cloud-like shape or a lavender spiky shape. These two images were displayed on the screen. For half of the participants in each condition, the cloud-like shape was associated with the /œ/ half of the continuum (1–4) and the spiky shape was associated with the /o/ half of the continuum (5–8). For the other half of the participants, the mapping was reversed. Participants pressed “5” to select the cloud-like shape and “8” to select the spiky shape. After making their response, immediate feedback was provided in the form of a “correct” or “incorrect” displayed on the screen. No practice trials were provided. Following the feedback screen, there was a 1000 ms delay before the next trial began. Each training block took approximately five minutes to complete. Participants completed four training tasks spanning two days to allow for consolidation and multiple training blocks on each day.

Each training task involved listening to 48 stimuli from the eight-step synthesized continuum in random order. The number of times each step was presented varied based on condition. The distributions of the bimodal and unimodal conditions are depicted in Figure 1 and match those from previous distributional learning studies. For the bimodal condition, participants heard steps 1, 4, 5, and 8 three times, steps 3 and 6 six times, and steps 2 and 7 twelve times. In the unimodal condition, participants heard steps 1, 2, 7, and 8 three times, steps 3 and 6 six times, and steps 4 and 5 twelve times. As with previous studies, steps 1, 3, 6, and 8 were heard the same number of times in both the unimodal and bimodal conditions.

Vowel perception tasks

Two perception tasks were included in the current study: an ABX discrimination task and an identification task. The discrimination task provided a way to explore both within and across-category perception before, during, and after training. As we were also interested in the degree of categoricity that subjects acquired in each condition, we also administered an identification task. An identification task reveals the steepness of a perceptual identification curve, but must be administered after training has begun (when the categories have thus already begun to form).

In the ABX discrimination task, participants heard three stimuli drawn from the acoustic continuum with a 750 ms inter-stimulus interval. After presentation of the third stimulus, participants were asked whether the final stimulus (X) was the same as the first (A) or the second (B) stimulus. An ABX task, rather than AX, was used to minimize response bias (Best et al., 2001). Participants pressed “5”, which was labeled “1st,” or “8”, which was

labeled "2nd" to make their response. After making their selection, there was a 1000 ms delay before the next trial began. No feedback was provided. To ensure that participants understood the task, a practice block was administered in which participants performed the task with /i/ and /ɛ/. The entire task took approximately eight minutes to complete.

The stimuli for the discrimination task were selected pairs from the 8-step synthesized continuum. As in Baese-Berk (2010), we included contrasts from within a sound category (stimulus points 1–3 and 6–8) and contrasts across the category boundary (stimulus points 1–8, and 3–6), which are the stimuli that both groups heard the same number of times. An additional pair of stimuli (4–5) was also included to test whether additional practice on these ambiguous stimuli (differing by only one step along the continuum) in the unimodal condition would result in better discrimination than the bimodal condition. For each comparison, four possible orders were created (e.g., 1–3-1, 1–3-3, 3–1-1, 3–1-3). Each ABX task included three blocks of 20 trials (5 pairs * 4 orders) that were presented in a fully random order, resulting in 60 trials. For the across-boundary comparisons, steps 1 and 8 were considered easy to discriminate as they were maximally different, steps 3 and 6 were considered moderate, and steps 4 and 5 were considered hard to discriminate.

Participants also completed two identification tasks in which each of the eight steps along the continuum was presented five times in random order. The setup of the task was identical to training, but no feedback was provided. To remind listeners of the labels, four practice trials of the two endpoint stimuli were provided. The identification task took approximately four minutes. Since the mapping of shapes to categories was established in the first training task, the first administration of the identification task immediately followed the first training. To capture as much learning as possible from the beginning to the end of the training, a second identification task was administered immediately after the fourth (final) training.

Working memory tasks

To control for individual differences in working memory, participants completed four tests believed to tap into different components of working memory.

The Forward Digit Span is associated with the Phonological Loop which permits short-term storage of auditory information and verbal rehearsal (Baddeley, 2000). In this task, participants listened to pre-recorded lists of digits from the Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV) (Wechsler, 2008) and were asked to repeat them. Lists varied in length from two to ten digits, with two lists at each length. The criterion for termination of the test was two incorrectly repeated lists at one length. The total number of correctly repeated lists was used as each participant's score.

The Backward Digit Span is associated with the Central Executive, which controls attention and manipulation of information (Baddeley, 2000). This task was like the Forward Digit Span task, but participants were asked to repeat the numbers in the reverse order.

The ability to listen to and repeat sentences is associated with the Episodic Buffer, which serves as an intermediary store that interfaces with long-term memory (Baddeley, 2000). This was assessed with the Recalling Sentences subtest of the Clinical Evaluation of

Language Fundamentals - Fourth Edition (CELF-4) (Semel et al., 2003). Participants listened to pre-recorded sentences and were asked to repeat them exactly as they heard them. As this test was normed to an adult population, we used the published standard scores with a mean of 10 and a standard deviation of 3.

Finally, a visual recall task was used as a measure of the Visuospatial Sketchpad, the visual analogue of the Phonological Loop (Baddeley, 2000). A computerized version of the Corsi block-tapping task was used (Berch et al., 1998; Corsi, 1972). In this task, nine white squares with black outlines were displayed on a white screen. The squares were selected by turning entirely black one at a time in a particular sequence and participants were asked to tap the squares in exactly the same order on a touch-screen computer. Following a brief practice, the main task included three sequences of each length ranging from two to nine squares. The criterion for termination was incorrect recall of all three sequences at a particular length. Similar to the Digit Span tasks, the Corsi block-tapping task was scored by counting the total number of correctly remembered lists.

Repetition task

Before any training had occurred and after all training was finished, participants completed a repetition task in which they listened to /dVt/ tokens and repeated them. Each repetition task involved three practice trials (/det/, /dat/, /dut/) followed by random presentation of four repetitions of each of the two endpoint stimuli (Steps 1 and 8) of the synthetic continuum and four repetitions of each of the nine natural French stimuli (/œ/, /o/, /i/, /y/, /e/, /ɛ/, /a/, /ɔ/, and /u/). Participants were told that they would be hearing some words in another language and that they should repeat what they heard. They were not explicitly told to imitate the tokens. As training occurred with the synthesized stimuli, analyses in the current study were on the repetitions of the synthesized tokens only. After each stimulus was presented, a 1500 ms delay allowed sufficient time for participants to repeat each stimulus and prepare for the next trial.

Acoustic analysis of the repetition data

Trained research assistants marked the boundaries of the vowels in the /dVt/ repetitions, using the same measurement and formant extraction process used for the training stimuli. The frequency maximum was set to five formants in 5000 Hz for males and five formants in 5500 Hz for females. For each participant per vowel target, any F1 and F2 values that were more than two standard deviations from that participant's mean were hand checked. These hand-checked formant values were used to replace the initial, automatic formant measures if they differed from these initial measures by more than 10 Hz. For Step 1, four F1 values and three F2 values were checked and of these only one F1 value and no F2 values were corrected. For Step 8, eight F1 values and four F2 values were checked and of these only one F1 value and one F2 value were corrected.

For each participant, average F1 and F2 formant values were calculated at pre-training and at post-training for repetitions of Steps 1 and 8. For each participant, the Euclidean distance between these two vowels was then calculated based on Equation (1). A larger Euclidean

distance between Step 1 and Step 8 represents a greater acoustic distinction between the two vowels, suggesting a more native-like production.

$$Euclidean\,distance = \sqrt{\left(\mathrm{F1}_{\mathrm{Step1}} - \mathrm{F1}_{\mathrm{Step8}}\right)^2 + \left(\mathrm{F2}_{\mathrm{Step1}} - \mathrm{F2}_{\mathrm{Step8}}\right)^2} \quad (1)$$

Statistical analysis

For both the discrimination task and the identification task, generalized logistic mixed effects models were fit to the accuracy data with the glmer function in the 'lme4' package (Bates et al., 2015) in R. All categorical predictors were sum-coded and all models were fit with bound optimization by quadratic approximation. When main effects and interactions needed to be examined, marginal means were estimated and the significance of the relevant marginal contrasts was evaluated using the emmeans function in R (Lenth, 2019). When marginal means were estimated, an effect size and standard error of the difference (or difference of differences) were provided. All p-values in emmeans were adjusted using Holm's method for multiple comparisons.

Two participants were eliminated from all analyses. One participant was excluded from the unimodal condition based on performance in which the participant marked too many identical responses in a row on the discrimination task (17/60, 14/60) and on the identification task (18/48, 13/48, 16/48), which suggested inattention to the tasks. Another participant was excluded from the bimodal condition based on perfectly inaccurate discrimination performance (0% correct) on the easiest (1–8) difficulty level, suggesting that this participant had misunderstood the task instructions to select the item that matched the third stimulus.

Results

First, we examined performance on the working memory tasks to ensure that participants in the two conditions did not differ in these abilities. Independent samples t-tests confirmed no differences between the two training conditions on any of the working memory measures (see Table 4).

Second, to ensure that the participants in the two conditions did not differ in their perceptual ability prior to training, generalized logistic mixed-effects models were fit to the discrimination data at Time 1. The full model included fixed effects for Difficulty (across-category easy [1–8], across-category moderate [3–6], across-category hard [4–5], and within-category [1–3 and 6–8]), Condition (bimodal, unimodal), and their interaction. The model also included random slopes for Difficulty by Participant and a random intercept for Participant.

Model comparison between models with and without the interaction between Difficulty and Condition revealed no difference in model fit ($\chi^2(3) = 1.34$, $p = 0.721$), thus the model without the interaction was used as the base model. Results of this model at Time 1 revealed no significant difference between the two conditions ($\beta = -0.034$, $SE = 0.061$, $z = -0.56$, $p =$

0.576), suggesting that listeners in both groups had similar pre-training perceptual abilities. The results of the full model can be found in Appendix A.

We calculated least-mean squares on Difficulty using the emmeans function, which revealed significant differences for all comparisons, as shown in Table 5. Performance was significantly different across all levels of Difficulty with performance as follows: easy across-category better than moderate across-category better than within-category better than hard across-category (as can be seen in Figure 2 in the next section). Due to these differences, subsequent analyses of the discrimination data involved separate models for each level of Difficulty.

Discrimination

Four models were fit to the discrimination data, one for each level of Difficulty with Time (1,2,3,4), Condition (unimodal, bimodal), the interaction between Time and Condition, and Order (mapping the purple spiky shape to the /o/ end of the continuum or to the /œ/ end of the continuum). We also included random slopes for Time by Participant and a random intercept for Participant. Figure 2 presents the means and confidence intervals at all four Time points.

For the easy across-category trials (discriminating Steps 1 and 8, left panel of Figure 2), model comparisons between a model with and without the interaction between Time and Condition revealed no difference in model fit ($\chi^2(3) = 0.54$, $p = 0.909$), thus the model without the interaction was used as the base model. This model without the interaction revealed a significant effect of Time, and a subsequent emmeans analysis revealed significant improvement in accuracy between Time 1 and Time 3 ($p = 0.018$), as seen in Table 6.

For the moderate across-category trials (discriminating Steps 3 and 6, second panel of Figure 2), model comparisons between a model with and without the interaction between Time and Condition revealed no difference in model fit ($\chi^2(3) = 1.97$, $p = 0.580$), thus the model without the interaction is used as the base model. As with the easy across-category model, there was a significant effect of Time, in which an emmeans analysis on Time revealed significant improvement in accuracy between Time 1 and Time 4 ($p = 0.007$), as seen in Table 7.

For the hard across-category trials (discriminating Steps 4 and 5, third panel of Figure 2), model comparisons between a model with and without the interaction between Time and Condition revealed no difference in model fit ($\chi^2(3) = 2.91$, $p = 0.407$), thus the model without the interaction is used as the base model. The model without the interaction revealed no significant effects. As can be seen in Figure 2, performance on the hard across-category trials is near chance.

For the within-category trials (discriminating Steps 1 and 3 and discriminating Steps 6 and 8, right-most panel of Figure 2), model comparisons between a model with and without the interaction between Time and Condition revealed no difference in model fit ($\chi^2(3) = 1.75$, $p = 0.625$), thus the model without the interaction is used as the base model. No predictors

contributed significantly to the model fit. As can be seen in Figure 2, performance on the within-category trials is also near chance.

Importantly, none of the models had significant effects of Condition (all $p > 0.356$) or of the interaction between Time and Condition. Together, these results suggest that participants in both conditions learned to discriminate similarly. Additionally, Order also was not a significant predictor of discrimination accuracy in any of the models (all $p > 0.355$), indicating that across all sessions and Difficulty levels, participants who completed the training with mapping A (mean = 67.7%, sd = 0.47) did not differ in discrimination accuracy from those who were trained with mapping B (mean = 67.6%, sd = 0.47). The full set of results for all models can be found in Appendix B.

Identification

To assess perceptual learning on the identification tasks, we examined overall accuracy to provide a broad picture of performance, as was done in several of the original studies of second-language sound acquisition. An additional reason to look at overall accuracy in these data is that after a single training block, some participants had not learned to label the two categories with any consistency. As a result, these participants did not have a pattern of responses that would allow a categorization curve to be fit well. To avoid spurious slopes (e.g., those where the fitted curves were positive instead of negative), we elected to examine the raw data (accuracy) so that participants who did not achieve sufficient categorical performance in the first identification task (and thus whose categorization curves could not be fit) could remain in the analysis.

It is specifically reasonable to examine raw accuracy in our identification task because participants were trained to map the first four steps onto one category and map the last four steps onto the other category. Due to the U-shaped nature of the identification accuracy curves, we fit a model that included both linear and quadratic terms for Step. Step was treated as a continuous variable because it is based on a continuous change in F2. First, Step was centered (subtracting the mean) and then the quadratic term for Step (Step^2) was generated from the centered version of Step. This allowed us to examine the convexity of the curve at the point where Step is 0 (the center of the distribution). A generalized logistic mixed-effects model was fit to the identification accuracy data with Time (1, 2), Condition (unimodal, bimodal), scaled and centered Step, scaled and centered Step^2, Order, and all possible interactions between Time, Condition, and the linear and quadratic terms for Step. The two Step terms were not included as interactions with each other. We also included a random intercept for Participant, random slopes for Time, Step, and Step^2 by Participant, and random slopes for the interaction of Time and Step by Participant and of Time and Step^2 by Participant. Accuracy data by Condition, Step, and Time and the fitted quadratic curves are plotted in Figure 3.

Model comparisons between a model with and without the two three-way interactions (Time:Condition:Step and Time:Condition:Step^2) revealed no significant difference in model fit ($\chi^2(2) = 0.207$, $p = 0.902$). Thus, the model with only the two-way interactions is used as the base model. The full model output for this model with all the 2-way interactions is in Appendix C. This model revealed a significant effect of Time ($p < 0.001$), with greater

accuracy at Time 2 than Time 1. There was no significant effect of Step (linear), because the slope at the center of Step is essentially flat. As expected, there was also a significant effect of Step^2 ($p < 0.001$), where accuracy is greater at the two ends than in the middle. Importantly, the model also revealed two significant interactions with Step^2 indicating differences in the degree of convexity. Curves that are less convex are flatter and indicate less difference in performance between Steps in the middle versus endpoints of the continuum. The interaction between Time and Step^2 ($p < 0.001$) indicates that the quadratic curve is more convex at Time 2 than Time 1. The interaction between Condition and Step^2 ($p = 0.001$) indicates that the quadratic curve is more convex in the bimodal than the unimodal condition. Both of these interactions appear to be driven by an increase in accuracy at the endpoints of the continuum, rather than a decrease in accuracy at the middle of the continuum. Indeed, a post-hoc analysis using a model examining performance for Steps 1 and 8, with Time, Step, Condition, Order, and the interaction between Condition and Time as fixed-factors and random slopes for Time by Participant revealed significantly better performance for the bimodal than the unimodal group on these endpoint stimuli (estimate = 0.93, SE = 0.43, z = 2.15, p = 0.0319).

Production

Two analyses were conducted on the production data. First, we conducted an analysis of the Euclidean distance to examine change over time for the two groups. To address this question, a repeated-measures analysis of variance (ANOVA) was run on Euclidean distance because each participant only contributed two data points, one at pre-training and another at post-training. A paired-samples t-test confirmed that the groups did not differ at pre-training ($t(30) = 1.54$, $p = 0.134$). The ANOVA included Condition (unimodal, bimodal) as a between-subjects factor and Time (pre, post) as a within-subjects factor, as shown in Figure 4. Results revealed a main effect of Time ($F(1, 30) = 7.56$, $p = 0.009$, Cohen's $d = 0.346$). Neither the main effect of Condition ($F(1, 30) = 1.38$, $p = 0.249$) nor the interaction ($F(1, 30) = .67$, $p = 0.418$) reached significance. The changes in Euclidean distance are driven primarily by changes in the second formant, as expected. Across all participants, F1 values for the two target stimuli changed by less than 4 Hz. In contrast, F2 for Step 1 was raised (fronted) on average by 22 Hz and for Step 2 was lowered (backed) by 54 Hz. Thus, participants produced a greater difference between the vowels both by fronting the /œ/ target and backing the /o/ target.

Second, we explored whether individual differences in perception are associated with production post-training. To do this, we compared a series of 16 linear regression models similar to the model comparison performed in Campbell et al. (2018) to determine *which* perceptual measure and interaction structure would best predict production post-training. All models included two structural variables: Euclidean distance at pre-training as a measure of baseline production and Condition (unimodal, bimodal). Each model examined one of four possible measures of perception: average accuracy on the ABX discrimination task pre-training (Time 1, *ABX pre*) and post-training (Time 4, *ABX post*), and average accuracy on the identification task across all eight steps at Time 1 (*ID pre*) and Time 2 (*ID post*). For each of these four perception measures, a set of four models was considered that differed based on the interaction structure: (i) no interaction term, (ii) an interaction between the

perception measure and Condition, (iii) an interaction between the production measure (Euclidean distance at pre-training) and Condition, or (iv) the interaction between the perception measure and the production measure (Euclidean distance at pre-training). All continuous variables were scaled and centered. Figure 5 provides a visual summary of these 16 models. Before creating the models, we confirmed that there were no significant correlations between the production measure (Euclidean distance at pre-training) and any of the four perception measures (all $r < 0.164$, all $p > 0.369$).

The best among the 16 models was selected based on both the Akaike Information Criteria (AIC) (Akaike, 1974) and Bayesian Information Criteria (BIC) (Schwarz, 1978), where lower values indicate a better-fitting model. AIC and BIC can be used to identify the model that best explains the variation in the outcome measure. Both AIC and BIC penalize the log-likelihood of the data by accounting for the cost of estimating the parameters in each model, but BIC penalizes models with more parameters more than AIC. All regression models were fit using the lm function in the 'lme4' package in R (Bates et al., 2015).

Table 8 provides the AIC and BIC values for each of the 16 models. The model that included post-training identification with no interactions was selected as the best-fitting model based on both AIC and BIC.[1] In this best model, a significant effect was found for Euclidean distance at pre-training (β =149.59, SE = 27.49, t = 5.44, $p < 0.0001$), but not for Identification at post-training (β = 52.85, SE = 26.47, t = 1.99, p = 0.055) or Condition (β = −13.59, SE = 26.47, t = −0.51, p = 0.614). The effect of Euclidean distance at pre-training indicates that participants with a larger Euclidean distance at pre-training also had larger Euclidean distance at post-training (see Figure 6).

Discussion

The current study examined perceptual learning for the non-native vowel contrast /œ/-/o/ by native speakers of American English and whether perceptual training would transfer to gains in production. Participants were either trained with a bimodal distribution or a unimodal distribution. Importantly, rather than completing a passive learning task, participants actively engaged with each stimulus by selecting one of two images (visual support) and receiving feedback regarding their accuracy (active learning). Perceptual learning was assessed with both a discrimination task and an identification task. No differences were found between the two conditions in the discrimination task (Question 1). On the identification task, participants in the bimodal condition had a more convex (less flat, narrower) curve due to an increase in accuracy at the endpoint stimuli relative to those in the unimodal condition (Question 2). With respect to Question 3 that asked whether training in perception would lead to gains in production, we found that production improved between baseline (pre-training) and post-training, as measured by the Euclidean distance between the vowel

[1]While the simple model (i.e., with no interactions) with post-training identification as the perception predictor had the lowest BIC value, the other simple models with the other perception predictors had the next lowest BIC values. To determine whether these slightly higher BIC values indicated meaningfully different models, we calculated Bayes Factors for each model based on Wagenmakers (2007) and Raftery (1995). The Bayes Factors suggested weak evidence for the simple model with post-training identification compared to the model with pre-training identification (BF_{01} = 0.675), post-training ABX (BF_{01} = 0.738), and pre-training ABX (BF_{01} = 0.698). This suggests that there is only weak evidence to support one perceptual predictor over the other in this set of simple models.

productions, for both groups of participants. Contrary to our expectations, we did not find that an individual's perceptual skills predicted degree of transfer to production (Question 4). Together, these results provide evidence for the benefit of exposure to a bimodal distribution, corroborating findings from previous studies. Additionally, these results offer support for our methodological strategy of enhancing perceptual learning in the unimodal condition. We address each of these conclusions in turn below.

Data from the identification task supports previous findings of an advantage for learning novel sound contrasts when exposed to a bimodal distribution. In our analysis, we found that the two groups differed in their accuracy on stimuli drawn from the endpoints of the acoustic continuum. While this finding provides additional evidence for the benefit of a bimodal distribution, it critically shows that this advantage persists even when participants are provided with additional supports in the form of images to reinforce the contrast and in the form of an active learning paradigm. As mentioned in the introduction, previous work with adults that had used a distributional learning paradigm typically did not include images to reinforce the contrast in the unimodal group: Baese-Berk (2010) provided the bimodal participants with two images, but unimodal participants with only one image. Although Hayes (2003) did include a condition where the unimodal group saw two images, no direct comparison was made between the unimodal group with two images and the bimodal group with two images. Thus, our findings from the identification task provide new and additional support for the benefits of a bimodal distribution. Furthermore, our finding that the difference between the two groups was characterized by changes in performance on the endpoint stimuli fits nicely with a recent study that found performance on endpoint stimuli to be especially useful for comparing across groups, in their case for individuals with and without reading impairments (O'Brien et al., 2018).

There are two possible explanations for the found pattern of results in the identification analysis. First, the differences at the endpoints may truly reflect a benefit for learning in the bimodal condition. Alternatively, the difference may stem from the attention paid to frequent versus infrequent stimuli and the corresponding acoustic and perceptual distances between the emphasized stimuli. The unimodal condition draws attention to the steps in the center of the continuum, whereas the bimodal condition draws attention to near-endpoint stimuli (stimuli 2 and 7). Thus, the difference in accuracy at the endpoints may have been due to inherent condition-specific differences in direction of attentional focus. This may reflect a Perceptual Magnet Effect (Kuhl & Iverson, 1995) in which listeners in the bimodal group, who have extensive exposure to stimuli 2 and 7, may have perceived stimuli 1 and 8 as similar to these frequent stimuli. That is, stimuli 1 and 8 are acoustically similar to stimuli 2 and 7 (1 and 2 are only 0.32 Bark apart; 7 and 8 are only 0.31 Bark apart).

While the identification task provided some evidence for an advantage in the bimodal condition, the remaining tasks and analyses showed no differences between the two groups. With the exception of the identification task, these results differ from previous work that showed a learning advantage in the bimodal condition across perception and production tests following training (Baese-Berk, 2010; Hayes, 2003; Maye & Gerken, 2000, 2001; Maye et al., 2002). In the present study, it is likely that overall performance was enhanced due to the modifications in the design of the training portion of both the unimodal and bimodal

conditions, specifically the inclusion of two images to reinforce a contrast and the inclusion of accuracy feedback. Furthermore, it may be that these supports increased performance in the unimodal condition relatively more than has been demonstrated in previous studies. Future research should explore this possibility by comparing these results to another group trained passively on the same stimuli.

The design of the current study is not equipped to tease apart which support helped the unimodal group the most and allowed them to perform at levels similar to those in the bimodal group. We speculate that the inclusion of contrasting images in the unimodal condition allowed participants to begin building new categories for sounds not present in their L1. While the addition of contrasting images has been shown to improve performance for participants in bimodal conditions (Baese-Berk, 2010; Hayes, 2003), previous research has not included analysis of two images applied to a unimodal condition. In addition, we speculate that the addition of active learning (identification with feedback) also aided learning. Previous research has already demonstrated that the inclusion of feedback is associated with gains in non-distributional training studies (Baese-Berk & Samuel, 2015; Bradlow et al., 1997; Goudbeek et al., 2008; McCandliss et al., 2002). The current study was specifically designed to combine these two modifications to provide the most support for participants in the unimodal condition, which had previously been found to either not result in learning or to actually inhibit learning of a novel sound contrast. Future studies will be needed to tease apart the relative contribution of these two supports, and to determine whether these supports aid learning above and beyond the benefits of a bimodal distribution alone. Given the finding from Harmon et al. (2019) that no differences are found for the unimodal and bimodal groups when accuracy feedback is also provided, we cautiously suggest that the active engagement with the stimuli was partially responsible for the learning found in both groups, as well as the increased learning found in the unimodal group compared with previous studies.

One additional possibility is that the distributions of the stimuli are not perfectly bimodal and unimodal if we consider exposure during both the training blocks and the testing blocks. As described, stimulus presentation during the training blocks was carefully controlled to provide participants with exposure following the pattern in Figure 1. However, interleaved throughout the training blocks were testing blocks (discrimination, identification, repetition) in which participants were exposed to additional tokens of steps 1, 3, 4, 5, 6, and 8. Although participants were not provided with feedback following their interaction with the stimuli (as was done during the training blocks), it is possible that the additional exposure altered the distributions to be more similar to one another than intended. Although this is a possibility, our study is not the first to include testing blocks interleaved throughout training. Indeed, Baese-Berk (2010) also included discrimination and identification using steps 1, 3, 6, 8 and still found differences in category learning between the unimodal and bimodal groups in a passive learning paradigm with contrasting images only in the bimodal condition. Given this precedent for interleaving trainings with the same stimuli, we do not believe that this additional exposure underlies the lack of a difference between the two groups in our study.

While we interpret our findings to indicate that the additional supports aided learning, we acknowledge that it is also possible that aspects of the particular stimuli used in this task were responsible for the learning pattern we observed. Most previous studies using a distributional learning paradigm have used a temporal consonant (VOT) contrast, whereas the current study used a spectral vowel contrast. It is possible that vowels and consonants are learned differently, in such a way that distributional information is less important or used less efficiently for learning a vowel contrast. Therefore, the results of the current study may reflect a phenomenon specific to vowels, such that the added supports may only improve learning of a non-native vowel contrast. We leave exploration of this possibility to future research.

In addition to examining the benefit of distributional learning (and the added supports) for perception, we also examined whether this perceptual training paradigm would lead to gains in production. Based on Euclidean distances in acoustic space between the two vowels in pre-training and post-training repetition tasks, the bimodal and the unimodal groups both demonstrated gains in production. This indicates that perceptual training not only leads to gains in perception, it also leads to gains in production. These results are in line with past studies such as Bradlow et al. (1997), Baese-Berk (2010), and Rvachew (1994) that showed gains in both perception and production following perceptual training. We acknowledge that our outcome measure (Euclidean distance) reflects the subject-specific acoustic distinctness between the two trained vowels instead of a more direct index of accuracy, such as native listener ratings.

To address whether individual differences in perceptual leaning predict production, we compared a series of models with different perceptual measures and interaction terms. We expected that listeners with better perceptual skills would also demonstrate better production following training. Contrary to this expectation, we found that the perceptual measure in the best-fitting model did not significantly predict production performance. This finding differs from Baese-Berk (2010), who found that post-training discrimination significantly predicted gains in production. However, in line with our expectations, we did find that participants who produced the most acoustically distinct target vowels prior to any perceptual training also produced the most acoustically distinct target vowels following training. In other words, participants who started out with a larger division between the two sounds maintained a larger division between the two sounds after all training was over.

Conclusions

The primary focus of the current study was to test whether additional supports would result in learning a non-native contrast in participants exposed to both a bimodal and unimodal distribution. We asked whether the previously found disadvantage for a unimodal distribution relative to a bimodal distribution could be mitigated by including contrasting images and accuracy feedback in both learning conditions. Results of the discrimination task indicated that listeners in the unimodal condition can learn as well as those in a bimodal condition, which we interpret as evidence for the benefit of these additional supports. We found some evidence for a benefit of a bimodal distribution for the endpoint stimuli in the identification task. Importantly, the analysis of the production data revealed a transfer effect,

where the Euclidean distance between the two target vowels increased following perceptual training. However, no relationship was found between perceptual skill and production performance. The current study opens several lines of additional inquiry, including the question of which added support was responsible for the relatively enhanced learning found in the unimodal condition. In particular, the addition of an active training paradigm, which in this study relies on the presence of two images and accuracy feedback, may enhance learning and be useful to second language learners. While the original studies of distributional learning showed that both infants and adults can learn a contrast through passive exposure to a bimodal distribution, the current study shows that active learning with a unimodal distribution leads to as much learning as with a bimodal distribution. This deemphasis of the role of distribution in favor of learner supports has implications for research in second language learning in adults.

Acknowledgements

The authors would like to thank Duncan MacConnell, who synthesized the stimuli, and Ashley Quinto, who helped with data collection. We are also grateful to Erika Levy and Tara McAllister for their valuable feedback, and to Scott Seyfarth and Daphna Harel for statistical support. Many thanks to all participants for their cooperation in this study.

Funding Details

This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Grant F31DC018197 (PI: H. Kabakoff).

Appendix

Appendix A

Output of the logistic mixed effects models for the discrimination task at Time 1

		estimate	SE	z value	p-value
(Intercept)		0.85	0.089	9.52	**<0.001**
Condition1		−0.034	0.061	−0.56	0.576
Difficulty1 (easy: 1, within: −1)		1.28	0.18	7.23	**<0.001**
Difficulty2 (hard: 1, within: −1)		−0.81	0.11	−7.12	**<0.001**
Difficulty3 (medium: 1, within: −1)		0.030	0.11	0.28	0.776
Random effects	Type	Variance	Std.Dev.		
	Participant (intercept)	0.12	0.35		
	Difficulty1 (by-participant slope)	0.27	0.52		
	Difficulty2 (by-participant slope)	0.11	0.33		
	Difficulty3 (by-participant slope)	0.02	0.15		

Appendix

Appendix B

Output of the logistic mixed effects models for the discrimination task by Difficulty level

Model output for the easy, across-category contrast				
	estimate	**SE**	**z value**	**p-value**

(Intercept)	2.81	0.26	10.75	**<0.001**
Time1 (Time 1: 1; Time 4: −1)	−0.68	0.21	−3.22	**0.001**
Time2 (Time 2: 1; Time 4: −1)	−0.13	0.25	−0.52	0.603
Time3 (Time 3: 1; Time 4: −1)	0.77	0.34	2.25	**0.025**
Condition1	0.02	0.20	0.11	0.911
Order1	0.01	0.21	0.03	0.976

Random effects	Type	Variance	Std.Dev.
	Participant (intercept)	1.42	1.19
	Time1 (by-participant slope)	0.11	0.33
	Time2 (by-participant slope)	0.18	0.42
	Time3 (by-participant slope)	0.44	0.67

Model output for the moderate, cross-category contrast

	estimate	SE	z value	p-value
(Intercept)	1.27	1.14	8.86	**<0.001**
Time1 (Time 1: 1; Time 4: −1)	−0.37	0.12	−3.14	**0.002**
Time2 (Time 2: 1; Time 4: −1)	0.05	0.14	0.37	0.711
Time3 (Time 3: 1; Time 4: −1)	0.02	0.14	0.12	0.902
Condition1	−0.12	0.13	−0.92	0.356
Order1	0.12	0.13	0.92	0.355

Random effects	Type	Variance	Std.Dev.
	Participant (intercept)	0.50	0.71
	Time1 (by-participant slope)	0.07	0.27
	Time2 (by-participant slope)	0.10	0.32
	Time3 (by-participant slope)	0.19	0.43

Model output for the hard, across-category contrast

	estimate	SE	z value	p-value
(Intercept)	0.07	0.06	1.12	0.264
Time1 (Time 1:1; Time 4: −1)	−0.03	0.09	−0.30	0.767
Time2 (Time 2:1; Time 4: −1)	−0.01	0.09	−0.06	0.955
Time3 (Time 3:1; Time 4: −1)	0.04	0.09	0.42	0.678
Condition1	0.01	0.06	0.21	0.830
Order1	0.00	0.06	−0.01	0.993

Random effects	Type	Variance	Std.Dev.
	Participant (intercept)	0.04	0.19
	Time1 (by-participant slope)	0.01	0.08
	Time2 (by-participant slope)	0.03	0.17
	Time3 (by-participant slope)	0.00	0.07

Model output for the within-category contrast

	estimate	SE	z value	p-value
(Intercept)	0.44	0.06	6.92	**<0.0001**
Time1 (Time 1: 1; Time 4: −1)	−0.10	0.07	−1.35	0.177
Time2 (Time 2: 1; Time 4: −1)	−0.05	0.07	−0.76	0.449

Time3 (Time 3: 1; Time 4: −1)		0.00	0.07	−0.03	0.974
Condition1		−0.05	0.06	−0.85	0.393
Order1		0.00	0.06	−0.06	0.952
Random effects	Type	Variance	Std.Dev.		
	Participant (intercept)	0.08	0.29		
	Time1 (by-participant slope)	0.03	0.18		
	Time2 (by-participant slope)	0.01	0.12		
	Time3 (by-participant slope)	0.01	0.07		

Appendix

Appendix C

Output of the generalized logistic mixed effects model for identification accuracy

	estimate	SE	z value	p-value
(Intercept)	2.48	0.24	10.43	**<0.001**
Condition1	0.39	0.19	2.09	0.037
Time1	−0.73	0.14	−5.14	**<0.001**
Order1	0.01	0.08	0.10	0.919
Step	0.18	0.15	1.20	0.232
Step^2	1.83	0.20	9.32	**<0.001**
Condition1:Time1	0.00	0.08	0.00	1.000
Time1:Step	−0.08	0.13	−0.62	0.537
Condition1:Step	0.10	0.12	0.90	0.368
Time1:Step^2	−0.69	0.13	−5.26	**<0.001**
Condition1: Step^2	0.50	0.15	3.26	**0.001**
Random effects	Type	Variance	Std.Dev.	
	Participant (intercept)	1.409	1.187	
	Time1 (by-participant slope)	0.349	0.591	
	Time1:Step (by-participant slope interaction)	0.235	0.485	
	Time2:Step (by-participant slope interaction)	1.197	1.094	
Time1:Step^2 (by-participant slope interaction)		0.430	0.655	
Time2:Step^2 (by-participant slope interaction)		1.607	1.268	

Appendix

Appendix D

Output of the generalized logistic mixed effects model for identification accuracy for only the near and far stimuli.

	estimate	SE	z value	p-value
(Intercept)	2.237	0.212	10.52	<0.001
Condition1	0.176	0.207	0.853	0.393
Time1	−0.424	0.147	−2.88	**0.003**

	estimate	SE	z value	p-value
Distance1	0.594	0.098	6.04	**<0.001**
Order1	0.211	0.193	1.09	0.273
Condition1:Time1	0.175	0.136	1.28	0.199
Condition1:Distance1	0.366	0.097	3.74	**<0.001**
Time1:Distance1	−0.108	0.098	−1.10	0.269
Condition1:Session1:Distance1	0.119	0.097	1.22	0.220
Random effects	Type	Variance	Std.Dev.	
	Participant (intercept)	0.994	0.997	
	Time1 (by- participant slope)	0.247	0.497	

Reference List

Akaike H (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.

Baddeley A (2000). The episodic buffer: a new component of working memory? Trends Cogn. Sci, 4(11), 417–423. [PubMed: 11058819]

Baese-Berk MM (2010). An examination of the relationship between speech perception and production. (Dissertation), Northwestern University, Evanston, IL.

Baese-Berk MM, & Samuel AG (2015). Listeners beware: Speech production may be bad for learning speech sounds. J. Mem. Lang, 89, 23–36.

Bates D, Maechler M, Bolker B, & Walker S (2015). Fitting linear mixed-effects models using ‘lme4’. J. Stat. Softw, 67(1), 1–48. doi: 10.18637/jss.v067.i01.

Berch DB, Krikorian R, & Huha EM (1998). The Corsi block-tapping task: Methodological and theoretical considerations. Brain Cogn, 38(3), 317–338. [PubMed: 9841789]

Best CT (1991). The emergence of native-language phonological influences in infants: A perceptual assimilation model. Haskins Laboratories Status Report on Speech Research, SR-107/108, 1–30.

Best CT, McRoberts GW, & Goodell E (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. J. Acoust. Soc. Am, 109(2), 775–794. [PubMed: 11248981]

Boersma P, & Weenink D (2019). Praat: doing phonetics by computer (Version 6.0.50) [Computer program]. Retrieved from www.fon.hum.uva.nl/praat/.

Bradlow AR, Akahane-Yamada R, Pisoni DB, & Tohkura Y. i. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. Percept. Psychophys, 61(5), 977–985. [PubMed: 10499009]

Bradlow AR, Pisoni DB, Akahane-Yamada R, & Tohkura Y. i. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. J. Acoust. Soc. Am, 101(4), 2299–2310. [PubMed: 9104031]

Campbell H, Harel D, Hitchcock E, & McAllister Byun T (2018). Selecting an acoustic correlate for automated measurement of American English rhotic production in children. Int. J. Speech Lang. Pathol, 20(6), 635–643. doi: 10.1080/17549507.2017.1359334. [PubMed: 28795872]

Corsi PM (1972). Human memory and the medial temporal region of the brain. (Ph.D. Dissertation), McGill University, Montreal, Canada.

Earle FS, Landi N, & Myers EB (2017). Sleep duration predicts behavioral and neural differences in adult speech sound learning. Neurosci. Lett, 636, 77–82. [PubMed: 27793703]

Escudero P (2005). Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization Netherlands Graduate School of Linguistics.

Escudero P, Benders T, & Wanrooij K (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. J. Acoust. Soc. Am, 130(4), EL206–EL212. [PubMed: 21974493]

Flege JE (2003). Assessing constraints on second-language segmental production and perception In Schiller NO & Meyer A (Eds.), Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities (Vol. 6, pp. 319–355). Berlin, Germany: Walter de Gruyter.

Fry DB, Abramson AS, Eimas PD, & Liberman AM (1962). The identification and discrimination of synthetic vowels. Lang. Speech, 5(4), 171–189.

Goldinger SD (1998). Echoes of echoes? An episodic theory of lexical access. Psychol. Rev, 105(2), 251. [PubMed: 9577239]

Goudbeek M, Cutler A, & Smits R (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. Speech Commun, 50(2), 109–125.

Gulian M, Escudero P, & Boersma P (2007). Supervision hampers distributional learning of vowel contrasts In Trouvain J & Barry WJ (Eds.), Proceedings of the 16th International Congress of Phonetic Sciences (pp. 1893–1896). Saarbrücken: University of Saarbrucken.

Harmon Z, Idemaru K, & Kapatsinski V (2019). Learning mechanisms in cue reweighting. Cognition, 189, 76–88. [PubMed: 30928780]

Hayes RL (2003). How are second language phoneme contrasts learned? (Dissertation), University of Arizona, Tuscon, AZ.

Hayes-Harb R (2007). Lexical and statistical evidence in the acquisition of second language phonemes. Second Lang. Res, 23(1), 65–94.

Jamieson DG, & Rvachew S (1992). Remediating speech production errors with sound identification training. J. Speech Lang. Pathol. Audiol, 16(3), 201–210. doi: 1993-27184-001.

Kong EJ, & Edwards J (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. J. Phon, 59, 40–57. [PubMed: 28503007]

Kuhl PK, & Iverson P (1995). Linguistic experience and the “perceptual magnet effect” In Strange W (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 121–154). Baltimore, MD: York Press.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, & Lindblom B (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. Science, 255(5044), 606–608. [PubMed: 1736364]

Labov W, Ash S, & Boberg C (2006). The Atlas of North American English: Phonetics, Phonology and Sound Change: A Multimedia Reference Tool. Berlin, Germany: Walter de Gruyter.

Lenth RV (2019). Estimated marginal means, aka least-squares means (Version 1.3.3). Retrieved from https://cran.r-project.org/web/packages/emmeans/emmeans.pdf.

Levy ES (2009a). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. J. Acoust. Soc. Am, 125(2), 1138–1152. [PubMed: 19206888]

Levy ES (2009b). On the assimilation-discrimination relationship in American English adults’ French vowel learning. J. Acoust. Soc. Am, 126(5), 2670–2682. [PubMed: 19894844]

Levy ES, & Strange W (2008). Perception of French vowels by American English adults with and without French language experience. J. Phon, 36(1), 141–157.

Manis FR, McBride-Chang C, Seidenberg MS, Keating P, Doi LM, Munson B, & Petersen A (1997). Are speech perception deficits associated with developmental dyslexia? J. Exp. Child Psychol, 66(2), 211–235. [PubMed: 9245476]

MathWorks Inc. (2000). MATLAB (Version 6.1) [Computer program]. Natick, MA Retrieved from https://www.mathworks.com/products/matlab/.

Maye JC, & Gerken L (2000). Learning phonemes without minimal pairs. Proceedings of the 24th annual Boston University Conference on Language Development 2, 522–533.

Maye JC, & Gerken L (2001). Learning phonemes: How far can the input take us? Proceedings of the 25th annual Boston University Conference on Language Development, 1, 480–490.

Maye JC, Werker JF, & Gerken L (2002). Infant sensitivity to distributional information can affect phonetic discrimination. Cognition, 82(3), B101–B111. [PubMed: 11747867]

McBride-Chang C (1996). Models of speech perception and phonological processing in reading. Child Dev, 67(4), 1836–1856. [PubMed: 8890511]

McCandliss BD, Fiez JA, Protopapas A, Conway M, & 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. Cogn. Affect. Behav. Neurosci, 2(2), 89–108. [PubMed: 12455678]

McGregor KK (2014). What a difference a day makes: Change in memory for newly learned word forms over 24 hours. J. Speech Lang. Hear. Res, 57(5), 1842–1850. [PubMed: 24845578]

Mitterer H, & Ernestus M (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. Cognition, 109(1), 168–173. [PubMed: 18805522]

O'Brien GE, McCloy DR, Kubota EC, & Yeatman JD (2018). Reading ability and phoneme categorization. Sci. Rep, 8(1), 16842. [PubMed: 30442952]

Polka L (1995). Linguistic influences in adult perception of non-native vowel contrasts. J. Acoust. Soc. Am, 97(2), 1286–1296. [PubMed: 7876448]

Polka L, & Werker JF (1994). Developmental changes in perception of nonnative vowel contrasts. J. Exp. Psychol. Hum. Percept. Perform, 20(2), 421. [PubMed: 8189202]

Quené H (2014). 'hqmisc' package in R: Miscellaneous convenience functions and dataset (Version 0.1–1).

Raftery AE (1995). Bayesian model selection in social research. Sociol. Methodol, 25, 111–164.

RStudio Team. (2017). RStudio: integrated development for R (Version 1.0.136) [Computer program]. Boston, MA: RStudio, Inc. Retrieved from https://www.rstudio.com/products/rstudio/.

Rvachew S (1994). Speech perception training can facilitate sound production learning. J. Speech Lang. Hear. Res, 37(2), 347–357. doi: 10.1044/jshr.3702.347.

Rvachew S, Nowak M, & Cloutier G (2004). Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. Am. J. Speech Lang. Pathol, 13(3), 250–263. [PubMed: 15339234]

Schwarz G (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464.

Semel E, Wiig EH, & Secord WA (2003). Clinical Evaluation of Language Fundamentals, 4th Ed (CELF-4). Toronto Ontario: Pearson: The Psychological Corporation.

Shockley K, Sabadini L, & Fowler CA (2004). Imitation in shadowing words. Percept. Psychophys, 66(3), 422–429. [PubMed: 15283067]

Strange W, Levy ES, & Law II FF (2009). Cross-language categorization of French and German vowels by naive American listeners. J. Acoust. Soc. Am, 126(3), 1461–1476. [PubMed: 19739759]

Trehub SE (1973). Infants' sensitivity to vowel and tonal contrasts. Dev. Psychol, 9(1), 91.

Trehub SE (1976). The discrimination of foreign speech contrasts by infants and adults. Child Dev, 466–472.

Wagenmakers EJ (2007). A practical solution to the pervasive problems ofp values. Psychon. B. Rev, 14(5), 779–804.

Wanrooij K, Boersma P, & Benders T (2015). Observed effects of "distributional learning" may not relate to the number of peaks. A test of "dispersion" as a confounding factor. Front. Psychol, 6.

Wechsler D (2008). Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV). San Antonio, TX: NCS Pearson.

Werker JF, & Tees RC (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav. Dev, 7, 49–63.

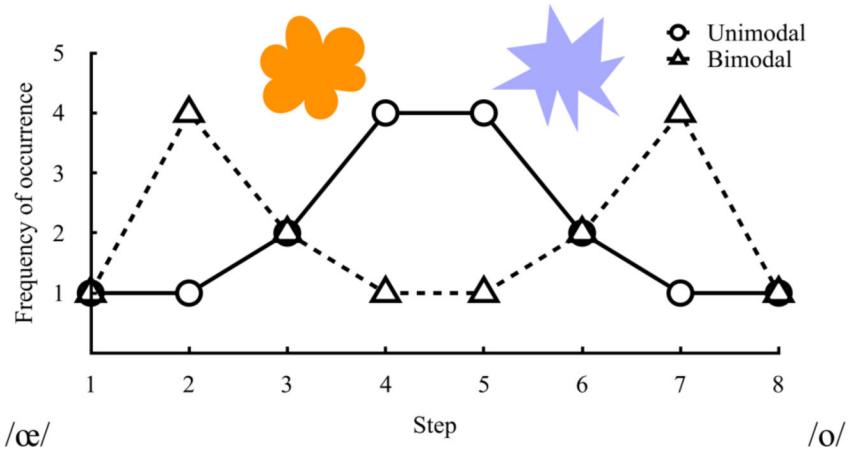

Figure 1.
Sample distributions of the unimodal and bimodal conditions.

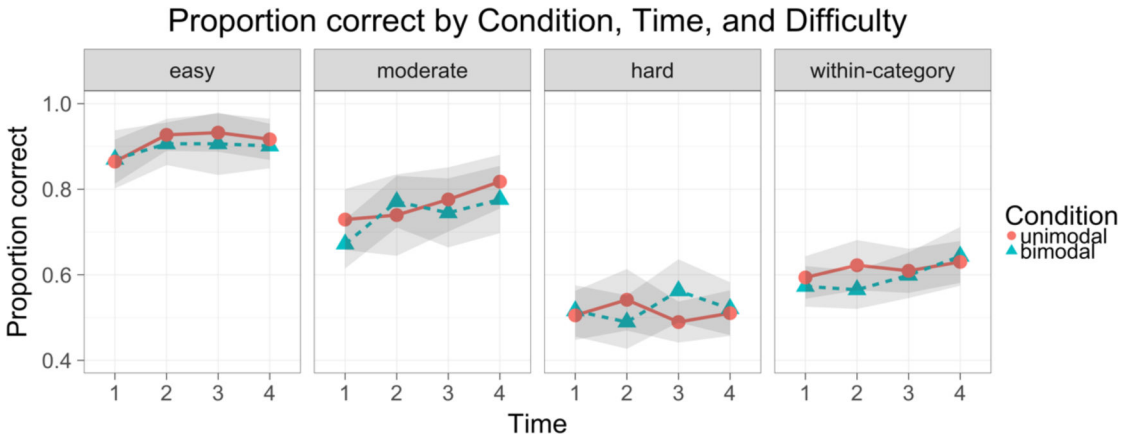

Figure 2.

Discrimination accuracy across the four Time points split by level of Difficulty with standard error of the mean bands.

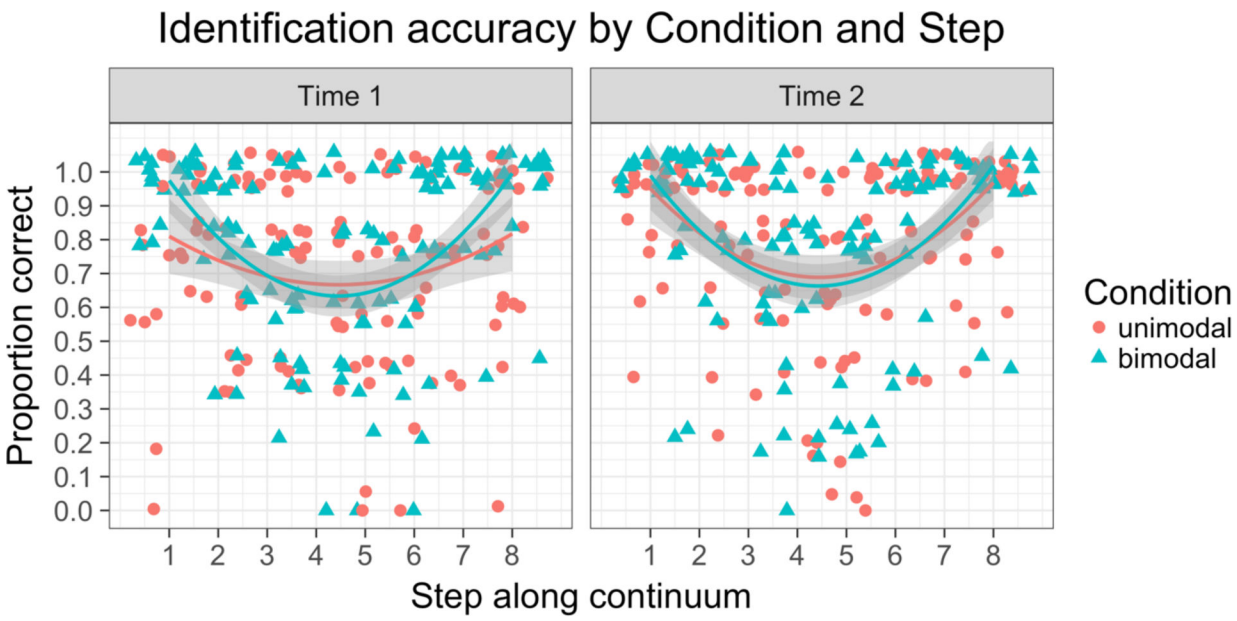

Figure 3.

Identification accuracy and fitted quadratic curves for each Step at each Time point split by Condition. Horizontal and vertical jitter are added to the plot.

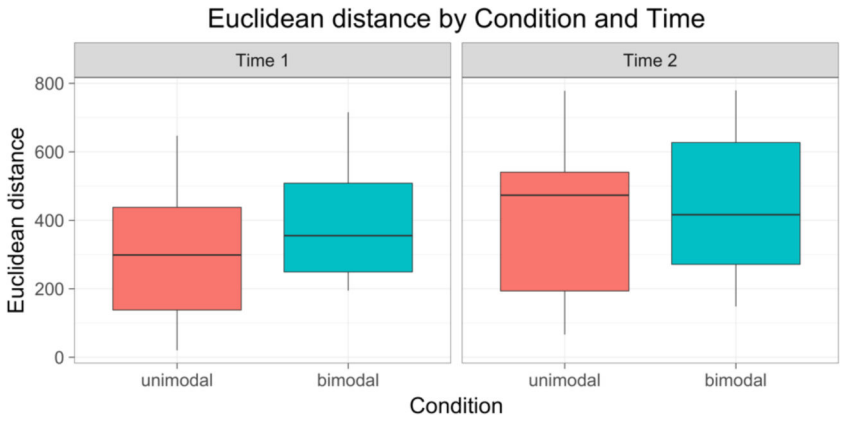

Figure 4.
Box plots of Euclidean distance between /œ/ and /o/ at pre-training and post-training, grouped by Condition.

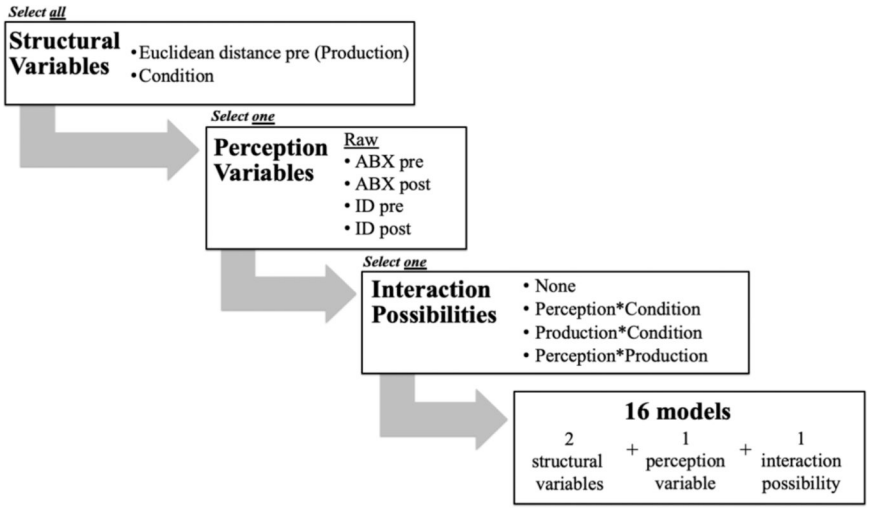

Figure 5.

All models predicted the Euclidean distance at post-training from two structural variables, one of four perception variables, and one of four possible sets of interactions, for a total of 16 models.

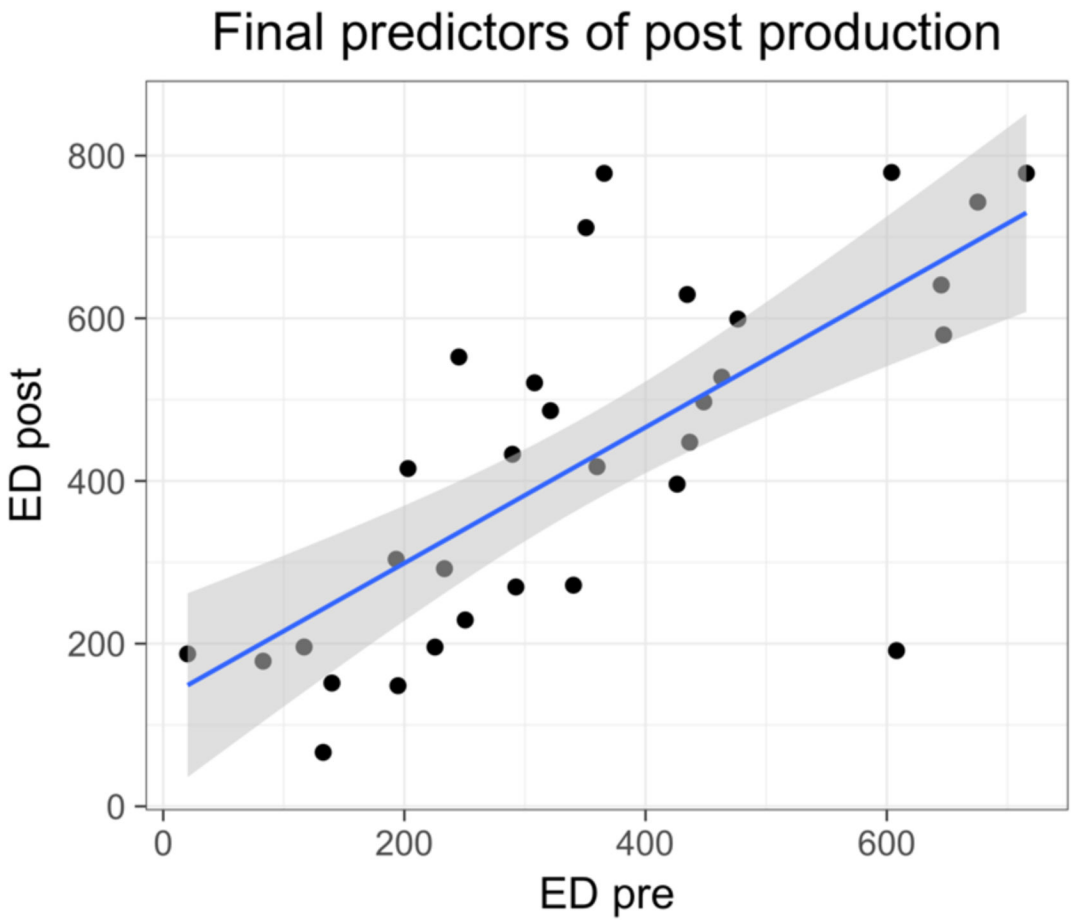

Figure 6:
Scatter plot of Euclidean distance (ED) pre-training (x-axis) and the Euclidean distance at post-training (y-axis).

Table 1.

Age breakdown by Condition and Order

	N	Age range	Age mean (SD)
Bimodal	16	18–29	22.4 (3.1)
A	9	19–25	22.0 (2.5)
B	7	18–29	23.0 (3.9)
Unimodal	16	18–30	22.2 (2.8)
A	8	18–30	22.0 (3.6)
B	8	20–25	22.5 (1.8)
Total	32	18–30	22.3 (2.9)

Table 2.

Formant frequencies of original and synthesized stimuli

Step	Ideal F2 (Bark)	Actual F2 (Bark)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Duration (ms)
original /œ/	-	**10.95**	**401**	**1467**	**2400**	**140**
1	10.95	10.94	400	1465	2399	140
2	10.63	10.63	403	1398	2408	140
3	10.33	10.33	406	1335	2413	140
4	10.02	10.03	407	1273	2419	140
5	9.72	9.73	408	1215	2422	140
6	9.41	9.44	409	1161	2422	140
7	9.10	9.08	410	1095	2425	140
8	8.79	8.79	410	1045	2425	140
original /o/	-	**8.79**	**410**	**1045**	**2465**	**138**
/i/	-	-	292	1936	2952	129
/y/	-	-	307	1762	2383	119
/e/	-	-	393	1999	2466	218
/ɐ/	-	-	501	1741	2548	131
/a/	-	-	651	1451	1451	133
/ɔ/	-	-	531	1359	2518	133
/u/	-	-	306	1083	2405	126

Table 3.

Order of tasks

Day 1	**Day 2**
Discrimination 1	Discrimination 3
Production 1	Training 3
Training 1	*Corsi block-tapping*
Identification 1	*Recalling Sentences*
Hearing screening	Training 4
Forward/Backward Digit Span	Identification 2
Training 2	Discrimination 4
Discrimination 2	Production 2

Table 4.

Descriptive statistics for working memory measures for participants in two conditions

WM measure	Bimodal		Unimodal		Student t-test comparing conditions	Cohen's d
	Mean	SD	Mean	SD		
Forward Digit Span	10.3	2.30	9.94	1.98	t(30) = 0.49, p = 0.63	0.167
Backward Digit Span	6.94	2.54	6.56	2.22	t(30) = 0.49, p = 0.63	0.159
Recalling Sentences	10.50	2.25	10.25	2.59	t(30) = 0.49, p = 0.63	0.103
Corsi block-tapping	14.75	2.18	13.56	2.34	t(30) = 0.49, p = 0.63	0.528

Table 5.

emmeans comparison between the four levels of Difficulty at Time 1 in the ABX discrimination task. P-values are adjusted using Holm's method.

contrast	estimate	SE	z ratio	p-value
easy - hard	1.85	0.18	10.137	<**0.0001**
easy - moderate	1.03	0.19	5.509	<**0.0001**
easy - within	1.55	0.17	9.251	<**0.0001**
hard - moderate	−0.82	0.15	−5.376	<**0.0001**
hard - within	−0.30	0.13	−2.361	**0.0182**
moderate - within	0.52	0.13	3.872	**0.0002**

Table 6.

emmeans comparison between the four levels of Time for the easy across-category contrasts in the ABX discrimination task. P-values are adjusted using Holm's method.

contrast	estimate	SE	z ratio	p-value
Time 1 - Time 2	−0.55	0.33	−1.68	0.337
Time 1 - Time 3	−1.44	0.49	−2.97	**0.018**
Time 1 - Time 4	−0.72	0.33	−2.17	0.150
Time 2 - Time 3	−0.90	0.52	−1.73	0.335
Time 2 - Time 4	−0.17	0.38	−0.44	0.659
Time 3 - Time 4	0.73	0.49	1.49	0.335

Table 7.

emmeans comparison between the four levels of Time for the moderate across-category contrasts in the ABX discrimination task. P-values are adjusted using Holm's method.

contrast	estimate	SE	z ratio	p-value
Time 1 - Time 2	−0.42	0.20	−2.12	0.171
Time 1 - Time 3	−0.39	0.21	−1.87	0.245
Time 1 - Time 4	0.67	0.21	−3.26	**0.007**
Time 2 - Time 3	0.03	0.24	0.14	0.890
Time 2 - Time 4	−0.25	0.22	−1.16	0.595
Time 3 - Time 4	−0.28	0.22	−1.29	0.595

Table 8

AIC and BIC for all 16 models predicting the Euclidean distance at post-training. Lowest AIC and BIC values of the 16 models are marked in bolded text.

Perception measure included in model	**Main effects Only**		**Main effects + Perception* Condition**		**Main effects + Production* Condition**		**Main effects + Perception* Production**	
	AIC	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
ABX pre-training	416.87	424.20	416.62	425.42	418.34	427.14	417.69	426.48
ABX post-training	417.27	424.60	419.11	427.91	418.49	427.29	419.15	427.94
ID pre-training	416.66	423.99	418.34	427.13	418.52	427.31	418.63	427.43
ID post-training	**415.19**	**422.52**	416.59	425.39	417.03	425.83	417.18	425.98