

THE UNIVERSITY OF CHICAGO

Asymmetrical Disruption:
Bimodal Interaction in Nonnative Contrast
Acquisition

By

Xutong Zhang

July 2025

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Melissa M. Baese-Berk

Preceptor: Natalie Dowling

Abstract

Previous studies have indicated that on the first days of second language (L2) phonemic acquisition, the immediate repeating of contrasting nonnative phonemes after hearing them could inhibit the acquisition of the sensitivity in the perception of these nonnative contrasts. This finding challenges the well-established interaction between the two modalities, which suggests a position correlation. This paper examines the effects of the implementation of production training following a perceptual identification task on the learning outcomes for the perceptual sensitivity of L2 contrasting phonemes and the production of them. 26 undergraduate students at the University of Oregon performed ABX tasks and tested for their acquisition achievement in perception and production. The paper identified a negative correlation between the production training and the perceptual learning result, while finding no correlation between the training and the production learning result. The paper argues that at the earliest phase of L2 phonemic acquisition, the simultaneous practice in perception and production could be counterproductive and focusing on the perceptual training would provide optimal learning outcomes for the acquisition of perception and production.

Keywords: Second language acquisition, phonemic acquisition, perception, production

Asymmetrical Disruption: Bimodal Interaction in Nonnative Contrast Acquisition

A common challenge faced by second language (L2) learners is the inability to distinguish specific phonetic contrasts. For example, native Japanese speakers whose language includes the phoneme “l” (IPA: /l/) but not “r” (IPA: /r/) struggle to perceive the difference between the English words “lead” and “read” or to correctly produce the word “read” (Kuhl et al., 1997). Since L2 learners can address this obstacle through extensive training in listening and speaking, language educators and researchers have traditionally supported models that suggest either independent development or a mutual reinforcement between the perception and production of contrasting nonnative phonemes (Flege, 1995; Best & Tyler, 2007). Accordingly, many language classes introduce the practice of repeating after hearing the letters, words, and sentences produced by native speakers, expecting that such training could facilitate students' development of listening and speaking abilities in the new language. However, more recent research by Baese-Berk & Samuel (2016) indicates that in the earliest stages of L2 phonemic acquisition, training focused on the production of nonnative contrasts may disrupt learners' ability to acquire the perception of those contrasts. This study will build upon Baese-Berk & Samuel's (2016) conclusions and expand their results on the acquisition of Hindi phonemic contrasts by native English speakers. We are interested in whether practicing the perception and production of unfamiliar Hindi phonemic contrasts together would interfere with the acquisition of these skills for native English speakers.

Literature Review***Acquisition of Contrastive Phonemes***

The acquisition of speech sounds involves categorizing continuous acoustic input, in the form of sound waves, into independent phonemes within a given language. Phonemic

categorization enables speakers to disregard subtle variations in speech sounds pronounced by different individuals, thereby facilitating effective communication among speakers. This task involves identifying distinctive acoustic properties from sound waves, excluding confounding features resulting from speakers' vocal characteristics or background noise, and recognizing evidence that indicates a segment of sound is one of the meaningful units in the listener's language (Holt & Lotto, 2010). However, despite the considerable demand for phonemic categorization in terms of labor and meticulousness, normal-speaking humans can perform this work with precision in the measure of milliseconds. This remarkable capability is partly due to the inborn nature of the human brain, which prefers to reorganize the hodgepodge of information nature presents to it into orderly displayed representations (Damper & Harnad, 2000), and the heightened sensitivity to meaningful acoustic traits speakers of languages acquired along with their language development (Livingston et al., 1998).

Phonemic categorization is a fundamental aspect of first language (L1) acquisition, occurring as children develop their native language abilities. For example, infants as young as 2 to 3 months old possess sensitivity to the categorical distinction between [ɪ] and [I] (Eimas, 1975). The researchers found that infants significantly increased their sucking rate of a baby pacifier by 11 times per minute after hearing a categorical switch from [ɪ] to [I], while a within-categorical shift in auditory input led to either a comparatively minor increase or decrease in infants' sucking rate. Moreover, a decline in sensitivity for the [ɪ]-[I] contrast appeared for infants developing Japanese as their first language at the age of 9-11 months (Lovčević & Tsuji, 2024). In summary, heightened sensitivity to phonemic contrasts and the desensitization of nonnative contrasts can both occur in infants within the first year of life. Several possible advantages empower infants to achieve efficient phonemic acquisition (Kuhl, 2004). These factors include infants' awareness of the statistical distribution for

possible acoustic items to appear on a continuum of sounds, the presence of active talkers who provide reinforcing social cues, and the neural structure that commits to the acquisition of a native language while filtering out nonnative noises.

In contrast, for L2 learners encountering nonnative contrasts, pre-established phonemic categories from their L1 may hinder the perception of these contrasts. When encountering nonnative contrasts that L1 generalized as one phoneme, a learner will need to restore the sensitivity to distinguish these long-abandoned acoustic features. This process requires learners to absorb a significant amount of L2 acoustic input of nonnative contrasts to reconstruct the statistical distribution of phonemic exemplars and reorganize their mental representations of phonemes. As a result, adapting native categories to accommodate the phonemic system of the L2 requires persistent practice (Flege, 1980; Eckman et al., 2003).

Perception and Production Development in L2 Phonemic Acquisition

The acquisition of nonnative contrast encompasses both the perception and production of contrastive phonemes. Two L2 acquisition models, the Speech Learning Model (SLM) and the Phonetic Assimilation Model for Second Language (PAM-L2), posit that during the process of L2 phonemic acquisition, perception and production develop independently, although they may also enhance one another (Flege, 1995; Best & Tyler, 2007). Flege (1995) reported that native Japanese speakers' ability to perceive and produce English phoneme [ɹ] improved as their exposure to and time spent learning English increased, indicating a simultaneous development of these two modules. Similarly, another research report a separate acquisition outcome for native Japanese speakers in the perception and production of the English [ɹ]-[l] contrast. With learning periods ranging from 5 to 12 years, the Japanese speakers tested achieved over 70% mean accuracy in both the perceptual identification and

the production rating of the English [ɹ]-[l] contrast, suggesting a correlated acquisition between these two modalities (Hattori & Iverson, 2010).

One issue associated with these findings is that they were insufficient to demonstrate the underlying mechanism and the potency of the correlation between perception and production during L2 acquisition. As Flege (1999) acknowledged, previous studies on the correlation between perception and production have only found a moderate positive correlation between the two modalities. The lack of motivation for L2 speakers to refine their production, as well as age-related production difficulties, could be the reason for a slower L2 production acquisition rate. Hattori & Iverson (2010) also reported the elimination of correlation between the two modalities when comparing the Japanese speakers' perceptual sensitivity to specific acoustic cues, such as F2 and F3, and their production intelligibility.

A limitation of previous studies is that their findings focused on the ultimate outcomes of L2 phonemic categorization, as the participants in the study had been learning English for a minimum of three years (Flege, 1995). Instead, researchers captured a different interaction pattern between perception and production in L2 acquisition by investigating the first days of the phonemic acquisition trajectory.

Perception-Production Interaction During Early L2 Phonemic Acquisition

Baese-Berk & Samuel (2016) observed that during two days of laboratory-based training, learners who engaged in practicing both the perception and production of nonnative contrasts simultaneously demonstrated significantly lower performance in perceptual learning compared to those who focused exclusively on perception training. Moreover, they found that substituting production training with a more straightforward task and prolonging the waiting time between perception and production training led to improved perceptual outcomes

(Baese-Berk & Samuel, 2016, 2022). In a related study, Baese-Berk (2019) further demonstrated that during the very first days of exposure, while specific training patterns effectively predicted L2 learners' perceptual learning, the results of perceptual learning were not reliable predictors of their L2 production learning outcomes.

In general, recent studies focusing on the initial stages of phonemic acquisition suggest that production training has a significant disruptive effect on L2 learners' perceptual acquisition, whereas the acquisition of production remains independent mainly of perceptual learning. These findings contradict the previously recognized interaction pattern between the two modalities during L2 acquisition, thus motivating further validation. Since the suggested underlying mechanism for this intervention is not language-specific, we expect to reproduce similar effects for the acquisition of another language with different phonetic properties.

Current Study

This study explores the relationship between auditory and articulatory training in learning nonnative speech sounds. The observation that practicing the production of nonnative contrasts can hinder the acquisition of the ability to distinguish these contrasts aurally has motivated attempts to replicate this finding with different languages. Moreover,

Specifically, building on the effect identified by Baese-Berk & Samuel (2016, 2022) and Baese-Berk (2019), which suggested that simultaneous introduction of trainings in perception and production asymmetrically inhibits the acquisition of phonological perception, this study aims to replicate these effects on the acquisition of Hindi voiceless dental-retroflex stops /ʈ/ - /ʈ/ among native English speakers. We hypothesize that the immediate repetition of sounds after hearing them will decrease learners' perceptual results in the subsequent test, indicating impaired perceptual acquisition caused by early training in production.

Furthermore, we anticipate no significant predictive relationship between the learning results in perception and production, as we expect the development in production to be independent of the influence of perceptual learning at the earliest stage of L2 phonological acquisition.

Counterevidence to the asymmetrical inhibition effect would be an equivalent or superior perceptual performance for the subjects who underwent immediate production training. Similarly, inhibited production results for subjects with worse perception learning results would also challenge the hypothesis. Such results would reflect more on the mutually beneficial model or confound the asymmetrical inhibition effect by invalidating the hypothesized unidirectional interference exerted by production training on perceptual acquisition.

Materials and Methods

Participants

Participants were 26 undergraduate students at the University of Oregon who were either monolingual English speakers or multilingual native English speakers whose L2 did not contain phonemic retroflex stops. This sampling approach allowed the generalization of the participants' learning ability while minimizing potential interference from prior exposure to the target contrasts.

Materials

This study employed an ABX paradigm to present experimental stimuli (Stevenson, 1979; Repp, 1981). Hindi dental-retroflex contrast samples were recorded from a native speaker's production of open syllables [ɖ̪a] and [ʈa]. A six-step continuum was then synthesized by varying the acoustic features that distinguish the target stops, primarily the

bunching or clustering of F2 and F3, and the lower-frequency burst spectrum for the retroflex stop (Pruitt, 2006). Step 1 should best represent [t̪a], and step 6 should best represent [ta]. For each ABX test trial, the template stimuli pair A and B were selected from the continuum in consecutive steps, forming five pairs: 1-2, 2-3, 3-4, 4-5, and 5-6. The test stimulus X matched either A or B. During one trial, the sample pair A and B, along with the test stimulus X, were played in the ABX sequence, and participants were required to decide whether X was A or B. The order for the two steps with a stimulus pair to appear as A or B was in either an ascending or descending manner; as a result, AB sequences of both 3-4 and 4-3 were possible. Each ABX test block contained 72 trials. In each test block, we randomized the likelihood of each stimulus pair appearing and the sequence in which they appeared within the stimulus pairs.

Procedures

Participants were required to complete a two-day experiment series. Each day of experiments included a pre-test, a training phase, and a post-test phase. Participants first completed a pre-test to familiarize themselves with the ABX test structure and provide their baseline perceptual performance of the Hindi contrast. The pre-test contained one ABX test block without correct answers provided to the participants. Subsequently, during the training phase, participants completed 7 ABX test blocks and received feedback on the correct answers for each trial they responded to, allowing them to become accustomed to the Hindi dental-retroflex contrast and adjust their phonological categories accordingly. During this phase, we divided participants into two groups based on the different tasks they performed. The Perception-Production group repeated the test token X immediately after hearing it and then made a perceptual choice. The Perception-Only group would make only a perceptual

choice after hearing the test token X without additional tasks. Lastly, participants completed a post-test consisting of one ABX test block, which evaluated their perceptual learning results.

Additionally, we implemented a production test in both the pre-test and post-test phases of perceptual training for each group on each experimental day. Each production test consisted of 54 trials, and within each trial, participants were required to repeat an auditory stimulus randomly selected from the six steps created for the ABX test. The participants' production was recorded for acoustic analysis.

Data Analysis

Perceptual learning results were automatically recorded as either correct or incorrect during the post-test. We expected the correct rate achieved by the two task groups for trials with the boundary-crossing stimuli pair 3-4 to be the most indicative of participants' perceptual acquisition, as it represented phonological categorization of the Hindi dental-retroflex contrast. Production learning results were analyzed using the phonetic analysis software Praat to capture the phonemes participants produced. The degree of similarity to native samples and differences among participants' productions were considered indicators of acquisition in production. Linear mixed-effects modal analysis was conducted using R, with training, contrast types, task groups, and their interactions as predictors of the acquisition results in perception and production.

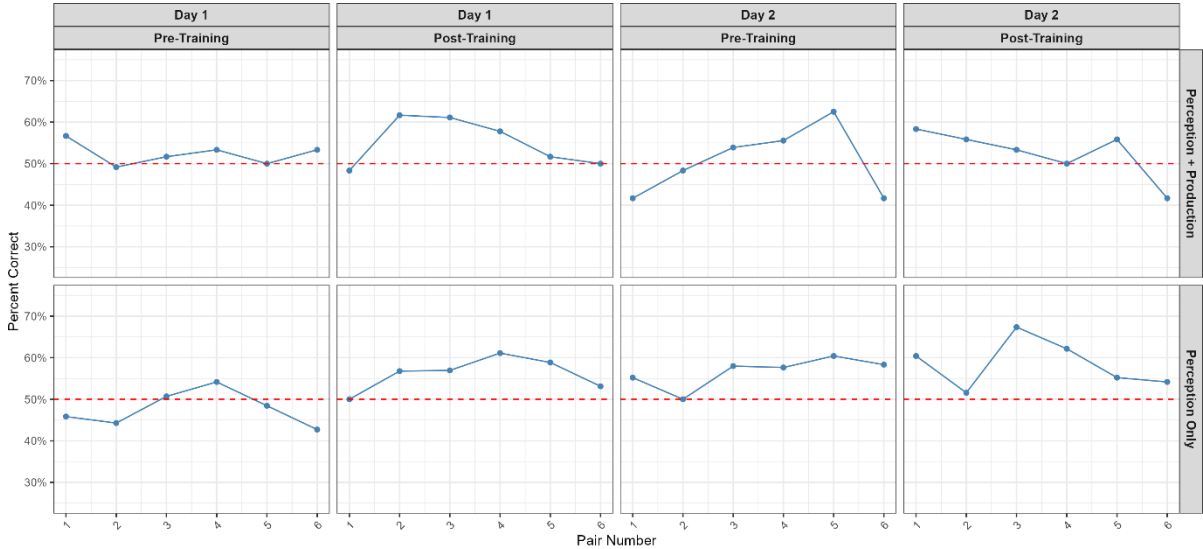
Results

Figure 1 shows the perceptual training results for the Perception + Production and Perception Only groups. For both training groups, there was a general improvement in perceptual discrimination accuracy at the cross-boundary position after the training phase on either experiment day. For statistical analysis, since either the mixed-effects model or logistic

regression, including interactions among predictor variables, exhibited more explanatory effects, we report only logistic regression with main effects here. Logistic regression results corroborate the visual representation of Figure 1 (see Table 1). Experiment Day also exhibited a noticeable effect, especially for the Perception-only group, where day 2 discrimination accuracy was higher than that of day 1. Statistical analysis provided a clearer picture, suggesting a significant effect of the experiment day (see Table 1). Conversely, training types offer no convincing evidence for significantly improving participants' discrimination accuracy visually. Statistical analysis supports a near-significant effect of training types: the Perception + Production group obtained a lower accuracy result compared to the Perception Only group (see Table 1). In summary, perceptual acquisition results indicate that perceptual accuracy is positively correlated with training completions and training days, while negatively correlated with the inclusion of syllable productions during perceptual training.

Figure 1

Correct Rate by Training Type, Faceted by Experiment Day and Training Phases



Note: Accuracy on the ABX discrimination task. Top left two panels: pre- and post-training test results for the Perception + Production Group on Experiment Day 1. Top right two

panels: pre- and post-training test results for the Perception + Production Group on Experiment Day 2. Bottom left two panels: pre- and post-training test results for the Perception Only Group on Experiment Day 1. Bottom right two panels: pre- and post-training test results for the Perception Only Group on Experiment Day 2.

Table 1

Model Summary with estimates and confidence intervals given in log

	<i>b</i>	<i>se</i>	95% CI	<i>z</i>	<i>p</i>
Intercept	-0.04	0.10	-0.23, 0.15	-0.42	0.67
Training Type	-0.08	0.05	-0.18, 0.01	-1.82	0.07
Perception +Production: Perception Only					
Phase	0.19	0.05	0.09, 0.27	3.83	< .001
Post: Pre					
Experiment Day	0.10	0.47	0.01, 0.19	2.16	<.05
Day 2: Day 1					

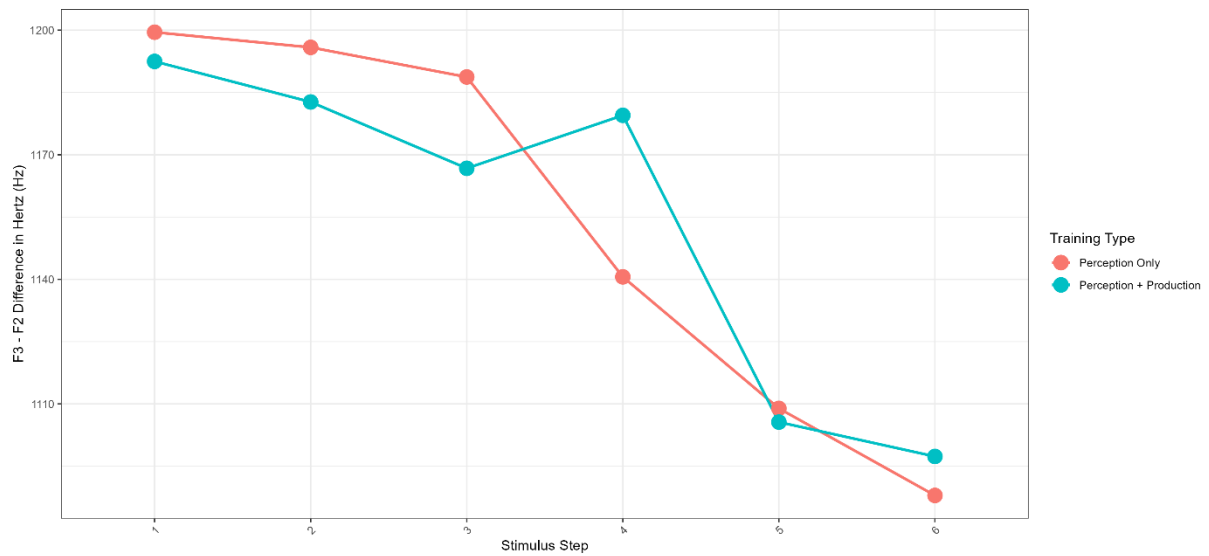
Production acquisition results contradicted our hypothesis, as we were unable to identify any successful acquisition of retroflex stops after auditory inspection and comparing the spectrum with that of native speakers. The predominant production acquisition result for the participants was the production of a velar stop [k] when presented with retroflex stop stimuli. However, we noticed that the convergence in velar stop [k] production could be a potential indicator of the acquisition of retroflex stop production. The distance between F2

and F3 for both the retroflex and velar stop was approximately 500 to 800 Hz shorter compared to the F2–F3 distance for the dental stop, due to the comparatively lower F3 shared by the retroflex and velar stop. As a result, reporting on this property retained the indicative capability of participants' acquisition of the retroflex stop production.

Figure 2 illustrates the production performance of participants in the Perception + Production and Perception Only groups. Specifically, Figure 2 suggests that there was a clear drop in F2–F3 distance at the cross-boundary point, and a continuous decline in formant distance as the representativeness of the retroflex stop of the target syllables increased. A follow-up linear mixed-effect regression, which reported significant effects for target steps 4, 5, and 6, supported this observation (See Table 2). Like the perceptual acquisition results discussed previously, experiment days and within-day training were significant predictors of production learning (see Table 2). Although training types failed to elicit significant main effects, a significant interaction was observed between training types and target steps at step 4 (see Table 2). At boundary–crossing step 4, the Perception + Production group produced a larger F2–F3 distance compared to the Perception Only group, indicating a weaker acquisition of retroflex production. Unlike the perception result, the inclusion of production training during the perceptual acquisition phase did not significantly influence participants' acquisition of nonnative phonemic production.

Figure 2

Mean F3 – F2 Difference by Pair Number and Condition



Note: By step F3 - F2 differences in production training for Perception Only and Perception + Production Groups.

Table 2

Model Summary with estimates and confidence intervals given in Hertz (Hz)

	<i>b</i>	<i>se</i>	95% CI	<i>df</i>	<i>t</i>	<i>p</i>
Intercept	1229.71	37.82	1155.57, 1303.84	21.99	32.51	< .001
Training Type	13.53	66.73	-117.25, 144.31	21.56	0.20	0.84
Perception + Production: Perception Only						
Phase	-21.39	5.27	-31.71, -11.07	4289.21	-4.06	< .001
Post: Pre						
Experiment Day	-46.76	5.44	-57.43, -36.10	4289.16	-8.59	<.001
Day 2: Day 1						
Target Step	-4.13	11.02	-25.72, 17.47	4289.00	-0.38	0.71

Step 2: Step 1						
Target Step	-11.25	10.99	-32.79, 10.28	4289.00	-1.02	0.31
Step 3: Step 1						
Target Step	-59.55	11.01	-81.11, -37.98	4289.00	-5.41	<.001
Step 4: Step 1						
Target Step	-90.39	11.00	-111.94, -68.84	4289.01	-8.22	<.001
Step 5: Step 1						
Target Step	-110.88	11.00	-132.44, -89.32	4289.01	-10.08	<.001
Step 6: Step 1						
Interaction:	43.18	19.70	4.56, 81.80	4289.01	2.19	<.05
Training Type × Target Step: Step4						
<hr/>						
	<i>Std.Dev.</i>					
<hr/>						
Random effect:	142.50					
Subject (Intercept)						
Residual	172.90					
<hr/>						

Discussion

One aim of this study is to replicate the effect identified by previous studies (Baese-Berk & Samuel, 2016, 2022; Baese-Berk, 2019; Baese-Berk et al., 2024), which is the inhibition of perceptual acquisition of nonnative phonemes under a learning paradigm that simultaneously practices the perception and production of the target phonemes. Our study partially concurred with the previous findings, as the statistical results represent a near-significant effect. This finding revalidates the pattern of interactions between the acquisition of perception and the production of nonnative phonemic contrasts and extends its practical implications.

The inhibitory relationship between L2 perception and production acquisition provides a more detailed explanation of the correlation between the two modalities. While contradicting the long-term positive relationship documented by a series of previous studies, this finding introduces a potential reason for the lack of a stronger positive correlation between the acquisition of perception and production. For instance, some learners who constantly practice producing a novel phoneme after listening to it may suffer from the inhibitory effect, while others who do not repeat the sound may achieve better perceptual learning results. Consequently, the different acquisition strategies employed by L2 learners could be a source of disruption to the long-term correlation between L2 perception and production acquisition. A longitudinal study analyzing the interaction between the two modalities on learners using different strategies could provide a more comprehensive account of the influence of learning strategies.

A potential explanation for the inhibitory effect that the Perception + Production learning paradigm exerts is the additional cognitive load required by the sequential implementation of two types of training (Baese-Berk et al., 2024). Neuroimaging studies

could help specify the cognitive load exerted by the simultaneous execution of the acquisition of perception and production of nonnative contrasts. For example, studies investigating the participation of the speech motor area during speech perception using electromyography and functional magnetic resonance imaging techniques have shown the activation of speech motor areas during speech perception (Watkins et al., 2003; Callan et al., 2014). Moreover, Callan et al. (2014) reported a significantly enhanced stimulation of the speech motor area when participants were required to perceive nonnative contrasts. Based on the available evidence, we speculate that the speech motor area also participates in the phonetic acquisition process of nonnative contrasts and contributes to the development of perception and production of new phonemes. Consequently, the simultaneous implementation of training in perception and production led to competition for a limited speech motor planning capacity, resulting in the inhibition of perceptual acquisition.

The other purpose of this study was to identify how the simultaneous processing of perception and production training would interact with the acquisition of nonnative phoneme production. Although we observed a significant effect of the training paradigm when participants repeated a retroflex stimulus (step 4), we were reluctant to accept this result as a definitive rebuttal to our hypothesis. Firstly, the production test design did not involve pairwise comparisons, unlike the perceptual test, which potentially increased the likelihood of misperception by providing near-boundary auditory stimuli. As a result, whether training paradigms influenced the acquisition of nonnative phoneme production could not be determined. Secondly, the effect of the training paradigm at step 4 explained an augmentation in F2–F3 distance by 43.18 Hz (See Table 2), which is a minor influence that could be attributed to the noise in the data.

This finding suggests that at the beginning of L2 phonemic acquisition, the immediate implementation of production training following the perception of target phonemes does not significantly improve learners' production compared to those who focus only on perceptual training. Given that the introduction of production training could hamper the perceptual acquisition of nonnative contrasts, we argue that during the first days of L2 phonemic acquisition, the exclusion of production training, especially the repetition of the phonemes, would provide a better learning outcome in the sensitivity to nonnative contrasts.

However, we could not thoroughly falsify the presence of a negative correlation between the inclusion of production training during perceptual acquisition. When this correlation is further proven, we would argue that in the fuzzier middle area of the Hindi [ɖ]-[t] continuum, the simultaneous practice in perception and production of the [ɖ]-[t] contrast directly inhibits learners' acquisition of the production of the retroflex stop. Another explanation would be that the inhibited perceptual acquisition of the retroflex stop subsequently impedes learners' acquisition of its production.

Interestingly, this production training result revealed a learning outcome in nonnative phoneme production for learners attempting to self-study a new language with only written and auditory learning materials, without receiving feedback from a native speaker or a professional instructor. When encountering nonnative phonemes close to the well-established phonological categories in L1, learners were inclined to assimilate the pronunciation of these nonnative phonemes to the corresponding L1 phonemes or substitute the nonnative phonemes directly in this case. An open question is whether the participants produced velar stops when hearing the retroflex stop stimuli because they perceived the retroflex stop as a velar stop, or whether they recognized the difference between a velar stop and a retroflex stop, but produced the retroflex stop since they could not produce the retroflex stop without speech

motor instructions and sufficient practice. Future studies could include post-experimental questionnaires on this issue.

Conclusion

This study aims to replicate the inhibitory effect of concurrent practice in perception and production on nonnative phonemes for the perceptual acquisition of these phonemes, as identified by previous studies (Baese-Berk & Samuel, 2016, 2022; Baese-Berk, 2019; Baese-Berk et al., 2024). Additionally, we investigated this effect on the acquisition of nonnative phonemes. The experiment results confirmed our hypothesis, suggesting that simultaneously implementing training in the perception and production of a nonnative phoneme inhibited the perceptual learning outcome, while the learning outcome for production was not influenced.

This study expanded the practicality of the inhibitory effect of production training on the perceptual acquisition of nonnative phonemes to Hindi retroflex stops, suggesting the unidirectionality of this inhibition. Furthermore, we indicated that subsequent studies investigate the involvement of speech motor areas in this inhibitory effect and actualize the representation of nonnative phonemes for naive listeners.

Reference

- Kuhl, P. K., & Meltzoff, A. N. (1997). Evolution, nativism, and learning in the development of language and speech. *The inheritance and innateness of grammars*, 7-44.
- Flege, J. E. (1995). Second language speech learning theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*/York.
- Best, C. T., Tyler, M., Bohn, O., & Munro, M. (2007). Nonnative and second-language speech perception. *Language experience in second language speech learning*, 13-34.
- Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, 89, 23-36.
- Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, 72(5), 1218-1227.
- Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception & psychophysics*, 62(4), 843-867.
- Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(3), 732.
- Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [rl] distinction by young infants. *Perception & Psychophysics*, 18(5), 341-347.
- Lovčević, I., & Tsuji, S. (2024). The developmental pattern of native and non-native speech perception during the 1st year of life in Japanese infants. *Infant Behavior and Development*, 76, 101977.

- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, 5(11), 831-843.
- Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: An investigation of English/r/-/l/perception and production by adult Japanese speakers. In *Interspeech workshop on second language studies: Acquisition, learning, education and technology* (pp. 22-24). Tokyo: Waseda University.
- Flege, J. E. (1999, August). The relation between L2 production and perception. In *Proceedings of the XIVth International Congress of Phonetics Sciences* (Vol. 2, pp. 1273-1276). Berkeley: Department of Linguistics, Univ. of California at Berkeley.
- Flege, J. E. (1980). Phonetic approximation in second language acquisition 1. *Language learning*, 30(1), 117-134.
- Eckman, F. R., Elreyes, A., & Iverson, G. K. (2003). Some principles of second language phonology. *Second Language Research*, 19(3), 169-208.
- Baese-Berk, Melissa M., and Arthur G. Samuel. (2022). "Just give it time: Differential effects of disruption and delay on perceptual learning." *Attention, Perception, & Psychophysics* 84.3 (2022): 960-980.
- Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, 81, 981-1005.
- Baese-Berk, M. M., Kapnoula, E. C., & Samuel, A. G. (2024). The relationship of speech perception and speech production: It's complicated. *Psychonomic Bulletin & Review*, 1-17.

Repp, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli.

Bulletin of the Psychonomic Society, 18(1), 12-14.

Stevenson, D. C. (1979). Categorical perception and selective adaptation phenomena in speech.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America, 119*(3), 1684-1696.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989–994.

Callan, D., Callan, A., & Jones, J.. (2014). Speech Motor Brain Regions are Differentially Recruited During Perception of Native and Foreign-Accented Phonemes for First and Second Language Listeners. *Frontiers in Neuroscience, 8*.