Just give it time: Differential effects of disruption and delay on perceptual learning

Melissa M. Baese-Berk[a,b,*] & Arthur G. Samuel[b,c,d]

[a]Department of Linguistics

1290 University of Oregon

Eugene, OR 97403

541-346-4899

mbaesebe@uoregon.edu

[b] Basque Center on Cognition Brain and Language

Paseo Mikeletegi 69, 2nd Floor

Donostia-San Sebastian 20009 Spain

[c] IKERBASQUE

Basque Foundation for Science

Bilbao 48011 Spain

[d] Department of Psychology

Stony Brook University

Stony Brook, NY 11794-2500

[*] Corresponding author

Abstract

Speech perception and production are critical skills when acquiring a new language. However, the nature of the relationship between these two processes is unclear, particularly for non-native speech sound contrasts. Although it has been assumed that perception and production are supportive, recent evidence has demonstrated that, under some circumstances, production can disrupt perceptual learning. Specifically, producing the to-be-learned contrast on each trial can disrupt perceptual learning of that contrast. Here, we treat speech perception and speech production as separate tasks. From this perspective, perceptual learning studies that include a production component on each trial create a task switch. We report two experiments that test how task switching can disrupt perceptual learning. One experiment demonstrates that the disruption caused by switching to production is sensitive to time delays: Increasing the delay between perception and production on a trial can reduce and even eliminate disruption of perceptual learning. The second experiment shows that if a task other than producing the to-be-learned contrast is imposed, the task switching component of disruption is not influenced by a delay. These experiments provide a new understanding of the relationship between speech perception and speech production, and clarify conditions under which the two cooperate or compete.

Keywords: Language production, language comprehension, second language acquisition, task-switching

Public Significance Statement

This study suggests that when learning a new language the relationship between listening to that language and producing it may be more complex than previously thought. Further, the relationship is affected by timing on very short time scales (i.e., a few seconds).

Relationship between speech perception and production

The relationship between speech perception and production has long been a focus of research. Because both skills are required for successful communication, a common assumption has been that perception and production share similar mental representations and rely on similar processes (e.g., Best, 1995; Fowler, 1986; Liberman et al., 1952, 1967; Liberman & Mattingly, 1989). However, in fluent speech processing by adults, it is often difficult to investigate the relationship between the two modalities because both perception and production are typically rapid and accurate. Therefore, many researchers have turned to learning of non-native language material as a test-bed for examining the relationship between these modalities.

The results of studies examining the relationship between perception and production during second language learning have been mixed. Some studies have found that perception and production work synergistically (Hopman & MacDonald, 2018; Wang et al., 2003; Zamuner et al., 2016). Others have demonstrated no relationship between the two modalities during learning and training (Bent, 2005; de Jong et al., 2009; DeKeyser & Sokalski, 1996; Flege, 1993). Perhaps surprisingly, some studies have shown that perception and production can exhibit an antagonistic relationship during learning. That is, under some circumstances, speech production disrupts perceptual learning of speech sounds (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Warker et al., 2008, 2009; Zamuner et al., 2017, 2018) and words (Kaushanskaya & Yoo, 2011; Leach & Samuel, 2007). This result may be surprising given the widely held belief that production and perception are two sides of the same coin. Understanding the source of this disruption is critically important for understanding the nature of both perception and production. In turn, this can clarify when perception and production are mutually supportive, and when they instead are antagonistic.

In the present study, we ask how two aspects of producing tokens – task switching and the timing of the required tasks – impact the disruption of perceptual learning, in order to better understand why production disrupts perceptual learning in some circumstances[1]. To situate our two experiments, we briefly review some previous research that has examined the relationship between perception and production in both children and adults, across a variety of linguistic structures. We then review relevant literature on task-switching and the role of timing during learning, across a variety of domains. While we are specifically investigating interactions between perception and production in the present study, we do so within the broad framework of known factors that impact learning.

Interactions between speech perception and production during learning

As noted above, second language learning has provided a very compelling test-bed for investigations of the relationship between speech perception and production[2]. In previous work in second language learning, some studies have shown cooperation between perception and production. For example, Zamuner et al. (2016) demonstrated that participants recognize words produced during training more quickly than words that were only heard. Further, they are better able to resolve conflicting information (e.g., mispronunciations) if they produced words during training, instead of just hearing them. Similarly, Hopman and MacDonald (2018) examined language learning in adults who were assigned to either just-comprehension or production-and-comprehension training conditions. Participants whose training included production

[1] Our focus in the present work is on perceptual learning, as perceptual learning is a complex phenomenon influenced by many factors, including the production of the to-be-learned items. Here, we ask how production during training may impact perceptual learning, but do not focus on production learning itself. Speech production is extremely complex in its own right, and is the focus of substantial research. Here, we focus on perceptual learning, and leave the question of learning in production, and how that correlates with learning in perception, for future studies.

[2] Another potentially fruitful test-bed for investigation would be first language acquisition; however, previous work has suggested that the two modalities are quite tightly yoked during first language acquisition (see e.g., Hearnshaw et al., 2019).

demonstrated more robust learning of novel words and of grammatical dependencies among these new words. Indeed, a number of studies have demonstrated a "production effect" in language learning (Icht & Mama, 2015; Kaushanskaya & Yoo, 2011; MacLeod et al., 2010).

In the domain of speech sound learning, Bradlow and colleagues demonstrated that, at the group level, learners improve in production after training in perception alone (e.g., Bradlow et al. 1999; Bradlow et. al 1997). Although there was robust learning at the group level, individual performance was highly variable: some participants demonstrated robust learning in production without improvement in perception, and vice versa. Other studies have shown similar individual variation (Sheldon & Strange, 1982), or have failed to show a significant correlation between the two modalities (Bent, 2005; de Jong et al., 2009; Flege, 1993; Rochet & Strange, 1995).

Intriguingly, under some circumstances, producing tokens during training can actually disrupt perceptual learning (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007; Zamuner et al., 2018). That is, when learners repeat the to-be-learned token during each perceptual training trial, they show less robust learning than individuals who are not required to complete the production component of training. Figure 1, based on results of Baese-Berk and Samuel (2016), illustrates this effect. This figure shows performance on a discrimination task before training (left panel) and after (center and right panels). Each point on the x-axis is a pair of stimuli along a continuum (e.g., "Pair 1" is the comparison between the first stimulus on the continuum and the second; see Procedure below for more details). The dotted line in the panel is at 50% — chance performance. If participants are able to discriminate pairs that cross a category boundary, we expect a peak in the center of the distribution (i.e., an inverted V-shape). If they are unable to discriminate across a category boundary, we expect a flat function. The center panel has a peak, showing that participants in the "Perception-Only" group developed two

categories, whereas the right panel has no peak, showing that the "Perception+Production" group did not learn the category distinction. Because this study is critical for and intersects with the current study, we describe it in detail here.

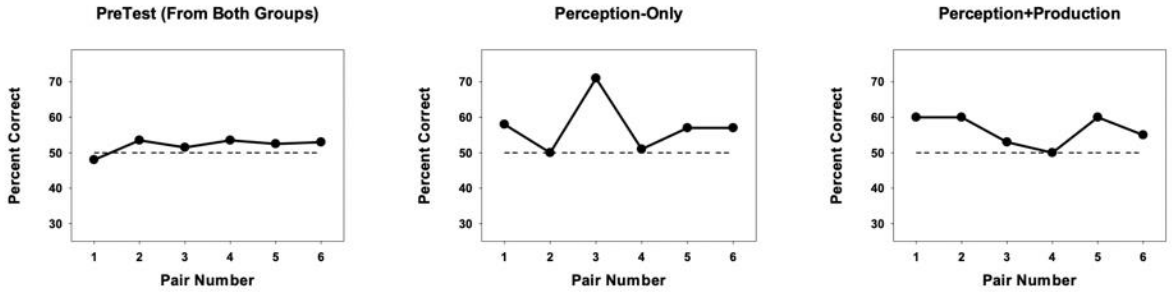


Figure 1: Left panel shows the pre-test data from the two training groups from Baese-Berk and Samuel (2016). Middle panel shows post-test data from the Perception-Only training group with a robust discrimination peak in the middle of the continuum. Right panel shows post-test data from the Perception+Production training group with no discrimination peak after training.

The participants in Baese-Berk and Samuel (2016) were native Spanish speakers who were being taught the distinction between "s" and "sh"; Spanish has "s" but does not have "sh". Using a 19-step continuum of syllables that ranged between /sa/ ("sah") and /ʃa/ ("shah"), each training trial presented two different tokens in an ABX discrimination task. Participants heard the two distinct tokens (A and B) and then a repetition of one of those tokens (X). Participants were asked to determine whether this third token was the same as the first or as the second token; feedback was given on each training trial. When participants started the experiment (i.e., at pre-test), because they only had one category (all tokens were heard as "s"), discrimination should be at chance, and it was (see the left panel of Figure 1).

Critically, one group of participants just did this perceptual task during training, while a second group also was required to produce the "X" token before making their ABX decision. For the Perception-Only training group, listeners learned the second category ("sh"), and as a result their discrimination function after training took on the "inverted-V" shape that is typically

found for speech sounds (usually discussed in the context of categorical speech perception) – see the middle panel of Figure 1. This higher discrimination score near the middle of the continuum demonstrates that, for pairs of A and B stimuli that cross a category boundary (i.e., the boundary between "sah" and "shah"), listeners are able to differentiate stimuli better than for pairs of stimuli that do not cross a category boundary. That is, they are sensitive to an acoustic distinction when that distinction is meaningful (i.e., signals two different categories), but not when it is not meaningful (i.e., is variation within a single phonological category). Indeed, this is the location of the discrimination peak for native listeners who are able to differentiate across, but not within, the two categories (/s/ and /ʃ/). Therefore, the emergence of this peak was taken as a demonstration of the emergence of categorical perception, a consequence of learning to distinguish between these novel categories[3].

Critically, even though participants in the Perception+Production condition experienced exactly the same perceptual training, their production during training blocked their learning of the perceptual distinction, leaving them with the same flat discrimination function that they had before training – there was no hint of improved discrimination near the middle of the continuum, the hallmark of there being two contrasting phonemic categories (see the right panel of Figure 1).

Our first paper demonstrated the basic phenomenon and suggested that extensive experience with a contrast and with task-switching could reduce the disruption to perceptual learning. The current study builds on those findings, investigating additional factors that may impact learning, and comparing new training groups to our original training groups. Deepening

[3] In the present analyses, we focus on inter-category learning, which is often taken as a hallmark of categorical perception and novel phoneme learning. Of course, it is possible that participants are improving their intra-category perception in addition to this inter-category learning. However, because our current and previous results do not demonstrate strong evidence for this, our focus here is on learning between categories. It is important to note, though, that speech perception likely includes both categorical and continuous properties and the type of performance displayed by participants may be influenced by the task being completed (see, e.g., McMurray et al., 2002).

our understanding of how production during training impacts perceptual learning is critically important because previous studies examining the interplay between the two modalities have been mixed.

Factors influencing the relationship between perception and production

There are many factors that vary from study to study, and understanding the pattern of results in the literature requires a careful consideration of the particular experimental details of each study. It is possible that one cause for the inconsistent results in the literature is the type of linguistic information that is being learned (e.g., syntactic, morphological, phonological, or phonetic information). For example, Hopman and MacDonald (2018) trained participants on syntactic patterns and found benefits in perception for producing to-be-learned constructions. Learning abstractions like syntactic rules generally will benefit from explicit practice, and production provides such practice, helping perceptual learning. Note that the syntactic rules will be the same, whether one is producing a sentence or listening to it. In contrast, for information that is more specific (e.g., what a particular speech sound is like) different factors may dominate. For example, for some contrasts the cues that are used to differentiate two sounds in perception are the same as those cues in production (e.g., geminate stops rely on duration cues in both perception and production). However, in other cases, the cues differ between the two modalities (e.g., tongue placement vs. changes in F3 for English /ɹ/ and /l/). Indeed, the relationship between speech perception and production may differ for these two types of contrasts (Kato & Baese-Berk, 2020).

Further, there may be multiple routes to perception and different types of perception may be used for different tasks (e.g., Scott, 2005; Scott & Johnsrude, 2003). For example, brain regions that are active for phoneme judgments may not be used for lexical access (Krieger-

Redwood et al., 2013). Therefore, perception to be used for eventual production and perception used for lexical access (or similar tasks) may not engage identical processes.

Some recent evidence suggests that, even though production can disrupt perceptual learning for novel speech sounds, participants who produce tokens during learning do show robust learning in the production modality. That is, participants in some cases demonstrate improvement in the accuracy of their productions (either repetition or naming) from pre- to post-test, even when they do not demonstrate improvement in perception. Importantly, performance across the two modalities is not correlated (Baese-Berk, 2019). These results constrain theories of the relationship between perception and production: Improvement in one modality does not necessarily entail improvement in the other modality, and engaging one modality may have an adversarial effect on the other. Given these results, delineating when production will help learning versus when it will hurt learning is far from simple (see Zamuner et al., 2017).

Although the existing literature suggests that there is a complex relationship between perception and production, a true understanding of each will require understanding the circumstances under which production helps versus harms perceptual learning. The inconsistent results in the literature make it likely that multiple factors are at work, and in fact previous research has suggested a number of candidates. For example, Baese-Berk (2019) demonstrated that individuals who were more variable in their productions during training developed less robust perceptual learning than individuals whose productions were less variable. It is important to note that the finding of production disrupting perceptual learning in some circumstances has been demonstrated across multiple targets of learning (e.g., English word learning for adults, Leach & Samuel, 2007; English word learning for children, Zamuner et al, 2018; Spanish word learning for adults, Kapnoula & Samuel, submitted; Basque fricatives learned by native Spanish

speakers, Baese-Berk & Samuel, 2016; prevoiced and short-lag stops learned by native English speakers, Baese-Berk, 2019), suggesting that the effect is relatively robust, even if many find it to be surprising.

In our previous work, we investigated two factors that may impact the disruption of perceptual learning: long-term familiarity with a contrast and production of the target token per se. We demonstrated that late-learners of a language demonstrate a smaller disruption to perceptual learning than naïve listeners (Baese-Berk & Samuel, 2016), suggesting that having even an imperfect initial perceptual representation can reduce the disruption of perceptual learning caused by production. Though all groups were at chance before training, participants in the late-learners group demonstrated more learning than those in the naïve group. However, even for the late learners, participants trained in perception alone demonstrated substantially larger improvements in perceptual learning than participants trained with both perception and production.

Our previous work (Baese-Berk & Samuel, 2016) has also shown that while production of the target token per se plays a role in the disruption to perceptual learning, it is likely that switching between two tasks during learning, regardless of the nature of those tasks, also contributes to the disruption. Specifically, when listeners were asked to produce an unrelated item (e.g., to name a letter unrelated to the training tokens) during training, there was a smaller, but still significant disruption to perceptual learning. This finding suggests that the disruption to perceptual learning is generated by at least two components: a linguistic component and a more general cognitive component. The more general cognitive component may reflect the challenge of switching between two tasks. A training paradigm that includes both perception and production is inherently a dual-task situation. Participants need to learn the perceptual

representation of the new speech sound (i.e., Task 1) and the production representation of the same sound (i.e., Task 2). From this perspective, the outcome of perceptual learning will depend on the participant's ability to switch successfully between these two tasks. In the following section, we briefly discuss some of the relevant findings from research on task switching.

The role of task switching during learning

Task switching, or changing from one cognitive task to another, has long been known to influence processing (Jersild, 1927). When switching between two tasks, individuals are slower and less accurate than when performing a single task (see Monsell, 2003 for a review). Task switching has several components. In general, 'switch' trials are slower than 'non-switch' trials; however, this effect is partially alleviated by preparation (e.g., Fink & Goldrick, 2015; Monsell et al., 2003): Knowing a switch is coming allows a participant to effectively prepare for the switch. However, the cost of switching is typically not eliminated by a delay (Kimberg et al., 2000; Sohn et al., 2000). Further, switch costs are typically higher when the retrieval demands of the second task are higher (Mayr & Kliegl, 2000). Given that speech production of an unfamiliar sound is a demanding task, the cost of task switching may be heightened by requiring individuals to produce such an unfamiliar token.

Often, task switching has been examined through paradigms that ask the participant to respond to different elements of a stimulus across trials. For example, on one trial a participant may be asked to classify a digit as high vs. low, whereas on another trial the judgment is even vs. odd. To respond appropriately on a given trial, participants must ignore some information that they have already processed about the stimulus during a previous trial (i.e., task-set inertia). A common language-based task switching paradigm involves code-switching, or asking bilingual speakers to switch between their two languages on any given trial. This work has demonstrated

"switch-costs" associated with producing speech in one language and then switching to produce in another language (e.g., Broersma et al., 2016; Kirk et al., 2018). These switches are typically unpredictable and cued, though voluntary switches have also been investigated (e.g., de Bruin et al., 2018; Gollan & Ferreira, 2009). Even when switching is voluntary, switching has a measurable cost.

In the present study, participants are asked to complete a perceptual judgment task *and* to produce a token on each trial. Previous work has demonstrated costs of mixing tasks in a block, not just switching on a given trial (e.g., Koch et al., 2005), reflecting the need to maintain two types of preparation simultaneously. Studies of switching costs and mixing costs suggest that participants who are trained in both perception and production may face the cognitive challenge of switching between two tasks on a given trial or in a given block. At least for switching costs, one factor that can affect the results is the timing of the required change from one task to the other – longer delays allow the participant to "shut down" the first task, and initiate the second, with a reduced switching cost.

The role of timing during learning

While it is clear that long timescales play a role in learning (i.e., learners improve over long stretches of time, and age of acquisition can influence learning; Archila-Suerte et al., 2012; Guion et al., 2000; Huang & Jun, 2011; Mackay et al., 2006), much less attention has been paid to the role of shorter timescales during language learning (though see the many studies that focus on relatively short-term laboratory-based training studies; e.g., Bradlow et al., 1997). These shorter timescales are potentially critical in understanding the varying effects that have been found in studies that combine perceptual learning with production requirements. However, there has not been any systematic investigation of timing in such studies yet. If, as suggested above,

task switching is time-sensitive, one might expect robust effects of delaying a task in our present work. If participants are able to delay a response, they may be able to successfully complete perceptual processing and learn the new distinction. On the other hand, if participants are forced to stop perceptual processing to initiate a second task, the cost of switching could be substantial. Given the lack of systematic investigation of the influence of short-term time scales on non-native speech sound learning, we begin our investigation here looking at delays around 2 seconds and around 4 seconds. We chose these values based on the literature on echoic memory which suggests that detailed acoustic information can be maintained for approximately this length of time (Darwin et al., 1972).

The hypothesis that timing may be critical in perceptual learning stems from previous speech perception studies that have found timing to be a driving factor. For example, the timing and order of stimuli within a trial of a discrimination test can affect the results. In an ABX task, participants must decide whether a token is the same as the first or second sound they heard. In this task, perception tends to be categorical. However, the same stimuli in a different task result in more continuous performance. That is, in a 4IAX task, which asks participants to choose which of two pairs of stimuli contain different sounds and which of the pairs contains the same sounds, participants tend to perform less categorically (Pisoni & Lazarus, 1974). Further, temporal separation of tokens within a trial can alter perception performance (e.g., Schouten et al., 2003). Given that short-term timing affects the perception of speech sounds, we certainly would expect it to affect the perceptual encoding of to-be-learned non-native speech sounds. Moreover, in perception/production learning studies, the learner is required to repeat the token aloud, entailing perception followed by production. A potentially critical factor is the timing between executing the first task – perceiving the token – and the second one – producing it

aloud. That is, speech perception, and perceptual learning, may be especially sensitive to timing, especially when coupled with another, particularly demanding, speech related task. It is also possible that this sensitivity to timing is reduced when the other task is less demanding or does not also entail speech processing. That is, the disruption we have observed in previous studies could be driven by two separable mechanisms – producing the tokens per se and task switching. Alternately, it could be that timing and task-switching are intertwined such that timing affects learning regardless of the second task.

Present study

In the present study, we investigate the source of the disruption to perceptual learning when producing tokens. We propose that many studies in the literature, including our previous work, use paradigms that result in dual-task situations. From this perspective, we conduct two experiments to investigate factors that influence performance on one of the two tasks – perceptual learning. In Experiment 1, we manipulate the timing of the two tasks to investigate whether temporal separation impacts the size of the disruption of perceptual learning. In Experiment 2, we keep the perceptual task constant, but manipulate the nature of the second task to ask whether the properties of the dual-task impact perceptual learning. In both experiments, the participants are native Spanish speakers who undergo training to learn the distinction between /sa/ (“sah”) and /ʃa/ (“shah”). Recall that for native Spanish speakers there is no such distinction because Spanish only includes /s/, not /ʃ/. We tested participants in a region of Spain in which exposure to languages with this distinction (e.g., Basque or English) is low. Thus, when the participants start the study, all of the stimuli that they hear sound like “s”, as is evident in the left panel of Figure 1. The question we examine is what factors determine how well the listeners learn to hear two separate categories.

Experiment 1 – Delayed production

As we have noted, a situation in which production is required during perceptual learning is effectively a dual task learning condition. From this perspective, giving learners time to do one task (perceptual learning) before they must do the other (production) should improve performance on the first task compared to testing that does not provide this undisturbed perceptual processing period. To test this hypothesis, in Experiment 1 we include one condition with a 2 second delay before the production demand, and a second condition with a 4 second delay. We compare these two delay conditions to the Perception-Only condition (i.e., no second task) and the Perception+Production condition (i.e., dual task, with no delay) of our previous study (Baese-Berk & Samuel, 2016). If the hypothesis is correct, then providing a delay should produce results in between these two endpoints, with the results also informing us about the amount of time needed for undisturbed perceptual learning.

Method

Participants

45 participants completed this experiment. 20 participated in the Short-Delay condition, and 25 were in the Long-Delay condition (see below)[4]. These participants were compared to the participants in the Perception-Only (No-Delay; n=15) and Perception+Production (No-Delay; n=15) conditions previously reported in Baese-Berk & Samuel (2016). All participants were native speakers of Spanish living in Murcia[5], with limited experience with Basque, English (i.e.,

[4] A variable number of participants were included across conditions. In general, we recruited to the "floor" number of participants. Once that number was reached, we continued running other participants who had already signed up for the experiment. Further, the total number fluctuated because some participants were unable to complete the second day of training within the specified time frame. Thus, their data were not included in the analysis.

[5] All participants in both experiments presented in the current manuscript lived in Murcia in southern Spain, and thus did not have significant community exposure to Basque (spoken primarily in the north of Spain)

self-rated proficiency in either language was <3 on a 9-point scale; did not report frequent use of either language), and other non-native languages. Participants in all groups were between 18-40 years old. No participant reported a history of speech, language, or hearing disorders. Our sample size was determined using effect sizes from Baese-Berk (2019). Using the simr package in R (Green & MacLeod, 2016*),* we calculated that a sample size of 20 participants is appropriate to detect the effect size reported in that paper at .8. All participants were recruited from the same population and around the same time as those originally reported in Baese-Berk and Samuel (2016).

Stimuli

Stimuli used in the experiments presented here were identical to those used in Baese-Berk and Samuel (2016). Here, we provide a description of these stimuli; a more complete description is available in that paper. A native Basque speaker recorded Basque sibilant fricatives and affricates in a sound-treated room using a Sennheiser ME65 microphone. A 19-step continuum from /sa/ to /ʃa/ was created using a mixing algorithm that shifted from one fricative to the other using a weighted-average method (see Kraljic & Samuel, 2005, 2006, 2007; Leach & Samuel, 2007 for a similar method)[6]. Stimuli were all 406 msec in duration; the consonantal portion was 213 msec and the vocalic portion was 193 msec.

The continuum was pilot tested with native Basque speakers whose responses confirmed that the mixing algorithm resulted in stimuli that induced categorical perception for native speakers. These native speakers produced a typical speech contrast discrimination function, with

[6] Basque has a three-way place distinction between apico-alveolar /s̻/, lamino-dental /s̪/, and post-aleolar /ʃ/. A similar place distinction exists for the homorganic affricates (/ts̻/, /ts̪/, and /tʃ/). These recordings served as endpoints for three continua (/s̻a/-/s̪a/, /s̻a/-/ts̻a/, and /s̻a/-/ʃa/)[6]. As in the Baese-Berk and Samuel (2016) paper, only the /sa/-/ʃa/ continuum was used in the data analyzed here.

poor discrimination within each category and a peak near the center of the continuum (pair 3; see the "peak" in the middle panel of Figure 1).

Procedure

The basic training and testing procedures were identical to those used in Baese-Berk and Samuel (2016). As in those experiments, all participants completed a pre-test, training, and a post-test. While the tasks included in the training differed across each of the experiments presented below, the pre-test, post-test, and perception exposure during the training were identical for all experiments. Further, the pre- and post-tests were identical to each other for all participants in the current studies and in Baese-Berk & Samuel (2016). The hardware, software, and peripherals (e.g., headphones, speakers, and microphones) were identical to those used in the previous study. A schematic for the training and testing paradigm is presented in Figure 2.

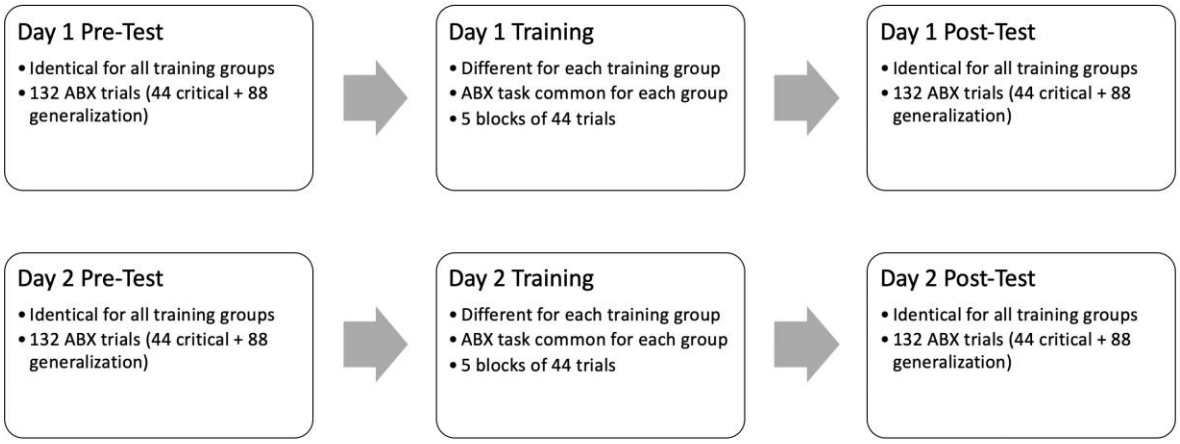


Figure 2: Schematic for training and testing paradigm for all participants. The training groups described in Experiments 1 and 2 performed different training tasks (middle panel of each row) but the testing and the general type of task during training were identical for all groups.

Following many previous studies, training was conducted across two days (Baese-Berk & Samuel, 2016; Baese-Berk, 2019; Earle & Myers, 2014, 2015a). The second day of training took

place a maximum of 48 hours after the first[7]. On each day, participants completed a pre-test, a training period (i.e., 5 blocks of 44 trials with feedback), and a post-test. Participants also completed a language background questionnaire after the experiment.

During the pre- and post-tests, participants heard 44 trials from the target continuum and 88 trials from each of the non-target (i.e., generalization) continua, for a total of 132 trials. The task was an ABX test without feedback. Each trial consisted of three tokens. Participants heard two tokens that were different from each other (A and B) and then were presented with a third token (X) that was identical to either token A or token B. After presentation of the third token, they were asked to determine whether this third token was the same as the first or second token they heard. They responded by pushing one of two buttons. On each trial, the tokens were presented with a 300 msec interstimulus interval, and participants had 3 seconds to respond after the last token was presented. The "A" and "B" portion of the trial consisted of pairs of tokens that were each four steps apart on the original 19-step continuum (i.e., tokens 1-5, 5-9, 7-11, 9-13, 11-15, and 15-19[8]). Training used the same ABX task, but with feedback ("correct" or "incorrect" message). Order of presentation was randomized.

Training consisted of 5 blocks of 44 trials each day for a total of 440 training trials. Presentation of the trials was randomized for each participant. Training tokens were presented in a symmetric distribution, such that participants heard each pair of stimuli (e.g., tokens 1-5, 5-9, etc.) equally often during training. While the pre-tests and post-tests included stimuli from all

[7] In Experiments 1 and 2, ~85% of participants completed the experiment 24 hours after their first session. The other 15% completed the experiment more than 24 hours later, but within 48 hours after the first session; these participants were roughly equally distributed across conditions. Participants' performance was similar in terms of accuracy on the ABX task regardless of the time between the first and second session, so their data are collapsed in the analyses described below.

[8] In addition to these 4-step pairs, training and testing also include 6-step pairs on the continuum (i.e., 1-7, 5-11, 7-13, 9-15, 13-19), which are not analyzed here. The previous report of our results (Baese-Berk & Samuel, 2016) did not accurately report this aspect of the training and testing paradigm. However, this does not impact the results presented here, as in all experiments, both those reported previously and those reported here, we only analyzed performance on the 4-step pairs.

three of the continua (see footnote 7), the training used stimuli only from the /sa/-/ʃa/ continuum. As in Baese-Berk and Samuel (2016), we examine only learning on the trained continuum (i.e., /sa/-/ʃa/). Participants in all training groups completed a basic ABX discrimination test. However, as mentioned above, the specific details of the timing of this task varied across experiments. These details are described below and schematics for each of the training types and within trial timing are presented below in Figure 3.

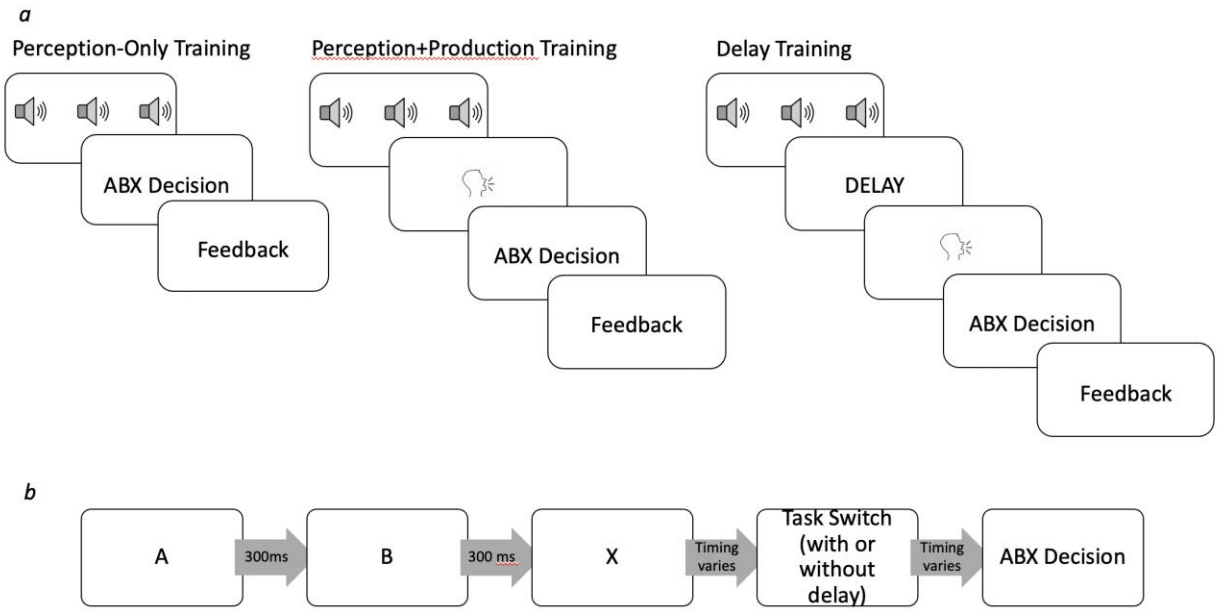


Figure 3: *Panel a* - Schematic for each trial. *Panel b* – Timing within each trial

Recall that Baese-Berk and Samuel (2016)'s study included a Perception-Only condition, and a Perception+Production (here, we will call this "No-Delay") condition. To compare to these two groups, Experiment 1 included two new groups, a short delay Perception+Production condition ("Short-Delay") and a long delay Perception+Production condition ("Long-Delay"). For participants in all three Perception+Production conditions (No-Delay, Short-Delay and Long-Delay) participants were required to say the X item aloud before they made their forced-choice perceptual judgment.

In the No-Delay group, participants were instructed to repeat the X token immediately after hearing it (i.e., as quickly as possible). For the two new delay groups, we manipulated the duration of the interval between presentation of the X item and the participants' repetition. Instead of immediately repeating token X of the ABX task, participants were asked to delay their response until they received a visual cue to produce the token. As such, they were also asked to delay their response to the ABX perception task. Both tasks needed to be delayed in order to better understand the effect of this delay on the previously observed disruption to perceptual learning. In the Short-Delay condition this delay was jittered within participants to be between 500 and 3500 msec. On a given trial, the jitter was instantiated by randomly and equiprobably selecting a delay of 500, 1000, 1500, 2000, 2500, 3000 or 3500 msec. In the Long-Delay condition, the delay was jittered between 2500 and 5500 msec (randomly and equiprobably selected among 2500, 3000, 3500, 4000, 4500, 5000 or 5500 msec). These two delay conditions were chosen because the literature on echoic memory suggests that detailed acoustic information can be maintained for a timeframe between 2 seconds (i.e., the average delay in the Short-Delay condition) and 4 seconds (i.e., the average delay in the Long-Delay condition; Darwin et al., 1972). We chose to jitter the response delay to induce uncertainty in listeners about how long they would need to delay their response. This uncertainty should reduce any strategic behaviors in response planning or execution.[9] In all cases, both for the participant groups reported here and all those reported in Baese-Berk & Samuel (2016), the "X" decision (i.e., the perception task decision) always occurs after the non-perception task (i.e., producing the token in this experiment). That is, a trial would follow this schema: A token, B token, X token, non-

[9] Note that the jitter was only included on training trials, not on test-trials. Therefore, a continuous analysis investigating performance after a specific delay is not possible given the current data.

perception task/production, ABX decision. The delay occurred between the X token and the non-perception task/production.

Analysis

We followed the same analysis procedure used in Baese-Berk and Samuel (2016). We constructed linear mixed-effects models using the lme4 package (Bates et al., 2014) within R (R Development Core Team, 2014). Performance on the ABX post-test[10] was the dependent variable. Fixed effects included training group and continuum pair and their interaction. Training group was Helmert coded to compare various effects. First, we compared the Perception-Only training group to all other training groups (i.e., the three groups in which production was required). Second, we compared the effect of a delay in production to no delay (i.e., the No-Delay Perception+Production group from Baese-Berk & Samuel to both the Short-Delay and Long-Delay groups here). Finally, we examined the effect of length of delay (i.e., comparing the Short-Delay and Long-delay groups to one another).

Continuum pair was categorically coded and models were run using the center of the continuum (Pair 3) as the referent level – that is, each step was compared to the step which is the "peak" of the discrimination function for native speakers. Random effect structure was the maximum that would allow the models to converge, and included only random intercepts for participants. Significance of factors was determined by model comparisons using the anova function in R, reported in the text below. Note that all experimental procedures (e.g., location, participant sampling, computer hardware, software, and peripherals, and stimuli) were identical

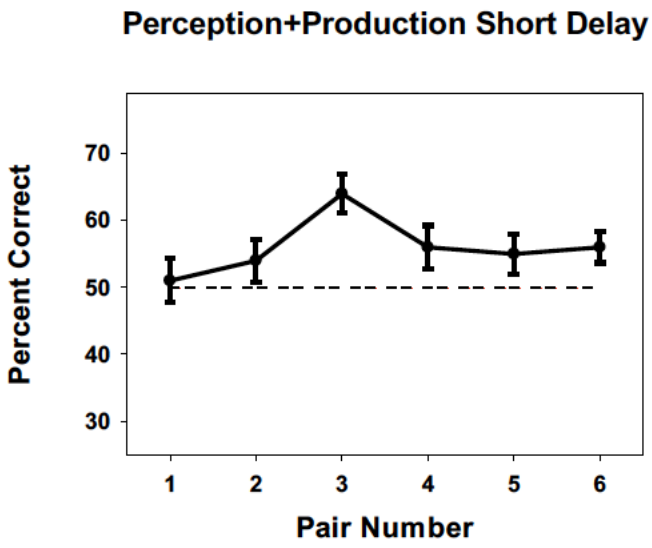
[10] We present only data from the Day 2 post-test. While data was also collected at post-test on Day 1 and pre-test on Day 2, previous investigations of these intermediate tests have not demonstrated robust learning for any training group, so here we focus on the Day 2 post-test. Further, we do not use performance on the pre-test as a covariate as it did not significantly improve model fit. Similarly, we chose not to investigate a measure of change from pre-to post-test as pre-test scores did not vary systematically (i.e., participants were guessing at pre-test, thus they demonstrated chance performance)..

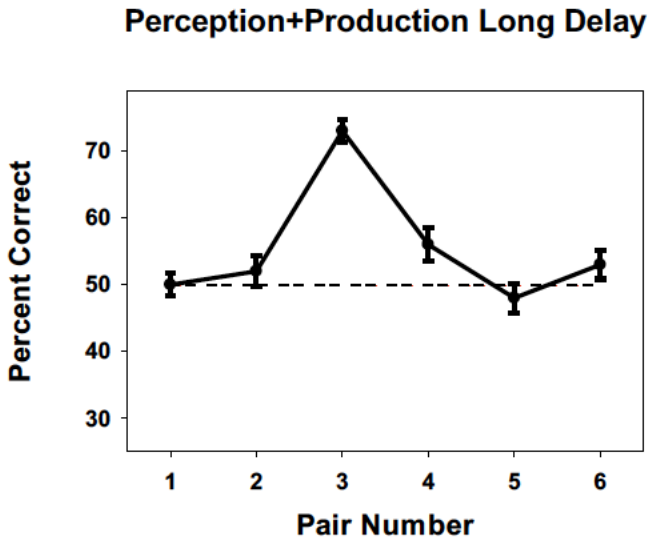
across training groups, including those originally presented in Baese-Berk & Samuel (2016) – the only differences were those induced by the particular training conditions (e.g., delay between the X token and its repetition). Results of the full models can be found in Appendix 2.

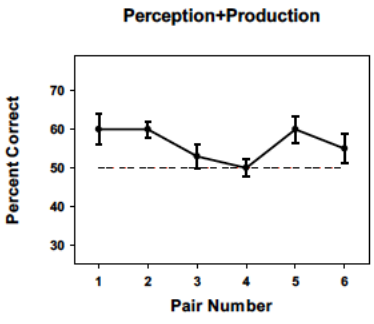
Results and Discussion

Before training, no groups differed from each other and all groups performed at chance (all $\chi^2 < 2$, $p > .1$) (see Appendix 1). Figure 4 presents the average results at post-test for the training groups. The two smaller panels at the bottom of the figure are copies of full-size panels from Figure 1, included here to simplify visual comparisons for the reader. The top panel shows the post-test scores for the Short-Delay training group, and the middle panel shows the post-test scores for the Long-Delay training group. These figures show that the disruption to perceptual learning demonstrated by the No-Delay Perception+Production training group is alleviated when participants delay their response to the stimuli. Specifically, the Short-Delay training group shows a small discrimination peak and the Long-Delay training group shows a large discrimination peak. The peak in the Long-Delay condition is at least as high as that for the Perception-Only condition in Baese-Berk and Samuel (2016), demonstrating that if the second task (production) is separated sufficiently from the first (perception), the between-task

antagonism is eliminated.







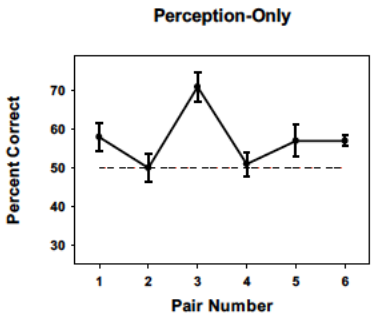


Figure 4: Bottom-left panel shows the post-test data from the Perception+Production training group from Baese-Berk and Samuel (2016), with a flat discrimination function after training. The bottom-right panel shows the post-test data from the Perception-Only training group from Baese-Berk and Samuel (2016), with a robust discrimination peak after training. The top panel shows post-test data from the Short-Delay training group, with a small, but significant, discrimination peak in the middle of the continuum. The middle panel shows post-test data from the Long-Delay training group, with a robust discrimination peak after training. The error bars show the standard error of the mean.

The results of the regression analyses support these visual observations. The key statistic in these analyses is the interaction between Training Group and Continuum Pair, as this tests how "peaked" the functions are as a function of training conditions[11]. The interaction between production of tokens and continuum step (i.e., Perception-Only vs. the other training groups) was not significant ($\chi^2 = 5.064$, $p = .408$). This indicates that the peaks present in the Short-Delay and the Long-Delay conditions are sufficient to make the average function not significantly different in shape from the Perception-Only condition. Importantly, the other two interactions were significant. First, the interaction between inclusion of a delay and continuum pair was significant ($\chi^2 = 12.433$, $p = .029$). This indicates that for specific continuum pairs, the No-Delay training group differed from the two groups with a delay. As Figure 4 shows, the No-Delay group's function does not show a discrimination peak, whereas the two delay groups do. Second, the interaction between length of delay and continuum was also significant ($\chi^2 = 11.03$, $p = .048$). This interaction reflects the smaller discrimination peak for the Short-Delay training group than for the Long-Delay training group.

Turning our attention to the main effects in the model, the main effect of continuum step was significant ($\chi^2 = 15.963$, $p = .007$). Recall that for this comparison the center of the continuum (pair 3) is compared to each other pair on the continuum, demonstrating that the other

[11] It is possible, of course, that main effects across groups could also be informative. However, given that most of the points are expected to be (and are) at chance, the interaction is our primary test of interest.

points on the continuum are different from the "peak". As Figure 4 shows, this peak occurs for three of the four training groups, driving the main effect. The main effects of training group were not significant, as expected, since groups should not differ on the points other than the center point (i.e., all should be at chance). Specifically, (a) the effects of producing tokens (Perception-Only vs. the other three training groups that involved production), (b) the effect of delayed production (No-Delay Perception+Production vs. the two groups that involved delay) and (c) the length of that delay (Short- vs. Long-Delay) were not significant ($\chi^2 = 2.9965$, $p = .0835$, $\chi^2 = 0.1934$, $p = .6601$, and $\chi^2 = 2.5438$, $p = .1115$ respectively). These null effects reflect the near-chance within-category performance for all conditions; significant effects, when they are present, are on the between-category peak, as expected.

The results show that delaying production responses during the ABX task reduced the disruption of perceptual learning. The observed pattern is exactly as predicted on the assumption that speech perception and speech production present the learner with two tasks, and switching from the perception to the production task disrupts perceptual learning. If perceptual processing is given sufficient time to proceed before the task switch is initiated, perceptual learning succeeds; if the task switch is imposed too soon, perceptual learning fails. The extremely similar functions for the 4-second delay condition and the Perception-Only condition indicate that the processes supporting perceptual learning can complete their operations within this time period. The intermediate results with a 2-second delay suggest that on some trials the perceptual learning processes successfully finished, but on others they did not. Note that with the jittered delays employed here to reduce strategic responses, the available processing time on some trials in the Short-Delay condition were as short as 500 msec (versus a minimal delay of 2500 msec in the Long-Delay condition).

We will return to the implications of these results in the General Discussion. We turn our attention now to another factor that our prior work suggested could be implicated in the disruption of perceptual learning when production is required. By imposing a production requirement, experimenters are necessarily imposing a second task, and it is important to tease apart how much of the disruption of perceptual learning is just due to there being a second task per se (i.e., simply switching and/or mixing costs) versus how much of the disruption is specifically due to an incompatibility between developing a production code for a novel sound and developing a perception code for it.

Baese-Berk and Samuel (2016) demonstrated that production of an unrelated token reduced the disruption of perceptual learning compared to when the production was of the to-be-learned sound. In Experiment 2, we will examine whether another second task that is unrelated to speech production similarly disrupts perceptual learning. Critically, this experiment also tests whether a delay manipulation like that in Experiment 1 produces the same modulation of the disruption in perceptual learning that we saw when the second task was producing the to-be-learned sound.

Experiment 2 – Nonspeech Task Switch

Baese-Berk and Samuel (2016) tested a condition in which participants spoke the name of a printed letter that was presented on a computer screen when the X item of an ABX trial occurred. This condition thus was like the No-Delay Perception+Production condition in terms of saying a speech sound while completing the perceptual judgment, but differed in that the token being spoken was unrelated to the to-be-learned sound. On each trial participants heard the ABX triplet and before making their decision, they named a letter presented on the screen in front of them. Then, they made their response to the ABX task. This procedure tested whether

the disruption of perceptual learning is attributable to the activation of the speech production system in general, or is instead more a function of task switching per se. The letter-naming task produced significant disruption of perceptual learning, but the disruption was significantly smaller than what was caused by producing the to-be-learned sound itself. This result suggests that the disruption is not simply a matter of task switching, and that some of the effect is grounded in the similarity between the sound to be learned and the sound that is produced.

In Experiment 2, participants also are presented with letters to identify, but they do so by pushing buttons, rather than by saying them out loud. Comparing this condition to the conditions originally presented in Samuel and Baese-Berk (2016) provides a more direct test of whether the partial disruption was due to engaging the speech production system. If it was, the new button-pushing test, with no speech production requirement, should result in reduced interference (see MacLeod et al., 2010 for another example of manual responses disrupting or diminishing the effect of production on learning). This condition can be seen as a further titration of the production interference effect: Producing the to-be-learned speech sound is most disruptive, while producing a different speech sound is somewhat disruptive; the new condition matches the second one in terms of the items to identify, but they are responded to with a button push rather than by engaging the speech production system. Each condition is designed to peel away one aspect of the original effect at a time. Experiment 2 also tests whether delaying the button-pushing task reduces any negative impact it has on perceptual learning, as we found in Experiment 1 when the second task was to produce the to-be-learned sound.

Method

Participants

45 participants completed this experiment. 21 participated in the "Button Push" condition, and 24 in the "Delayed Button Push" condition (see below). These participants were compared to the participants in the Perception-Only (No-Delay; n=15), Perception+Production (No-Delay; n=15), and Spoken Letter (No-Delay; n=20) conditions previously reported in Baese-Berk & Samuel (2016). All participants were native speakers of Spanish, with very limited experience with Basque, English, and other non-native languages. No participant reported a history of speech, language, or hearing disorders. None of the participants in the two new training groups had participated in Experiment 1, or any other experiments in this broader research program. All participants were recruited from the same subject population and were recruited around the same time as the participants in Baese-Berk and Samuel (2016).

Stimuli

Stimuli for this experiment were identical to those used in the previous experiment.

Procedure

As in Experiment 1, participants completed a pre-test, a training period, and a post-test on each of two days, separated by no more than 48 hours. As in that experiment, the training was conducted with an ABX task. In the Button-Push group, a letter was presented on a screen (L, M, N, O, P, Q, or R) when the X item was played, and participants pressed a button that corresponded with that letter[12]; participants then made their ABX response (using a different pair of buttons) and were given feedback on this response. Participants in the Delayed Button-Push training group completed the same task. However, as in the Short-Delay training group in

[12] An open question, and one not examined in the present work, is how the content of this distractor task may affect learning. By using a task with orthography (or a task that required production of a letter in Experiment 3 from Baese-Berk & Samuel, 2016), participants may still be engaging in phonological processing, at least more than in a purely visual task. This was intentional – we wished to keep this aspect of the situation constant across the current experiments. Shifting to a purely visual task would potentially allow further titration of the source of the interference with perceptual learning, peeling away one additional aspect of the original effect.

Experiment 1, they delayed their response until they were prompted following a delay. The jittered delay times on a given trial were the same as in the Short-Delay condition of Experiment 1 (500, 1000, 1500, 2000, 2500, 3000, or 3500 msec).

The results for these two Button-Push conditions will be compared to three conditions from Baese-Berk and Samuel (2016): the Perception-Only training group, the (No-Delay) Perception+Production training group, and the Spoken-Letter training group (described above). The first condition provides a no-interference baseline, the second provides a maximal-interference baseline, and the third provides a speech-required comparison to the no-speech-required conditions tested in the current experiment. As in the case of Experiment 1, all tests were run with similar populations during a similar timeframe.

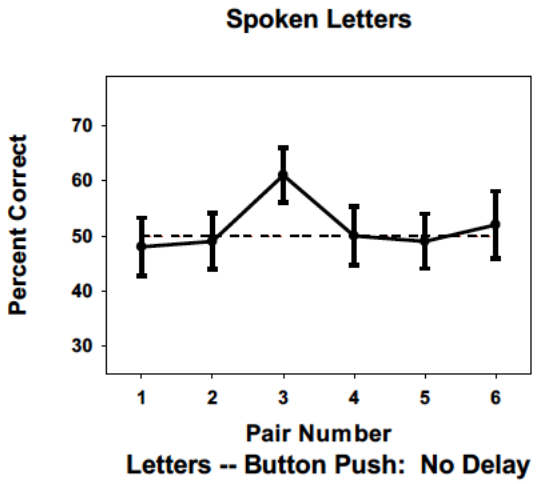
Analysis

Analyses were done as in Experiment 1 -- fixed effects included Training Group and Continuum Pair and the interaction of these effects. As in Experiment 1, the key statistic is the interaction of these two factors, as it again reflects the degree to which perceptual learning enhanced the peak in the ABX discrimination function. Training Group was Helmert coded to compare: (1) the effect of multiple tasks (i.e., the Perception-Only training group vs. the four other training groups which all involved a second task); (2) the effect of producing the to-be-learned token per se (i.e., the No-Delay Perception+Production training group vs. the three other "distractor task" training groups which all involved a task that was different from producing the target contrast); (3) the effect of saying a letter name vs. pressing a button (i.e., the Spoken-Letters training group vs. both the Button-Push and Delayed Button-Push training groups) and (4) the effect of delay (i.e., comparing the Button-Push group to the Delayed Button-Push group).

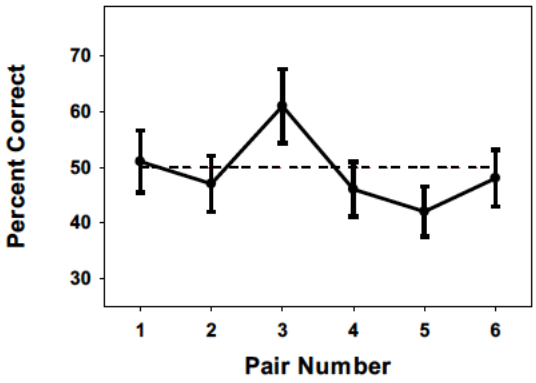
Continuum Pair was categorically coded and models were run using the center of the continuum (Pair 3) as the referent level, as before. Random effect structure was the maximum that would allow the models to converge, and included only random intercepts for participants. Significance of factors was determined by model comparisons using the anova function in R to compare the full model to models without each fixed effect, reported in the text below. Note that as in Experiment 1, all experimental procedures (e.g., location, participant sampling from the broader population, computer hardware, software, and peripherals, and stimuli) were identical across training groups, including those originally presented in Baese-Berk and Samuel (2016) – the only differences were those induced by the particular training conditions (i.e., delay between the X token and repetition). Before training, all participants were at chance (see Appendix 1 for figures of pre-test performance). Results of the full models can be found in Appendix 2.

Results and Discussion

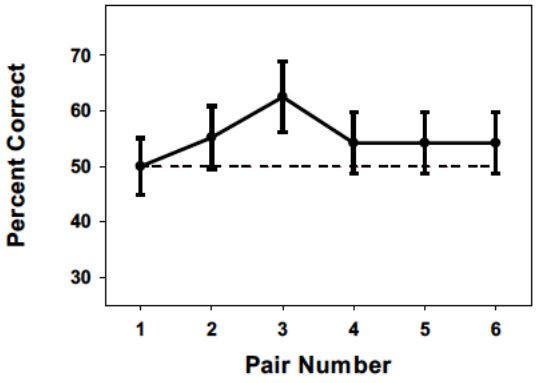
Before training, no groups differed from each other and all groups performed at chance (all $\chi^2 < 2$, $p > .1$; see Appendix 1). Figure 5 shows the average post-test performance for the three training groups that included reporting a letter (either by saying it, or by pushing the corresponding letter key). As before, small versions of panels from Figure 1 are provided for visual comparisons. Examining the figures, it appears that all groups except for the No-Delay Perception+Production group show a discrimination peak; however, the peak appears to be smaller for the three "distractor task" conditions than for the Perception-Only condition. In addition, the peak for all three "distractor task" groups is at the same intermediate level (approximately 61%), with all of the non-peak points hovering around chance.

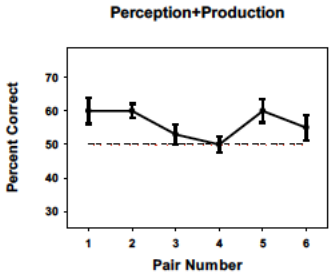


Letters -- Button Push: No Delay



Letters -- Button Push: Short Delay





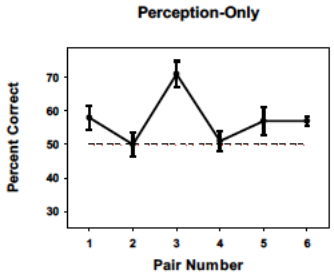


Figure 5: Bottom panels show data from the Perception+Production (No Delay) and Perception-Only conditions. The error bars show the standard error of the mean. Top panel shows the post-test data from the Spoken Letters training group from Baese-Berk and Samuel (2016) with a small, but significant discrimination peak after training. Second panel shows post-test data from the Button Push No-Delay training group with a small, but significant, discrimination peak in the middle of the continuum. Third panel shows post-test data from the Button Push Short-Delay training group and the same small discrimination peak after training as the other two groups.

The results of the mixed-effects regression support these observations. As in the previous experiment, the critical comparisons are the interactions between Training Group and Continuum Pair. The interaction between the comparison of single vs. multiple tasks (i.e., Perception-Only vs. all other groups) and Continuum Pair was significant ($\chi^2 = 10.357$, $p = .031$). Similarly, the interaction between the comparison of producing the to-be-learned token and doing a different "distractor task" and Continuum Pair was significant ($\chi^2 = 9.0456$, $p = .042$). Reflecting the very similar peaks for all three of the "distractor task" conditions, the interactions between Training Group and Continuum Pair were not significant (Spoken-Letter vs. Button-Push groups: $\chi^2 = 0.5887$, $p = .9966$; Button-Push vs. Delayed Button-Push: $\chi^2 = 7.3634$, $p = .289$).

As in Experiment 1, the main effect of Continuum Pair was significant ($\chi^2 = 12.836$, $p = .025$), reflecting the peak for Pair 3 for four of the five training groups examined here. Recall that Pair 3 serves as the reference level to which each of the other pairs is compared. The main effect of the comparison between a single task and two tasks during training was not significant (i.e., Perception-Only vs. all other training groups: $\chi^2 = 1.3659$, $p = .4978$). The main effect of the comparison between producing the target token per se and other distractor tasks was also not significant (i.e., Perception+Production vs. the three "distractor task" groups; $\chi^2 = .1704$, $p = .6798$). Further, the main effect of the comparison between the Spoken Letter group and the two Button Push groups was not significant ($\chi^2 = 0.0456$, $p = .831$). However, the main effect of the

comparison between the two Button push groups was significant, unexpectedly ($\chi^2 = 4.5269$, $p = .033$). Examining the figure, it appears that this is because performance away from the peak is higher for the Delayed Button-Push group. That is, participants perform slightly below chance in this condition for some of the within-category comparisons; we caution against over-interpretation of this particular finding as it is difficult to know the source of this below chance performance. Overall, the non-significant effects in almost all of the conditions again reflects the near-chance performance for within-category comparisons, as expected.

The presence of a (small) peak in the top three panels of Figure 5 shows that the three tasks based on letter names did not totally block learning of the new category, unlike the effect of producing the to-be-learned sound. While caution is always called for in interpreting a null effect, the non-effect of the delay manipulation for the button-push task reinforces the idea that these cases are different than those involving production of the to-be-learned sound. In the latter case, giving learners extra time before they had to execute the second task enhanced learning. Below, we discuss the implications of these results for our understanding of the source of the previously demonstrated disruption to perceptual learning and for our understanding of the relationship between perception and production, more broadly.

General Discussion

In this paper, we present an extension of our previous work that investigates the influence of producing tokens during training on perceptual learning. In the Introduction, we suggested that in studies examining the relationship between perception and production, experimenters implicitly create a dual-task paradigm for their participants, requiring them to learn how to perceive a new stimulus (Task 1) and how to produce it (Task 2). Our two experiments followed from this framing: We manipulated the timing of the two tasks as well as the nature of the second

task to examine how these factors influence the disruption of perceptual learning. In Experiment 1, we introduced a temporal separation between the two tasks, and found that the disruption of perceptual learning was partially alleviated by a brief delay and entirely alleviated after a longer delay between the two tasks. In Experiment 2, we investigate the second component – the nature of the second task. We demonstrated that when there was a switch between the perception task and an unrelated button pushing task, perceptual learning was better than when the second task was producing the to-be-learned sound, but still not as good as when no second task was imposed. Notably, delaying the unrelated task did not affect its disruptive effect, which suggests that there is a qualitative difference when producing the sound that is being learned perceptually.

The role of timing (and timescales) during learning

The results of this study highlight the importance of timing on multiple timescales when learning novel speech sounds. The role of timing on a fine-grained level has rarely been considered or manipulated in training studies in second language learning. Our results, along with the results of previous work, indicate that not only does timing matter, but one must consider timing on multiple time scales. There are multiple demonstrations that very long timescales (i.e., years) impact perceptual learning (e.g., Archila-Suerte et al., 2012; Baese-Berk & Samuel, 2016; Guion et al., 2000; Mackay et al., 2006). A growing body of work also demonstrates the importance of considering shorter timescales as well. Sleep, for example, can lead to overnight consolidation of perceptual learning (e.g., Dumay & Gaskell, 2007; Earle & Myers, 2014, 2015a, 2015b). That is, in addition to how much time has elapsed between training sessions and testing sessions, recent studies have considered *when* training and testing occur. Myers and colleagues have demonstrated that sleeping between training sessions can

significantly improve perceptual learning. Further, in some cases delayed feedback has been shown to result in more robust learning (Foerde & Shohamy, 2011).

In the present study, we shift to an even shorter timescale, looking at manipulations on the order of several seconds instead of days or years. The robust effect in Experiment 1, taken with previous findings, illustrates the importance of multiple timescales for learning novel speech sounds: Each timescale is nested within a larger timescale, and each can shape the relationship between perception and production. This idea of nested timescales and the dynamics of these timescales is not novel. In fact, a growing body of research in neuroscience examines the role of multiple timescales on cortical dynamics. For example, Honey et al. (2012) examined the temporal receptive windows of a number of brain regions and demonstrated that both fast and slow timescales are represented in the brain via fluctuations in activity. It is important to note that the crucial factor affecting perceptual learning in the current study may be allowing time to switch tasks, rather than the time elapsed per se. This suggests that in order to understand learning, we must consider multiple timescales, the interaction of timing with multiple tasks, and the impact of both of these factors on learning.

Working memory and timing

Delaying the second task (producing the X token; naming a printed letter; pushing a button to identify a printed letter) provides additional time to process the sounds in the ABX perception task. This potentially allows the learner to focus more on the first task before switching to the second, but it also increases memory load. While we found that the delay improved learning in the case of producing the target token, one could imagine a scenario in which the delay resulted in the opposite pattern – a further disruption of learning. A listener might be unable to hold the perceptual stimuli in memory while simultaneously preparing a

response. It is interesting that we find the opposite effect – participants demonstrate more learning after a jittered delay between perception and production.

One explanation for this increase in learning is that participants processed the perceptual stimulus more deeply because they were required to hold it in memory for an extended period. Depth of processing has been hypothesized to improve memory across a wide range of circumstances (Craik & Lockhart, 1972; Craik & Tulving, 1975); there is recent evidence that depth of processing may be sensitive to time pressure during response (Gross & Dobbins, under review). That is, by requiring listeners to hold their production and perception responses in memory for a longer time, they may process a stimulus more deeply than if they were able to respond to it immediately. It would be informative to test the effect of delay on a task other than production of the X item. For example, learners could be asked to make an additional perceptual judgment, such as whether that item matches a probe sound presented after varying delays. If delay per se increases the depth of processing, perceptual learning should benefit from delay, as it did in Experiment 1. If instead the effect of additional time is specifically related to maintaining the item in a form that can be used for production, then the effect of delay may be less beneficial when no such production is required.

Interestingly, the results in the present study are similar to a finding from the "Production effect" literature. In a study examining novel word learning, Mama and Icht (2018) asked participants to delay their response for 3 seconds during a recall task. Participants in this "delayed response" condition demonstrated superior performance to those in the immediate response condition. Mama and Icht explain this result with a "desirable difficulty" account (Bjork, 1994, Bjork, Little, & Storm, 2014). That is, the initial task is made more difficult because of the delay, as participants must hold the target information and/or their response in

memory for an extended period of time. The present work converges with that of Mama and Icht to suggest that with increasing difficulty of the initial task, via delay, learners demonstrate superior learning.

The relative timing of the perception and production tasks is also of interest. In the present task, participants must hold their perception decision in mind while producing the token. It is possible that the order of these tasks could be a driving factor in our results; this could be tested by switching the order of the perception decision and production. However, we believe this alone is unlikely to drive the results presented here. Baese-Berk (2019) found a similar disruption to perceptual learning with no perceptual task during training. In that set of experiments participants were asked to produce the token immediately and were not asked to make any perceptual judgment of the stimulus they heard during training. Further, Wright (2021) demonstrated a similar disruption to perceptual learning when a perceptual categorization task occurs before the production task during training (i.e., the opposite of the order presented here). Therefore, we believe that the timing of the perception and production tasks relative to one another is unlikely to be a strong driving force in the present study.

Task-switching

Experiments 1 and 2 produced an intriguing dissociation: Introducing a delay between perceiving the ABX triad and producing the X token was beneficial to learning, but a similar benefit was *not* shown for delays when the perception task was followed by an unrelated button-pushing task. Participants in the Delayed Button-Push condition of Experiment 2 demonstrated neither a disruption nor an improvement in perceptual learning as compared to participants who were asked to respond by pressing a button immediately after the letter's presentation; performance in both cases was equivalent to the condition in which participants said the letter

name aloud. While the differences between these two experiments must be interpreted with caution, as they include different participants, it is intriguing that a delay modulates performance in one case but not in the other.

The null effect of delay may initially seem counterintuitive. If depth of processing accounts for the effect of delay, shouldn't a delay affect perceptual learning regardless of what the second task is? To answer this question, it is important to note an important difference between the two experiments here: Experiment 1 provides listeners with sensory input in the form of their own production, and Experiment 2 does not. It is possible that this difference, regardless of the acoustic content of the production, results in different learning patterns across the two experiments. For example, immediate productions and short delays may still allow participants to access the sensory representation of the ABX tokens with some level of acoustic resolution that can be "masked" by their produced speech. However, in the longer delay condition, participants may be required to encode the ABX tokens (or at least a decision about the X token) in a more categorical way, which may be less impacted by the production of the target token. Experiment 2 does not introduce the production "masker", and thus delay may not affect learning in the same way.

Taken together, the pattern of results suggests that the disruption of perceptual learning when the training regime includes production has two separable components. The first component is related to the production of the target token per se. In Baese-Berk and Samuel (2016), participants demonstrated less robust perceptual learning when they were asked to produce the target token than an unrelated item (a letter name), suggesting that one component of the disruption is due to production of the target token itself. However, participants still demonstrated a decline in perceptual learning when they were asked to produce the unrelated

item compared to participants who were not asked to produce (or do some other task). The most plausible driver for this second component is task-switching.

If there really are two distinct components disrupting perceptual learning in the Perception+Production situation, it should be possible to show that they have different properties - other factors should influence them differently. In fact, the results of the delay manipulation provide such evidence. The component of the disruption that is attributable to production of the target token per se is sensitive to the delay (Experiment 1); the component of the disruption that is attributable to task switching is not (Experiment 2). In Experiment 2, it may be the case that distraction from the target task is the primary factor influencing the disruption, and additional manipulations do not further impact that distraction. Given that previous work (e.g., Mayr & Kliegl, 2000) has demonstrated that difficulty of retrieval affects switch costs, further dissociations might be obtained by examining how other factors (e.g., an increase in cognitive load due to task demands) impact learning across these two types of second tasks.

If there are two such components impacting learning in the Perception+Production situation, this could explain why producing tokens is worse for perceptual learning than pressing a button. Producing a token compounds both disruptive elements described here – task switching, as well as producing the token per se. Some disruption of perceptual learning occurs by forcing a switch away from the task of perceptual processing. In addition, producing a token may cause participants to not only stop perceptual processing but also alter that processing by engaging the newly-learned representation for another task, in this case, producing the novel sound. Combining both of these aspects into a single task (i.e., task-switching involving production of the to-be-learned sound) could compound the challenges for perceptual learning.

One unresolved question is whether the results of the present study are truly driven by task switching, or are instead driven by limiting the time allowed for the perceptual task. In the present work, these two factors co-vary (i.e., the time allotted to the perceptual task is driven by the requirement to switch to another task). These two possibilities could be teased apart by disrupting perceptual processing via other means that do not require a task switch (e.g., playing speech babble after the ABX trial is presented). It is possible that both aspects impact perceptual learning; the present findings cannot differentiate between these possibilities.

Implications for production learning

While most previous psycholinguistic studies examining non-native speech sound learning have focused on the perception modality, there are several studies that have directly examined production learning. It is clear that learners can improve their pronunciation of unfamiliar speech sounds and that training that focuses on production can result in large gains in production, even when listeners do not demonstrate gains in perception (e.g., Hattori & Iverson, 2008). However, performance in both perception and production, and the relationship between modalities, varies widely across learners (e.g., Bradlow et al., 1997).

In previous work, we have demonstrated that learners are able to improve in production after training in perception alone or after training in perception and production (Baese-Berk, 2019). Specifically, even learners in the Perception+Production training group who did not demonstrate perceptual learning demonstrated improvement in production from pre- to post-test. It should be noted that, in that study, production accuracy (i.e., how much of an acoustic difference participants produced between endpoint tokens they were learning) did not predict perceptual performance for participants in the Perception+Production training group. That is, one cannot explain the disruption solely by appealing to an account which states participants are

producing “bad” tokens or intermediate productions along some continuum. Rather, the act of producing was disruptive, regardless of the content of those productions.

Here, we do not directly investigate production data, focusing instead on perceptual learning. Of course, there are very important questions regarding production learning as well. It would be particularly interesting to investigate whether the content of a learner’s productions might influence their perceptual learning. Given the results discussed above, that there is more to the disruption than simply producing “bad” tokens, we think this is rather unlikely. That said, it is possible that learners whose productions do not exactly match the presented distributions may show a more diffuse peak, or an off-center peak in their discrimination functions. Further, individual differences in performance and/or learning strategies are likely to impact learning (e.g., Chandrasekaran et al., 2014; Perrachione et al., 2011), and perhaps even the interactions between perception and production. For example, one could imagine small-scale differences in the timing between perception and production responses that might impact learning.

Privileging one modality (e.g., producing tokens necessarily requires learners to focus on production) presumably also impacts learning, with learners performing better in that modality than the other at test. This is plausible, given that previous work has demonstrated that having learners focus attention on one dimension of a stimulus often results in them learning about that dimension to the detriment of other dimensions (Pederson & Guion-Anderson, 2010; Wright et al., 2010). However, it is also possible that the training that has been used has simply not been sufficient to demonstrate robust learning in both modalities. For example, while attention to one dimension of a stimulus (e.g., frequency of a tone) does not result in learning of another dimension of the stimulus (e.g., duration of a tone), shifting attention between two dimensions across blocks of training results in learning of both dimensions (e.g., Wright et al., 2010).

Further, shifting between active practice on a target task and passive exposure during an unrelated distractor task during learning results in as much learning on the distractor task as passive exposure alone (Wright et al., 2015). Therefore, it is possible that shifting may result in more robust learning, but learning that is divided across two modalities.

Given these previous results, it is likely that production learning may also depend on the details of the dual-task situation. For example, the Long-Delay condition in Experiment 1 might provide an ideal training scenario. Individuals in this condition demonstrate quite robust perceptual learning and, given their extensive production practice, they may well also enjoy substantial learning in production. If a delay allows robust learning in both perception *and* production, this has important implications for classroom teaching and our understanding of second language speech sound learning. Examining changes in production will be critically important for developing a complete understanding of how perception and production interact with one another.

Implications for our understanding of the relationship between perception and production

The results of Experiments 1 and 2 add to our understanding of the complex relationship between speech perception and production. The relationship between these two modalities has been of great interest to researchers for decades, in part because of the intuition that these two modalities should be very closely linked. A large body of evidence suggests that perception and production have a close, positive connection in many circumstances (Goldinger, 1998). However, there are also many studies that have not found any relationship between the two modalities or that have demonstrated an antagonistic relationship between the two (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007; Zamuner et al., 2018).

Our work (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007) has identified a clear negative relationship between perception and production. In the present study, we have taken the next step: The results of Experiments 1 and 2 explain some of the factors underlying production's disruption of perceptual learning, and begin to delineate the parameters that can explain several inconsistencies in previous results. As we noted in the Introduction, it is likely that the to-be-learned linguistic property impacts the relationship between the two modalities. That is, for something like syntactic patterns (e.g., Hopman & MacDonald, 2018), explicit production practice (especially including generation) may be beneficial, whereas for something less rule-based (like speech sounds), the task switching and task mixing costs of including production may outweigh any gains.

Further, it is important to note that the relationship between the two modalities may shift over time. For example, Kapnoula and Samuel (under review) demonstrate that during word learning, production results in improved learning early on, but becomes disruptive as exposure increases. Within the domain of second language speech sound learning, longitudinal studies have suggested that, when considering longer-term learning, the link between the two modalities may shift over time (Nagle, 2018; Nagle & Baese-Berk, 2021). By examining the link between perception and production in a dynamic way and incorporating both linguistic and non-linguistic factors in this examination, we will better be able to characterize the link between the two modalities.

One outstanding question raised by the present study is whether we might observe different results if we switched the order of the perception decision and the production task. In the present study and in Baese-Berk & Samuel, 2016, participants are asked to "hold" their perception response until after they have produced the token of interest. It is possible that the

specific challenges of this dual task (i.e., having to retain one response while preparing and executing a second response) could be alleviated if participants made the perception decision first and then produced the target immediately following execution of the decision. We believe, however, that the order of the two tasks during training is unlikely to explain the present set of results. In Baese-Berk (2019), participants do not complete any perceptual task and a disruption to perceptual learning is still observed. Further, in Wright (2021), the production task does occur after the perceptual task is complete, and the disruption to perceptual learning remains robust.

The present work also provides evidence for some accounts of the mechanisms underlying the perception-production link. We note that our work has demonstrated a negative relationship between the two modalities, which is different than claiming that there is no relationship. Our findings are consistent with an account that suggests that the two modalities, at a minimum, share processing resources (see, e.g., Baese-Berk, 2019). Our results, and those of other studies, help to flesh out this fundamental statement. If one assumes that the goal of speech production is to achieve a specific acoustic target, given a variety of biomechanical constraints (e.g., Tourville & Guenther, 2011), the act of speech production would involve selecting an acoustic target and then creating an articulatory plan to achieve such a target. The negative perception-production relationship that we and others have found could arise if an early learner of an L2 selects an acoustic target for production that competes with the acoustic information presented during perception; this competition could disrupt perceptual learning.

This scenario can be formalized in an exemplar model of phonology (e.g., Pierrehumbert, 2001) in which a learner selects an exemplar to serve as a target for production, but this target may not be closely connected to the acoustics of the sound that was heard. The production target would compete with the acoustic representation of the sounds that have been heard. One might

expect that, in successful learning, production representations would fade as perceptual representations strengthen, in part because the speaker would become better at choosing the "correct" exemplar for production. However, this explanation predicts that producing "bad" tokens would be more disruptive to perceptual learning than producing "good" tokens, and previous work (Baese-Berk, 2019) has suggested that accuracy of production is, in fact, not strongly correlated with perceptual learning for individuals trained in both perception and production. Instead, variability in the speaker's productions appears to be the more critical factor. An exemplar-based explanation that relies on acoustic targets of speech production might still be able to account for this pattern of results, given that more dispersed acoustic representations may result in a less clear delineation between two novel categories, but this remains to be shown.

The current findings enrich our understanding of the relationship between speech perception and speech production, and our understanding of the relationship between perception and action more broadly. Taken together with our previous work, the results implicate a major role for timing and experience in modulating this relationship. The more time given, on both short and long timescales, the smaller the disruption to perceptual learning. Further, the nature of the specific tasks used during training clearly will affect this relationship. As noted above, the collective pattern of data indicates that the disruption has both a general component (i.e., disruption due to distraction and task switching) and a specific one (i.e., disruption due to production of the to-be-learned sound). It is likely, in fact, that the antagonism of production and perception has multiple component parts. Additional manipulations (e.g., different second tasks, various types of load on the system, etc.) can elucidate our understanding of how perception and production interact with one another and interact with other cognitive processes, both during learning and during more established cognition.

Conclusion

Taken together, the results of our research program suggest that requiring production of novel tokens when learning novel speech sounds is best characterized as a dual-task situation: learners are obliged to focus on two separate tasks on each trial. The dual-task nature of such training is a likely driving factor in the disruption of perceptual learning when accompanied by production. The current study has identified two aspects of this disruption: (a) the temporal organization of the two tasks and (b) the content of the second task. We take these results as evidence for the complex relationship between perception and production, influenced by a number of factors, including timing of the two tasks, depth of processing, and attention. The results begin to delineate those factors that impact the relationship between perception and production. This delineation can provide the basis for reconciling inconsistent results across previous studies, as well as having potentially important implications for best practices in classroom-based teaching of second languages.

Acknowledgements

The authors would like to thank Larraitz Lopez and Marisa Abril for their assistance with data collection.

Funding

This work was supported by the National Science Foundation Grants BCS-1734166, BCS-1941739, by Economic and Social Research Council (UK) Grant #ES/R006288/1, Ministerio de Ciencia E Inovacion (Spain) Grant # PSI2017-82563-P, by Ayuda Centro de Excelencia Severo Ochoa (Spain) SEV-2015-0490, and by Grant PIBA18-29 from the Basque Government.

Open Sceince Practices

None of the experiments reported here were preregistered. Data and materials for experiments will be available following peer review.

References

Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism: Language and Cognition*, *15*(1), 190–201. https://doi.org/10.1017/S1366728911000125

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception & Psychophysics. 81*(4), 981-1005.

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. https://doi.org/10.1016/j.jml.2015.10.008

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*.

Bent, T. (2005). Perception and Production of Non-Native Prosodic Categories. *Linguistics.Northwestern.Edu*. Retrieved from http://www.linguistics.northwestern.edu/people/recent_grads/dissertations/bentDissertation.pdf

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171–204). Timonium, MD: York Press.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A.Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, *3*(3), 165–170

Broersma, M., Carter, D., & Acheson, D. J. (2016). Cognate Costs in Bilingual Speech Production: Evidence from Language Switching. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01461

Chandrasekaran, B., Koslov, S. R., & Maddox, W. T. (2014). Toward a dual-learning systems model of speech category learning. Frontiers in Psychology, 5(825), 1–17. https://doi.org/10.3389/fpsyg.2014.00825

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*(2), 255-267.

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28–43. https://doi.org/10.1016/j.jml.2018.07.005

de Jong, K., Hao, Y.-C., & Park, H. (2009). Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics*, *37*(4), 357–373. https://doi.org/10.1016/j.wocn.2009.06.001

DeKeyser, R. M., & Sokalski, K. J. (1996). *The differential role of comprehension and production practice*. *46*(4), 613–642. https://doi.org/10.1111/j.1467-1770.1996.tb01354.x/abstract

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Represenationts of Spoken Words. *Psychological Science*, *18*(1), 35–39.

Earle, F. S., & Myers, E. B. (2014). Building phonetic categories: an argument for the role of sleep. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01192

Earle, F. S., & Myers, E. B. (2015a). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *Journal of the Acoustical Society of America*, *137*(1), EL91–EL97. https://doi.org/10.1121/1.4903918

Earle, F. S., & Myers, E. B. (2015b). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1680–1695. https://doi.org/10.1037/xhp0000113

Fink, A., & Goldrick, M. (2015). Pervasive benefits of preparation in language switching. *Psychonomic Bulletin & Review*, *22*(3), 808–814. https://doi.org/10.3758/s13423-014-0739-6

Flege, J. E. (1993). Production and perception of a novel second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*(3), 1589–1608.

Foerde, K., & Shohamy, D. (2011). Feedback timing modulates brain systems for learning in humans. *Journal of Neuroscience*, *31*(37), 13157-13167.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*. Retrieved from http://files.eric.ed.gov/fulltext/ED274022.pdf#page=144

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279.

Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost–benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 640.

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498.

Guion, S., Flege, J. E., Liu, H.-M., & Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Journal of the Acoustical Society of America*, *21*(2), 205–228. https://doi.org/10.1121/1.1970569

Hattori, K., & Iverson, P. (2008). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*, *125*(1), 469–479. https://doi.org/10.1121/1.3021295

Hearnshaw, S., Baker, E., & Munro, N. (2019). Speech perception skills of children with speech sound disorders: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3771-3789.

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., … Hasson, U. (2012). Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. *Neuron*, *76*(2), 423–434. https://doi.org/10.1016/j.neuron.2012.08.011

Hopman, E. W. M., & MacDonald, M. C. (2018). Production Practice During Language Learning Improves Comprehension. *Psychological Science*, *29*(6), 961–971. https://doi.org/10.1177/0956797618754486

Huang, B. H., & Jun, S.-A. (2011). The Effect of Age on the Acquisition of Second Language Prosody. *Language and Speech*, *54*(3), 387–414. https://doi.org/10.1177/0023830911402599

Icht, M., & Mama, Y. (2015). The production effect in memory: a prominent mnemonic in children. *Journal of Child Language*, *42*(5), 1102–1124. https://doi.org/10.1017/S0305000914000713

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *14, 89*, 81–81.

Kapnoula, E.C., & Samuel, A.G. (under review). Reconciling the contradictory effects of production on word learning: Production may help at first, but it hurts later.

Kato, M., & Baese-Berk, M. M. (2020). The effect of input prompts on the relationship between perception and production of non-native sounds. *Journal of Phonetics*, *79*, 100964.

Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*(1), 121–148. https://doi.org/10.1080/01690965.2010.486579

Kimberg, D. Y., Aguirre, G. K., & D'Esposito, M. (2000). Modulation of task-related neural activity in task-switching: an fMRI study11Published on the World Wide Web on 5 May 2000. *Cognitive Brain Research*, *10*(1), 189–196. https://doi.org/10.1016/S0926-6410(00)00016-1

Kirk, N. W., Kempe, V., Scott-Brown, K. C., Philipp, A., & Declerck, M. (2018). Can monolinguals be like bilinguals? Evidence from dialect switching. *Cognition*, *170*, 164–178. https://doi.org/10.1016/j.cognition.2017.10.001

Koch, I., Prinz, W., & Allport, A. (2005). Involuntary retrieval in alphabet-arithmetic tasks: Task-mixing and task-switching costs. *Psychological Research*, *69*(4), 252–261. https://doi.org/10.1007/s00426-004-0180-y

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141–178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, *13*(2), 262–268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15. https://doi.org/10.1016/j.jml.2006.07.010

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *Journal of Cognitive Neuroscience*, *25*(12), 2179-2188.

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

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, *65*(4), 497–516.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, *243*(4890), 489–494.

Mackay, I. R. A., Flege, J. E., & Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*. Retrieved from http://journals.cambridge.org/abstract_S0142716406060231

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: General*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Mama, Y., & Icht, M. (2018). Production on hold: delaying vocal production enhances the production effect in free recall. *Memory*, *26*(5), 589-602.

Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1124–1140. https://doi.org/10.1037/0278-7393.26.5.1124

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33-B42.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, *31*(3), 327–342. https://doi.org/10.3758/BF03194391

Nagle, C. L. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, *68*(1), 234-270.

Nagle, C. & Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*.

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *Journal of the Acoustical Society of America*, *127*(2), EL54–EL59. https://doi.org/10.1121/1.3292286

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. Journal of the Acoustical Society of America, 130(1), 461–472. https://doi.org/10.1121/1.359336

Pierrehumbert, Janet (2001). Exemplar dynamics, word frequency, lenition, and contrast. In *Frequency Effects and the Emergence of Linguistic Structure*, Joan Bybee and Paul Hopper (eds.), 137–157. Amsterdam: John Benjamins.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328–333. https://doi.org/10.1121/1.1914506

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org

Rochet, B., & Strange, W. (1995). *Perception and production of second-language speech sounds by adults*.

Scott, S. K. (2005). Auditory processing—speech, space and auditory objects. *Current Opinion in Neurobiology*, *15*(2), 197-201.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100-107.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(03), 243–261. https://doi.org/10.1017/S0142716400001417

Sohn, M.-H., Ursu, S., Anderson, J. R., Stenger, V. A., & Carter, C. S. (2000). The role of prefrontal cortex and posterior parietal cortex in task switching. *Proceedings of the*

National Academy of Sciences, *97*(24), 13448–13453. https://doi.org/10.1073/pnas.240460497

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952-981.

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

Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition*, *112*(1), 81-96.

Warker, J. A., Dell, G. S., Whalen, C. A., & Gereg, S. (2008). Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1289–1295. https://doi.org/10.1037/a0013033

Wright, B. A., Baese-Berk, M. M., Marrone, N., & Bradlow, A. R. (2015). Enhancing speech learning by combining task practice with periods of stimulus exposure without practice. *Journal of the Acoustical Society of America*, *138*(2), 928–937. https://doi.org/10.1121/1.4927411

Wright, B. A., Sabin, A. T., Zhang, Y., Marrone, N., & Fitzgerald, M. B. (2010). Enhancing perceptual learning by combining practice with periods of additional sensory stimulation. *The Journal of Neuroscience*, *30*(38), 12868–12877.

Wright, J. M. (2021). Factors affecting the incidental formation of novel suprasegmental categories. [Unpublished doctoral dissertation] University of Oregon.

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. A. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, *89*(C), 55–67. https://doi.org/10.1016/j.jml.2015.10.003

Zamuner, T. S., Strahm, S., Morin-Lessard, E., & Page, M. P. A. (2018). Reverse production effect: children recognize novel words better when they are heard rather than produced. *Developmental Science*, *21*(4), e12636. https://doi.org/10.1111/desc.12636

Zamuner, T. S., Yeung, H. H., & Ducos, M. (2017). The many facets of speech production and its complex effects on phonological processing. *British Journal of Psychology*, *108*(1), 37–39. https://doi.org/10.1111/bjop.12220

Appendix 1: ABX discrimination before training

Figure A1: Experiment 1

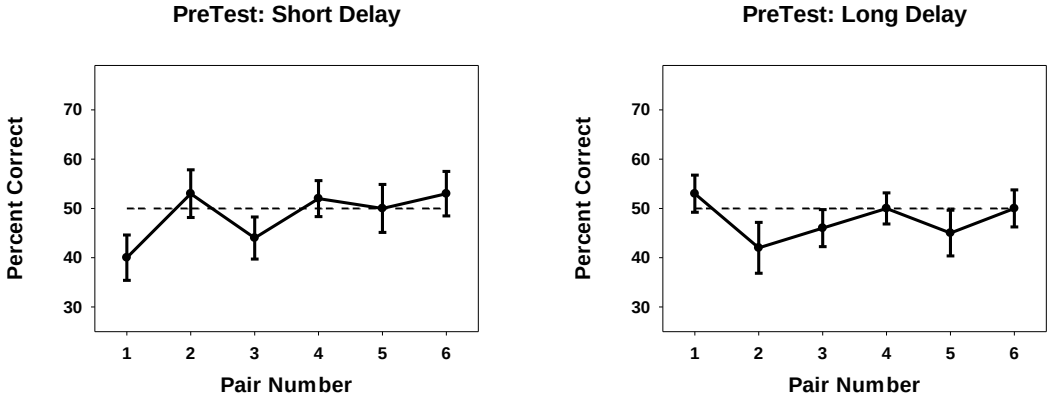
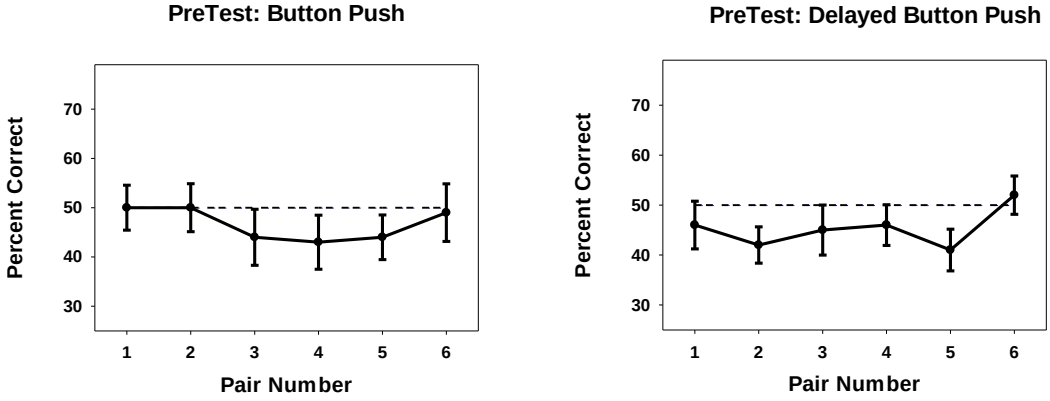


Figure A2: Experiment 2



Before training, performance was at chance for all groups of native Spanish speakers, as expected.

Appendix 2:

Table A1: Results of full mixed effects model for Experiment 1

lmer(Discrim ~ step.contrasts * percOnlycomp + step.contrasts * delay + step.contrasts * length +(1|Participant), data=delayprod)

Effect	Estimate	Std. Error	t-value
Intercept	0.532718	0.012881	41.358
Production (Y/N)	-0.041000	0.023368	-1.755
Delay (Y/N)	0.019145	0.043027	0.445
Length (Short/Long)	0.054615	0.033868	1.613
Step 1 vs. 3	0.018395	0.057604	0.319
Step 2 vs. 3	0.081341	0. 057604	1.412
Step 3 vs. 4	0.029577	0. 057604	0.513
Step 3 vs. 5	0.047314	0. 057604	0.821
Step 3 vs. 6	-0.049274	0. 057604	-0.855
Step 1 vs. 3 * Production	0.024806	0.104505	0.237
Step 2 vs. 3 * Production	-0.027633	0.104505	-0.264
Step 3 vs. 4 * Production	0.035987	0.104505	0.344
Step 3 vs. 5 * Production	-0.100122	0.104505	-0.958
Step 3 vs. 6 * Production	0.187906	0.104505	1.798
Step 1 vs. 3 * Delay	0.168547	0.192424	0.876
Step 2 vs. 3 * Delay	0.029060	0. 192424	0.151
Step 3 vs. 4 * Delay	0.275214	0. 192424	1.430

Step 3 vs. 5 * Delay	0.220855	0. 192424	1.148
Step 3 vs. 6 * Delay	-0.089915	0. 192424	-0.467
Step 1 vs. 3 * Length	-0.009231	0.151462	-0.061
Step 2 vs. 3 * Length	-0.107692	0. 151462	-0.711
Step 3 vs. 4 * Length	-0.246154	0. 151462	-1.625
Step 3 vs. 5 * Length	-0.087692	0. 151462	-0.579
Step 3 vs. 6 * Length	-0.006154	0. 151462	-0.041

Table A2: Results of full mixed effects model for Experiment 2

distract.lmer<-lmer(Discrim ~ step.contrasts * multtask + step.contrasts * prodtoken +

step.contrasts * push + step.contrasts * delay +(1|Participant), data=distract)

Effect	Estimate	Std. Error	t-value
Intercept	0.5302857	0.0111060	47.747
Multitask (Y/N)	-0.0434323	0.0246903	-1.759
Produce Token (Y/N)	-0.0094186	0.0234594	-0.401
Push Button (Y/N)	0.0076389	0.0358049	0.208
Delay (Y/N)	-0.0642361	0.0.0306708	-2.094
Step 1 vs. 3	0.0197844	0.0459606	0.430
Step 2 vs. 3	0.0629004	0. 0459606	1.369
Step 3 vs. 4	-0.0055133	0. 0459606	-0.120
Step 3 vs. 5	0.0361164	0. 0459606	0.786
Step 3 vs. 6	-0.1419074	0. 0459606	-3.088

Step 1 vs. 3 * Multitask	0.0261946	0.1021770	0.256
Step 2 vs. 3 * Multitask	-0.0460740	0.1021770	-0.451
Step 3 vs. 4 * Multitask	0.0008969	0.1021770	0.009
Step 3 vs. 5 * Multitask	-0.1113195	0.1021770	-1.089
Step 3 vs. 6 * Multitask	0.0952721	0.1021770	0.932
Step 1 vs. 3 * Produce	0.0455648	0.0970829	0.469
Step 2 vs. 3 * Produce	-0.0447720	0.0970829	-0.461
Step 3 vs. 4 * Produce	0.0524032	0.0970829	0.540
Step 3 vs. 5 * Produce	0.0659788	0.0970829	0.680
Step 3 vs. 6 * Produce	-0.2142433	0.0970829	-2.207
Step 1 vs. 3 * Push	-0.0672619	0.1523114	-0.442
Step 2 vs. 3 * Push	0.0232143	0. 1523114	0.152
Step 3 vs. 4 * Push	0.0740079	0. 1523114	0.486
Step 3 vs. 5 * Push	-0.0398810	0. 1523114	-0.262
Step 3 vs. 6 * Push	-0.0275794	0. 1523114	-0.181
Step 1 vs. 3 * Delay	-0.1522827	0.1269262	-1.200
Step 2 vs. 3 * Delay	0.0500992	0. 1269262	0.395
Step 3 vs. 4 * Delay	0.0262897	0. 1269262	0.207
Step 3 vs. 5 * Delay	0.0471230	0. 1269262	0.371
Step 3 vs. 6 * Delay	-0.0927579	0. 1269262	0.731

Table A3: Means and Standard Deviations (in parentheses) for new participant groups in Experiment 1 (across 6 continuum pairs)

Group	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Short-Delay	.51 (.29)	.53 (.29)	.63 (.22)	.55 (.29)	.54 (.26)	.56 (.20)
Long-Delay	.5 (.19)	.52 (.17)	.69 (.17)	.55 (.23)	.49 (.19)	.52 (.29)

Table A4: Means and Standard Deviations (in parentheses) for new participant groups in Experiment 2 (across 6 continuum pairs)

Group	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Button-Push	.51 (.20)	.46 (.19)	.60 (.22)	.45 (.18)	.41 (.18)	.46 (.18)
Delayed-Button Push	.5 (.19)	.55 (.20)	.63 (.21)	.54 (.23)	.54 (.22)	.54 (.24)