Listeners Beware: Speech Production may be Bad for Learning Speech Sounds

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

Spoken language requires individuals to both perceive and produce speech. Because both processes access lexical and sublexical representations, it is commonly assumed that perception and production involve cooperative processes. However, few studies have directly examined the nature of the relationship between the two modalities, particularly how producing speech influences speech perception. In a series of experiments, we examine the counter-intuitive finding that learning perceptual representations can be disrupted by producing tokens during training. We investigate whether this disruption can be alleviated by prior experience with the speech sounds, and whether the cause of the disruption is production of the particular sound being learned, or is a more general conflict between the production system and the system that develops new perceptual representations. Our results paint a more competitive relationship between perception and production than might be assumed and suggest that both demands inherent to production and cognitive demands modulate this relationship.

Introduction

Spoken language is a communication system that involves the interaction of production and perception: Each person produces a series of words, and perceives those produced by another person. Because the intention is to transfer a message from one person to another, and the medium is a series of words that are each made up of a series of sublexical sounds, it is natural to assume that perception and production are cooperative processes using common elements. This assumption may be natural, but it may also be wrong. In the current study, we examine whether speech perception and speech production are in fact cooperative processes when a listener is learning a new phonemic contrast.

The existing literature indicates that the relationship between speech perception and speech production is more complex than one might assume, particularly during learning. Studies have demonstrated that although perception and production are usually correlated during novel speech sound learning, individual performance differs greatly (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Wang, Jongman, & Sereno, 2003). In learning novel speech sounds, perception frequently improves without production learning, and vice versa (Bradlow et al., 1997; de Jong, Hao, & Park, 2009; Flege, 1993; Sheldon & Strange, 1982). In the present paper, we report the results of a series of studies examining how producing speech sounds while learning those sounds can influence how well the sounds are learned. A rather counter-intuitive finding has emerged: When training on perception alone, the improvement in perception of the

sounds is rather robust. However, when training includes both a perception and a production component, the perceptual improvement is disrupted.

Laboratory training of speech sounds

One of the classic hallmarks of speech perception in a listener's native language is categorical perception. In many early studies of speech perception (see Libermann, Cooper, Shankweiler, & Studdert-Kennedy, 1967 for examples) researchers created sets of syllables in which an acoustic parameter was varied in such a way that the syllable at one end of the continuum was heard in one way (e.g., /ba/), and the syllable at the other end in a different way (e.g., /pa/). For many speech continua, perception seemed categorical: Listeners heard a few items as one category (e.g., /ba/) and then things abruptly changed, with the remaining items heard as the other category (e.g., /pa/). Moreover, the ability to discriminate different tokens on the continuum depended on where the tokens were: Two tokens from within the same category were discriminated at near-chance levels, whereas tokens that crossed the category boundary were well discriminated. The categorical tendency in perception was strongest for stop consonants, somewhat weaker for other consonants (e.g., fricatives), and weaker still for vowels (see Repp, 1984 for an excellent review of the categorical perception literature)[1].

When listening to speech sounds from a non-native language, however, the picture can be quite different. It is not uncommon for listeners to be unable to correctly divide tokens from the continuum into two categories. These listeners are typically

[1] Recent work on categorical perception (e.g., Gerrits & Schouten, 2004; Schouten, Gerrits, & van Hessen, 2003) has suggested that it is more complex than some of the classic work portrayed it. Specifically, categorical perception can be triggered by certain tasks. In the current study, we are examining (and reporting) "classic" categorical perception data using tasks that have been shown to trigger such perception.

unable to discriminate among tokens at any point on the continuum, even if these tokens do cross a category boundary in the non-native language. The best known example of such non-native perception comes from Japanese learners of English, who have a notoriously difficult time categorizing and discriminating between tokens on an /r/-/l/ continuum (e.g., Goto, 1971; Strange & Dittman, 1984), tasks which are not difficult for English listeners. It is hypothesized that this is because Japanese does not have two distinct categories for /r/ and /l/. Flege (1995) and Best (1995) have argued that the perception of non-native phonemes is reliant on the category structure of the listeners' native language. That is, when learning their native language, a listener not only learns what variability is important, but also what variability is not. Therefore, listeners must not only learn the categories of their language, but also learn to not use irrelevant information in categorizing speech stimuli. The task of the non-native learner, then, is to learn to attend to variability which is not important for categorization in their native language.

Many previous studies have demonstrated that even within a listener's native language category boundaries can be shifted with experience (Kraljic & Samuel, 2005; 2006; 2007; Maye, Aslin, & Tanenhaus, 2008; Norris, McQueen, & Cutler, 2003), suggesting substantial flexibility in the perceptual system and its representations. Within a non-native language, dozens of previous studies have demonstrated that individuals are capable of learning novel phonological categories in the laboratory (Strange & Dittman, 1984). A number of language backgrounds and many different segmental and suprasegmental contrasts have been examined including English listeners' perception of a three-way voicing continuum (McClaskey, Pisoni, & Carrell, 1983; Tremblay, Kraus, & McGee, 1998), Spanish and German listeners' perception of English vowels (Iverson &

Evans, 2009), English listeners' perception of German vowels (Kingston, 2003), French listeners' perception of English interdental fricatives (Jamieson & Morosan, 1986; 1989) Japanese listeners' perception of English /r/ and /l/ (Logan, Lively, & Pisoni, 1991), and English listeners' perception of Mandarin tone (Wang, Spence, Jongman, & Sereno, 1999). In many cases, listeners are able to learn to perceive contrasts that do not exist in their native language with a relatively small amount of laboratory-based training. Taken with the results from perceptual learning within a native language, this suggests that listeners' perception is relatively flexible and can be changed through experience.

The relationship between perception and production during learning

Most models of speech perception and production suggest a close tie between the two modalities. The two modalities are frequently discussed as being two ends of a single process, as in Denes and Pinson's speech chain (Denes & Pinson, 1963). Researchers have frequently cited perception-oriented changes in production such as the Lombard effect (Lane & Tranel, 1971; Lombard, 1911) as evidence for a necessarily tight connection between the two modalities. Other such effects include shadowing of various phonetic properties such as voice onset time (Goldinger, 1998), shifts in vowel production as a result of shifted perceptual input (Houde & Jordan, 1998; 2002), and phonetic accommodation to a conversation partner's speech (Kim, Horton, & Bradlow, 2011; Pardo, 2006). These findings have been interpreted as demonstrating that perception and production must be tightly coupled, or as Casserly and Pisoni (2010) state "… the two processes must at some point even deal in the same linguistic currency." These results all conform to the basic fact, stated at the beginning of this article, that

production and perception must be compatible enough to allow communication to occur: Speakers produce in order that perceivers will understand a message.

Most prior work showing this naturally cooperative relationship has looked at perception and production within a relatively steady-state native language system. How the two modalities relate during learning of non-native contrasts is less clear. Most of the previous studies examining learning of novel phonemes have focused on perceptual learning. However, in order to successfully learn a language, one must master both perception and production of difficult contrasts. A relatively small number of studies have explicitly examined the relationship of perception and production during the learning of novel sound contrasts, and the results of these studies have been inconsistent. For example Bent (2005) did not find a correlation between perception and production of Mandarin tones by naïve, native English speakers whereas Wang et al. (2003) demonstrated a correlation between the two skills after a brief period of training. Several studies have demonstrated that production skills often precede perceptual learning of the novel contrasts (Flege, 1993). However, many other studies have failed to show a correlation between perception and production skills when learning a novel contrast (de Jong et al., 2009; DeKeyser & Sokalski, 1996). Those studies that have shown significant correlations between the two skills after some training have demonstrated that there is substantial individual variability both in each of the skills independently and their relationship to one another within a learner (e.g., Bradlow et al., 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Wang et al., 2003).

Given that the examinations of perception and production during learning of novel phonological contrasts have often produced contradictory results, it is clear that

there is no simple relationship between perception and production during learning. In the following section, we explore one line of evidence that suggests that we must re-consider the commonly held notion that perception and production are naturally cooperative processes that utilize many common elements. Specifically, we examine two previous studies that have suggested that producing tokens during training can actually disrupt perceptual learning rather than enhance it via cooperation. The three experiments of the current study continue this approach to examining the relationship of perception and production.

Disruption in perceptual learning

In the first study showing disruption of perception due to production, Leach & Samuel (2007) examined how adults learn new words. Native English-speaking college students were taught new English "words" such as "bibershack" and "nomemsoly". Each student was taught a dozen new words of this sort in a set of five training sessions held over the course of five days. In total, each new word was presented 120 times. In the critical training regime, each time a word was played, two pictures of odd objects were shown on a screen, and the subject had to indicate which of the two pictures was associated with that word. For example, the word "bibershack" was presented as the name of one unfamiliar object, while "nomemsoly" was associated with a different odd object. The subjects quickly learned what picture was associated with each new word.

The focus was on how *functional* the new words had become over the course of training, using two tasks. One task measured how well listeners could report words when they were presented under very noisy conditions: If an item has developed a functional lexical representation, listeners will be better able to hear the word in noise than if there

is no such functional representation. The second measure was how well the words could produce "perceptual recalibration" of speech sounds (Kraljic & Samuel, 2005; 2006; 2011; Norris et al., 2003). Perceptual recalibration is necessary for optimal comprehension because speakers vary in how they produce speech sounds – one person's version of a particular vowel or consonant can be rather different than another person's. Listeners must adjust to the idiosyncracies of each speaker, and recent evidence shows that native speakers use their detailed knowledge of sounds and words to do so. This was first shown by Norris et al. (2003), who played listeners words in which a particular sound was intentionally produced in an idiosyncratic way. For example, whenever a word should have had the sound of "s", it actually had a sound that was ambiguous, made from a mixture of "s" and "f". Other listeners heard the same ambiguous mixtures, but in words that should have had the sound of "f" in that position. Listeners recalibrated their perceptual categories for the speech sounds, using the lexical context to learn that a given ambiguous sound was either "s" or "f". For example, if they heard the ambiguous sound in words like "sheri*" (with the ambiguous sound represented here with "*"), they learned that such ambiguous sounds are variants of "f", and they recalibrated their phonetic categorization process accordingly (this was assessed in a post-test that measured how listeners heard sounds like "s" and "f").

Leach and Samuel (2007) used the ability of the newly-learned words to drive perceptual recalibration as an index of how functional the words were: If well-established words can direct the recalibration, and if items like "nomemsoly" have become fully established, then an ambiguous pronunciation of "s" sounds in such newly learned words should also lead to shifts in listeners' phonetic categorization. In fact, by the second day

of training, presentation of such words did produce significant perceptual recalibration. The line plotted with open circles in Figure 1 shows the size of the shift for the first two days of the five-day experiment, and for the last two days. The newly-learned words in this Perception Only condition functioned as real words do, producing significant shifts even in the first two sessions, and very robust recalibration by the last two sessions.

The critical result for the purpose of the present study comes from a second group of subjects who had exactly the same training as just described, with one critical change: In this condition, after the subjects selected the correct picture for a word heard during training, they had to repeat the word they had just heard. When Leach and Samuel created this condition, they expected that the production requirement would enhance learning of the words, an expectation grounded in the common intuition that perception and production are cooperative processes. This intuition was wrong: Subjects in the Perception + Production condition showed much worse perceptual recalibration. Rather than helping to develop fully functional perceptual representations of words that were being learned, the production requirement dramatically undermined such learning. The weak recalibration during the first two sessions was not statistically different from zero, and by the end of training the effect was in fact essentially zero (see Figure 1).

(Insert Figure 1 here)

In the Leach and Samuel (2007) study, the negative impact of production on the perceptual functionality of newly learned speech was observed at the lexical level, as the words being learned did not acquire all of the useful perceptual properties of a lexical representation. A study conducted by Baese-Berk (2010) demonstrates that this problem

is not limited to the lexical level, but instead seems to be quite general. In this project, listeners were taught to distinguish speech sounds that were not differentiated in their native language, as reviewed above.

A "voiced" – "prevoiced" distinction, which occurs in many languages including Hindi, was taught to native English speakers across two sessions held on consecutive days. A set of stimuli forming a continuum was constructed; the syllables at one end of the continuum were prevoiced, where vocal fold vibration precedes the release of the stop closure, while those at the other end were voiced (as in English), with a relatively synchronized onset of voicing and release of stop closure. Recall that for the native speakers of American English who served as subjects, all of these sounds are heard as members of the same phonetic category. Thus, at the beginning of training, listeners could not distinguish among the members of the continuum.

One group of subjects received Perception Only training. They heard the syllables many times, and the presentation was structured in order to help them organize the syllables into the two different categories. Following Maye & Gerken (2000; 2001) participants were exposed to tokens along the prevoiced-voiced continuum. The stimuli were presented in one of two different distributional patterns. For subjects receiving a unimodal distribution, tokens near the center of the continuum were presented many times and tokens at the ends were presented relatively few times, creating a single category for the listener. For subjects receiving a bimodal distribution, tokens near the ends of the continuum were presented many times and tokens near the center were presented relatively few times, creating two distinct categories for the listener. Another two groups of listeners received the same training, but as in the Leach and Samuel (2007)

study, these groups were instructed to say each syllable aloud before moving on to the next one (thus, Perception + Production training).

Again, in designing this condition, the expectation was that production would help the subjects to learn the distinction, assuming a cooperative relationship between perception and production. After training, both groups were tested on their ability to distinguish pairs of sounds. Some test trials involved pairs that would be in different categories (one voiced, the other prevoiced) if the category distinction had been learned; other pairs would be from within a single category, regardless of whether or not two categories had emerged. The categorical perception literature, reviewed above, leads to the prediction of above-chance discrimination for pairs straddling a phonemic category boundary, versus near-chance performance for pairs within a category. As expected, pairs that came from within a category produced chance discrimination, regardless of the training conditions. Critically, for the across-category pairs, the Perception Only subjects demonstrated significantly better discrimination than those in the Perception + Production group. In fact, the Perception + Production group did not demonstrate significant improvement from pre- to post-test, suggesting that learning was fully disrupted in this condition. Thus, saying the to-be-learned sounds aloud impaired learning of this phonemic contrast, just as production had impaired lexical development in the Leach and Samuel (2007) study.

The present study

The similarity of the effects at the lexical level (Leach & Samuel, 2007) and the phonetic level (Baese-Berk, 2010) led to the present study. In Experiment 1, we examine whether naïve Spanish listeners show a similar pattern of learning when exposed to a

Basque contrast that they are unfamiliar with. In Experiments 2 and 3, we examine some aspects of the apparent disruption to perceptual learning, and test whether this disruption can be alleviated under some circumstances. In Experiment 2, we tested native Spanish listeners who had some prior experience with Basque to investigate whether the disruption is influenced by experience with the language. In Experiment 3, we tested whether the disruption of perceptual learning is driven by production of the particular new sound being learned, or is instead a consequence of simply engaging the production system while trying to establish a new perceptual contrast.

Experiment 1

Methods

Participants 30 native speakers of Spanish participated in this experiment. Half of the individuals participated in the Perception Only training and half participated in the Perception + Production training. The participants had minimal experience with Basque, English, and other non-native languages. None reported any speech or hearing deficits.

Materials and Apparatus

A native Basque speaker recorded Basque sibilant fricatives and affricates in a sound-treated room using a Sennheiser ME65 microphone. The signal was input to a Marantz PMD-671 digital recorder at a 44.1 kHz sampling rate. Any background noise was then filtered using Goldwave software and individual syllables were extracted using Praat software (Boersma & Weenik, 2015) by a linguist with phonetic training. The sibilant fricative and affricate inventories in Basque and Spanish are quite different. Spanish only utilizes an alveolar fricative and a post-alveolar affricate. Basque, on the other hand, uses a three-way place contrast (i.e., lamino-dental, apico-alveolar, and post-

alveolar) for both fricatives and affricates. Thus, in addition to the three-way place contrasts, there are also fricative-affricate contrasts at each of these places of articulation. Table 1 illustrates the quite different sibilant fricative/affricate inventories in Basque and Spanish.

(Insert Table 1 here)

Three continua were created (/s̻a/- /ʃa/, /s̻a/-/ts̻a/ and /s̻a/-/s̪a/). Though all three continua were presented during the test phases of the experiments, only the /s̻a/- /ʃa/ continuum was presented during training, and only the results from this continuum at test are discussed in the present paper. Below, the procedures for creation of the /s̻a/- /ʃa/ continuum are presented in more detail. Similar procedures were used for the other two continua. A 19-step continuum was created for the /s̻a/- /ʃa/ continuum using a mixing algorithm that shifted slowly from one fricative to the other, a technique that has been used for stop consonants (Repp, 1981; Stevenson, 1979) and for a similar fricative contrast in English (McQueen, 1991). The two base tokens were equated in duration since previous acoustic measures of these two phonemes do not show substantial durational differences between them. Then, the consonantal portions of the two base tokens were mixed using a simple weighted-average method. For example, the most extreme /s̻a/ was made by giving each point in the /s̻a/ base token a weight of 0.95, and giving the points in the /ʃa/ base token a weight of 0.05. For the next member of the continuum, the weights were changed to 0.90 and 0.10. The 19 steps were constructed by changing the weightings in 5% increments (see Kraljic & Samuel, 2005, 2006, 2007; Leach & Samuel, 2007, for examples of the successful use of this method for this contrast). For /s̻a/ and /ʃa/, the critical dimensions that differentiate the two tokens are the

first four spectral moments (i.e., center of gravity, standard deviation, skew, and kurtosis), which determine the shape of the spectrum of the fricatives. The primary dimension on which the two fricatives differ is center of gravity, or the mean spectral energy of the fricative. /s̪a/ has a much higher center of gravity than /ʃa/ (in the case of the present tokens: 4777 Hz vs. 3847 Hz). Therefore, when mixing the two tokens together, the center of gravity shifted across the continuum. The center of gravity for each step on the continuum differed by roughly 49 Hz. These continua were piloted with native Basque speakers. The change from /s̪a/ to /ʃa/ was roughly at the center of this continuum, between steps 7 and 10.

Procedure

All participants completed a pre-test, training period and post-test. In the pre- and post-tests participants heard 72 trials consisting of tokens from /s̪a/- /ʃa/, /s̪a/-/ʦ̪a/ and /s̪a/-/s̪a/ continua. These pre- and post-tests used an ABX task without feedback to assess discrimination of pairs of items from each continuum; therefore, each trial consisted of three tokens. Participants heard two different tokens. Then they were presented with a third token and were asked to determine whether the third token was the same as the first or second. A total of 72 tokens were included from each of the three continua, counterbalancing for order of the three tokens. In each trial, each of three tokens was presented with a 300 msec interstimulus interval. Participants had 3 seconds to respond after the last token was presented. Participants heard pairs of tokens that were three steps apart on the continua. That is, training and test pairs were tokens 1-4, 4-7, 7-10, 10-13, 13-16, and 16-19 from the original 19-step continuum. Training of the participants also was based on an ABX task, but during training feedback was provided on each trial.

Participants heard 5 blocks of 72 trials for a total of 390 training trials, again counterbalancing for the order of the tokens presented. Feedback was presented 1 second after the participants responded. Training tokens were presented in a flat distribution, with each training pair being presented equally often. Training was only on stimuli from the /s̪a/- /ʃa/ continuum, and in the current study we focus on the results for the trained continuum (significant generalization of learning to the untrained continua was not found in any condition). During Perception Only training, participants heard sets of tokens and made a perceptual judgment (i.e., an ABX discrimination judgment). During Perception+Production training, participants heard sets of tokens, repeated the final token, and made the same perceptual judgment that the other group made. Training was conducted on two days, separated by at most 48 hours. On each day, the procedure was as described above, beginning with a pre-test, moving to training, and concluding with a post-test.

Analysis

Results were analyzed using linear mixed-effects models using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) within the R computing program (R Development Core Team, 2014). Performance on the ABX task was the dependent variable. Fixed effects included in the final model were training group (Perception Only vs. Perception + Production) and continuum pair (categorically coded). Models were run using the center of the continuum (pair 3) as the referent level. Random effect structure was the maximal structure that would allow the model to converge, and included random intercepts for participants and items. Significance of factors was determined using model comparisons, and these values are reported in the text below.

Results and Discussion

Figure 2 shows discrimination performance across the /s̪a/- /ʃa/ continuum on the pre- and post-tests, with "Pair Number" referring to where the "A" and "B" tokens came from along the test continuum. For example, pair number 1 consisted of the first and fourth tokens on the continuum. If participants learn two non-native categories, they should show a discrimination peak near the center of the continuum after training, but not on the pre-test. If participants are unable to discriminate between the two non-native categories, we expect to see discrimination near chance (i.e., 50%) for the entire continuum.

At pre-test, as shown in the top panels of Figure 2, both the Perception Only and Perception + Production groups demonstrate flat discrimination functions, indicative of an inability to discriminate between pairs of tokens on the /s̪a/- /ʃa/ continuum. Both groups are near chance with no discrimination peak. No factors emerged as significant predictors of performance on the ABX test (all t-values <1; p-values > 0.5). This flat pattern confirms that before the training, our subjects were unfamiliar with the contrast, a contrast that is not present in their native Spanish inventory.

After training, participants in the Perception Only group (Figure 2, bottom-left panel) learned the distinction and thus performed well above chance when the members of a test pair came from the two different categories ("s" versus "sh") that they had been exposed to during training. The peak near the middle of the graph for the Perception Only group illustrates the learning for this group. In contrast, the Perception + Production group (Figure 2, bottom-right panel) did not learn the distinction, yielding chance performance on all pairs of syllables on the discrimination test held after training. The peak near the middle of the continuum is conspicuously missing for the Perception +

Production group. That is, producing tokens during training disrupted perceptual learning.

(Insert Figure 2 here)

The mixed-effects model analysis supports these observations. Beta values, standard errors and t-values for the best fitting models are included in Appendix A. Training group and continuum pair were both significant predictors of model fit (χ^2=5.26, p<.05 and χ^2=6.17, p<.05, respectively). The interaction between these two factors was significant (χ^2=4.15, p=.05). Examining the results of the mixed model, it is clear that pair 3, the referent level, differs from all other values These results demonstrate that when naïve listeners do not produce the to-be-learned tokens, they can learn to discriminate the new contrast. However, when such listeners do produce the tokens during training, learning the perceptual distinction is disrupted.

The findings in Experiment 1 replicate those found in Baese-Berk (2010), using a novel fricative contrast (/ʒa/- /ʃa/) rather than a pre-voiced – short-lag contrast, and with native Spanish speakers rather than native speakers of American English. In both situations, development of perceptual discrimination was disrupted by producing tokens during training. The perceptual disruption in the current study is particularly striking because the subjects here received active training with explicit feedback, unlike the listeners in Baese-Berk (2010), whose category learning was based on a simple distributional exposure paradigm.

Experiment 2

The experiments demonstrating a disruption of perceptual learning after Perception + Production training have all tested naïve learners. That is, listeners have had very limited prior experience with the trained sounds or words. However, it is clear that eventually learners are able to match their perceptual and production-oriented representations, resulting in successful language learning. In Experiment 2, we asked whether the disruption in developing functional perceptual codes was modulated by experience with the non-native contrast being learned. We trained native Spanish listeners with some prior exposure to Basque on the native Basque contrasts. Because Basque includes the /s/- /ʃ/ contrast we are testing, these listeners have at least some initial familiarity with the distinction.

If the disruption is attenuated with experience, then after training we could see a discrimination peak emerge for the Perception + Production group. By comparing the Perception + Production training group in Experiment 1 to the corresponding group in Experiment 2, we can test whether experience with the language modulates the effect of producing tokens during training. By comparing the Perception + Production group in Experiment 2 to a Perception Only training group with similar prior exposure to Basque, we test whether increased language experience modulates the disruptive effect of production on perceptual processing during learning.

Methods

Participants Thirty native Spanish speakers were recruited to participate in the experiment. These participants were not native Basque speakers but self-reported an intermediate proficiency in Basque (ranging between 4-7 on a scale of 10 for listening, speaking, reading and writing skills). All participants reported multiple years of

experience with Basque. Fifteen participants completed Perception Only training and fifteen participants completed Perception+Production training.

Materials

The materials for Experiment 2 were identical to those in Experiment 1.

Procedure

The procedure for Experiment 2 was identical to that in Experiment 1.

Results and Discussion

The top panels of Figure 3 show the pre-test performance for the two training groups. Before training, listeners demonstrated essentially flat discrimination functions. Neither continuum pair nor training group emerged as significant predictors of performance (all t-values <1; p-values >.5). These flat functions, and in particular the absence of a mid-continuum discrimination peak, demonstrate that this population's prior Basque experience was insufficient (in the absence of any explicit training) to support categorization of the members of the test continuum.

The post-test results for the Perception Only participants were quite similar to those for the Perception Only group in Experiment 1. As the bottom-left panel of Figure 3 shows, we again see a peak in performance for pairs that straddle the category boundary, accompanied by chance-level performance on the within-category pairs. These listeners, like their counterparts in Experiment 1, appear to have learned the perceptual contrast.

The results for the participants in the Perception+Production group (Figure 3, bottom-right panel) are intermediate. There is a small peak at the category boundary, with chance-level performance for all of the other pairs. This pattern is consistent with the idea

that some experience with the contrast may reduce the disruptive effect that production has on developing functional perceptual representations.

(Insert Figure 3 here)

In order to examine this statistically, we used a linear mixed effects model to examine the data. As in Experiment 1, continuum pair was a significant predictor of model fit (χ^2=10.8, p<.05). This significant effect shows that across the two training sessions there was enough category learning to produce a significant discrimination peak at the category boundary. However, unlike Experiment 1, both group, and the interaction between continuum pair and group were not significant (χ^2<1 for both comparisons). Beta values, standard errors and t-values for the best fitting models are included in Appendix A. Examining the results of the mixed model, it is clear that pair 3 differs from the other pairs, which do not differ from one another.

We examined the data from Experiments 1 and 2 together in a mixed effects model, using language background (Experiment), training method (Perception Only versus Perception + Production), and continuum pair as fixed factors, as well as all interactions between these factors. Continuum pair emerged as a significant predictor of model fit (χ^2=16.48, p<.05), reflecting the overall tendency for the emergence of a discrimination peak at the category boundary. No other fixed effects or interactions were significant. Model parameters for the simple fixed effects are shown in Appendix A. This analysis suggests that although experience can partially alleviate the disruption of perceptual learning, there is not significant evidence that the disruption completely dissipates.

We also conducted an analysis comparing the Perception Only groups from Experiments 1 and 2. This analysis revealed that continuum pair was a significant predictor of model fit (χ^2=11.17; p<.05). However, neither experiment (i.e., language background) nor the interaction between continuum pair and experiment were significant predictors of model fit (χ^2<1 for both comparisons). The reliable effect of continuum pair, with no interactions, is a consequence of the discrimination peak found for both Perception Only groups.

A similar analysis of the Perception + Production groups from Experiments 1 and 2 reveals that neither continuum pair nor experiment were significant predictors in the model (χ^2<1 for both comparisons). The interaction between experiment and continuum pair was marginally significant (χ^2=2.64; p=0.1), which is consistent with the plots suggesting that language experience may provide some immunity to the disruption in perceptual learning, but is not sufficient to entirely eliminate it.

Experiment 3

Across the first two experiments, and in the prior work with native English speakers, we have substantial evidence that engaging in production during a learning trial can disrupt the development of functional perceptual representations. In Experiment 3, we begin to examine what the cause of this disruption is. Specifically, we ask whether production of the token being learned is the cause of the disruption, or whether production of non-target speech would also disrupt the development of perceptual representations. In other words, is the disruption tied to something about a particular item (segment, word), or is the

problem a more fundamental clash between the engagement of the production system and the system needed for laying down new speech units for perception?

If engaging the production system in any way is problematic at the moment that perceptual codes are to be established, then we should not see the between-category peak that was evident in the bottom-left panels of Figures 2 and 3. If instead the conflict is item-specific – the system cannot simultaneously work with the perception and production codes for a particular segment – then producing other sounds should not disrupt learning the contrast, and the peak will be found.

Methods

Participants 20 native Spanish speakers with no prior exposure to Basque participated in Experiment 3.

Materials

The speech materials were identical to those in Experiments 1 and 2.

Procedure

Participants underwent Perception+Production training, as in Experiments 1 and 2. As before, during training we presented ABX triplets to be discriminated, with feedback. However, rather than repeating the last token of a triplet, participants instead named a letter presented on the screen. On each trial, one of seven possible letters (L, M, N, O, P, Q, R; none of these include the critical sounds in their pronunciation) was randomly selected and presented on a screen at the end of the presentation of the ABX triplet. The subject named the presented letter out loud, and then pushed one of two buttons to indicate the response on the ABX trial.

Results and Discussion

The top panel of Figure 4 shows performance on the pre-test. As in Experiments 1 and 2, before any training participants show relatively flat discrimination functions (all t-values <1, all p-values>0.4).

The bottom panel of Figure 4 shows the discrimination performance after training that included a production component, but one that did not involve the production of the to-be-learned sounds. On the post-test participants demonstrate a discrimination peak at the category boundary. However, this peak is not as robust as in the Perception Only condition in Experiment 1.

(Insert Figure 4 here)

We compared the two training groups from Experiment 1 to the training group here; recall that the listeners in Experiment 1 and the listeners in the current experiment were all naïve to the contrast before training. Training groups were Helmert coded to allow comparisons among multiple levels within this factor without a loss of power. That is, the groups were split into two comparisons, rather than only one training group serving as the referent level. Again, the beta values, standard errors and t-values for the best fitting models are included in Appendix A. Continuum pair continued to be a significant predictor of performance (χ^2=5.51, p<.05). The group trained in Experiment 3 differed significantly from the Perception + Production training group in Experiment 1 (χ^2=8.34, p<.05), reflecting the discrimination peak found in Experiment 3 but not for that group in Experiment 1.

To focus most specifically on the effects at the category boundary, we looked at the interaction of continuum pair (Pair 1-6) with training condition (Perception Only vs Perception+Production vs Perception+LetterProduction). The interaction between

continuum pair and the comparison of the Perception + Production training group to the other two groups was a significant predictor of discrimination performance (χ^2=4.71, $p<.05$). Collectively, the main effects and the interactions suggest that the learning in Experiment 3 is truly intermediate between the lack of learning demonstrated in Experiment 1's Perception + Production training group and the relatively robust learning demonstrated by the Perception Only training group.

These results indicate that producing something other than the target during training does not disrupt development of perceptual representations as much as producing the target itself. However, the results also demonstrate that producing anything during training, even if it is completely unrelated to the training task, still results in some disruption of the robust perceptual learning achieved if training focuses on the perceptual task at hand. This suggests that the disruption is not simply due to "bad" productions of the target token, but rather that this disruption is caused by a combination of both task related factors (i.e., how one is engaging with the target stimuli during training by producing and perceiving them) and cognitive mechanisms such as selective attention and task switching.

General Discussion

The three experiments presented here examine a disruption in perceptual learning after producing tokens during training. The findings of Experiment 1 converge with previous findings (Baese-Berk, 2010; Leach & Samuel, 2007) that producing tokens during training can disrupt perceptual representations. In Experiment 2, we saw that the disruption of perceptual learning can be lessened with prior exposure to the distinction being learned, but that the disruption cannot be completely alleviated. This suggests that

while experience can be a mitigating factor, perceptual learning is still disrupted by producing tokens. Recall that although participants in Experiment 2 did have substantial experience with the Basque language and reported their proficiency to be at an intermediate level, their ability to discriminate between the target phonemes before training was at chance, demonstrating that they had not yet mastered this contrast. Given the partial alleviation we observed, it is possible that with even more prior language experience, a learner might not show a disruption in perceptual learning as a result of producing tokens during training. It is also possible that with longer periods of training (i.e., more than two days), the observed difference between the Perception Only and the Perception + Production groups would decrease. Baese-Berk (2010) found that adding a third day of training partially alleviated the disruption in perceptual learning. However, participants in that study trained in Perception Only also showed substantial improvements from day 2 to day 3, suggesting that Perception Only training was still more efficient in obtaining perceptual learning. Further, individual differences in learning persisted throughout training for listeners trained in Perception + Production regardless of the length of training. The data in the present study do not allow us to pinpoint when the discrimination peaks emerge for the groups that demonstrate learning. The exact timeline of the emergence of learning under these conditions should be examined in future work.

We should note that these results speak specifically to the development of perceptual representations after training that includes production. It is quite possible that the relationship is not equal in both directions – perceptual training might well aid production. In fact, in Baese-Berk (2010), listeners trained in perception alone improve significantly in their ability to produce tokens. This finding is relatively common in the

literature (see also Bradlow et al. 1997, Bradlow et al., 1999, Lametti et al. 2014, among others). Similarly, given that ultimately perception and production must be coordinated in order for communication to succeed, there should be situations in which production could aid perception. For example, Leach and Samuel (2007) found that subjects who produced the words being learned had an advantage when tested on those words presented in noise. Adank, Hagoort, and Bekkering. (2010) reported that production helped listeners who were trying to adapt to speech in which the experimenters had changed the normal pronunciations of the vowels in words. The phenomenon being studied here is the surprisingly negative impact of production when learning new words or phonetic contrasts, with Experiments 2 and 3 intended to start defining the domain in which this negative effect obtains.

A listener's state of readiness to learn a new contrast develops through years of experience with a language; the training regimes used here and in the related prior studies were conducted over the course of days. Within a trial, of course, the timing issues are on the order of seconds. In all of the cases in which production has been shown to disrupt the development of perceptual representations, participants have been given two tasks to do on each trial. That is, they are required to perceive and produce tokens in relatively rapid succession. As we start to explore the mechanisms that cause the observed disruption in learning, it will be important to understand the time course of the disruptive effect. For example, if we delay the production component by a matter of seconds (i.e., within a trial) or by a matter of minutes (i.e., across blocks of training), will we reduce or even eliminate the disruption? We are beginning such an investigation, and the results will clarify the time that is required to solidify the learning of perceptual representations.

The listener's ability to learn a novel contrast may also differ depending on the relationship of the target sounds to the learner's native language. In the present study, only one endpoint of the continuum is a novel token. That is, Spanish has the alveolar fricative /s/, similar to the Basque /s̪/. Therefore, the listener needs to learn the post-alveolar fricative /ʃ/, rather than learning two novel endpoints to a continuum. This situation mirrors that presented in Baese-Berk (2010), where listeners were asked to learn to distinguish pre-voiced stops, which do not exist in English, from short-lag stops, which do exist in English. It is possible that the disruption in perceptual learning may be modulated if listeners were asked to learn a contrast made of two novel sounds, though whether the disruption would be larger or smaller is unclear. Previous studies examining acquisition of non-native speech sounds have posited that the relationship between the target non-native sounds and the learner's native language will critically impact their ability to learn a contrast (e.g., Best, McRoberts, and Sithole, 1988). This suggests that a disruption in perceptual learning could be modulated by both the relationship of each sound to categories in the learner's native language, and also whether one or both endpoints is new to the learner.

The comparisons across Experiments 1 and 3 demonstrate that while production of the to-be-learned token itself results in the greatest disruption, production of unrelated material can also disrupt the establishment of perceptual representations. These results are particularly informative for two reasons. Had we seen no disruption for the training group in Experiment 3, a reasonable inference would be that the disruption is caused primarily by engaging with the target token in two different modalities in close temporal proximity. That is, if the disruption only occurred when related tokens were produced,

this would implicate an incompatibility in activating related representations in two modalities. Had we seen as large a disruption as in Experiment 1, the source of the disruption would be localized to the engagement of the production system itself. The actual results suggest a more complex and interesting picture. If the question is whether the disruption is caused by engagement of the production system in general or engagement of the target token in two modalities, the answer appears to be "both." While production per se disrupts the robust perceptual learning demonstrated after Perception Only training, production of the token being learned results in a larger disruption.

It is clear that timing, experience, and the nature and content of the intervening tasks influence the disruption seen here. However, these factors are not the only possible features driving the disruption. One logical possibility for a potential source of the disruption in perceptual learning is the content of the learner's productions. If they are presenting themselves with poor exemplars of the training categories, it is possible that learning is being disrupted because they are essentially "flattening" the distributions presented to them, collapsing the two categories. A preliminary analysis of a subset of the training tokens produced in the current study suggests that learners use a number of cues to differentiate the two categories and that these cues do not always match with those used by native speakers. For example, some speakers utilize durational differences rather than the combination of spectral cues (center of gravity, standard deviation, kurtosis, and skew) used by native speakers. Comparing Experiments 1 and 3, it is clear that producing the to-be-learned token increases the disruption in perceptual learning.

However, while this may be a factor, the available data suggest that this explanation cannot account for the full range of the data. Previous studies suggest that

self-productions do not necessarily influence perceptual learning. Baese-Berk (2010) demonstrated that the content of productions during training and during test did not predict the level of learning the novel phonological categories. Further, Kraljic and Samuel (2005) have shown that perceptual learning can be quite robust, even in the face of hundreds of tokens that could potentially have eliminated the effect. Some of the clearest evidence against the hypothesis that inaccurate productions are the driving force of the disruption we see comes from Experiment 3 in the current study. If it were the case that "bad" productions were the primary root of the problem, we would expect to see *no* disruption of perceptual learning when participants were producing an entirely unrelated token. Yet, a notable amount of disruption was observed. At the lexical level, the subjects in Leach and Samuel's (2007) study might have had a small amount of difficulty the first time they had to repeat an item like "bibershack", but most of the 120 productions should have been very accurate, suggesting that the disruption in that in that case is likely caused by attentional constraints, rather than production of "bad" tokens, specifically. Taken together, these findings make it extremely unlikely that the primary cause of the disruption in perceptual learning is exposure to "bad" tokens that the subjects themselves produce.

The results of the three experiments presented here suggest that there is likely not one root cause for the disruption. Learning is a complex process, and phoneme learning is no exception. In addition to linguistic skills, including discrimination and categorization of the target sounds, a learner must also bring to bear the cognitive system, including working memory and attention. The need to simultaneously utilize multiple cognitive skills potentially imposes a substantial cognitive load during training. The repeatedly

observed impairment of perceptual learning when production is required seems to be one type of cognitive load effect, and the results of Experiment 3 indicate that this effect has both a general aspect (some interference by production in general) and a more specific one (more interference from producing the to-be-learned token). In addition, the results of Experiment 2 suggest that pre-existing expertise can mitigate the cognitive load and reduce the magnitude of the disruption.

We believe that it is useful to think about these cognitive load effects as demonstrating a type of "effortful listening." Previous studies have shown that when decoding the speech signal is challenging, the subsequent availability of the content of that signal is impaired (Rabbitt, 1991). Most of the studies examining this phenomenon have focused on cases of elderly and/or hearing-impaired listeners (Pichora-Fuller, Schneider, & Daneman, 1995; Tun, McCoy, & Wingfield, 2009). However, this work has also been extended to non-native listeners (Kewley-Port, Nishi, Park, Miller, & Watson, 2009; Miller, Sillings, Watson, & Kewley-Port, 2009), a population comparable to those tested here. While this literature has primarily addressed disruption of higher-level processing (e.g., semantic processing or memorization) as a result of effortful listening, learning is also a higher-level process that may be disrupted when listening is an extremely challenging task. Requiring a production component during perceptual training may create an effortful listening situation in which perceptual learning is disrupted.

The disruption demonstrated here and in previous studies has important implications for our understanding of the relationship between comprehension and production. When designing the initial experiments which uncovered this disruption (Baese-Berk, 2010; Leach & Samuel, 2007), it was expected that production would

enhance perceptual learning. This expectation was based on the prevailing theories of speech perception and production that suggested that the two modalities are closely related. In fact, theories such as direct realism suggest that perception and production are not just closely tied, but rather that production gestures are the basis of speech perception (Best, 1995; Fowler, 1986). Even in cases where such a close relationship is not explicitly stated, most models of speech perception and production assume that representations are largely shared between the two modalities and that the processes underlying access of these representations are also highly similar. That is, speech perception and production are two sides of the same coin.

Given the results of the current study and those of the studies that motivated it, a more nuanced view of the relationship between perception and production is needed. If the two modalities do share virtually all of their representations and processes, we would not expect to see the observed disruption in both phonological and lexical perceptual learning when production is required. However, it is also clear that representations are not entirely separate across the two modalities. For example, previous work has demonstrated that relatively rapid responses in production are facilitated by prior perceptual exposure (Goldinger, 1998; Goldinger, 2000; Goldinger & Azuma, 2004; Nielsen, 2011; Pardo, 2006; Shockley, Sabadini, & Fowler, 2004). This body of work suggests that there must be some relatively close connection between the two modalities. And of course our own data provide evidence for a connection between the two modalities because if production did not engage the perceptual representations being formed during training, there would not be any disruption observed in the current studies.

Our results suggest that the relationship between the two modalities shifts over the

course of learning. At the earliest stages of learning, as in the case of the participants in Experiments 1 and 3, perceptual representations are relatively fragile. Engaging the production system disrupts the learning process because these representations are still being formed, with particularly strong interference generated when the item being produced directly competes with the to-be-learned sound. Experiment 2 suggests that when there is at least a nascent representation, the representations are less fragile, and engaging in production is less disruptive. The pattern is consistent with systems that are initially competitive, then less so, and perhaps eventually mutually supportive. This is clearly speculative at this point, but it is already clear that the relationship between perception and production is more complex than previous studies have suggested. Producing tokens during training can disrupt perceptual learning, and the cause of this disruption is multi-faceted. The results of the current study provide some preliminary information about the sources of this disruption.

References

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903-1909.

Baese-Berk, M. (2010). The relationship between perception and production when learning novel phonological categories, Unpublished doctoral dissertation, Northwestern University.

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, Unpublished doctoral dissertation, Northwestern University.

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

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human perception and performance*, *14*(3), 345.

Boersma, P., & Weenik, D. (2015). Praat: doing phonetics by computer.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception and Psychophysics*, *61*(5), 977–985.

Bradlow, A. R., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English/r/and IV: IV. Some effects of perceptual

learning on speech production. *JournaI of the Acoustical Society of America*, *101*(4), 2299–2223.

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 629–647.

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.

DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, *46*(4), 613–642.

Denes, P. B., & Pinson, E. N. (1993). *The Speech Chain: The Physics and Biology of Spoken Language*. Macmillan.

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.

Flege, J. E. (1995). *Second language speech learning: Theory, findings, and problems* (pp. 233–277). Speech perception and linguistic experience: Issues in cross-language research.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*(1), 3-28.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363-376.

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–278.

Goldinger, S.D. (2000). The role of perceptual episodes in lexical processing. Keynote Address. Published in Proceedings of the Workshop on Spoken Word Access Processes. Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. Pp. 155-158.

Goldinger, S., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review*, *11*, 716–722.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds. *Neuropsychologia*, *9*(3), 317–323.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor Adaptation in Speech Production. *Science*, *279*(5354), 1213–1216.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech ICompensation and Adaptation. *Journal of Speech, Language, and Hearing Research*, *45*(2), 295–310.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakersa). *Journal of the Acoustical Society of America*, *126*(2), 866–877.

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

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *43*(1), 88-96.

Kewley-Port, D., Nishi, K., Park, H., Miller, J. D., & Watson, C. S. (2009). Learn to

Listen (L2L): Perception training system for learners of English as a second language. *Journal of the Acoustical Society of America*, *125*(4), 2773.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125–156.

Kingston, J. (2003). Learning Foreign Vowels. *Language and Speech*, *46*, 295–349.

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.

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations, *Cognition, 121*, 459-465.

Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science, 25* (7), 1325-1336.

Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech, Language, and Hearing Research*, *14*(4), 677–709.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353.

Libermann, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431-461.

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

Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales Des Maladies De LOreille Et Du Larynx*, *XXXVII*(2), 101–109.

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

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us. Proceedings of the 25th annual Boston University conference on language development, 480–490.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science: a Multidisciplinary Journal*, *32*(3), 543–562.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training to a new linguistic contrast in voicing. *Perception and Psychophysics*, *34*(4), 323–330.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 433.

Miller, J. D., Sillings, R., Watson, C. S., & Kewley-Port, D. (2009). Speech Perception Assessment and Training System (SPATS-ESL) for speakers of other languages learning English. *Journal of the Acoustical Society of America*, *125*(4), 2755-2771.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*,

1–11.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*, 2382-2393.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*(1), 593–608.

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

Rabbitt, P. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica*, *111*(s476), 167–176.

Repp, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, *18*, 12–14.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in Basic Research and Practice, 10*, 243-335.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71-80.

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.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception and Psychophysics*, *66*(3), 422–429.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, *36*, 131–145.

Stevenson, D. C. (1979). *Categorical perception and selective adaptation phenomena in speech* Unpublished doctoral dissertation, University of Alberta, Edmonton, Alberta, Canada.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *NeuroReport*, *9*(16), 3556-3560.

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, *24*(3), 761–766.

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.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones: Transfer to production. *Journal of the Acoustical Society of America*, *106*(6), 3649–3658.

Appendix A: Summaries of mixed effects models

Experiment 1

Fixed Effects	Estimate	Standard Err	t-value
Intercept	0.69	0.063	9.985
Training Group	-0.211	0.09	-2.157
Pair 1	-0.116	0.09	-1.19
Pair 2	-0.192	0.09	-1.98
Pair 4	-0.173	0.09	-1.78
Pair 5	-0.115	0.09	-1.191
Pair 6	-0.117	0.09	-1.2
Training Group * Pair 1	.017	0.13	1.36
Training Group * Pair 2	0.27	0.13	1.96
Training Group * Pair 4	0.18	0.13	1.26
Training Group * Pair 5	0.19	0.13	1.4
Training Group * Pair 6	0.19	0.13	1.4

Experiment 2

Fixed Effects	Estimate	Standard Err	t-value
Intercept	0.700	0.071	9.913
Training Group	-0.103	0.09	-1.106
Pair 1	-0.2	0.1	-2.02
Pair 2	-0.201	0.1	-2.02
Pair 4	-0.25	0.09	-2.52
Pair 5	-0.15	0.09	-1.512
Pair 6	-0.155	0.09	-1.151
Training Group * Pair 1	.085	0.132	0.641
Training Group * Pair 2	0.103	0.132	0.787
Training Group * Pair 4	0.096	0.132	0.729
Training Group * Pair 5	0.092	0.132	0.699
Training Group * Pair 6	0.92	0.132	0.699

Experiments 1 & 2 Comparison

Fixed Effects	Estimate	Standard Err	t-value
Intercept	0.7	0.075	9.307
Training Group	-0.104	0.100	-1.04
Experiment	-0.008	0.100	-0.077
Pair 1	-0.200	0.105	-1.900
Pair 2	-0.200	0.105	-1.900
Pair 4	-0.250	0.105	-2.375
Pair 5	-0.151	0.105	-1.425
Pair 6	-0.153	0.105	-1.425

Experiment 1 & 2 Comparison (Perception-Only)

Fixed Effects	Estimate	Standard Err	t-value
Intercept	0.700	0.075	9.269
Language Background	-0.008	0.101	-0.077
Pair 1	-0.2	0.107	-1.873
Pair 2	-0.2	0.107	-1.873
Pair 4	-0.25	0.107	-2.341
Pair 5	-0.15	0.107	-1.404
Pair 6	-0.15	0.107	-1.494
Language Background * Pair 1	0.085	0.142	0.596
Language Background * Pair 2	0.008	0.142	0.054
Language Background * Pair	0.008	0.142	0.541

4			
Language Background * Pair 5	0.035	0.142	0.244
Language Background * Pair 6	0.035	0.142	0.244

Experiment 1 & 2 Comparison (Perception+Production)

Fixed Effects	Estimate	Standard Err	t-value
Intercept	0.595	0.066	9.062
Language Background	-0.116	0.093	-1.252
Pair 1	-0.115	0.089	-1.293
Pair 2	-0.096	0.089	-1.078
Pair 4	-0.154	0.089	-1.725
Pair 5	-0.058	0.089	-0.647
Pair 6	-0.057	0.089	-0.647
Language Background * Pair 1	0.173	0.126	1.372
Language Background * Pair 2	0.173	0.126	1.372
Language Background * Pair 4	0.154	0.126	1.219
Language Background * Pair 5	0.135	0.126	1.067
Language Background * Pair 6	0.135	0.126	1.067

Experiment 3

Fixed Effects	Estimate	Standard Err	t-value
Intercept	0.6	0.055	10.846
Production vs. Other Groups	-0.211	0.097	-2.180
Perception vs. Other Groups	0.092	0.088	1.047
Pair 1	-0.125	0.078	-1.614
Pair 2	-0.113	0.078	-1.453
Pair 4	-0.100	0.078	-1.291
Pair 5	-0.113	0.078	-1.453
Pair 6	-0.075	0.078	-0.968
Production * Pair 1	0.173	0.136	1.274
Production * Pair 2	0.269	0.136	1.982
Production * Pair 4	0.173	0.136	1.274
Production * Pair 5	0.192	0.136	1.415
Production * Pair 6	0.192	0.136	1.415
Perception * Pair 1	0.009	0.123	0.078
Perception * Pair 2	-0.079	0.123	-0.647
Perception * Pair 4	-0.073	0.123	-0.592
Perception * Pair 5	-0.003	0.123	-0.023
Perception * Pair 6	-0.040	0.123	-0.327

Figure and Table captions

Figure 1: Size of the perceptual learning shifts found by Leach and Samuel (2007), for the first two days of training and for the last two days of training.

Figure 2: Accuracy on the ABX discrimination task, before and after training, Experiment 1. Top-left panel: Pre-test results for the Perception Only group. Top-right panel: Pre-test results for the Perception+Production group. Bottom-left panel: Post-training results for the Perception Only group. Bottom-right panel: Post-training results for the Perception+Production group.

Figure 3: Accuracy on the ABX discrimination task, before and after training, Experiment 2. Top-left panel: Pre-test results for the Perception Only group. Top-right panel: Pre-test results for the Perception+Production group. Bottom-left panel: Post-training results for the Perception Only group. Bottom-right panel: Post-training results for the Perception+Production group.

Figure 4: Accuracy on the ABX discrimination task, before and after training, Experiment 3. Top panel: Pre-test results. Bottom panel: Post-training results.

Table 1: Phoneme inventories for Basque and Spanish sibilant fricatives and affricates.