Citation for published version:

Tania S. Zamuner, Elizabeth Morin-Lessard, Stephanie Strahm, and Michael P. A. Page, 'Soke word recognition of novel words, either produced or only heard during learning', *Journal of Memory and Language*, Vol. 89, August 2016, pp. 55-67.

DOI:

https://doi.org/10.1016/j.jml.2015.10.003

Spoken word recognition of novel words, either produced or only heard during learning

Tania S. Zamuner[1], Elizabeth Morin-Lessard[2], Stephanie Strahm[1], Michael P. A. Page[3]

[1]Department of Linguistics, University of Ottawa

[2]Department of Psychology, Concordia University

[3]Department of Psychology, University of Hertfordshire

Address for Correspondence:

Tania S. Zamuner

Dept. of Linguistics

University of Ottawa

Arts Hall, 70 Laurier Ave. East

Ottawa ON, Canada, K1N 6N5

tzamuner@uottawa.ca

Acknowledgements:

This research was supported by a SSHRC grant awarded to TSZ and an FQRSC Doctoral Fellowship awarded to Elizabeth Morin-Lessard.

Abstract

Psycholinguistic models of spoken word production differ in how they conceptualize the relationship between lexical, phonological and output representations, making different predictions for the role of production in language acquisition and language processing. This work examines the impact of production on spoken word recognition of newly learned non-words. In Experiment 1, adults were trained on non-words with visual referents; during training, they produced half of the non-words, with the other half being heard-only. Using a visual world paradigm at test, eye tracking results indicated faster recognition of non-words that were produced compared with heard-only during training. In Experiment 2, non-words were correctly pronounced or mispronounced at test. Participants showed a different pattern of recognition for mispronunciation on non-words that were produced compared with heard-only during training. Together these results indicate that production affects the representations of newly learned words.

Spoken word recognition of novel words, either produced or only heard during learning

Second language instructors will often ask their students to repeat new words when they are first introduced, in order to facilitate learning (Duff, 2000). Similar processes are seen in first language acquisition where children imitate the gestures and words of their caregivers (Bannard, Klinger, & Tomasello, 2013; Clark & Bernicot, 2008; Kuhl & Meltzoff, 1996; Meltzoff & Moore, 1977). Imitation productions have been examined from different perspectives in language processing, especially from the view of language development. For example, Bannard et al. (2013) found that the rate of 3-year-old children's imitative productions increased when a novel adjective had a contrastive or descriptive function. Other work has shown that there is a pragmatic function of imitation between children and adults that marks additional information in the common ground of the conversation (Clark & Bernicot, 2008). Imitation is not limited to human language, as it also plays a role in other communication systems such as in songbirds (Tchernichovski, Mintra, Lints, & Nottebohm, 2001). However, despite the extensive research on the topic of imitative speech productions in language processing and language development, few studies have controlled the effects of speech production on word learning. We investigated adults' spoken word recognition of non-words learned under two conditions, either heard-then-produced or heard-only at the time of training. The goal of the first experiment was to determine the effect of production on spoken word recognition for newly learned non-words. The goal of the second experiment was to investigate whether participants would show different abilities to detect mispronunciations of non-words depending on whether the non-words were heard-then-produced or heard-only during training.

Speech production serves a variety of functions in language processing, such as allowing for the creation of articulatory representations or the generation of auditory and sensori-motor feedback (see review by Hickok, 2014). These different functions of speech production are not mutually exclusive. One function of production is in its relationship to the short-term and long-term memory of word-forms, as described in Baddeley and Hitch's (1974) model of working memory. In this model, the phonological loop is a module of working memory directly involved in the processing of phonological input (Baddeley, 1986; 2012; Gathercole, Pickering, Ambridge, & Wearing, 2004; etc). The phonological loop is made up of a phonological store and a subvocal rehearsal system, and it is the module of working memory that is argued to be involved in the learning of novel words (Baddeley, 1986, 2012; Baddeley, Gathercole, & Papagno, 1998; Gathercole et al., 2004; Gupta & MacWhinney, 1997; Gupta & Tisdale, 2009; Majerus, Poncelet, Elsen, & Van der Linden, 2006; Page & Norris, 2009, etc.). The speech stimulus enters the phonological short-term store where it may be rehearsed using a (normally subvocal) rehearsal system. Young children are presumed to use the phonological loop for short-term storage of speech-based materials but may use general learning strategies other than rehearsal to complete working memory tasks. In both adults and children, repeated exposure to a given list for serial recall leads to longer-term learning (Hebb, 1961; Mosse & Jarrold, 2008; Page, Cumming, Norris, McNeil, & Hitch, 2013). Moreover, the longer-term learning that is evident in this so-called "Hebb repetition learning" has been closely related to the learning of phonological word-forms (Szmalec, Duyck, Vandierendonck, Mata, & Page, 2009; Szmalec, Page, & Duyck, 2012). Importantly for the work presented here, Hebb repetition learning is very much stronger when participants are asked to attempt recall of repeatedly presented lists (i.e., they attempt to produce them) as opposed to when they are simply exposed to particular lists on multiple occasions (Cohen & Johansson, 1967a; 1967b; Cunningham, Healy, & Williams, 1984; Oberauer & Meyer, 2009; Page, Cumming, Norris, Hitch & McNeil, submitted). If Hebb repetition learning is akin to lexical learning, therefore, we would hypothesize that such lexical learning would also be enhanced by an attempt to produce the word during learning, relative to passive exposure.

Production might influence the strength, or the nature, or both, of newly formed linguistic representations. Models differ in how words' articulatory or output gestures are stored, which has implications for learning and language processing. In some models, representations have multiple dimensions: phonetic details are stored bundled together with semantic and other lexical information. In exemplar and usage-based theories of speech perception and speech production (e.g., Goldinger, 1998; Pierrehumbert, 2001; 2002; 2003; Sosa & Bybee, 2008), the speaker keeps track of instances of a particular input to create a phonetic map or exemplar cloud. Each token leaves a unique memory trace, which includes detailed perceptual and conceptual information. In perception, traces are activated according to their similarity to the input stimulus. The average features of the traces with the highest activation provide the meaning to the incoming token (Goldinger, 1998). In speech production, a trace is selected randomly from the exemplar cloud and the target for production is a weight of the surrounding exemplars (Pierrehumbert, 2001; 2003). These models allow for gradation in articulation based on lexical factors, as found in studies such as that by Goldrick, Vaughn and Murphy (2013), who report that the production of voicing contrasts in words (measured word-initially with voice onset time and measured word-finally with vowel duration) was enhanced or reduced depending on the words' status as a minimal pair, with or without contrastive voicing. Working within this framework, the prediction is that newly formed lexical representations include detailed phonetic information, and that this additional representation would lead to faster word recognition for produced words relative to heard-only words. The assumption is that, as in Pierrehumbert (2002), the production and perception system are closely tied together, with the same levels of representations in both systems. Note that production effects may also be captured by recent hierarchical and connectionist models which allow for cascading activation from the lexical level to the phonological level and to the phonetic level (see Goldrick, Baker, Murphy & Baese-Berk, 2011 for a review). They are also consistent with recent neurocognitive models of speech production that include sensory-motor components (Dell, Schwartz, Nozari, Faseyitan, & Coslett, 2013; Hickok, 2012). Essentially, production effects on word recognition will emerge if the act of producing a nonword establishes some form of *bidirectional* link between a newly learned (lexical) input representation of that nonword and some corresponding output representation, either comprising a newly learned output lexeme or direct, newly learned connections to output phonology. Effects on word recognition would be contingent on this output system's being activated during word recognition (either automatically or otherwise), with activation flowing back to the input system via the learned bidirectional links. It should be noted, however, that most models of speech production lack detail regarding the processes by which new lexical items are learned and integrated in this way into the speech perception-production system. We now turn to describing previous research that has examined the role of production in word learning.

Most studies investigating the effects of production have focused on adults' learning of novel words using reading methods (Gathercole & Conway, 1988; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; Forrin, MacLeod, & Ozubko, 2012), or learning words in a second language (Dahlen & Caldwell-Harris, 2013; Ellis & Sinclair, 1996). For example, Gathercole and Conway (1988) performed a series of experiments in which adults were presented with lists of words and were asked to learn them in a number of conditions: read out loud, read and heard, heard only, mouthed, written, or read silently. Words that were read out loud were more often recognized at test than words in any other condition. Studies have also looked at the role of articulation in learning a second language. Ellis and Sinclair (1996) taught English-speaking adults words and phrases in Welsh and found that participants who repeated the training utterances performed better than participants who were not able vocally to rehearse the Welsh words, on a variety of linguistic tasks including grammaticality judgments, knowledge of mutated nouns, and translation ability. Similar results have been found for adults learning Turkish nouns, where vocal or subvocal rehearsal resulted in increased recall and recognition compared with silent listening or when auditory feedback was masked by white noise (Dahlen & Caldwell-Harris, 2013). Studies with adults indicate, therefore, that participants remember and recognize words better when they are practiced orally.

This effect of better recall and recognition for overtly produced words has been termed the *production effect* (Hopkins & Edwards, 1972; MacLeod et al., 2010). This effect occurs in studies with within-subject, mixed-list designs in which some words are read aloud and others are read silently. The production effect is also found when mouthing words compared to silent reading, but not with other types of overt responses such as pressing a button for half the words or saying "yes" following half the words and not the other half. Most comparable to the conditions in the current experiments, MacLeod (2011) also found that the production effect is strongest when participants generate the words themselves, compared with an experimenter or another participant generating the word. Existing lexical representations are not thought to be necessary for a production effect to occur, as evidenced by a production effect with non-word stimuli (MacLeod et al., 2010). The production effect is also found in other linguistic modalities, so that words that are spelled out loud, written, or typed, all improve recognition for words at test compared with those read silently (Forrin et al., 2012), though words read aloud are recognized better than in other conditions. Recognition in these studies was defined as participants' ability to correctly indicate with a keyboard response whether or not a word was presented during the study phase. It is argued that the production effect stems from a distinct record in episodic memory for produced words, making them easier to retrieve from memory during a recall or recognition task. According to Forrin et al. (2012), "Reading out loud facilitates recognition more than these other productions because the act of speaking is more likely to be incorporated into the processing record, and hence is more distinct. Speaking involves processing along an extra dimension relative to many other productions, processing that is more active and more embodied" (p. 1054). Exactly how production is represented is not clear, though previous studies were not specifically designed to test this. In the current research, we explore how the production effect relates to linguistic representations and processing, while controlling, more carefully than previously, the number of exposures to each word.

There is some evidence that shows a differential effect of production depending on the linguistic characteristics of the stimuli. Kaushanskaya and Yoo (2011) found that adults were better at recognizing non-words that were produced at training compared with non-words that were silently rehearsed, but only when these non-words contained phonemes within the native language of the participants. The opposite effect was found if the non-words contained non-native sounds. It is argued that it is easier to learn non-words by producing them when they contain similar sounds to the L1 because speakers' experience with the sounds means they do not have to focus on articulation and can instead concentrate on associating the meaning of the non-words. When non-words contain unfamiliar sounds, participants are more focused on pronouncing the sounds correctly than associating the sounds with a meaning. Also, the participants' productions did not match the auditory target, which may have disrupted learning. This leads to an important point, namely, that production may not always show beneficial or improved learning effects, and this will depend on the linguistic characteristics of the stimuli. We return to this issue in the discussion.

As a starting point, we first tested how hearing-then-producing versus only-hearing non-words might affect spoken word recognition of newly formed lexical representations. In Experiment 1, participants were trained on non-word/nonce-animal pairs and either repeated the non-word or remained silent. During the test phase, participants were presented with two images and asked to look at a target. If participants learn the novel associations, they were expected to look longer at the target compared with the distractor. Importantly, based on the research reviewed in the previous sections, it was also predicted that participants would be more accurate at recognizing non-words that were produced during training as opposed to heard-only.

Experiment 1

Method

Participants.

Participants were 29 university students (4 males, mean age = 20 years) who received partial course credit for participating. All adult participants had English as their L1, and 13 reported some knowledge of an L2. All adult participants reported having normal or corrected-to-normal vision, normal hearing, and no history of language impairment. Five additional participants were tested but their data was not included because of: missing data on more than 3 out of the 8 test trials (2); no data in one of the conditions; equipment error (1); no video for off-line coding (1); audio quality too poor for off-line coding (1).

Stimuli.

Stimuli consisted of 16 CVC non-words (Table 1), which were divided into two sets of 8 CVCs, labeled Set 1 and Set 2, such that each CVC in Set 1 had a corresponding CVC in Set 2 with which it shared a rhyme (VC). Across the two sets, each CVC had a unique initial consonant, so each could be identified from the initial consonant. In Experiment 1, participants were tested on half of the stimulus set (8 items), that is, on either Set 1 or Set 2. The stimuli were recorded by a female native speaker of English and were normalized for amplitude (70 dB). The visual stimuli consisted of eight images of coloured nonce animals (Ohala, 1999). The association of the non-words and the nonce animals was counterbalanced across eight distinct stimuli lists to control for potential differences in the salience of the non-words.

----------TABLE 1 ABOUT HERE----------

Design.

There were two phases, both of which had the same design and timing for the trials: a Task Practice Phase followed by an Experimental Phase. Each phase consisted of training trials that were immediately followed by test trials. Participants began with the Task Practice Phase (six Practice Training trials, followed by six Practice Test trials). The Task Practice Phase was to familiarize participants with the task. For this Task Practice Phase, only one participant required a second run, due to calibration issues rather than difficulties with the task. The Practice stimuli consisted of six images of fruits (e.g. mango, apple, kiwi, etc.) and their corresponding name. Real words were used to avoid introducing additional non-words to the participants, which might increase the difficulty of the task. Following the Task Practice Phase, participants began the Experimental Phase, consisting of 16 Experimental Training trials, followed by 8 Experimental Test trials.

In the Experimental Training trials, half of the eight non-words were heard-then-produced (4 non-words, 2 trials each, henceforth referred to as produced) and half were heard-only (4 non-words, 2 trials each). Participants were presented with an auditory token of a non-word, while they saw the corresponding nonce animal image. After 2000 ms, a prompt image appeared below the nonce-animal, which depicted the appropriate response. When participants saw a picture of a finger pointing at them, the appropriate response was to repeat the non-word. When they saw a picture of a woman gesturing “shh” with her finger over her lips, the appropriate response was to remain silent and hear a second recording of the non-word (see Figure 1). This meant that participants did not know the assigned training condition (produced or heard-only) until after they saw the prompt-image appear.

----------FIGURE 1 ABOUT HERE----------

It is important to note that in both produced and heard-only conditions, participants were exposed to the non-word twice in each training trial. In the produced training trials, participants heard the item once from the computer and once when they produced it. In the heard-only training trials, the non-word was heard twice from the computer (the second presentation of the heard non-word was repeated by the computer 500 ms after the prompt-image appeared and was the same audio token in both instances). This was to ensure that each non-word had an equal amount of auditory exposure in each training condition. Controlling for the amount of input for non-words in the different conditions is important because of the known effect of frequency on learning (e.g., Gershkoff-Stowe, 2002; Pierrehumbert, 2001), particularly with reference to the Hebb repetition learning effect, in which learning is approximately linearly related to the number of repetitions. Figure 1 illustrates the sequence of produced and heard-only trials. Pre-randomization of training trials ensured that no more than two produced or two heard-only trials were presented consecutively. Participants’ responses during the training phrase were recorded with an external video camera. Once the training phase was completed, an instruction screen followed that indicated the start of the experimental test trials.

In the Experimental Test Phase, participants were presented with two images side-by-side simultaneously, and were instructed to look at the image that matched the non-word presented auditorily. On each trial, the two images were presented for 2000 ms in silence for a preview of the visual display, which was followed by the presentation of the target non-word (the same audio token that was used during training). After 1500 ms, a sentence prompting participants to continue looking was played (“Do you see it?” or “Do you like it?”). Prompts were included because the study was designed also to be suitable for children. Each trial ended with a blank screen for 2000 ms. Images were paired so that the items in a simultaneously presented pair were either both produced or both heard-only during the Experimental Training Phase. This was to avoid a possible effect of produced non-words being more salient in the pair than heard-only non-words or vice versa. Each non-word was used as a target once (for a total of eight test trials) and counterbalanced for which side the target appeared on (left or right). Images were horizontally centered on the screen, spaced 420 pixels apart and sized 400 x 400 pixels. Areas of interest were created around both the target and distractor using the Experiment Builder software.

Procedure.

The experiment was presented using Experiment Builder software (SR Research, Ottawa) installed on an Apple Mac Mini. Eye movements and looking times to both the target and the distractor were recorded using an Eyelink 1000 Remote eyetracker (SR Research, Ottawa), tracking participants’ dominant eye, with a sampling rate of 500 Hz. Participants were seated in a sound attenuated booth 500-600 mm from a monitor. Calibration was based on a 5-point calibration grid arranged in a 'plus sign' shape. Drift correction was performed between every trial in the form of a central fixation image (a round-shaped blue bird's face) to account for shifts in eye position. Additional auditory and visual recordings were made using a Zoom Q2HD Handy Video Recorder for offline coding of training responses. Participants were tested in one session, lasting approximately 10 minutes. At the end of the experiment, participants were informally asked how they found the task, and whether they guessed the purpose of the study. The majority of participants were not aware of which variables had been manipulated, but some reported that they thought they remembered the non-words that they produced better than the ones that they only listened to.

Coding of training

During training trials, participants were either to repeat the non-word or to remain silent. At times, participants did not provide the appropriate response (i.e., they failed to produce a non-word when prompted, or produced a non-word when unprompted). Two researchers coded the audio-video recordings from the training sessions off-line. The training trials on which participants gave incorrect responses and their corresponding test trials were deleted prior to data analyses. For instance, if a participant mistakenly produced the non-word /mig/ on a heard-only training trial, the corresponding experimental test trials in which /mig/ appeared as the target or distractor were removed. Other incorrect responses, such as mispronouncing the initial consonant of the target /hæs/ as /pæs/, also resulted in the removal of the corresponding experimental test trials. In total, 22 trials were removed from analyses (13 produced, 9 heard-only), which was an average of .76 trials per participant.

Results

The total looking time and proportions of time spent looking at the target and distractor areas of interest were extracted using DataViewer software (SR Research, Ottawa), analyzed in 100 ms time bins, starting from 200 ms after the onset of the target non-word as this is the time it typically takes for adults to program an eye movement (Matin, Shao, & Boff, 1993; Salverda, Kleinschmidt, & Tannenhaus, 2014). A visual inspection of the fixation data indicated that looks to the target non-word peaked around 1000 ms after the target non-word onset, indicating that reference had been resolved. This resulted in a window of analysis spanning from 200 to 1200 ms from the target non-word onset, with 10 time bins in total. A visual inspection of the data in Figure 2 shows differences in the proportion of looking to targets on produced vs. heard-only trials; which was a reliable difference at 200 and 300 ms into the trial (based on sliding *t*-tests on each 100 ms time bin).

The proportional data in the 200 to 1200 ms window were examined with a growth curve analysis (GCA) used for analyzing eyetracking in spoken language data (Mirman, Dixon, & Magnuson, 2008). GCA is a multilevel regression method allowing for the assessment of both the differences in time-spent-looking and in the steepness of a looking curve (i.e., in the rapidity of access to a target non-word over time), information that may be lost when performing traditional ANOVAs that collapse all potential changes in response patterns following the presentation of spoken language data (Barr & Frank, 2009).

Participants' proportional looking to the target was analyzed using a growth curve analysis performed in R (R Core Team) using the lmer() function from the lmer4 package (version 1.1-7; Bates, Maechler, Bolker & Walker, 2014). The empirical logit was calculated for each time bin as an approximation to log odds (Barr, 2008). The model contained Training (produced, heard-only) as the main predictor, which was deviation-coded. The model also included main effects of Training and Time (captured by orthogonal polynomials) and a Training*Time interaction. Random by-participant effects were included in the model to account for by-participant variation in their slopes and intercepts.

The overall time course of looking to the target was modeled with second-order (quadratic) orthogonal polynomials. Fixed effects of Training (produced vs. heard-only, a within-subjects variable) and how they interact with time were included as the main predictors of interest. By-participant random effects were also included in the model, but are not reported in the tables for simplicity. There was no statistically significant effect of Training (produced vs. heard-only) overall (*Estimate* = 0.83, *SE* = 0.47, p = .08), and no statistically significant effect of training on the linear term (*Estimate* = 0.20, *SE* = 0.55, p = .71). However, there was a statistically significant effect of Training on the quadratic term, indicating a difference in the shape of the looking curve between the conditions (*Estimate* = 1.13, *SE* = 0.54, p = .04). As seen in Figure 2, there is a rise in fixations in the Produced condition over time, and, by contrast, a rise and a fall in the Heard condition over time.

----------TABLE 2 ABOUT HERE----------

----------FIGURE 2 ABOUT HERE----------

Discussion

In Experiment 1, non-words that were either produced or heard-only during training were successfully learned by the adult participants, as indicated by the looking to the target image in both conditions. There was also a differential pattern of looking to targets that were produced compared with heard-only during training, with a rise in looking to produced targets, in spite of the fact that both produced and heard-only targets received exactly the same number of exposures during training.

In seeking to explain this difference, a question arises regarding the nature of the newly formed representations. It is possible that produced targets were recognized faster than heard-only targets, because their representations are qualitatively different in some way. Such a qualitative difference might, for instance, consist in the memory representations of produced non-words being embellished with a newly learned motor representation, while representations of heard-only non-words are not. If this were the case, then it would be necessary to explain how the addition of a motor representation results in speeded performance in a task that is ostensibly based on recognition rather than production. Alternatively, non-words in the different learning conditions may be qualitatively similar but quantitatively different, such as in the strength or robustness of their representation. In this case, it would be interesting to know why two "perfect" presentations of a given non-word (in the heard-only condition) resulted in a memorial representation that was less strong/robust than the representation of a non-word with one presentation and one production. It is, of course, possible that there are both qualitative and quantitative differences, as in the explanation of a similar result seen in the Hebb repetition-learning literature.

There are several models that relate Hebb learning to the learning of phonological word-forms (e.g., Burgess and Hitch, 2006; Gupta & Tisdale, 2009; Page & Norris, 2009). Space forbids a detailed comparison between these models (see Page & Norris, p.

3748-3749, for such a comparison); here, we discuss our results within the framework of the Page and Norris model with which we are most familiar. This framework suggests that the act of production leads to the learning of a new production-based representation (a qualitative addition), with that productive representation connected by a bidirectional link to the perception-based representation of the same item. The linked production-based representation can, by spreading activation between itself and the corresponding perceptual representations, then support perceptual activation even in a subsequent recognition task in which no production is required (a quantitative effect). This structure mirrors the distinction, and the connection pattern, between input and output lexicons found in classic models of the structure of auditory-verbal representation, such as that of Ellis and Young (1988). As in the Ellis and Young model, both the perceptual and production representations of a given item can also be connected bidirectionally to a corresponding semantic representation. The triangle of facilitatory connections between the perceptual, production, and semantic representations of a given item, can lead to enhanced activation of, say, the perceptual representation in an item recognition task as, by hypothesis, occurs in the experiment described above.

In Experiment 2, we sought to investigate whether participants would show different abilities to detect mispronunciations of non-words, that is, forms that are not a one-to-one match or a “perfect” representation of newly-learned non-words. As in Experiment 1, we also explored whether training (produced vs. heard-only) would affect recognition of correctly pronounced and mispronounced non-words at test. If production improves recognition at test compared with hearing-only, then we would expect to find similar results to those found in Experiment 1.

Experiment 2

In Experiment 2, we trained participants on non-words that were either produced or heard-only during training. At test, participants were presented with the trained non-words that were either correctly pronounced (CP) or mispronounced (MP). If production strengthens newly formed representations or access/processing of these representations, we predicted that participants would be more likely to detect mispronunciations on non-words that were produced compared with those that were heard-only during training.

Method

Participants.

Participants were 30 university students (3 males, mean age = 20 years) who received partial course credit for participating. All adult participants had English as their L1, and 16 reported some knowledge of an L2. As in Experiment 1, all adult participants had normal or corrected-to-normal vision, normal hearing, and no history of language impairment. Three additional participants were tested but not included because of equipment error (1) and no data in one of the conditions (2).

Stimuli.

Same as in Experiment 1.

Design.

The design of the Task Practice Phase was the same as in Experiment 1. The design of the 16 Experimental Training trials was also the same as in Experiment 1. In the training trials, participants were presented with 8 non-words that were either produced (4 non-words, 2 trials each, referred to as produced) or heard-only (4 non-words, 2 trials each). At test in Experiment 1, participants were presented with eight

Experimental Test trials in which all of the stimuli were correctly pronounced. In Experiment 2, by contrast, participants were also presented with both correctly pronounced (CP) and mispronounced (MP) tokens of all eight trained non-words (Set 1 was CP and Set 2 was MP, or vice-versa) for a total of 16 test trials. Each MP token differed from the target and distractor by a consonant-initial sound change, to maximize the likelihood that a mispronunciation would be detected, rather than differing only in one feature such as voicing (see Table 1). The amount of acoustic/phonetic overlap between the MP and target and distractor was not controlled, given the constraint that the stimuli be non-word minimal pairs, each beginning with a unique consonant.

Coding of training.

As in Experiment 1, the audio-video recordings from the training sessions were coded off-line. This was to identify trials on which participants gave incorrect responses during training, and the corresponding experimental test trials were removed prior to data analysis. A total of 49 trials were removed from the analyses (19 produced, 30 heard-only), which was on average 1.6 trials per participant.

Results

As in Experiment 1, participants' looks to the target were analyzed using a GCA performed in R. The empirical logit was calculated for each time bin as an approximation to log odds (Barr, 2008). The model contained Training (produced, heard-only) and Pronunciation (CP, MP) as the main predictors, which were both deviation-coded. The model thus included main effects of Training, Pronunciation, and Time (captured by orthogonal polynomials) and a Training*Pronunciation*Time interaction. Random by-participant effects were included in the model to account for by-participant variation in their slopes and intercepts.

A growth curve analysis was used to analyze looks to the target image in the same time window of 200 ms to 1200 ms after non-word onset as was used in Experiment 1. The overall time course of fixations to the target was modeled with a second-order (quadratic) orthogonal polynomial and fixed effects of Training (produced vs. heard-only, within-subjects variable), and Pronunciation (CP vs. MP, a within-subjects variable) on time. By-participant random effects were also included in the model (see Table 3 for all parameter estimates of the fixed effects). There was a statistically significant effect of Training (produced vs. heard-only) overall (*Estimate* = 0.58, *SE* = 0.21, $p < .01$), and a significant effect of Training on the linear term (*Estimate* = 0.99, *SE* = 0.51, $p = .05$), indicating more looking on produced compared with heard-only trials over time, and a steeper slope for Produced compared with heard-only trials. There was a statistically significant overall effect of Pronunciation (*Estimate* = -1.29, *SE* = 0.24, $p < .001$), and with Pronunciation on the linear term (*Estimate* = -2.01, *SE* = 0.36, $p < .001$), indicative of a more gradual slope for MP compared with CP non-words. That is, over the relevant time period the proportion of looks to the target grew faster for CP items compared with MP items (see Figure 3 for proportional looking data and Figure 4 for model predictions). There were no significant two-way interactions between Training and Pronunciation overall and no interactions with the orthogonal polynomials, indicating that the effect of Training was the same in both Pronunciation conditions (CP and MP).

---------- TABLE 3 ABOUT HERE----------

----------FIGURE 3 ABOUT HERE----------

----------FIGURE 4 ABOUT HERE----------

To determine whether the effect of Training was statistically significant in both Pronunciation conditions (CP and MP), two additional GCA analyses were performed to assess the effect of Training on the CP produced and CP heard-only trials, and then a separate analysis on only the MP produced and MP heard-only trials from Experiment 2 (Table 4). On CP trials, there was a near-significant effect of Training overall (*Estimate* = 0.59, *SE* = 0.31, *p* = .058) and a near-significant effect of Training on the linear term (*Estimate* = 0.93, *SE* = 0.48, *p* = .053), indicating that participants looked faster to CP produced targets compared with CP heard-only targets. On MP trials, there was a statistically significant effect of Training on the linear term (*Estimate* = 1.69, *SE* = 0.53, *p* < .01), indicating that participants looked faster to the target image in the MP produced trials than in the MP heard-only trials.

Discussion

In Experiment 2, participants were tested on their recognition of non-words that were either produced or heard-only during training, and on their recognition of non-words that were either correctly pronounced or mispronounced at test. As in Experiment 1, participants showed a different pattern of looking to targets that were produced compared with heard-only during training, with a faster rise in looking to produced targets. Moreover, participants in Experiment 2 also showed different abilities to detect mispronunciations of non-words, with more looking to mispronounced targets in the produced conditions compared with the heard-only condition.

General Discussion

In this research, adult participants were taught non-words which were either heard-then-produced (produced) or heard-only during training. At test, participants were tested on the recognition of these newly learned non-words. In Experiments 1 and 2, participants' proportion of looking to targets indicated that they successfully learned non-words that were either produced or heard-only during training. Despite this, participants looked faster to targets that were produced during training compared with targets that were heard-only during training, indicating that production during training influenced how participants recognized the newly learned words. Importantly, this was in spite of the fact that participants had been exposed an equal number of times to tokens of each class of stimulus. These results are consistent with previous findings in the literature on the production effect (Forrin et al., 2012; MacLeod et al. 2010). The results of the current study further show that the production effect is also found when comparing how targets are processed in a visual-choice task for which both images correspond to two previously produced or two previously heard-only non-words. Recall that in our design, participants were presented at test with two images corresponding to non-words that were either both produced during training or both heard-only during training, rather than two images where one was produced and the other one was heard-only during training. Put differently, in the current research, participants were faster at recognizing non-words that were produced during training in the presence of other non-words that were also produced during training. If the production effect stems solely from a distinct record in episodic memory for produced words, which makes them easier to retrieve from memory during a recognition task relative to words not produced, one would not expect to find a production effect in recognition between two produced words as seen in the current work.

Recognition in the current experiments was defined as participants looking to the target image, which required participants to map a non-word to a novel image. The experimental design does not give any reason to assume that there is any difference in the non-word to image mapping, that is independent of the strength of the non-word representation. Indeed, images in our design were equally well represented because the choice of a target at test is between two images of equivalent status. This leaves the strength of the non-word representation as the remaining candidate for any effect. However, it is possible that production during learning may itself influence the mapping from the non-word to the image in a manner that is somewhat independent of the strength of the non-word representation. For example, in the triangle arrangement described above, in which each of the perceptual, production, and semantic representations are positively connected to the other two, then activation of the perceptual representation will spread to activate the semantic representation by two routes: first, via the direct connection between perceptual and semantic representations; and, second, via activation of the production representation and onward activation of the semantic representation. This would result in faster and enhanced activation of the semantic representation under circumstances in which a productive representation has been established during learning. Furthermore, in this scenario, activation of connected productive and semantic representations will "back-activate" the perceptual representation, giving another reason to expect enhanced perceptual performance (i.e., in a recognition task) following an opportunity to establish a production representation during learning.

In Experiment 2, produced and heard-only non-words were either correctly pronounced (CP) or mispronounced (MP) at test. There was an overall advantage for produced over heard-only when taking CP and MP trials together (only CP trials reach near significance for produced versus heard-only, and only MP trials were significant for produced versus heard-only). For non-words that were mispronounced at test, there was a different effect of produced versus heard-only during training. When produced non-words were mispronounced at test, participants appeared to detect the mispronunciation initially, not looking strongly towards either image; however, they were able eventually to resolve the referent by looking to the target later in the trial, presumably by virtue of its (correctly pronounced) rhyme (see below). By contrast, when heard-only non-words were mispronounced at test, participants' looking seemed hardly to depart from chance performance over the course of the whole trial.

As a starting point, we indicated that production might serve a variety of functions, which are not necessarily mutually exclusive. However, the current findings give more insight into the nature of production representations and how they relate to linguistic representations and processing. Specifically, the current findings are consistent with exemplar theories in which phonetic details are stored together with semantic and other lexical features. This extra phonetic information stored for previously produced non-words leads to faster recognition for produced non-words, and more detailed representations from which a mispronunciation mismatches the signal. A key assumption here is the link between the production and perceptual systems (also see Dell et al., 2013; Hickok, 2012; 2014). The data are also consistent with certain non-episodic theories, such as that proposed by Page and Norris (2009) in relation to Hebb repetition learning and, by extension, word learning. In their framework, the production of a non-word, just like the recall of a list for immediate serial recall, links a production-based representation (akin to an output lexeme) to the input-based representation that itself results from only hearing a new word or an unfamiliar list. The linking of input and output representations during those early presentations accompanied by production (i.e., recall), allows for a strengthened recognition response in subsequent tests. The experiments presented here, while demonstrating and replicating an advantage for produced items over heard-only items, do not allow us to distinguish between these episodic and non-episodic explanations.

There are a number of avenues for future research to address the nature of production in word learning. To examine the robustness of the production effect, one manipulation would be to include testing over multiple days to determine if one condition leads to better retention of the new items, or whether there are different types of consolidation effects for produced versus heard-only words over time. Experiments 1 and 2 consisted of a short practice phase, immediately followed by the testing phase. This is similar to the experimental design used in the production effect literature (MacLeod et al., 2010); however, this is different from other psycholinguistic studies that used a delayed testing phase. For example Davis, Di Betta, Macdonald and Gaskell (2009) found different competition effects for newly learned non-words words tested one day after learning versus after immediate testing. Previous research with production has also shown varying learning rates for first-order versus second-order artificial phonotactic constraints, acquired through repeating strings of syllables. While first-order phonotactic constraints (typically restrictions on sounds in a specific position, e.g., /f/ does not occur in coda position) are learned relatively quickly (Dell, Reed, Adams, & Meyer, 2000), participants learn second-order phonotactic patterns (restriction dependent on another property, e.g., /k/ only occurs before the vowel /ɪ/) only after multiple days of exposure (Warker & Dell, 2006). Studies looking at long-term effects of production, would allow one to assess the impact of production on long-term representations, how representations are integrated into existing representations, and whether effects of production may be enhanced or degraded over time.

All of these manipulations can also be done comparing data from children at different stages of development with the adult data (e.g., Gathercole, Adams, & Hitch, 1994); such work is currently in progress. Recall that production does not always show beneficial or improved learning effect, and this depends on the linguistic characteristics of the stimuli (Kaushanskaya & Yoo, 2011). Effects of production are also expected to depend on the developmental stage of the learner and the difficulty of the task as seen in other domains of language development; these effects are modeled in PRIMIR (Processing Rich Information from Multi-dimensional Interactive Representations; Curtin, Byers-Heinlein, & Werker, 2011; Werker & Curtin, 2005). As children learn language, they spontaneously produce speech, and the impact of these productions has also been investigated in language learners (Curtin & Zamuner, 2014). In the current study, we manipulated whether participants produced the non-words during training. However, the more common approach in first language acquisition has been to manipulate the sound patterns of new words so that they either contain sounds within or outside of a specific child's production repertoire. Early in development, there is an advantage for words and non-words with sounds in the child's production repertoire (Keren-Portnoy, Vihman, DePaolis, Whitaker, & Williams, 2010; Leonard, Schwartz, Morris, & Chapman, 1981; Schwartz & Leonard, 1982; Schwartz, Leonard, Leob, & Swanson, 1987). Infants also show differential looking behavior in perception tasks to passages with non-words containing sounds within their production repertoire (DePaolis, Vihman, & Keren-

Portnoy, 2011; DePaolis, Vihman, & Nakai, 2013; Majorano, Vihman, & DePaolis, 2014). Similarly, young children are better able to recall and produce the plurality of non-words that contain sounds within their production repertoire (Ettlinger, Lanter & van Pay, 2014). It has been suggested that production experience can facilitate word learning by supporting the memory of those words, such that they require less processing resources to link words and their referents (Keren-Portnoy et al., 2010; Vihman, DePaolis, & Keren-Portnoy, 2014). Interestingly, our preliminary results indicate that children show the opposite effect to the patterns reported for adults in Experiment 1 and look more to the target on heard-only trials compared with produced trials. These results seem to indicate age-based and possibly task-based effects in the effect of production in development (Zamuner, Morin-Lessard, Strahm, & Page, 2015).

One final issue concerns what is assumed to be the correct referent when target is presented with a mispronunciation (MP). After all, the mispronounced non-words above did not correspond to any referent *per se*, even though our adult participants did seem able to resolve the reference, in produced conditions at least. Studies with children show that when they hear an auditory target 'vaby', they will consistently look at the image of a 'baby' over a distractor image 'dog' (Swingley & Aslin, 2000). However, it is less clear what adults consider an acceptable target given such mispronunciations. For instance, when participants in our studies were presented with a mispronounced target (e.g., /kɛl/) and the visual scene consisted of two images corresponding to trained non-words (e.g., /zɛl/ and /bos/), it may have been unclear to adult participants what the correct target should be, assuming that they realized that the mispronounced non-word did not specifically refer to any target in the visual scene. Recent work with bilingual adults has shown that they consider certain mispronunciations to be acceptable given the linguistic environment (Samuel & Larraza, 2015). In Experiment 2, the correct answer might be looking at the closest sounding target (i.e., /zɛl/, that at least shares a rhyme); alternatively, the appropriate response under these circumstances might be reflected by chance levels of looking to the target (i.e., looking equally at the two images). Moreover, the "correct" response might vary by participant. Based on previous research in this paradigm (e.g., McMurray, Tanenhaus, & Aslin, 2002; White & Morgan, 2008), one may also expect that looking would vary depending on the amount of acoustic/phonetic overlap between the mispronounced target (e.g., /kɛl/), the correctly pronounced target (e.g., /zɛl/) and the distractor (which might be something like /bos/ or /gʌb/). For example, upon hearing the initial /k/ in the mispronounced target /kɛl/, participants should look more at the distractor if it were /gʌb/, because /g/ matches in both place of articulation and manner of articulation with /k/, and participants should look less at the distractor if it were /bos/ because the /b/ does not overlap in place of articulation with /k/. Although the amount of acoustic/phonetic overlap across the mispronunciation between the target and distractor was not controlled in this study given other constraints on the stimuli, an analysis of the looking data (not represented here) did suggest that participants were somewhat sensitive to the degree of sub-phonemic overlap, though a balanced experimental design would be needed to address this issue directly.

Notwithstanding the issue of what is deemed to be the appropriate response in cases of mispronunciation, in Experiment 2 participants' looking to the target in the MP condition was different depending on whether it had been produced or heard-only during training. Our interpretation is that in the MP produced condition, participants were able to resolve the intended referent, probably by reference to its rhyme, as seen by their looking at the target at the end of the trial (see similar findings by Swingley & Aslin, 2000). In the MP heard-only condition, participants were not able to resolve the reference, indicated by approximately 50% looking to the target image throughout the trial. This is because of either weaker representations and/or due to slower access to, or processing of those representations. This further predicts that if participants are presented with a visual scene in which there are three referents (two previously named images and one novel and unnamed), participants will resolve the MP trials differently in the produced or heard-only conditions. Studies to test this possibility are currently ongoing.

Conclusion

Our experiments have shown that the spoken word recognition of newly learned words depends on how the non-words are learnt. Under the produced learning condition, participants were faster to look at targets than in the heard-only learning condition. Spoken word recognition also varied for how mispronounced items were recognized in the produced versus heard-only learning condition. When produced non-words were mispronounced at test, participants were eventually able to resolve the referent; however, when mispronunciations occurred on heard-only non-words, participants did not look consistently to the target or distractor as the word unfolded. These results are consistent with both exemplar models of speech perception and of production, and with other frameworks based on a structural distinction between linked representations separately based on perception and on production.

References

Baddeley, A. D. (1986). *Working Memory*. New York: Oxford University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29.

Baddeley A. D. & Hitch G. J. (1974). Working memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (47-89). New York: Elsevier

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158-173.

Bannard, C., Klinger, J., & Tomasello, M. (2013). How selective are 3-year-olds in imitating novel linguistic material? *Developmental Psychology, 49*, 2344-2356.

Barr, D. J., & Frank, A. F. (2009, March). Analyzing multinomial and time-series data. In *Workshop on ordinary and multilevel modeling at 2009 CUNY conference on sentence processing, UC Davis*.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of memory and language*, *59*(4), 457-474.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

Burgess, N. & Hitch, G. J. (2006) A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627–652.

Clark, E. V., & Bernicot, J. (2008). Repetition as ratification: How parents and children place information in common ground. *Journal of Child Language* 35, 349-371.

Cohen, R. L., & Johansson, B. S. (1967a). The activity trace in immediate memory: A re-evaluation. *Journal of Verbal Learning and Verbal Behavior, 6*, 139-143.

Cohen, R. L., & Johansson, B. S. (1967b). Some relevant factors in the transfer of material from short-term to long-term memory. *Quarterly Journal of Experimental Psychology, 19*(4), 300-308.

Cunningham, T. F., Healy, A. F., & Williams, D. M. (1984). Effects of repetition on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 575-597.

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics, 39,* 492-504.

Curtin, S., & Zamuner, T.S. (2014). Understanding the developing sound system: Interactions between sounds and words. *WIREs Cognitive Science*, *5*, 589-602.

Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Languages Journal, 97*(4), 902-916.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*(4), 803-820.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1355–1367.

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *3*, 380-396.

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants*? Infant Behaviour and Development*, *34*, 590-601.

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behaviour and Development*, *36*, 642-649.

Duff, P. (2000). Repetition in foreign language classroom interaction. In J. K. Hall & L. S. Verplaerse (Eds.), *The development of second and foreign language learning through classroom interaction* (109-138). Mahwah, NJ: Lawrence Erlbaum.

Ellis, A., & Young, A. (1988). Human cognitive neuropsychology. Hove, East Sussex: Lawrence Erlbaum Associates.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49a(1),* 234-250.

Ettlinger, M., Lanter, J., & Van Pay, C. K. (2014). Learning to remember by learning to speak. *Developmental Psychology*, *50*(2), 431-438.

Forrin, N. D., MacLeod, C. M. & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory and Cognition, 40*, 1046-1055.

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to better retention. *Memory and Cognition, 16,* 110-119.

Gathercole, S. E., Adams, A.-M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory and Cognition, 22*, 201-207.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177-190.

Gershkoff-Stowe, L. (2002). Object naming, vocabulary growth, and the development of word retrieval abilities. *Journal of Memory and Language, 46*, 665-687.

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

Goldrick, M., Baker, H.R., Murphy, A., & Baese-Berk, M. (2011). 2011 Interaction and representational integration: Evidence from speech errors. *Cognition*, *121*, 58-72.

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America, 134*, EL172-EL177.

Gupta, P. & MacWhinney, B. (1997) Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain and Language, 59*, 267–333.

Gupta, P. & Tisdale, J. (2009) Word learning, phonological short-term memory, phonotactic probability and long-term memory: towards an integrated framework. *Philosophical Transactions of the Royal Society B, 364*, 3755–3771.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. *Brain Mechanisms and Learning*, 37-46.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*, 135–145.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29*(1), 2-20.

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behaviour, 11*, 534-537.

Kaushanskaya, M. & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*, 121-148.

Keren-Portnoy, T., Vihman, M. M., DePaolis, R. A., Whitaker, C. J., & Williams, N. M. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research*, *53*, 1280-1293.

Kuhl, P. K., & Meltzoff, A. (1996). Infant vocalization in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100*, 2425-2438.

Leonard, L. B., Schwartz, R. G., Morris, B., & Chapman, K. (1981). Factors influencing early lexical acquisition: lexical orientation and phonological composition. *Child Development*, *52*, 882-887.

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*(6), 1197-1202.

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: Learning, Memory, and Cognition, 36*(3), 671-685.

Majerus, S., Poncelet, M., Elsen, B. & Van der Linden, M. (2006) Exploring the relationship between verbal short term memory for serial order and item information and new word learning in adults. *European Journal of Cognitive Psychology, 18*, 848–873.

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014) The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10,* 179-204.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372-380.

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

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75-78.

Mirman, D., Dixon, J.A., and Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475-494.

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *The Quarterly Journal of Experimental Psychology*, *61*(4), 505-514.

Oberauer, K., & Meyer, N. (2009). The contributions of encoding, retention, and recall to the Hebb effect. *Memory, 17*(7), 774–781.

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1536), 3737–3753.

Page, M. P. A., Cumming, N., Norris, D., Hitch, G. J. & McNeil, A. M. (submitted). Exploring the importance of list recall in the Hebb repetition effect.

Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*, 506-526.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (137-157). Amsterdam: John Benjamins Publishing Company.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology 7*, 101-140.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46(2-3),* 115-154.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145-163.

Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception?. *Journal of Memory and Language*, *81*, 51-71.

Schwartz, R. G., & Leonard, L. B. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, *9*, 319-336.

Schwartz, R. G., Leonard, L. B., Loeb, D. M. F., & Swanson, L. A. (1987). Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance. *Journal of Child Language*, *14*, 411-418.

Sosa, A. V., & Bybee, J. (2008). A cognitive approach to clinical phonology. In M. J. Ball, M. R. Perkins, N. Miller, & S. Howard (Eds.), *The Handbook of Clinical Linguistics*. (480-490) Oxford UK: Blackwell Publishing Ltd.

Swingley, D., & Aslin, R.N. (2000) Spoken word recognition and lexical representation in very young children. *Cognition, 76,* 147-166.

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Quarterly Journal of Experimental Psychology, 62*(3), 435–443.

Szmalec, A., Page, M. P. A., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language, 67*(3), 342–354.

Tchernichovski, O., Mitra, P.P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science, 291*, 2559-2561.

Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, *64*, 121-140.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 387.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*, 197-234.

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*, 114-132.

Zamuner, T.S., Morin-Lessard, E., Strahm, S., & Page, M. (2015, June). *Developmental differences in the effect of production on word-learning*. Paper presented at the International Child Phonology Conference, St. John's Newfoundland.

Highlights

- We investigate the effect of production on novel word learning.
- Participants learned new words that were either produced or heard-only during training.
- Adults were faster to recognize new words that were produced during training.
- Mispronunciations were recognized differently depending on training condition.

Table 1. Non-word stimuli in IPA

Set 1	Set 2
/kɛl/	/zɛl/
/vup/	/tup/
/nɪs/	/wɪs/
/rem/	/pem/
/dæs/	/hæs/
/fʌb/	/gʌb/
/mig/	/jig/
/los/	/bos/

Table 2. Experiment 1 GCA results. Empirical logit GCA for the effect of Training on looking data for a window of analysis 200-1200ms after the target non-word onset. The values in each cell represent parameter estimates.

Predictor	Estimate	*SE*	*t*	*p*
Intercept	1.82	0.28	6.65	<.001
Training (produced vs. heard-only)	0.83	0.47	1.78	.08
Linear Term	2.59	0.58	4.46	<.001
Quadratic Term	-1.02	0.30	-3.37	<.001
Training * Linear Term	0.20	0.55	0.38	.71
Training * Quadratic Term	1.13	0.54	2.07	.04

Note. SE = standard error. GCA structure in R: lmer(elog ~ Training * (Linear + Quadratic) + (1 + Training + Linear + Quadratic | Participants). *p* = *p*-values, calculated in R using a normal approximation, based on the assumption that the *t*-distribution converges to the *z*-distribution.

Table 3. Experiment 2 GCA results. Empirical logit GCA for the effect of Training and Pronunciation on the looking data, for a window of analysis 200-1200 ms after the target non-word onset. The values in each cell represent parameter estimates.

Predictor	Estimate	*SE*	*t*	*p*
Intercept	1.55	0.22	7.07	<.001
Training (produced vs. heard-only)	0.58	0.21	2.72	<.01
Pronunciation (CP vs. MP)	-1.29	0.24	-5.33	<.001
Linear Term	3.04	0.33	9.21	<.001
Quadratic Term	-0.59	0.34	-1.69	.09
Training * Pronunciation	-.0.76	0.22	-1.19	.23
Training: Linear Term	0.99	0.51	1.97	.05
Training: Quadratic Term	0.05	0.51	0.10	.92
Pronunciation * Linear Term	-2.01	0.36	-5.72	<.001
Pronunciation * Quadratic Term	0.20	0.36	0.56	.58
Training * Pronunciation * Linear Term	0.80	0.71	1.11	.27
Training * Pronunciation * Quadratic Term	-0.48	0.72	-0.67	.51

Note. SE = standard error. GCA structure in R: model <- lmer(elog ~ Training * Pronunciation * (Linear + Quadratic) + (1 + Training + Pronunciation + Linear + Quadratic | Participants). *p* = *p*-values (calculated in R as described under Table 2).

Figure 1. Training for produced and heard-only trials.

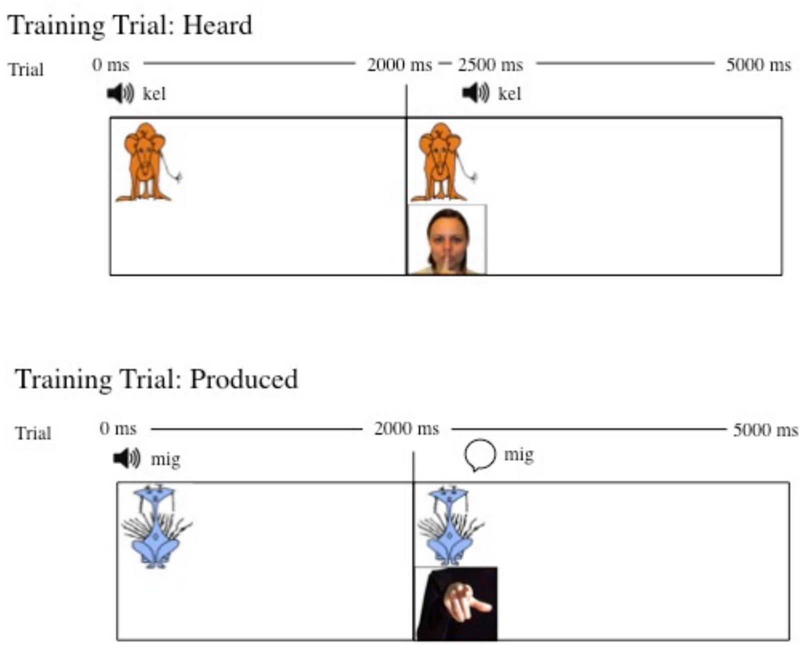

Figure 2. Experiment 1 proportion of looking graph and model predictions. (Left) Time course of proportion of looking to the target after non-word onset for non-word from produced and heard-only training conditions. The horizontal dotted line corresponds to chance (0.50), and the vertical line throughout each point represents the standard error of the mean. (Right) Model predictions for Training.

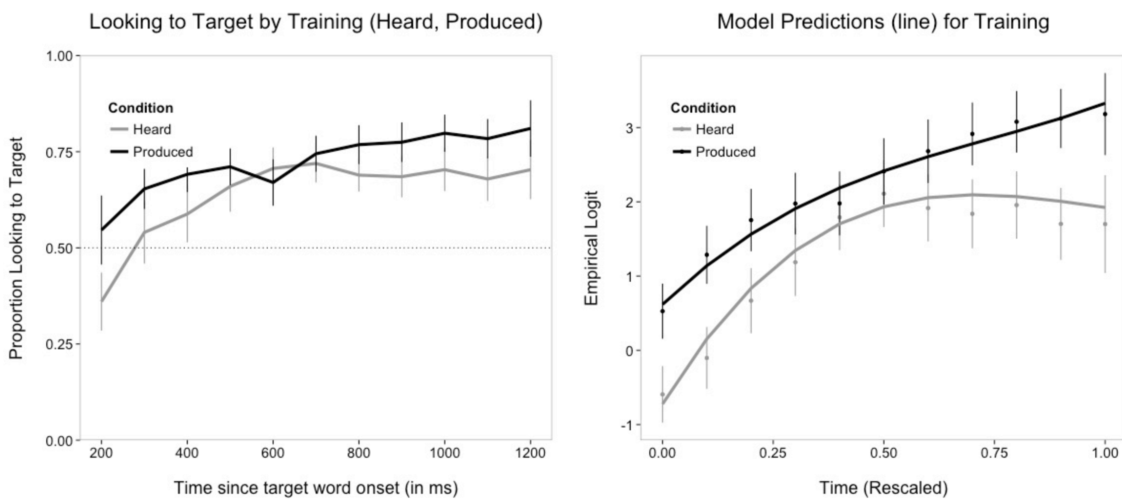

Figure 3. Experiment 2 proportion of looking graphs. Time course of proportion of looking to the target after onset for non-words produced (Left) and for non-words heard-only (Right) at training by Pronunciation (CP, MP). In all graphs displayed, the horizontal dotted line corresponds to chance, and the vertical line throughout each point represents the standard error of the mean.

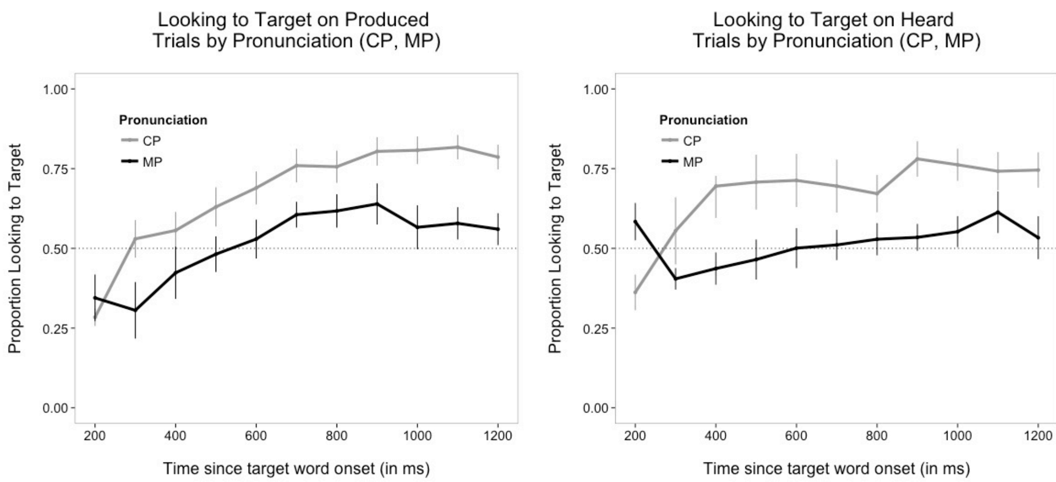

Figure 4. Experiment 2 overall model prediction graphs for Training and Pronunciation. (Left) Model predictions for Training. (Right) Model predictions for Pronunciation.