Published in final edited form as:

Cogn Psychol. 2007 December ; 55(4): 306–353. doi:10.1016/j.cogpsych.2007.01.001.

Lexical Configuration and Lexical Engagement: When Adults Learn New Words

Laura Leach and **Arthur G. Samuel**

SUNY Stony Brook

Abstract

People know thousands of words in their native language, and each of these words must be learned at some time in the person's lifetime. A large number of these words will be learned when the person is an adult, reflecting the fact that the mental lexicon is continuously changing. We explore how new words get added to the mental lexicon, and provide empirical support for a theoretical distinction between what we call lexical *configuration* and lexical *engagement*. Lexical configuration is the set of factual knowledge associated with a word (e.g., the word's sound, spelling, meaning, or syntactic role). Almost all previous research on word learning has focused on this aspect. However it is also critical to understand the process by which a word becomes capable of lexical engagement – the ways in which a lexical entry dynamically interacts with other lexical entries, and with sublexical representations. For example, lexical entries compete with each other during word recognition (inhibition within the lexical level), and they also support the activation of their constituents (top-down lexical-phonemic facilitation, and lexically-based perceptual learning). We systematically vary the learning conditions for new words, and use separate measures of lexical configuration and engagement. Several surprising dissociations in behavior demonstrate the importance of the theoretical distinction between configuration and engagement.

Keywords

Lexical configuration; lexical engagement; learning new words

According to Herodotus, the experimental study of lexical development can be traced to an Egyptian Pharaoh, Psammetichus. In an effort to discover which language was innate, the Pharaoh arranged for two newborns to be raised without any spoken language input. When their first spoken word was reported to be “bekos”, the Pharaoh concluded that Phrygian was in fact the pre-packaged language of humans, because “bekos” referred to a kind of bread in Phrygian.

Although one must admire Psammetichus for his willingness to report an experimental result that ran counter to his hypothesis (he of course had predicted that Egyptian was the natural language of humans), a more plausible view is that each human acquires the words of whatever language(s) the environment presents. There is now a substantial literature that tracks the first words learned by very young children, and the development of increasingly long utterances.

Address correspondence to: Arthur G. Samuel, Dept of Psychology, SUNY Stony Brook, Stony Brook, NY 11794-2500, asamuel@ms.cc.sunysb.edu.

In the current study, we are concerned with much later lexical development: The addition of new words to the lexicon of an adolescent or adult language user. According to Nation and Waring (1997), this is a surprisingly frequent event. They estimate that people add about a thousand word forms (a word and its close morphological relatives) per year, up to an asymptote of about 20,000 word forms. This translates to about three word forms per day, every day of the year, which entails more lexical change than one might expect.

Adding a word to the mental lexicon is presumably at least partially incremental, given all that is potentially involved in a fully specified lexical entry. At a minimum, such an entry should include a word's input form (phonological and orthographic), its meaning (which will typically be multidimensional and often context-specific), and its syntactic role(s). In most cases, the language user will get some of this information initially, and more of it over time. For our current purposes, we will refer to all of this information as the “Configuration” of a lexical entry. The lexical configuration is the set of factual information that one knows about a word. As we have noted, this information develops over a rather long time course, perhaps measured in weeks, months, or even years. From this perspective, at any particular moment, lexical configuration can be thought of as relatively static.

We can distinguish lexical configuration from what we will call lexical “Engagement”. By lexical engagement, we mean how the lexical representation behaves dynamically, over a much shorter time scale. For example, most current theories of word recognition talk about the activation of lexical entries by input signals, and the effect that such activation can have on other lexical entries, and on other types of representations. Perhaps the most established example of one type of lexical engagement is semantic priming: Activating the representation for “doctor” leads to an increase in the activation of another lexical representation, the one for “nurse” (Meyer & Schvaneveldt, 1971). There are also hypothesized inhibitory or competitive functions of lexical representations (e.g., if “cat” is activated, it can compete with and possibly inhibit “cap”) in most theories, beginning with Cohort theory (e.g., Marslen-Wilson, 1990), and including models such as Trace (McClelland & Elman, 1986), Shortlist (Norris, 1994), and NAM (Luce & Pisoni, 1998). Interactive models such as Trace also posit lexical influences on sublexical representations, with activation of a lexical entry (e.g., “cat”) producing increased activation of its component phonemes (e.g., /k/).

We believe that in order to understand the lexicon and its development, it is important to keep the distinction between lexical configuration and lexical engagement clear. In examining the existing literature on how people add words to the lexicon, we found that it was difficult to discern a coherent story, because different studies seemed to tap into different lexical properties, using different experimental techniques. Thus, the goal of the current study is to take a more systematic approach, both theoretically and experimentally. We will first attempt to establish plausible measures of both lexical configuration and of lexical engagement, so that in the following experiments we can see how each of these aspects of lexical representations may develop. With these measures in hand, we will then examine three different learning procedures for new words. Finally, we then bring together language production and language perception, to see the effect of their interaction on new word learning. This approach allows us to provide a much more cohesive understanding of how new lexical representations get established than has been available previously.

Lexical Configuration

It is instructive to examine some of the prior studies of new word learning to see how each has given us bits of the big picture. For example, Salasoo, Shiffrin, and Feustal (1985) conducted an interesting and thorough examination of new word learning in the visual modality. Each training trial consisted of a word (e.g., “pelican”) or nonword (e.g., “shargat”) briefly presented on a computer monitor, immediately followed by a mask; each stimulus was presented thirty times during the course of training. The task was simply to report what was presented. In several experiments, participants either typed in their response or orally identified the word. In such a test of perceptual identification under challenging sensory conditions, Salasoo et al. predicted words should be named more quickly and more accurately than nonwords. Initially this was in fact the case, but after just six presentations, participants' responses to nonwords matched their speed and accuracy of the real words, suggesting that lexical representations were developed rather quickly. These representations were apparently quite well developed, because they produced comparable results in a test conducted a year later, without intervening training. From our perspective, this is an excellent example of lexical configuration development, because the orthographic information was learned and available for recognizing the words; there is no information provided by this study about the possible development of lexical engagement, as there was no measure of more dynamic effects of these representations on any other representations (i.e., we have no evidence that the activation of “shargat” had any effect on anything other than its own lexical representation).

There are a number of studies of spoken word acquisition that share many properties with the Salasoo et al. approach, with a number of differences as well. For example, Storkel (2001) examined how children (ages three to six years old) acquired new lexical entries as a function of repetition and semantic reference. Children heard new words in the context of a paragraph, along with a drawing depicting the story. The new words were repeated in the story from one to seven times. Storkel asked the children to identify the new word by selecting from three recorded nonwords, and by naming the word when a picture of it was presented. The participants' performance increased through such training, though not as much as in the Salasoo et al. study. Note that Storkel's study differed in two important ways. First, the new words were given a semantic referent (both through the story context, and the drawings). Second, the representations were probed with two different but related tasks (naming, and forced choice recognition). Nonetheless, this study is still only assessing lexical configuration, as both tasks call for retrieval of an individual lexical item's phonological codes.

Gupta's (2003; Gupta, Lipinski, & Abbs, 2004) examination of new word learning also falls into this category. Gupta's primary interest was in the relationship between learning new words and measures of memory span, since learning a new word can be viewed as learning an ordered series of sounds or letters. For the current purpose, the similarities to Storkel's (2001) study are most relevant: The (college-age) participants were taught names for novel animal pictures (Experiment 1) or drawings of “aliens” (Experiment 2), by pairing the names and the pictures. For the aliens, participants also learned various semantic properties (e.g., each alien's favorite food). Later testing showed that the participants could both produce the name, given a picture, and could provide the relevant semantic information. These are the sorts of factual properties that we have categorized as lexical configuration information.

Lexical Engagement

The Salasoo et al. (1985), Storkel (2001), and Gupta (2003) studies are all well designed investigations of the acquisition of new lexical representations. They are also typical of most studies in this literature, in that they assess lexical development by relatively direct measures of “factual” information, such as the ability to report a word given a brief orthographic input or a picture, or the ability to retrieve information associated with a newly learned word. These studies were not designed to assess the development of more dynamic lexical engagement.

There are two recent papers that have begun the investigation of lexical engagement. Magnuson, Tanenhaus, Aslin, and Dahan (2003) trained participants on sets of either 16 disyllabic new words (Experiments 1 and 2), or 20 monosyllabic ones (Experiment 3). Each word was paired with a meaningless visual pattern formed by randomly filling eight contiguous cells in a 4×4 grid. Participants were taught to associate each word with a particular pattern by presenting two (or later, four) of the patterns visually, and providing feedback after the participant selected the pattern that matched a given word. In the experiments of primary interest, the new words came in sets of four phonologically similar items (e.g., /pibo/, /pibu/, /dibo/, and /dibu/). In the later phases of the experiment, Magnuson et al. used an eyetracker to determine which visual patterns were being considered as a participant heard a word. The central theoretical question was whether the newly learned words would show a form of lexical competition typically found with normal words: Would listeners look at the "pibu" relatively often, when they were hearing "pibo"? Note that lexical competition is an example of lexical engagement, as it potentially involves the activation of one lexical item affecting the activation of another one (competitively). Magnuson et al. did in fact find that newly learned words showed looking patterns comparable to those for more developed lexical entries.

However, it is not clear that these results tell us about "normal" lexical development. A central goal for Magnuson et al. (2003) was the establishment of an encapsulated artificial lexicon, because such a small-scale system could be used to test various hypotheses without the immense messiness of a real lexicon. They demonstrated that their procedures had succeeded in this respect, as the new words did not seem to be affected by properties of the normal English lexicon (e.g. the presence or absence of many similar words in English). Presumably, the use of organized item sets, and the general experimental procedures, helped to seal off the newly learned words from the existing lexicon. Although this was methodologically attractive, it potentially limits the conclusions one can draw about lexical engagement, as the latter is by definition based on interactions among representations in the lexicon.

Gaskell and Dumay (2003) took an approach to studying new word learning that was essentially the opposite of Magnuson et al.'s (2003): Rather than try to design conditions that would create an encapsulated artificial lexicon, Gaskell and Dumay created conditions that maximally linked new words to existing ones. For example, for the real word "cathedral", Gaskell and Dumay created the new word "cathedruke". Over the course of five days, participants were repeatedly exposed to such nonwords in the context of a phoneme monitoring task. In this task, on each trial participants were given a target (e.g., "p") and then heard the nonword, and judged whether the nonword contained the sound. This task was purely a vehicle for exposure to the items. Each day, the participants also completed a lexical decision task that included real words (e.g., "cathedral") that were similar to the newly learned nonwords (e.g., "cathedruke"). If a functional lexical entry for "cathedruke" has been created, it should compete with the entry for "cathedral" in a lexical decision task, slowing responses to such similar words (compared to controls without new competitors). By the third day of training (i.e., after approximately 30 learning trials for each new word), Gaskell and Dumay found exactly this pattern. This is the first clear evidence for the emergence of lexical engagement. Although there were of course many differences between the stimuli and procedures of Gaskell and Dumay's study and those of Salasoo et al. (1985), it may be instructive to note the much longer learning period (30 exposures) needed to see evidence of lexical engagement than that needed for the emergence of lexical configuration (6 exposures).

The Current Study

The existing literature thus provides some information about the constraints and time course of learning the configuration of a new word (Gupta et al., 2003; Salasoo et al. 1985; Storkel, 2001), and some preliminary results about lexical engagement (Magnuson et al., 2003; Gaskell & Dumay, 2003). However, each study provides a rather isolated look at some aspect of lexical development, without a systematic comparison of configuration versus engagement, or of the effects of different word learning conditions. The current study is designed to provide a much more comprehensive and systematic analysis of these two factors. Across experiments, we examine different learning procedures. Within each experiment, we compare measures of configuration learning to measures of lexical engagement. We begin by adopting Gaskell and Dumay's training regime (exposure through phoneme monitoring). In Experiment 1, we assess lexical development with four tasks: Two were chosen to be natural measures of lexical configuration, and two were appropriate measures of lexical engagement. In addition to providing the first comparison of the development of lexical configuration versus lexical engagement, Experiment 1 also serves a methodological purpose. The results will be used to select one good measure for each aspect of lexical representations, which will allow us to then compare how different learning conditions affect the development of lexical configuration and lexical engagement (Experiments 2-5).

EXPERIMENT 1

Gaskell and Dumay's (2003) stimuli were constrained by their measure of lexical engagement – lexical competition. This measure required them to study new words that were close variations of existing words (e.g., cathedruke-cathedral). Magnuson et al.'s (2003) stimulus set was also strongly affected by their desire to study shared onsets (e.g., /pibo/-/pibu/) and rimes (e.g., /pibo/-/dibo/). In the current study, our measurements did not place such strong constraints on our stimulus selection. We therefore created new words that were not tied to particular existing words, and that were not systematically related to each other, which is probably more typical of normal lexical additions. Because previous research has shown that relatively long words generally produce stronger lexical effects in various tasks (e.g., Davis, Marslen-Wilson, & Gaskell, 2002; Pitt & Samuel, 2006), our stimulus set consisted of three- and four-syllable words. Table 1 shows the twelve critical items. As the table shows, six of the words included an /s/ (e.g., “gatersy”), and six included an /ʃ/ (e.g., “wickoshah”). This contrast was needed for one of the measures of lexical engagement (see below).

Participants in Experiment 1 each learned all twelve new words. The learning took place in the context of making many phoneme monitoring judgments for each word over the course of five days, just as in Gaskell and Dumay's (2003) study. On each day, participants also did two other tasks, one of which was intended to measure the current state of lexical configuration, and the other of which was designed to assess lexical engagement. By using both types of measurement each day, we can see whether the two aspects of lexical representations develop together, or separately.

As we have noted, a methodological goal of Experiment 1 is to help us select a good measure of lexical configuration, and a good measure of lexical engagement. We chose two tasks that by their nature seem to primarily tap lexical configuration, and two which should depend on lexical engagement:

Lexical Configuration Measures—One task used to assess the state of lexical configuration was a three alternative recognition judgment (Storkel, 2001; see also Gaskell & Dumay, 2003, for a two-alternative version). For each of the critical words, we recorded two very similar items, and we tested how well a word was learned by a participant's ability to identify which word was the one that had been heard during training. For example, for the critical word “penivasher”, participants would hear “penivasher”, “penikasher”, and “penivawer”, and they would have to indicate which one of these three stimuli was the one that they had heard previously. This task clearly depends on the establishment of the kind of factual information (in this case, primarily phonological form) that defines lexical configuration.

The second measure of lexical configuration was a “word-in-noise” task, a type of perceptual identification test (Salasoo et al., 1985). The critical words (and many filler words and nonwords) were initially presented to participants buried under a high level of white noise. The stimulus was then played repeatedly, with each iteration including lower and lower amplitude noise. The participant was told to hit a button to stop the presentations when the speech could be identified (the participant was then required to write down what the word or nonword was). Pilot work shows that participants can identify the speech sooner (i.e., at higher noise levels) when a real word is presented than when a nonword is the target. To the extent that a word has developed a strong lexical configuration, it should be recognized at higher noise levels, much as Salasoo et al. showed for visually-presented stimuli.

Lexical Engagement Measures—The defining feature of lexical engagement is the ability of a lexical representation to affect the activation of another lexical representation, or the activation of a sublexical representation. The phonemic restoration effect seems to be a good example of such lexical functioning. Phonemic restoration occurs when part of an utterance is removed, and is replaced by an extraneous sound. In the initial report of this phenomenon, Warren (1970) recorded a sentence, and then spliced out a phoneme from one of the words, replacing that segment with the sound of a cough. When the result was played to listeners, they consistently reported that none of the speech was missing; they heard the cough as co-occurring with an intact utterance. In a series of experiments, Samuel (1981a, 1996) has shown that the strength of this effect depends on the amount of lexical support: When the replacement occurs within an utterance that should provide strong lexical activation, the restoration of missing phonemic information is strongest. Phonemic restoration thus appears to involve a lexical representation's activation of a sublexical phonemic representation. As such, it is a good example of lexical engagement. We therefore tested the ability of each of the trained words to support phonemic restoration. To the extent that the training enables lexical engagement, the words should be better able to support restoration of missing phonemes.

Norris, McQueen, and Cutler (2003) recently reported another perceptual effect that also seems to depend on lexical engagement – lexically-driven perceptual learning. Participants in the Norris et al. study first did a lexical decision task in which they categorized 100 words and 100 pseudowords. Among the 100 words were 20 critical items. For half of the participants, the critical items were words that end in /f/ (e.g., sheriff); for the other half, the critical items ended in /s/ (e.g., Paris). Norris et al. created a sound that was midway between /f/ and /s/ by mixing the two sounds together. This ambiguous sound replaced all of the final /f/ or final /s/ sounds in the critical items. After the lexical decision task, participants identified items from an /Ef/-/Es/ continuum, labeling each syllable as either /f/ or /s/. Norris et al. found that exposure to the ambiguous sound produced substantial changes in how listeners identified the test syllables. Participants who heard the sound in a lexical context that called for /s/ expanded their categorization of /s/; the reverse was true for those who had /f/ context. The results show that lexical items can cause substantial changes in the definition of sublexical categories, a clear case of lexical engagement (no such perceptual learning occurred when the ambiguous sound was placed in nonwords). Thus, we can use the ability of our trained words to produce perceptual learning as an index of the development of lexical engagement.

METHOD

Participants—Forty Stony Brook University students participated. Eight participants were replaced after Day 1, based on their failure to reach criterion performance on the phoneme monitoring test (85% correct). All were native speakers of American English and reported no hearing problems. They were either given credit toward a requirement in a psychology course or paid for their participation. No individual participated in more than one experiment in this study.

Apparatus—Participants were tested in a sound-shielded booth that accommodated three individual testing stations. Responses were entered on a 4-button response board, with appropriate labels on the buttons used in each task. The experiment was controlled by a Pentium PC, on which the responses were stored for later analysis.

Stimulus recording was done in a sound shielded booth, using a high quality microphone. The speech was digitized on a Pentium PC (12 bit analog-to-digital conversion, 16 kHz sample rate). Individual tokens were edited from the original digital recordings using Goldwave sound processing software. Stimuli were output through a digital to analog converter (12 bit; 16 kHz rate), low-pass filtered (7.8 kHz), amplified, and presented binaurally over high quality stereo headphones.

Design—Each participant was given a phoneme monitoring training task every day for the first four days. On all five days, each participant was given one test of configuration (either words-in-noise or three-alternative recognition) and one test of engagement (either phonemic restoration or perceptual learning). There were four groups of participants, incorporating the 2×2 combination of the two configuration tasks with the two engagement tasks. Within each group, order of the two tasks was counterbalanced across participants. On Day 1, the phoneme monitoring training preceded the two test tasks. On Days 2-4, the training came after the two tests. This provided a way to look for any change in performance between Days 1 and 2 due to consolidation without any further training (cf. Dumay & Gaskell, 2006; Gaskell & Dumay, 2003; Storkel, 2001). However, because there were no interesting consolidation effects found here, this manipulation will not be considered further. On Day 5, there was only testing, not training.

In the phoneme monitoring training task, the participants received six words (e.g., marfeshick) with /ʃ/ sounds, and six words (e.g. “figondalis”) with /s/ sounds, as shown in Table 1. In the testing phase, the twelve critical items were divided such that each participant heard six of the items in a configuration testing task, and the other six items in a task measuring lexical engagement. The tasks and the stimuli used in them are described below.

Phoneme Monitoring (training): In the phoneme monitoring training task, listeners responded to the nonword items presented one at a time. Each day the twelve critical items were presented a total of 24 times each, with a new randomization on each pass. A target letter was displayed on a computer screen for 1500 msec before the playing of each nonword. The participant was given three seconds to respond. Participants were asked to press a button labeled “yes” if the target sound was present anywhere in the stimulus they heard, or to press a button labeled “no” if the target was not present. We instructed them to respond as quickly and as accurately as possible. Latencies were recorded at button push, measured from the onset of the stimulus. The order of the trials was randomized for each participant.

Target letters were chosen so that each nonword had twelve target present and twelve target absent trials. For each, three target present target letters were repeated four times, and three target absent target letters were repeated four times. Target letters that were target present trials for some nonwords were target absent trials for other nonwords. All targets appeared in three different nonwords in both target present and target absent cases.

Three-Alternative Recognition: For each critical nonword we recorded two foils that differed from the original nonword by one phoneme. Table 2 presents the stimuli used in this task. The three items (critical nonword plus two foils) were played in a random order. Buttons on a response board were labeled with “1”, “2”, and “3”. Participants were instructed to identify which of the three items they had heard in the training task by pressing the corresponding button.

Threshold Discrimination: Words-in-Noise: Sixty test items were presented in a random order: six of the trained nonwords, thirty word fillers, and twenty four nonword fillers. Each stimulus was played with a mask of white noise completely covering the item. The item was then played repeatedly; with each repetition we reduced the noise level by ten percent. Participants were instructed to press a button at the earliest point that they could recognize the word or nonword. They then wrote down the item on an answer sheet. We recorded the level of noise at which the participant responded, and scored the written answers for accuracy.

A response was coded as correct if it was within one phoneme of the original stimulus. For example, if a participant responded with “fibershack” and the correct word was “bibershack”, this was coded as correct, whereas a response of “fiberdack” would be incorrect.

Phonemic Restoration: The stimuli included six of the trained nonwords, thirty six filler words and thirty six filler nonwords. The test followed the signal detection methodology used by Samuel (1981a). Critical phonemes (primarily fricatives) were selected in each word or nonword, chosen to be near the middle or end of the item. Half of the items were played with noise replacing the selected phoneme, and half of the items were played with noise merely added on top of it. On each trial, the participant first heard a clear version of the item and then a version containing (added or replacing) noise. The task was to discriminate between the added and the replaced versions by selecting “added” or “replaced” on the button board. Strong lexical support for a missing phoneme produces strong perceptual restoration, and in such cases the “replaced” stimulus should be perceived as intact, yielding poor discriminability.

Perceptual Learning: As in Norris et al.'s (2003) perceptual learning study, a categorization task followed an initial exposure task. Their exposure task was lexical decision, which we did not use here because we did not want to ask our participants to make any explicit judgments about the lexical status of the words they were learning. Instead, we used an old-new recognition task. Participants first heard a list of twenty items – ten words and ten nonwords, and were told that a second list would follow. Participants were to call any item on the second list “old” if it had been played in the first list; otherwise, the correct response was “new”.

The first list included six of the critical new words, but with one sound slightly mispronounced (see below). For half of the participants, the six critical items were new words that had been presented during training with an /s/; for the other half, the six critical items were based on new words that had had /ʃ/. Each list was balanced in the sense that if it had six critical /s/ words, there were also six (real, untrained) words that had an /ʃ/; if the critical items were based on /ʃ/, the balancing words had /s/.

We used the second list to expose participants to slightly mispronounced versions of six of the words that they were being taught. For half of the participants, the mispronunciations were of /s/; for the other half, the mispronunciations were of /ʃ/. For example, one group of participants heard the critical nonword “gatersy” during the training phase, and heard an ambiguous version of that nonword (“gater?y”) in the second list. All mispronunciations were generated by mixing an /s/ and an /ʃ/ taken from recordings of the critical item (e.g., “gatersy” and “gatershy”), and splicing the ambiguous mixture in place of the original /s/ or /ʃ/. As noted above, each list was balanced by including an equal number of (real) words with the opposite sound.

Participants were exposed to the six critical items (and the six balancing real word items) three times each, for a total of eighteen mispronunciations of either /s/ or /ʃ/. This total was designed to be similar to the twenty critical items used by Norris et al. (2003). The critical items were randomly mixed in with sixty filler items (thirty words and thirty nonwords) that did not contain either an /s/ or an /ʃ/. All filler items were presented twice over the course of the second list.

After completing the old/new task, participants categorized members of a seven-step /asa-aʃa/ continuum. Participants identified each sound as either an /s/ or an /ʃ/ by pressing one of two labeled keys on a button board. Ten different randomizations of the seven sounds were presented. The /s/-/ʃ/ continuum was created by first recording clear tokens of /asa/ and /aʃa/, and then creating various mixtures (a weighted average of the waveforms) of the /s/ and /ʃ/ sounds. The sounds were digitally mixed in 5% increments, e.g., 5% /s/ mixed with 95% /ʃ/, 10% /s/ mixed with 90% /ʃ/, etc. Based on pilot testing, five steps were selected: 85% /s/ mixed with 15% /ʃ/, 75% /s/+ 25% /ʃ/, 65% /s/ + 35% /ʃ/, 55% /s/ + 45 % /ʃ/, and 45% /s/ + 55% /ʃ/. The original /asa/ and the original /aʃa/ were included as the endpoints of this seven-step series.

RESULTS

Training Task

Phoneme Monitoring: Figure 1 shows average reaction times and accuracy rates across the four days of training. An ANOVA revealed that there was a significant difference in response times across the four days, $F(3,117) = 135.90$, $p < 0.001$. Accuracy also improved across the four days, $F(3,117) = 20.07$, $p < 0.001$. The training data indicate that participants were learning the nonwords well enough to know the sounds within them.

Lexical Configuration Measures

Recognition: The responses for the recognition task showed that participants were moderately accurate on Day 1 (56%) and reached 71% accuracy by Day 5 (see Figure 2). This trend was marginally significant, $F(4,76)=2.25$, $p=.07$. The recognition foils were very similar to the newly-learned words, and the developing representations are apparently only moderately detailed: They are sufficient to generate performance well above chance (33%) but not yet precise enough for extremely fine distinctions.

Threshold Discrimination: The other configuration test, threshold discrimination, produced more systematic results. Figure 3 shows average accuracy rates and noise thresholds (given in the percentage of the signal that was noise, rather than speech). Both the accuracy and noise threshold improved systematically across the five days. Accuracy on Day 1 averaged 69%, and by Day 5 performance reached 94%, $F(4,76) = 16.02$, $p<.001$. Listeners were also able to recognize the trained items in increasingly louder noise across the five days, $F(4,76) = 20.01$, $p< 0.001$. In conjunction with the recognition and phoneme monitoring results, these data show that lexical configuration can be established through the type of training used in Experiment 1.

Lexical Engagement Measures

Phoneme Restoration: We conducted a one-way ANOVA, examining the differences in the participants' ability to distinguish between added and replaced versions of the critical nonwords. Participants were not able to successfully tell the difference between added and replaced stimuli for these critical nonwords, $F < 1$. The task was clearly not useful in measuring lexical engagement, due to the essentially chance level of performance. In retrospect, it is clear that using mostly fricatives as the critical phonemes made the task much too difficult (Samuel, 1981b).

Perceptual Learning: For each participant, we calculated the average choice of “sh” for each of the seven items on the /asa/-/aʃa/ test series. Figure 4 shows the group average each day, broken down by whether the Old-New exposure task should have increase “sh” report (solid lines) or “s” report (dashed lines). As Figure 4 shows, on each day there was a small trend towards a perceptual learning effect, with the solid curve consistently lying above the dashed one.

For statistical analysis, we focused on the middle three members of the test continuum (Samuel, 1986; Bertelson, Vroomen, & de Gelder, 2003), where any effects would be most evident. For each subject, we calculated the average percentage of "sh" choices for these three items, and then compared these averages in an ANOVA with Day of testing as one factor, and Exposure condition (favoring "sh" vs "s" perceptual learning) as the other. Despite the consistent 4-5% shift in the direction favored by perceptual learning, Exposure condition did not produce a reliable effect (F<1), and there was no interaction of the two factors (F<1). The consistency of the pattern suggests that there might be a small amount of perceptual learning going on, but clearly any such effect is too weak to imply any significant engagement by the representations for the words being learned.

DISCUSSION

During the training phase, phoneme monitoring accuracy was close to ceiling, and the participants grew faster across the four days of training. The steady increase in speed and accuracy suggests that the training regime accomplished the goal of giving listeners many attended-to exposures to the novel words.

The impact of such exposure could clearly be seen on the tasks that measured lexical configuration as a function of training. The results of the recognition task showed that participants were able to recognize the words at levels well above chance, improving from Day 1 (56%) to Day 5 (71%). The results of the threshold discrimination task were more systematic: There were large and significant performance gains across days of training. The difference in sensitivity between the two tasks may be related to the coding criterion used for accuracy in each task. For the recognition task, the foils generally only differed from the targets by only one phoneme. The participant had to get 100% of the sounds correct to make a correct response. In contrast, in the threshold discrimination task, responses were coded as accurate if they were within one phoneme of the original word. Taken together, the results of these two tasks showed that with phonological training, participants developed a good (though not perfect) configuration for each new word. Of course, the configuration was limited to form information, because no higher level (semantic, syntactic, pragmatic) information was available.

The development of lexical engagement was quite different than what we observed for lexical configuration. We will limit our discussion of lexical engagement to the ability of newly learned words to sustain perceptual learning, as the phonemic restoration data were uninformative because of our use of mostly fricatives as the critical phonemes.

If the training had produced fully functional lexical representations, then the new words should have supported perceptual learning: Hearing ambiguous versions of /s/ or /ʃ/ during the Old-New exposure task should have caused listeners to expand their category for the target phoneme. Although there was a 4-5% trend in the appropriate direction on each of the five days, the effect was very small and showed no hint of growing across days of training. Overall, the responses were not statistically different for the /ʃ/-trained group than the /s/-trained group.

Taken together, the results of the Threshold task and the Perceptual Learning task provide a striking contrast. The training regime produced very strong evidence for lexical configuration, but very weak evidence of lexical engagement. This dissociation underscores the importance of the theoretical distinction between knowing the "facts" for a word, and for that word's representation actually behaving as a true lexical representation.

This dissociation suggests that certain information about each new word was being represented, but that the representations had not achieved some criterion needed for behaving as "true" lexical items. This lack of full lexical engagement is especially interesting in light of the fact that we used the same training task as Gaskell and Dumay (2003) – phoneme monitoring – and they found evidence for lexical competition, which is one type of lexical engagement.

One important difference between Experiment 1 and the Gaskell and Dumay (2003) study is the nature of the new words to be learned. As we pointed out previously, the words to be learned in Gaskell and Dumay's study were necessarily variants of existing words (e.g., “cathedral-cathedruke”), whereas our new words were not closely linked to any members of the lexicon. One interpretation of the different results is that it may be easier to grow an offshoot of an existing lexical representation than to establish a fully functional one with no such links. That is, it is possible that “cathedruke” inherited some of its lexical functioning by virtue of its obvious ties to “cathedral”.

The other important difference between Experiment 1 and the Gaskell and Dumay study is the measure of lexical engagement. It is possible to devise models of lexical competition in which there is actually no real lexical engagement – the lexical representations do not directly affect each other's activation levels. For example, simply having two similar lexical items could cause a slowdown in any decision process, if that process needed to consider both entries, even if the entries themselves had no communication with each other. In contrast, for perceptual learning to occur, it is absolutely necessary for the lexical representation to inform the sublexical representation.

We noted that a methodological goal of Experiment 1 was to establish one good measure of lexical configuration, and one good measure of lexical engagement. Based on the patterns that we observed, we believe that the Threshold task provides a very orderly measure of lexical configuration; both accuracy and the level of noise at which a listener responded improved very systematically over training. The Perceptual Learning task seems promising as a measure of lexical engagement. It seems to be sensitive to lexical engagement (there were very consistent, though small, trends), and it seems to absolutely require the type of lexical function we wish to understand. Thus, in the following experiments, we will use these two tasks as our measures.

In the Introduction, we reported that several recent studies of the establishment of lexical representations have used a training technique in which participants learn to associate each new word with some visual stimulus (e.g., Gupta et al., 2003; Magnuson et al., 2003; Storkel, 2001). Although these studies only assessed lexical configuration, the training technique seems to be one that could potentially lead to both lexical configuration and engagement. In Experiment 2, we will use a variation of this training method, and examine the growth of lexical configuration and engagement.

EXPERIMENT 2

In Experiment 2 (and in the experiments that follow) we will use the same basic design that was used in Experiment 1: Participants will learn the same twelve new words over the course of five days, and we will track changes in the words' lexical configuration and engagement. In the current experiment, instead of using phoneme monitoring to present the new words, we will associate each new word with the picture of an unusual object (see Figure 5, below, for the pictures). Participants will be given the same number of training trials as before, and we will use the Threshold and Perceptual Learning tasks to assess changes in lexical configuration and lexical engagement, respectively.

METHOD

Participants—Twenty Stony Brook University graduate and undergraduate students participated in the five day experiment; four were replaced after Day 1 due to their failure to reach the criterion of 75% accuracy on the picture choice task (see below). All were native speakers of American English. They were either given credit toward a requirement in a psychology course or paid for their participation.

Design—Each participant received picture identification training every day for the first four days. Each participant was given one test of lexical configuration (threshold) and one test of lexical engagement (perceptual learning) every day for five days. The two tests were identical to the tests in Experiment 1.

Stimuli—The same twelve critical items used in Experiment 1 were used here.

Training Procedure—Each training session consisted of hearing the nonwords repeatedly in a picture identification task. We arbitrarily associated each nonword with the picture of an unusual object. Pictures were shown in color, and were collected by searching the Internet for hard to recognize objects. The twelve pictures are shown in Figure 5.

On each training trial, pictures of two objects were presented side by side on a computer screen. Each picture was presented as approximately the size of a normal playing card, making each very easy to see. At the same time that the pictures appeared, one of the critical words was presented over headphones. The participant was asked to identify which of the two pictures represented the word by pushing one of two buttons (left button = left picture, right button = right picture). As soon as the participant responded (correctly or incorrectly), the picture that was NOT associated with the word disappeared, leaving only the correct picture. This provided feedback to the participant about the correct association of words and pictures. Before participants had learned the associations, they simply guessed. In each training session (i.e., each day), each word was played 24 times with the picture that represented that word, each time paired with a randomly chosen picture from the remaining 11; the foil picture was randomly selected each time. Responses were coded for accuracy for the last 12 presentations on Day 1 (i.e., after the initial guessing period), and for all 24 repetitions on Days 2 through 4.

Testing Procedure—In the testing phase, each participant did the threshold task and the perceptual learning task. The twelve critical items were divided such that each participant heard six of the items in the threshold task, and the other six items in the perceptual learning task. The procedures were identical to those in Experiment 1, and the order of the two tasks was counterbalanced across participants.

RESULTS

Picture Identification: For the training task, responses were recorded for analyses of reaction times and accuracy. We conducted a one-way ANOVA on the reaction times, and found a significant effect for Day of training, $F(3,57) = 46.45$, $p< 0.001$. As Figure 6 shows, this effect reflects the much slower responses on Day 1 than on the other days. For accuracy, the second half of the responses on Day 1, and the complete set of responses on Days 2 through 4 were near ceiling, as can be seen in Figure 6. The 2% improvement over days (95% to 97%) did reach significance, $F(3,57)=2.91$, $p<.05$.

Lexical Configuration: The picture association training allowed participants to develop lexical representations that supported excellent and improving performance on the threshold task. Just as in Experiment 1, participants showed steady and substantial gains in both accuracy and the level of noise at which they responded. Figure 7 shows the data, and a comparison of these results to those in Figure 3 shows the similarity in lexical configuration under the two training regimes. With the picture association training, accuracy increased significantly over

Days of training (F(4,76)=20.86, p<.001), as did the noise level at response, F(4,76)=22.61, p<.001.

Lexical Engagement: As we have just noted, the two training regimes produced similar results on the measure of lexical configuration. The results for lexical engagement are dramatically different. Figure 8 shows the perceptual learning data for the participants who learned the new words as the names of unusual pictures. Recall that in Experiment 1, perceptual learning effects were small, and they did not increase across the five days of the experiment. In the current experiment, the effects were quite large, and they increased over the course of learning. Participants who heard ambiguous /s/ sounds in the old/new task were more likely to later categorize sounds on the /asa/-/aʃa/ continuum as /s/ sounds; the opposite was true for the participants who had heard ambiguous /ʃ/ sounds on the old-new task. The only way that listeners could know that a given ambiguous sound should be /s/ or should be /ʃ/ was by virtue of its appearing in one of the critical new words (e.g., the /s/ in “figondalis”, or the /ʃ/ in “bibershack”).

An ANOVA with Exposure type (ambiguous /s/ vs /ʃ/) and Day (1-5) confirmed the reliability of the perceptual learning effects. The average difference in “sh” report between the two groups was 23%, F(1,18)=11.94, p<.005. Although the size of the effect is numerically clearly larger later in training, the interaction of Exposure group and Day did not reach significance, F(4,72) =1.83, n.s. Another way to look at the size of the effect as a function of training is to assess the perceptual learning effect each of the five days. On the first day, the 16% difference in “sh” identification did not quite reach significance, F(1,18)=3.66, p=.07. On Days 2 and 3, the shifts became reliable: Day 2 (19%), F(1,18)=5.06, p<.05; Day 3 (18%), F(1,18)=6.15, p<.05. On the final two days, the effects were enormous: Day 4 (32%), F(1,18)=12.30, p<.005; Day 5 (30%), F(1,18)=10.54, p<.005. It is instructive to compare the size of these effects to perceptual learning shifts when real words were used to produce the effects (Kraljic & Samuel, 2005). The perceptual learning task and the test items were quite similar to those used here, but all of the perceptual learning took place within a single session. Across a half dozen conditions, the perceptual learning shifts consistently were in the 10-15% range, very similar to what was found on Day 1 here. The shifts of over 30% found on Days 4 and 5 of the current study are two to three times larger. This suggests that in addition to the new words becoming better represented, the perceptual learning task that was run each day provided an accumulating strength in the representation of the neutral fricative sound.

DISCUSSION

In thinking about the results of Experiment 2, it is useful to keep in mind how similar it is to Experiment 1. Participants learned exactly the same twelve new words, they had exactly the same number of training trials (24 per word per day, for four days of training), and they were tested on exactly the same measures of lexical configuration (threshold) and lexical engagement (perceptual learning). And, the pattern of lexical configuration performance looked quite similar to what was found when participants learned the words in a phoneme monitoring context: Listeners were quite good at pulling the words out of high levels of noise, and they improved quite a bit.

What is dramatically different, of course, is performance on the measure of lexical engagement. Participants who associated each new word with the picture of an odd (and presumably, unfamiliar) object showed clear evidence of having representations of these new words that were capable of lexical engagement. In particular, these new words could do what “real” words do – they can support perceptual learning in which the boundaries of phonetic categories get reshaped, with the lexical representations guiding the respecification of the sublexical units.

We believe that it is most parsimonious to attribute these results to the development of lexical representations for the critical trained items. As Norris, McQueen, and Cutler (2003) have demonstrated, ambiguous phonemes only generate perceptual learning when they are presented in real words, not in pseudowords. The ability of items like “bibershack” to support perceptual learning thus argues for their having taken on word-like properties. An alternative perspective is that the perceptual learning was generated by the existence of a learned part-whole relationship between the fricative (part) and an item such as “bibershack” (whole), without making any attribution of lexicality per se. Although this is logically possible, we believe that this leads to a theoretical dead end. In our stimuli, the “whole” is a phonotactically legal utterance; all to-be-learned real words have exactly this property. To arbitrarily segregate the current situation from lexical acquisition leads to a potentially fatal complication: somehow/sometime, there must be a magical moment in “normal” lexical development when a nonlexical “whole” becomes a lexical representation. It is extremely unclear how or when such a transformation could occur. If instead one adopts the view we have advocated, there is a natural growth of the lexical configuration, and a corresponding development of the ability of the developing lexical representation to engage other representations. In less technical terms, if something waddles like a duck and quacks like a duck, it is most parsimonious to assume that it is a duck.

Across Experiments 1 and 2, we thus see two types of dissociation. First, we see that it is possible to have lexical configuration without lexical engagement (the pattern in Experiment 1). Second, we see that different learning conditions can lead to different lexical engagement properties, even when lexical configuration seems similar (the pattern across Experiments 1 and 2). In a new research project that we have recently started, we have replicated this difference in the patterns across Experiments 1 and 2; Appendix A provides a brief report of this replication.

Given all of the similarities between experiments, the obvious question is, why is it that the participants in Experiment 2 developed representations with lexical engagement properties, while those in Experiment 1 did not? There are a number of differences in the learning conditions, but the most salient difference is that in the second experiment participants learned some kind of meaning to go with the sounds that they heard. Participants in the first experiment had no semantic information to go with the phonological codes that they clearly learned (as shown in the measure of configuration). In Experiment 3, we test whether having semantic information to go with the phonological will produce representations capable of lexical engagement, using a learning situation that in most other respects is quite different than the picture association regime of Experiment 2.

EXPERIMENT 3

As we have noted, an obvious difference between the learning conditions in the first two experiments is the presence of a referent for each new word in the picture association situation, and its absence in phoneme monitoring. There are other differences as well. The learning in Experiment 1 was more incidental than in Experiment 2, and all of the input was in one modality in the first case, versus having both auditory and visual input in the second.

In Experiment 3, the learning situation is purely auditory; it is less incidental than phoneme monitoring, but not as intentional as in the picture association task. Participants were exposed to the new words in the context of short passages that they listened to over headphones. Each passage was designed to provide a meaning for the word, without explicitly giving a definition. In this respect, the learning situation is probably most similar to the way that words are typically learned in the real world. The central question is whether hearing words in this kind of meaningful context will lead to representations capable of lexical engagement, as measured by their ability to support perceptual learning.

METHOD

Participants—Twenty Stony Brook University graduate and undergraduate students participated in the five day experiment. At the end of Day 1, participants were asked to give their idea what each of the six new words meant (they had heard each passage three times; see below), and three participants were replaced because they could not provide a reasonable gloss for a majority of the items. All were native speakers of American English. They were either given credit toward a requirement in a psychology course or paid for their participation.

Design—Each participant received training every day for the first four days, and each was given the test of lexical engagement (perceptual learning) every day for five days. Because we were specifically interested in whether semantic information is critical for developing lexical engagement, we did not use the threshold task of lexical configuration in this experiment.

Stimuli—The same twelve critical items used in Experiment 1 were used here. Half of the participants heard stories about the six items with /s/, and the other half of the participants trained on the six /ʃ/ items.

Training Procedure—Each participant learned six of the new words that had been used in the first two experiments. We only used six, rather than twelve, for two reasons. First, in the previous experiments six of the words were used for the threshold configuration task, which we did not run in this experiment. Second, the training procedure, using spoken passages, took more time than the other methods, and using twelve words would have made the sessions unreasonably long. Each training session consisted of a series of passages and questions presented over headphones. Each passage mentioned one of the critical items six times. Appendix B presents the stories used for each of the words. Two sample passages are:

> Joe's heart was failing, and his doctor told him he needed a replacement. The new bibershack seemed most promising. The bibershack had four valves, like the human heart. The survival rate with the bibershack was very encouraging. Moreover, the bibershack could be used with someone of Joe's age. Joe's brother had alerted him to the development of the bibershack. Based on both his brother's and his doctor's advice, Joe decided to have the bibershack implanted.
>
> Jack's car had broken down so he took it to the shop. The mechanic popped the hood, and immediately said, “yep, it's your figondalis that's fried. You're going to need a new figondalis”. Jack wanted a second opinion, so he took the car to an uncle, who knew a lot about cars. His uncle took a look at the figondalis and told Jack that he could probably get another 20,000 miles out of the old figondalis if he just repaired it. Jack really couldn't afford a new figondalis so Jack took the car back to the mechanic, got his figondalis fixed instead of replaced, and drove off the lot happy.

After each passage was played, it was followed by two “yes-no” questions that the participant answered by pushing appropriately labeled buttons. For example, for the first passage above, one of the questions might be “Would an expert be needed to fix a bibershack?” The questions were used to make sure that participants were paying attention to the story, and learning the meanings of the critical new words. The questions for all stories were based on a stock of general but targeted frames, and the particular questions for a given story varied with each presentation.

The questions also increased the number of presentations of the critical items. Specifically, each time a story and its questions were played, the participant was exposed to the critical item eight times (six times in the passage, plus once in each of the two questions). Within each training session, each of the six stories (and its questions) was played three times, for a total of 24 (3 × 8) presentations. Note that this procedure yielded the same number of training exposures for each word as was used in Experiments 1 and 2.

Testing Procedure—As noted above, participants were given the perceptual learning test of lexical engagement on each of the five days of the experiment. For half of the participants, the new words could potentially increase the report of /s/; for the other half, the potential increase was for /ʃ/.

RESULTS

Lexical Engagement: The results of the perceptual learning task were analyzed as in the first two experiments. Figure 9 shows the resulting curves. Inspection of the pattern suggests that the story-learning procedure produced lexical representations with stronger engagement than those from the phoneme monitoring situation, but not as strong as the picture-association procedure. An ANOVA with Exposure type (ambiguous /s/ vs /ʃ/) and Day (1-5) confirmed these impressions. The average difference in “sh” report between the two groups was 14%, F (1,18)=4.40, p=.05. As in the previous experiments, the interaction of Day and Exposure type was not significant, F<1. The pattern of shifts was relatively flat across training, though numerically the largest shift did occur on Day 5 (18%), F(1,18)=7.55, p<.05. The only other reliable single-day shift was on Day 2 (17%), F(1,18)=7.40, p<.05. The shifts on Day 1 (11%), Day 3 (14%), and Day 4 (9%) did not reach significance.

DISCUSSION

Hearing stories about a “bibershack” or a “figondalis” caused participants to establish lexical representations that were sufficient to engage sublexical representations and thereby produce perceptual learning. This suggests that having some kind of referent is important in generating lexical representations capable of such engagement; both hearing stories and seeing pictures of objects were successful in this respect, whereas merely hearing new words in the context of a phoneme monitoring situation did not produce significant engagement.

Although the story contexts did produce evidence of significant lexical engagement, the effects were generally smaller than those we observed in the picture association test, and there was less evidence for much growth in the strength of engagement over the five days of training. It may be that the more explicit learning demands of the picture association procedure played a role in creating such functional lexical representations. Although the story training clearly encouraged listeners to develop an idea of what each new word meant, it did not provide quite as explicit a mapping as the picture case.

The moderate level of lexical engagement for the story situation, and the very weak engagement shown after rather massive exposure in a phoneme monitoring context, led us to wonder whether the resulting representations might be missing something important. One such property is some kind of a production code for the new words. In many cases, when people learn a new word, they also produce it. This is certainly true in most explicit vocabulary teaching situations, both in a person's native language and in a foreign language learning context. There is also a growing literature (e.g., Schiller & Meyer, 2003) that suggests that some of the representations used in language production and perception may be shared, in which case establishing a “complete” lexical representation may require production information. The final two experiments pursue this idea.

EXPERIMENT 4

Experiment 4 used the same phoneme monitoring training regime that was used in Experiment 1, and included the measures of lexical configuration (threshold) and lexical engagement (perceptual learning) that proved to be diagnostic in the first two experiments. There was only one change in the training procedure: On each trial, after the participant made the target-present/target-absent button push, there was a production requirement: The participant repeated the stimulus aloud. This means that each day, the participants produced “bibershack”, “figondalis”, and the other ten critical items 24 times.

METHOD

Participants—Twenty Stony Brook University graduate and undergraduate students participated in the five day experiment; five were replaced after Day 1 due to sub-criterion performance on the phoneme monitoring task. All were native speakers of American English. They were either given credit toward a requirement in a psychology course or paid for their participation.

Design—Each participant received phoneme monitoring training, with a production requirement, every day for the first four days. Each participant was given one test of lexical configuration (threshold) and one test of lexical engagement (perceptual learning) every day for five days. The two tests were identical to the tests in Experiment 2.

Stimuli—The same twelve critical items used in Experiments 1 and 2 were used here.

Training Procedure—The training procedure was identical to the phoneme monitoring procedure used in Experiment 1, except that after each button-push response, the participant repeated the word that had been presented over headphones on that trial. The interval before starting the next trial was increased by 1000 msec to allow time for this spoken response.

Testing Procedure—In the testing phase, each participant did the threshold task and the perceptual learning task. The twelve critical items were divided such that each participant heard six of the items in the threshold task, and the other six items in the perceptual learning task. The procedures were identical to those in Experiments 1 and 2, and the order of the two tasks was counterbalanced across participants.

RESULTS

Lexical Configuration: Figure 10 shows the average accuracy and noise level at the point of responding, across the five days of testing. Performance clearly improved over the course of the five days, both in terms of accuracy (F(4,76=11.22, p<.001), and in noise level, F(4,76) =10.23, p<.001. We will consider these results further after we report the corresponding data in Experiment 5.

Lexical Engagement: The perceptual learning data were scored and analyzed as in the first three experiments. Figure 11 shows the resulting curves. It is obvious that adding a production requirement to the learning situation did not increase the level of lexical engagement for the resulting representation. In fact, the small but consistent 4-5% shift that had been found in Experiment 1 seems to have disappeared entirely here. Statistically, there is no hint of any perceptual learning: In an ANOVA with Exposure type (ambiguous /s/ vs /ʃ/) and Day (1-5), both the overall difference in Exposure group (less that 2%) and the interaction of the effect with Day of testing produced F<1. Moreover, for each of the five Days of testing, the Exposure factor also yielded F<1. Clearly, requiring participants to produce the new words many times in the context of a phoneme monitoring situation did not produce lexical representations that could engage sublexical ones to generate perceptual learning shifts.

DISCUSSION

The evidence for lexical engagement for words learned through phoneme monitoring without production (Experiment 1) was quite weak to start with. Perhaps it was unrealistic to expect that any production encoding would have an effect under these circumstances, as the lexical representations might just be too weak to have benefited. Still, it is curious that not only did the production requirement not increase evidence for lexical engagement, it actually seemed to reduce it.

We initially added the production aspect to the phoneme monitoring regime precisely because the representations had seemed to be weak enough to show a benefit from such converging cues. Given that this clearly did not occur, in the final experiment, we take a different approach: We add the production requirement to a training regime that demonstrably does produce lexical representations capable of engagement – the picture-association training method. The a priori expectation is that the additional lexical information will increase the strength of lexical engagement for the resulting representations. However, given the results of the current experiment, we must also consider the possibility that production during training could actually hinder the development of lexical engagement properties.

EXPERIMENT 5

Experiment 5 used the same picture-association training regime that was used in Experiment 2, with exactly the same stimuli, and included the same measures of lexical configuration (threshold) and lexical engagement (perceptual learning). As in Experiment 4, there was only one change made to the original training procedure: On each trial, after the participant made the left vs right picture choice, the test word had to be repeated aloud. The same visual feedback was provided as in Experiment 2.

METHOD

Participants—Twenty Stony Brook University graduate and undergraduate students participated in the five day experiment; after Day 1, two were replaced due to sub-criterion performance on the picture choice task. All were native speakers of American English. They were either given credit toward a requirement in a psychology course or paid for their participation.

Design—Each participant received the picture-association training, with a production requirement, every day for the first four days. Each participant was given the lexical configuration (threshold) test and the lexical engagement (perceptual learning) test every day for five days. The two tests were identical to the tests in Experiments 2 and 4.

Stimuli—The same twelve critical items used in the previous experiments were used here.

Training Procedure—The training procedure was identical to the picture-association procedure used in Experiment 2, except that after each button-push response, the participant repeated the word that had been presented over headphones on that trial. The interval before starting the next trial was increased by 1000 msec to allow time for this spoken response.

Testing Procedure—The testing procedures were identical to those in Experiments 1, 2 and 4.

RESULTS

Lexical Engagement: The central question in Experiment 5 is whether adding production to the picture-association learning situation will strengthen lexical engagement for the newly learned words. Figure 12 shows the critical perceptual learning data. As in Experiment 4, it is clear that saying words during the learning phase did not lead participants to create lexical representations with stronger lexical engagement properties. In fact, just as in Experiment 4, it appears that perceptual learning was weaker with the production component than without it.

Statistical analyses support these conclusions. Recall that in Experiment 2, picture association training led to robust perceptual learning, with a visible growth in the strength of the perceptual learning effect over the course of training (a 31% effect on the last two days, versus an 18% shift over the first three days). Exactly the opposite pattern can be seen in Figure 12: The effects are weak, and they seem to get weaker over time. Overall, the Exposure condition did not produce a reliable perceptual learning effect (7%), F(1,18)=1.00, n.s. As in Experiment 2, the apparent change over time in the size of the effect did not reach significance, F(4,72)=1.40, n.s.

We tested the difference in perceptual learning as a function of the difference in the training regime: with (Experiment 5) or without (Experiment 2) production. Overall, the 15% perceptual learning effect was reliable, F(1,36)=9.77, p<.005. There was also a marginally significant overall drop in the perceptual learning effect caused by the production requirement, F(1,36)=2.85, p=.10. Most interestingly, the three-way interaction of Experiment (with or without production), Day (1-5), and Exposure condition (/s/ vs /ʃ/ perceptual learning) was significant, F(4,144)=2.86, p<.05. This three-way interaction confirms the impression across Figures 8 and 12: Lexical engagement processes get stronger with training in the picture-association situation, but they get weaker with training when production is added to this regime.

The question, of course, is why saying a word while learning it would result in the development of lexical representations with reduced engagement. One plausible possibility is that the additional task requirement of production essentially exceeded the participant's processing capacity, leading to weaker lexical encoding for the to-be-learned words. One way to examine this possibility is to compare the picture association accuracy with (Experiment 5) versus without (Experiment 2) the production requirement. If participants are overwhelmed by the dual task requirement, we should see a drop in performance of the picture choice task. Table 3 shows the average picture choice accuracy for the four days of training, with and without the production requirement. For Days 2-4, all trials are included; for Day 1, only the second half of the trials (after the initial guessing period) are included. As the table shows, there was clearly some extra difficulty on Day 1 when production was required, with about a 9% drop in accuracy. However, after this initial cost, performance was at least as good in the dual-task case as in the single-task case. An analysis of variance confirmed the reliability of this pattern, with the main effect of Experiment (F(1,38)=5.65, p<.05) and of training Day (F(3,114) =46.28), p<.001) both being qualified by the interaction (F(3,114)=26.23, p<.001) caused by the cost being limited to Day 1.

The training task results thus suggest that after an initial cost, the production requirement does not have a significant impact on encoding the new words. We next consider another, and perhaps more focused way to see if an attentional overload due to the dual-task training procedure is indeed the cause of reduced lexical engagement. If this is the cause, we should also find evidence for similar costs in the second measure of lexical development through training – performance on the Threshold task.

Lexical Configuration: Figure 13 shows the average accuracy and noise level at the point of responding, across the five days of testing. Performance clearly improved over the course of the five days, both in terms of accuracy (F(4,76)=16.94, p<.001), and in noise level, F(4,76) =7.87, p<.001.

For our current purposes, the question is not just whether the participants showed improved evidence of lexical configuration across the five days of training. Rather, we need to determine if the extra attentional demands of the production situation caused a general impairment of lexical encoding, which is one possible explanation for the reduced perceptual learning effects we have seen in Experiment 5. To answer this question, we conducted an ANOVA on the threshold scores, using Day (1-5) and Experiment (2 vs 5 – no production vs production) as the factors. If the production requirement impairs lexical encoding, then the threshold performance should be lower in Experiment 5 than in Experiment 2. The left panel in Figure 14 shows the comparison, and it is evident that the general attentional account is not correct. In fact, participants were actually better at recognizing the newly learned words in noise when they had the production requirement during training, especially in the early non-asymptotic Days. The advantage is split between the main effect of Experiment (Noise Threshold: F(1,38) =4.20, p<.05; Accuracy: F(1,38)=5.71, p<.05), and the interaction reflecting the early difference, (Noise Threshold: F(4,152)=4.54, p<.005; Accuracy: F(4,152)=1.86, p=.12). These differences are in the wrong direction for a general attentional cost due to the need to produce the words during training.

Recall that in Experiment 4 we also saw an apparent decrease (to zero) in the perceptual learning trend that was shown in the first experiment, with phoneme monitoring training. It is instructive to compare the configuration measures across those two experiments, as we have just done for Experiments 2 and 5, to see whether there is a consistent effect of production on the threshold measure. The right panel of Figure 14 presents this comparison, and it is clear that the same (particularly early) improvement in the lexical configuration measure is found for production here as well. As in the picture-association training case, the production advantage is partly contributing to an overall increase in Experiment 4 over Experiment 1 (Noise Threshold: F (1,38)=1.76, n.s.; Accuracy: F(1,38)=4.78, p<.05), and partly to the interaction of Experiment with Day of testing (Noise Threshold: F(4,152)=2.86, p<.05; Accuracy: F(4,152)=2.90, p<.05).

DISCUSSION

In the current experiment, just as in the previous one, we tested the idea that adding a production component to developing lexical representations would lead to richer, more functional structures capable of greater lexical engagement. The results of Experiment 4 suggested that this was not the case, but the apparent drop in perceptual learning there was compromised by the very low level of engagement observed without the production component (Experiment 1). This was not an issue in the current experiment, given the very strong perceptual learning results for the picture association learning situation (Experiment 2). The results were clear: Participants who produced the new words as they were learning them developed representations that were less capable of lexical engagement.

Although this result seems counterintuitive, there are a number of ways that it might occur. For example, when participants produce an item after making their picture choice, the item that they produce does not contain one of the critically-ambiguous phonemes that the training items contain. Instead, the participants' utterances provide a “good” fricative, and one might imagine that hearing a good rendition could offset the training effects of the ambiguous token. However, there are three results from our laboratory that make this explanation unlikely. Kraljic and Samuel (2005), using similar fricative-based critical items (in real words), found that hearing a large number (approximately 100) of “good” tokens, after having trained on twenty ambiguous ones, left the perceptual learning effects essentially intact; the normally-produced fricatives did not lead listeners to reset their category boundaries. Of course, in the current experiment, the (experimenter-presented) ambiguous fricatives are alternating with (participant-produced) unambiguous ones, leaving open the possibility that the correcting feedback might be more effective if it co-occurs with the training tokens. However, Kraljic and Samuel (in preparation) tested a mixed presentation order of (experimenter-presented) good and (experimenter-presented) bad tokens, and found that it did not diminish the perceptual learning effect. There is also strong reason to doubt that the participants' productions can affect their own perceptual representations in this domain. Kraljic, Brennan, and Samuel (submitted) looked for any changes in participants' productions of their fricatives, before and after undergoing the perceptual learning task. Despite showing large changes on the perceptual side, the participants showed no changes in their own productions of these sounds, suggesting a decoupling of perception and production in this case. Thus, although we cannot definitively rule out the possibility that the participants' own productions in Experiments 4 and 5 served as counterevidence for perceptual learning, the results from our laboratory make this an unlikely explanation.

Recall that in Experiment 5, the analyses of lexical configuration demonstrated that the poor lexical engagement cannot be attributed to any simple attentional resource limitation: For both phoneme monitoring and picture association, learning regimes that required production led to stronger lexical configuration effects, rather than weaker ones. Thus, the production component of word learning had a positive effect on lexical configuration, coupled with a negative effect on lexical engagement. The implications of this pattern will be discussed shortly, in the General Discussion.

GENERAL DISCUSSION

We undertook this project because we believed that in order to understand how new words get incorporated into the mental lexicon, it is important to draw a theoretical and empirical distinction that has not been well-developed in the literature: Each word that a person learns includes a certain configuration of “facts”, and each fully functioning lexical representation is capable of engagement with other lexical representations as well as with sublexical ones. We therefore began by developing plausible measures of lexical configuration and lexical engagement, which allowed us to then examine how different conditions of learning new words might affect these two hypothesized aspects of lexical representations. Throughout this project, we observed dissociations that provide strong support for the importance of distinguishing between lexical configuration and lexical engagement; it is one thing to learn the form of a new word, and something else for the representation to take on all the “rights and responsibilities” of a fully functional lexical entry.

Two of the observed dissociations were particularly striking. The first such dissociation was revealed by the different patterns of performance on the threshold task versus the perceptual learning task, as a function of the initial training regime. Although participants showed similar and high levels of lexical configuration across different learning situations, the level of lexical engagement varied considerably. The representations formed in the phoneme monitoring training were apparently quite adequate in representing the phonological form of the newly learned words, but were not capable of engaging sublexical codes sufficiently to support perceptual learning. Picture association, in contrast, led to representations with both lexical configuration and engagement properties. Story context also generated representations with engagement properties, though their strength was not as great as that found with picture association.

The differing effect of phoneme monitoring versus picture association training on the resulting ability of the learned items to support perceptual learning was observed twice: The difference was seen in the initial sessions both with production requirements (Experiment 4 versus Experiment 5), and without (Experiment 1 versus Experiment 2). As described in Appendix A, we have also recently conducted a third comparison of the two training regimes (in the context of developing more efficient procedures for pursuing this research effort), and we again observed that the picture training association method leads to stronger perceptual learning than the phoneme monitoring training. Moreover, with the new procedures, we confirmed that there were no between-groups differences pre-training. Thus, this dissociation seems to be quite robust.

The second type of dissociation involved the effect of adding a production component to the learning situation. For both the phoneme monitoring case, and for picture association, adding production weakened the evidence for lexical engagement, but it enhanced the development of lexical configuration. Clearly, knowing the “facts” for a word, and having a word act like a word within the mental lexicon, are not the same thing.

The only previous work that has explicitly examined a distinction of this type is the research of Gaskell and Dumay (2003; Dumay & Gaskell, 2006; Dumay, Gaskell, & Feng, 2004), the research line that most directly inspired the current project. Gaskell and Dumay drew a distinction between simple “phonological learning” and “lexicalization”, which corresponds to our distinction between lexical configuration and lexical engagement. In a series of elegant and interesting experiments, these authors have shown that simple exposure to new words via phoneme monitoring is sufficient to produce good phonological learning as measured by the ability to choose the correct item, given a choice of a trained item versus a similar lure. However, many more repetitions were required before there was evidence that the developing representation was capable of engaging in lexical competition, Gaskell and Dumay's (2003) measure of “lexicality”. Moreover, Dumay and Gaskell (2006) have shown that there is a significant increase in the strength of such lexical competition if the participant is tested after getting a night's sleep than if the same amount of time passes without sleep. These results provide an excellent base for distinguishing between simply configuring certain phonological information, versus establishing a representation that can fully engage in lexical behavior.

Gaskell and Dumay (2003) noted that their procedure for teaching new words “was quite impoverished. No meanings were explicitly associated with the novel words, and no sentential context was available… This information scarcity was deliberate, and mirrors numerous recent developmental and theoretical studies that have focused on the informational content of the speech stream alone” (p. 125). Reducing the complexity of the experimental situation is a standard method scientifically, and does have the benefits that Gaskell and Dumay alluded to. However, the results of the current study show that comparing new word learning with versus without semantic anchoring will be critical to understanding how new words get added to the mental lexicon.

In fact, in a conference proceedings paper, Dumay, Gaskell, and Feng (2004) conducted a preliminary comparison of learning new words in a phoneme monitoring situation versus learning them through a kind of semantic context. In the latter case, each new word like “cathedruke” occurred in two sentences, such as “A cathedruke is a variety of vegetable”, and “The cook served the boiled cathedruke with a steak and baked potatoes”. The degree to which participants had acquired each new word's meaning (in this example, “vegetable”) was assessed by measuring how often the participant gave the meaning as the response in a free association test. “Lexicalization” was assessed in terms of slow responses to the base words (e.g., “cathedral”), as in the Gaskell and Dumay (2003) study, except in this case both the base words and the new words were included in the test. Although the base words always came earlier than their corresponding new words, participants may have delayed responses to the base words as they listened to and processed the endings to make sure that they were not being “tricked” by a new word (new words were to be responded to as nonwords, a conflicting response). Overall, the semantic condition was not very successful (e.g., on the three lexicalization tests, conducted on the training days plus a week afterward, there was never stronger interference for semantically-trained words than for those trained via phoneme monitoring), perhaps reflecting the apparently weak learning of the meanings (only 30-44% of the free associations were the target meaning). This may in turn reflect the need to learn a large number of new words (24), with weak semantic help, in only two training sessions.

In the current study, we used much stronger semantic manipulations, which resulted in essentially perfect learning of the meanings of the new words (we also taught people fewer new words, making the learning of each one easier, and we trained for four days). These procedures led to the first dissociation mentioned above: Training without meaning, via phoneme monitoring, led to very weak lexical engagement, whereas training with meaning (either picture association, or full stories) produced much stronger evidence of engagement. Note that Gaskell and Dumay (2003) reported that with about 30 phoneme monitoring trials, “lexicalization” was indicated by the impaired performance for lexical decisions on the related base words (e.g., “cathedral”). The fact that our measure of lexical engagement shows very weak effects for words learned through phoneme monitoring, even after almost 50 training trials, suggests that it may be easier for the system to develop “lexicalized” representations for new words that are close variants of existing words, than to build such representations from scratch; recall that our new words were constructed to have no such close relatives in the lexicon. Alternatively, it may be that lexical competition does not require the same level of lexical engagement that is required to produce perceptual learning. This seems at least plausible, given that perceptual learning entails restructuring sublexical categories, whereas lexical competition can be mediated simply through the existence of similar lexical entries.

We have focused our research, and the presentation here, on how adults add new words to their mental lexicons, rather than on the question of how this happens in children. The language acquisition literature is vast, with research made much more complicated by all of the developmental factors that are intertwined with the question we have taken on here. There are, however, a few studies of children's language acquisition that may provide some helpful convergence with our approach. For example, our results (and those of Gaskell & Dumay) suggest that rather different things are being learned when we consider lexical configuration than when we consider lexical engagement, with the latter benefiting more from semantic support than the former. Naigles (2002) observed that certain findings in the literature on children's language acquisition seemed paradoxical, until one considers a distinction somewhat like the one we have found here. The paradox is that very young infants seem to have the ability to abstract patterns over the items they are trained with, and even to distinguish normal English utterances from those that are not, whereas somewhat older children (toddlers) typically fail to demonstrate generalizations to new test items after being presented with training stimuli. Naigles' resolution to the paradox is given in the title of her paper: “Form is easy, meaning is hard”. Her analysis of the studies in the literature shows that the generalization shown by infants has typically been based on acoustic or phonetic patterns that do not actually require any understanding of the input. In contrast, the tests that toddlers typically fail are ones that explicitly or implicitly require semantic knowledge. Thus, the apparent paradox can be resolved if we assume that babies come equipped with efficient pattern tracking procedures, a mechanism that is sensitive to the form of the input. It takes considerable further development before semantic processing is available. The results of the current study suggest that this sequence plays out in the development of individual lexical representations as well: It is relatively easy to establish lexical configuration information that captures the phonological (and/or orthographic, see below) form, but it takes more time/training (and possibly sleep) to develop lexical structures based on meaning that are capable of lexical engagement.

One other study from the language acquisition literature seems particularly germane. Graf Estes, Evans, Alibali, and Saffran (in press) have very recently examined two aspects of lexical development that had not previously been investigated together. Earlier work (e.g., Saffran, Aslin, & Newport, 1996) showed that infants can use the statistical properties of the speech that they hear to extract word candidates. In such statistical learning studies, infants hear a stream of syllables with certain statistical properties. Potential words are defined by sequences of syllables that co-occur with high frequency (this procedure shares critical aspects with the phoneme monitoring training of our participants). Saffran et al. demonstrated that after relatively brief exposure periods to these statistical distributions, infants show listening preferences that reflect the “words” that they had been exposed to. In the Graf Estes et al. study, 17-month-olds first underwent this statistical learning procedure, and then participated in a task that required them to associate utterances with certain colored objects shown on a monitor (much like our picture-association training). The critical finding was that the participants were more able to associate an utterance with an object if the utterance was a “word” that had just been learned, rather than an utterance made up of other syllables that they had heard that did not have high co-occurrence properties. Thus, this study indicates that the initial exposure led to preliminary lexical configuration (the high-probability forms), and that the availability of this configuration enhanced the ability to then add meaning information (the objects to be associated with the phonological forms) to these developing lexical representations. What remains to be seen is whether these newly-developed lexical representations can engage other representations.

We turn now to the second major dissociation that we observed: Requiring participants to produce the words that they were learning significantly enhanced their ability to recognize the words under difficult perceptual conditions (the threshold task), whereas this requirement eliminated the ability of the new words to sustain perceptual learning (i.e., lexical engagement). This split in performance is striking, and underscores the importance of distinguishing between lexical configuration and engagement.

A reasonable interpretation of this pattern follows from the first dissociation – the difference in engagement as a result of learning the words through phoneme monitoring (very weak engagement) as opposed to picture association (very strong engagement). By its nature, the phoneme monitoring task focuses the learner's attention on the phonological properties of the new words: The task explicitly requires the listener to attend to phonemic properties. In contrast, the picture association task (and, similarly, story context) directs the learner to the meaning of each new word, or at least its referent. The perceptual learning results of the first two experiments show that phonological encoding provides poor support for later lexical engagement, while semantic encoding enhances such engagement. From this perspective, the production results can be viewed as another demonstration of the strong relationship between encoding conditions and the resulting lexical representations. In order to produce each new word as it is presented, the participant must focus on the phonological structure. Such a focus produces good lexical configuration, leading to excellent performance on the threshold task. However, this encoding, as we have seen for phoneme monitoring training, is poorly suited to produce the representations needed to support perceptual learning.

This analysis of the two major dissociations clearly shares many features with some of the classic memory findings of the last thirty years. Most directly, the correspondence between semantic information during new word learning and the subsequent evidence for lexical engagement is reminiscent of the memory literature on depth of processing (e.g., Craik & Tulving, 1975). This classic memory finding is that material learned with semantic processing is typically better recalled and recognized when explicit memory tests are given, than if the initial learning took place under “shallower” conditions, such as those that encourage attending to phonological or orthographic form. Although our results do have some similar characteristics, it is not clear that they fit quite so neatly into the memory literature. The complication is that our measure of lexical engagement is the ability of the newly learned word to support perceptual learning, which is very far removed from the type of explicit memory test that normally shows depth of processing effects. In fact, we believe that most memory theorists would consider the perceptual learning task to be one that is more similar to the kinds of perceptual/implicit tests that typically do not show depth of processing effects. Given that learning a new word is certainly a type of memory task, our results may provide a rather different set of constraints for memory theorists.

We noted that Gaskell and Dumay (2003) should be credited for introducing the kind of theoretical distinction that we have pursued here. They distinguished between “phonological learning” and “lexicalization”. We prefer to refer to lexical configuration and lexical engagement, for several reasons. With respect to the latter term, we believe that “engagement” is more specific, as it refers directly to the aspect of lexical activity that both our study and Gaskell and Dumay's attempted to investigate. The advantage of lexical “configuration”, conversely, is that it is more general than phonological learning, and we believe that such generality is called for.

The value of such generality can be seen in the way that this perspective can encompass a research program by McKague, Pratt, and Johnston (2001; Johnston, McKague, & Pratt, 2004), despite the fact that their investigation was undertaken with quite different goals and in a different modality. The focus of their work is the potential link between spoken words and orthography. In particular, they have tried to determine whether simply hearing a new word can cause a person to create an orthographic representation; the existence of such auditorily generated orthographic representations could be inferred if there are certain performance differences the very first time that a participant is exposed to a printed version of the word.

McKague et al. (2001) worked with children (approximately seven years old), and taught them 10 new words, either with or without a semantic context. The semantic context was provided by a story in which 10 “strange creatures” were mentioned several times by name, with pictures of each one presented several times (this method is reminiscent of the procedures used by Storkel, 2001, and by Gupta, 2003; Gupta et al., 2004). In the non-semantic condition, each new word was presented several times in a list-learning situation. In both conditions, the children repeated the new words on 5 of the learning trials. The measure of having a pre-established orthographic code for a new word was an accuracy advantage in reading the word the very first time it was presented, relative to control nonwords that had not been heard. A robust advantage of this type was observed, with children accurately reading words that they heard before about 60% of the time, versus about 30% accuracy for the control items. This pattern did not vary as a function of semantic versus non-semantic training. In a similar second experiment, the authors explored the role that production may have played in the results, and found that its role was minimal: The same training advantage occurred regardless of whether or not the child had produced the new words during training.

Although McKague et al. (2001) concluded that their results suggested that people do indeed create (imperfect) orthographic codes as a result of strictly auditory exposure to new words, they acknowledged that the evidence was not compelling, as the observed advantage could occur in a number of ways. To provide a more decisive test, Johnston et al. (2004) used a more stringent approach: They tested whether words that had been learned only through listening were capable of producing masked repetition priming, a phenomenon that has generally been observed for real words (i.e., those with lexical representations) but not for nonwords. In the masked priming paradigm, participants make a judgment (typically, lexical decision or naming) to a printed stimulus that is preceded by a very briefly flashed stimulus (the prime). There is generally a premask before the prime, and the prime is presented for less than 60 msec, a combination that usually prevents participants from being aware that there was any prime at all. In addition, primes are usually presented in lowercase (e.g., “pleach”), with targets in uppercase (e.g., PLEACH), to minimize simple visual overlap. It has been argued that masked priming is an automatic lexical process, because it is primarily or exclusively (depending upon which investigator one reads) found for real words, and because participants are completely unaware of the prime.

Johnston et al. taught adults a number of new words by presenting them in auditory form, along with definitions that were to be associated with them. Using appropriate control conditions to assess simple orthographic and simple phonological effects, the authors found evidence that masked priming could indeed be found under these conditions, leading them to suggest that when a literate person hears a new word (seven times, in this experiment), a lexical representation is created that includes a “broad” orthographic code. There were some aspects of the data that call for caution at this point, but the basic effect was relatively sound. In fact, when they discussed some of the limitations of their results, Johnston et al. said “The conditions required to make the representations of novel words as well specified as representations of familiar words remain unclear from [the] present data. One possibility is that classification of a target in a lexical decision task requires only a superficial analysis of its orthographic structure (i.e., targets are presented for a brief time and no semantic analysis is required). It is possible reading novel words in a meaningful context (i.e., in text) would increase the precision of representations after considerably fewer visual encounters.” (p. 306). This speculation by Johnston et al. seems quite prescient, in the context of our picture-association and story-context learning situations.

There are a handful of other studies sprinkled through the literature that also provide intriguing converging results, though most of these are based on recognition of printed words. Studies by Whittlesea and Cantwell (1987) and by Rueckl and Olds (1993) are examples of these. In both of these studies, participants were given some experience with one- or two-syllable printed pseudowords, and were later given a perceptual identification test, with these items (and various controls) presented under difficult recognition conditions (e.g., 30 msec exposures with postmasking). In both studies, a key manipulation was whether or not a pseudoword was given a meaning during the exposure phase. The consistent finding was that items that did have an attached meaning produced better perceptual identification. Moreover, Whittlesea and Cantwell showed that this advantage was not dependent on participants' memory for the pseudowords' meaning. This led them to argue, consistent with the position that we have advocated here, that the critical aspect was the processing that took place during the encoding of meaning itself.

As we noted, we believe that a full understanding of how a new word gets added to the mental lexicon requires a theoretical distinction between the accumulation of “facts” (how a word sounds, what it looks like, what it means, how it fits into sentences), and how the lexical representation supports perceptual and cognitive processes. This framework is quite general, and can be applied to the kind of methods and measures that we have used here, as well as to the kinds of methods and measures used by McKague and her colleagues, and of course to the many other studies of lexical acquisition (e.g., Dumay & Gaskell, 2006; Gaskell & Dumay, 2003; Gupta et al., 2004; Magnuson et al., 2003; Salasoo et al., 1985; Storkel, 2001). The results of the five experiments of the current study make it clear that one cannot simply ask “How is a word added to the mental lexicon?”. A more appropriate question would be “What learning conditions lead to the development of lexical configuration information, and what conditions promote lexical engagement?”. The results of the current study provide some preliminary answers to this question, and should help to frame further investigations of this important question.

Acknowledgements

This project was supported by NIMH grant R0151663, and by NSF Grant No. 0325188. We thank Donna Kat for her enormous technical assistance, and Suparna Rajaram for her very helpful advice. We also thank Anne Cutler for her quite helpful suggestions with respect to our critical terminology. Gareth Gaskell, Jeff Bowers, and an anonymous reviewer all provided very insightful suggestions during the review process.

Appendix A: A confirmation experiment for the difference between learning new words through phoneme monitoring and through picture association

After the completion of the project reported in this paper, we began a new research effort in which we have been developing improved procedures for examining new word learning. As part of this effort, we have taught new sets of participants the same new words as in the current study, using both the phoneme monitoring training method, and the picture association regime. In this Appendix, we report the results of a small initial experiment using the new methods, an experiment that provides a replication of the difference observed in Experiments 1 and 2 of the current study.

The procedures used in the new work were similar to those used in the current study, with several improvements intended to increase the sensitivity of the method. There were four modifications:

1. Each participant was only trained on six new words, rather than 12. This change allowed us to present each word 36 times in a session, rather than 24. This also allows participants to focus more on each word.
2. Data come from a single day, rather than five.
3. We collected a baseline identification function for each participant before any training had occurred, to determine if there were any pre-existing differences in how participants in the training groups identify the /asa/-/aʃa/ test series.
4. We used two sets of critical items. One set was the same group of 12 new words that were learned in the current study (e.g., “bibershack”, “figondalis”). The second set was the complement of this: the same 12 items, but with /s/ replacing /ʃ/, and vice versa (e.g., “bibersack”, “figondalish”). These complementary items were recorded by the same speaker, at the same time, as the items used in the current study. Testing these items provides a check for any (unlikely but conceivable) idiosyncratic effects that might be associated with the original stimulus set. Within each training group (phoneme monitoring; picture association), half of the participants learned six items from the original set, and half of them trained on the complementary versions.

We tested 64 participants from the same population as those in the current study, and after eliminating those who failed to follow instructions, or who could not produce clean labeling functions on the /asa/-/aʃa/ test series, data from 41 participants were analyzed.

Phoneme Monitoring Training

For each participant, we calculated the average “sh” report on the test series, before training, and after training (and the perceptual learning procedure). Ten of the participants underwent perceptual learning procedures that would potentially increase “s” report; for the other ten, the “old-new” task would potentially increase “sh” report. On the baseline measure of /asa/-/aʃa/ identification, there was no difference between these two subsets: The to-be-/s/ group labeled

61.9% of the test items as “sh”, and the to-be-/ʃ/ group identified 62.7%, F(1,18)=0.018, n.s. This confirms that there were no pre-existing differences. And, just as in the current study, using words learned through phoneme monitoring training to drive the perceptual learning effect was ineffective; there were also no differences in “sh” report after word learning and the “old-new” perceptual learning procedure (/s/ group: 59.2%, /ʃ/ group: 59.0%), F(1,18)=0.001, n.s.

Picture Association Training

Among those who learned the new words through picture association, twelve participants were in the /s/ group, and nine were in the /ʃ/ group. We first compared the labeling functions for the two groups. As with those who learned words via phoneme monitoring, there was no initial difference in how the /asa/-/aʃa/ test items were identified: /s/ group: 53.9%, /ʃ/ group 53.8%, F(1,19)=0.001, n.s. Critically, after picture association training, the words were capable of lexical engagement, as the two groups showed a significant difference on the perceptual learning measure: /s/ group: 66.8%, /ʃ/ group 47.1%, F(1,19)=5.315, $p < .05$.

Thus, the results of this additional experiment completely replicate the dissociation found in Experiments 1 and 2: When listeners learned new words in the context of phoneme monitoring, the words were unable to drive perceptual learning. In contrast, participants who learned the same words through picture association showed a robust 20% perceptual learning effect. The new procedures used in this experiment provide reassurance that the dissociation cannot be attributed to any pre-existing differences in fricative identification, and that they are not due to any idiosyncracies in the assignment of fricatives (/s/ versus /ʃ/) to particular critical words.

Appendix B: Stories and questions used in Experiment 3

1--benemshalow

They stood in the kitchen, closely following the recipe. Sandy said, “open the can of tomatoes with that benemshalow.” Sarah took the benemshalow, attached it to the can, and twisted the blue knob. The benemshalow opened the can with ease. Once the benemshalow had opened the can, they poured the tomatoes into the sauce. Sarah started to put the benemshalow back in the drawer but Sandy said, “keep the benemshalow handy – we may need it again.”

2--bibershack

Joe's heart was failing, and his doctor told him he needed a replacement. The new bibershack seemed most promising. The bibershack had 4 valves, like the human heart. The survival rate with the bibershack was very encouraging. Moreover, the bibershack could be used with someone of Joe's age. Joe's brother had alerted him to the development of the bibershack. Based on both his brother's and his doctor's advice, Joe decided to have the bibershack implanted.

3--figondalis

Jack's car had broken down so he took it to the shop. The mechanic popped the hood, and immediately said, “yep, it's your figondalis that's fried. You're going to need a new figondalis”. Jack wanted a second opinion, so he took the car to an uncle, who knew a lot about cars. His uncle took a look at the figondalis and told Jack that he could probably get another 20,000 miles out of the old figondalis if he just repaired it. Jack really couldn't afford a new figondalis so, Jack took the car back to the mechanic, got his figondalis fixed instead of replaced, and drove off the lot happy.

4--galasod

Judy ran into the mall in hopes of finding a galasod. She had a stubborn headache and was sure the galasod was the answer. She had never tried a galasod but she knew she had to have one. She ran into Brookstone, found the galasod and rushed to the cashier. After buying it, she ripped the galasod out of the box and immediately put it on her head. The galasod began massaging her headache away. Within minutes, she gave a sigh of relief and left the mall.

5--gatersy

Mary and Tom were getting married and had set up a gift registry. Jane wanted to buy them something special, perhaps a gatersy. She checked the registry and behold, the gatersy was on it! They must love modern art as much as I do, she thought. She went to the art museum gift shop and found a beautifully sculptured gatersy sitting behind the counter. The salesman told her that the gatersy was even on sale that week. So, she had it gift-wrapped, and took the gatersy home to wait for the wedding. But then temptation won out. She decided that she loved the sculpture so much that she kept the gatersy for herself.

6--komalsheum

The Italian cook always made his own pasta. The key to his success was his komalsheum. The komalsheum provided a place for all his pasta pieces to dry without sticking to each other. He would mix the dough, run it through the cutter, and hang it on the komalsheum to dry. When he took the pasta off the komalsheum, it would go straight into the water for the freshest pasta around. His restaurant was always full and the kids would rush to the window to see the komalsheum at work. At the end of the day, the cook was careful to scour the komalsheum to be sure it would be ready for the next busy day.

7--lifrisen

Ted got lost driving home one night. It was late, it was raining, and his lifrisen wasn't working. He needed the lifrisen to tell him how to get home, but it was acting up, telling him to turn right when he needed to turn left. The lifrisen was supposed to give him the shortest route home, but it was taking him on a detour. Maybe the lifrisen knew about some accident and was directing him around it, or maybe the lifrisen was broken. Much to his surprise, he suddenly saw the lights of his house, so the lifrisen had worked after all.

8--marfeshick

Charles had been walking in the desert for what felt like hours. His marfeshick was almost empty, and he was worried that he might dehydrate. He had bought one of the special marfeshicks that kept the water cool, but it would soon be no use to him. Why didn't he bring a second marfeshick, he asked himself. His marfeshick was down to the last drop when all of a sudden, Charles spotted a creek ahead. He ran to the creek, filled his marfeshick with water, and took a long drink from it. Now that his marfeshick was refilled, he could continue walking until he reached the station.

9--naronesay

Bob was perusing the antiques stores downtown when something special caught his eye. The naronesay was sitting on the shelf in the antique clock store. Bob collected naronesays and was in search of a specific one made in 1867. He ran to the naronasay, wound it up with the key, and listened as the naronsay ticked away the seconds. The year on the back said 1867, so he knew this was the naronesay he had been searching for. He took his special naronesay to the cash register and said, “I'll take it!”

10--nomemsoly

Samantha wanted the perfect thing to finish her garden with. She searched at Home Depot and Lowes, until she found the perfect nomemsoly. This nomemsoly spouted water from all angles, creating a circle of sprayed water. She could put flowers in the center, and have the nomemsoly water the surrounding area. The nomemsoly was just what her garden needed. The birds would also love the nomemsoly when the days were hot. Getting a brass nomemsoly would cost a bit more than a plastic one but Samantha decided it was worth it.

11--penivasher

The monkey danced on the street musicians shoulder as the musician gleefully played his penivasher. He squeezed and pulled the penivasher as it filled the air with melodies. People had begun to surround the musician, throwing coins into his hat. They cheered as he played the penivasher with his hands and the harmonica with his mouth. He always got more tips when he played the penivasher, he thought. The monkey jumped down on the ground and the musician let the monkey jump up and down on the penivasher. He then realized it was the monkey who really brought in the tips, not his own penivasher playing.

12—wikoshah

The archeologist had a good feeling about his site. As he dug, he hit something hard. So he dug a little bit deeper and saw the top of a wikoshah. He picked up the wikoshah and rubbed it clean to reveal a shiny gold reflection. The wikoshah was featured in many stories he had read, supposedly possessing many magical powers. He rubbed the side of the wikoshah, just like in the stories, but nothing happened. Disappointed, he pulled the lid of the wikoshah open and looked inside. A little label saying “made in China” made it clear that this wikoshah was not what he was looking for.

Can opener

YES: Is a benemshalow made of metal?

Can you find a benemshalow in the kitchen?

Does a benemshalow cost less than 10 dollars?

Would it be odd to get a previously owned benemshalow?

NO: Does a benemshalow need batteries?

Is a benemshalow edible?

Can you find a benemshalow in a car?

Would it be useful to own more than one benemshalow?

Artificial heart

YES: Does a bibershack move on its own?

Does a bibershack cost more than 10 dollars?

Would an expert be needed to fix a bibershack?

Does a bibershack save lives?

NO: Is a bibershack made of metal?

Can you find a bibershack in a house?

Does a bibershack cost less than 100 dollars?

Is water involved in a bibershack's function?

Car part

YES: Does a figondalis cost more than 10 dollars?

Would an expert be needed to fix a figondalis?

Is a figondalis made of metal?

Is a figondalis found in a car?

NO: Can you find a figondalis in a park?

Is a figondalis made of plastic?

Is a figondalis found in the kitchen?

Would a child play with a figondalis?

Sculpture

YES: Does a gatersey cost less than 10 dollars?

Can you find a gatersey in a house?

Would most people need training to create a gatersey?

Can you find a gatersy in a museum?

NO: Does the gatersey need batteries?

Is the gatersey used to entertain?

Can you find a gatersey in a park?

Does a gatersey make noise?

Head massager

YES: Does a galasod move on its own?

Does a galasod need batteries?

Would a galasod be used to relax?

Does a galasod cost more than 10 dollars?

NO: Is the galasod used to entertain?

Can you drink from the galasod?

Is a galasod edible?

Does the galasod keep time?

Pasta rack

YES: Does a komalsheum produce something edible?

Can you find a komalsheum in a kitchen?

Would it be useful to have more than one komalsheum?

Would a child like a komalsheum?

NO: Can you find a komalsheum in a park?

Does a komalsheum make noise?

Does a komalsheum move on its own?

Would a komalsheum be used to relax?

GPS system

YES: Can you find a lifrisen in a car?

Would you take a lifrisen on a hike?

Does a lifrisen cost more than 10 dollars?

Does a lifrisen give directions?

NO: Is a lifrisen edible?

Does a lifrisen keep time?

Is a lifrisen used to entertain?

Does a lifrisen cost less than 100 dollars?

Canteen

YES: Is water involved in the marfeshick's function?

Would you take a lifresin on a hike?

Can you drink from a marfeshick?

Is a marfeshick made of plastic?

NO: Is a marfeshick made of wood?

Does a marfeshick move on its own?

Is a marfeshick edible?

Does the marfeshick make noise?

Clock

YES: Does the naronesay cost more than 10 dollars?

Does the naronesay keep time?

Does a naronesay make noise?

Would it be useful to have more than one naronesay?

NO: Can you drink from a naronesay?

Does a naronesay need batteries?

Does the naronesay cost less than 10 dollars?

Would most people need training to be able to use a naronesay?

Water fountain

YES: Is water involved in a nomemsoly's function?

Would a child like a nomemsoly?

Would a nomemsoly be used to relax?

Does a nomemsoly cost more than 10 dollars?

NO: Is a nomemsoly made of plastic?

Can you find a nomemsoly in a house?

Is a nomemsoly used to keep time?

Would it be odd to receive a previously owned nomemsoly?

Accordion

YES: Is a penivasher used to entertain?

Does a penivasher make noise?

Would most people need training to be able to use a penivasher?

Would it be odd to receive a previously owned penivasher?

NO: Can you drink from a penivasher?

Can you find a penivasher in the kitchen?

Would it be useful to have more than one penivasher?

Is a penivasher edible?

Magic lantern

YES: Would a child like a wikoshah?

Is a wikoshah made of metal?

Does a wikoshah cost more than 100 dollars?

Will a wikoshah bring someone good luck?

NO: Does the wikoshah need batteries?

Does a wikoshah make noise?

Can you find a wikoshah in a car?

Would it be odd to get a previously owned wikoshah?

References

Bertelson P, Vroomen J, de Gelder B. Visual recalibration of auditory speech identification: A McGurk aftereffect. Psychological Science 2003;14:592–597. [PubMed: 14629691]

Craik FIM, Tulving E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General 1975;104:268–294.

Davis MH, Marslen-Wilson WD, Gaskell MG. Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance 2002;28:218–244.

Dumay N, Gaskell MG. Sleep-associated changes in the mental representations of spoken words. Psychological Science. 2006in press

Dumay, N.; Gaskell, MG.; Feng, X. A day in the life of a spoken word. In: Forbus, K.; Gentner, D.; Regier, T., editors. Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society; Mahwah, NJ: Erlbaum; 2004. p. 339-344.

Gaskell MG, Dumay N. Lexical competition and the acquisition of novel words. Cognition 2003;89:105–132. [PubMed: 12915296]

Graf Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science. 2006in press

Gupta P. Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. Quarterly Journal of Experimental Psychology: Human Experimental Psychology 2003;56A:1213–1236.

Gupta P, Lipinski J, Abbs B. Space aliens and nonwords: Stimuli for investigatin the learning of novel word-meaning pairs. Behavior Research Methods, Instruments & Computers 2004;36:559–603.

Johnston M, McKague M, Pratt C. Evidence for an automatic orthographic code in the processing of visually novel word forms. Language and Cognitive Processes 2004;19:273–317.

Kraljic T, Brennan SE, Samuel AG. Accommodating variation: Dialects, idiolects, and speech processing. Cognition. 2006submitted

Kraljic T, Samuel AG. Perceptual learning for speech: Is there a return to normal? Cognitive Psychology 2005;51:141–178. [PubMed: 16095588]

Kraljic T, Samuel AG. Even the perceptual system falls for first impressions and good excuses. 2006In preparation

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing 1998;19:1–36. [PubMed: 9504270]

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The time course of spoken word learning and recognition: Studies with artificial lexicons. Journal of Experimental Psychology: General 2003;132:202–227. [PubMed: 12825637]

Marslen-Wilson, W. Activation, competition and frequency in lexical access. In: Altmann, GTM., editor. Cognitive models of speech processing. Cambridge, MA: MIT Press; 1990. p. 148-172.

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology 1986;18:1–86. [PubMed: 3753912]

McKague M, Pratt C, Johnston MB. The effects of oral vocabulary on reading visually novel words: a comparison of the dual-route-cascaded and triangle frameworks. Cognition 2001;80:231–262. [PubMed: 11274984]

Meyer DE, Schvaneveldt RW. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology 1971;90:227–234. [PubMed: 5134329]

Naigles LR. Form is easy, meaning is hard: resolution of a paradox in early child language. Cognition 2002;86:157–199. [PubMed: 12435535]

Nation, P.; Waring, R. Vocabulary size, text coverage, and word lists. In: Schmitt, N.; McCarthy, M., editors. Vocabulary: Description, acquisition, pedagogy. New York: Cambridge University Press; 1997. p. 6-19.

Norris D. SHORTLIST: A connectionist model of continuous speech recognition. Cognition 1994;52:189–234.

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. Cognitive Psychology 2003;47:204–238. [PubMed: 12948518]

Pitt MA, Samuel AG. Word length and lexical activation: Longer is better. Journal of Experimental Psychology: Human Perception and Performance 2006;32:1120–1135. [PubMed: 17002526]

Rueckl JG, Olds EM. When pseudowords acqure meaning: Effect of semantic association on pseudoword repetition priming. Journal of Experimental Psychology: Learning, Memory, and Cognition 1993;19:515–527.

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science 1996;274:1926–1928. [PubMed: 8943209]

Samuel AG. Phoneme restoration: Insights from a new methodology. Journal of Experimental Psychology: General 1981a;110:474–494. [PubMed: 6459403]

Samuel AG. The role of bottom-up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance 1981b;7:1124–1131. [PubMed: 6457110]

Samuel AG. Red herring detectors and speech perception: In defense of selective adaptation. Cognitive Psychology 1986;18:452–499. [PubMed: 3769426]

Samuel AG. Does lexical information influence the perceptual restoration of phonemes? Journal of Experimental Psychology: General 1996;125:28–51.

Salasoo A, Shiffrin RM, Feustel TC. Building permanent memory codes: Codification and repetition effects in word identification. Journal of Experimental Psychology: General 1985;114:50–77. [PubMed: 3156946]

Schiller, NO.; Meyer, AS. Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities. Mouton de Gruyter; 2003.

Storkel HL. Learning new words: Phonotactic probability in language development. Journal of Speech, Language, and Hearing Research 2001;44:1321–1337.

Warren RM. Perceptual Restoration of missing speech sounds. Science 1970;167:392–393. [PubMed: 5409744]

Whittlesea BWA, Cantwell AL. Enduring influence of the purpose of experiences: Encoding-retrieval interactions in word and pseudoword perception. Memory & Cognition 1987;15:465–472.

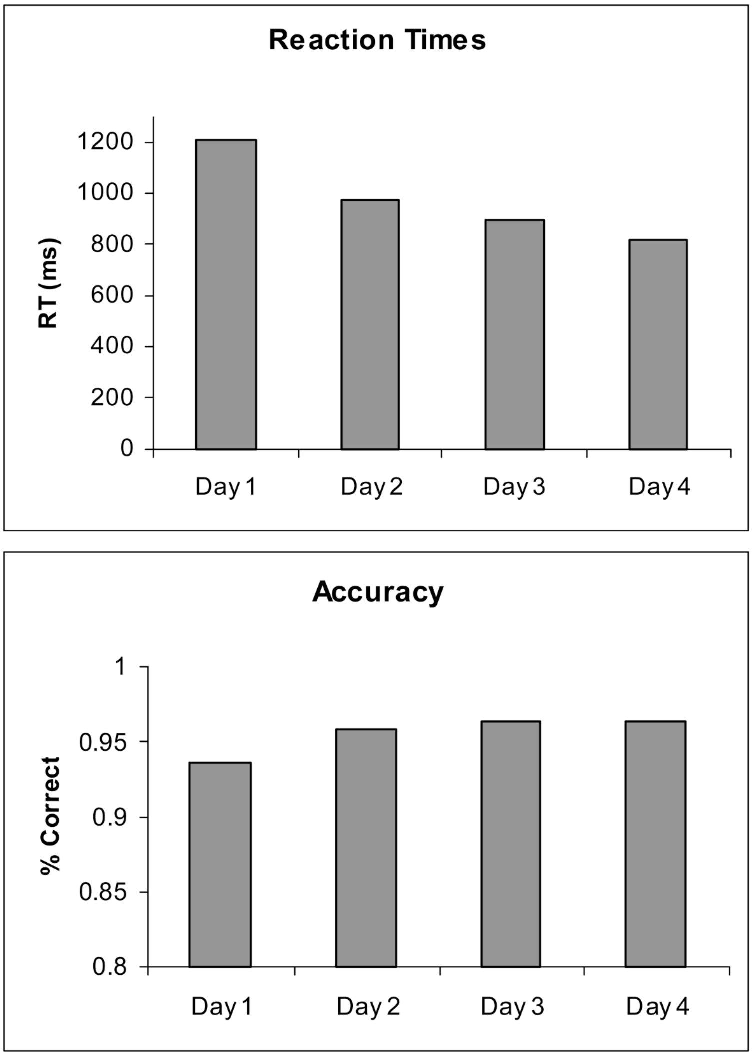


Figure 1.

Experiment 1: Phoneme monitoring reaction times and accuracy

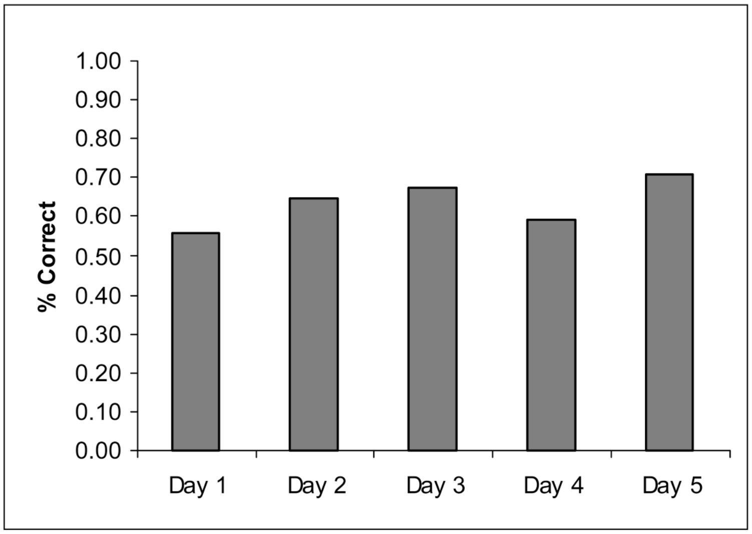


Figure 2.
Experiment 1: Three alternative forced choice recognition accuracy

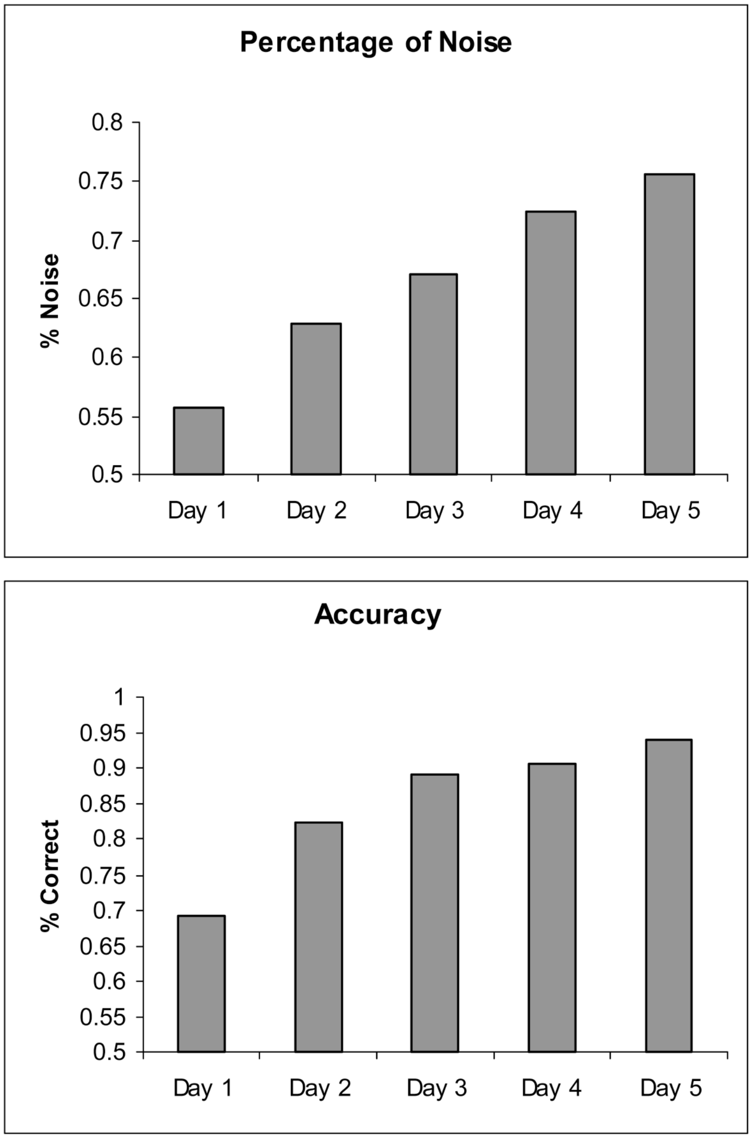


Figure 3.
Experiment 1: Threshold task, average noise level and average accuracy for recognizing items in noise

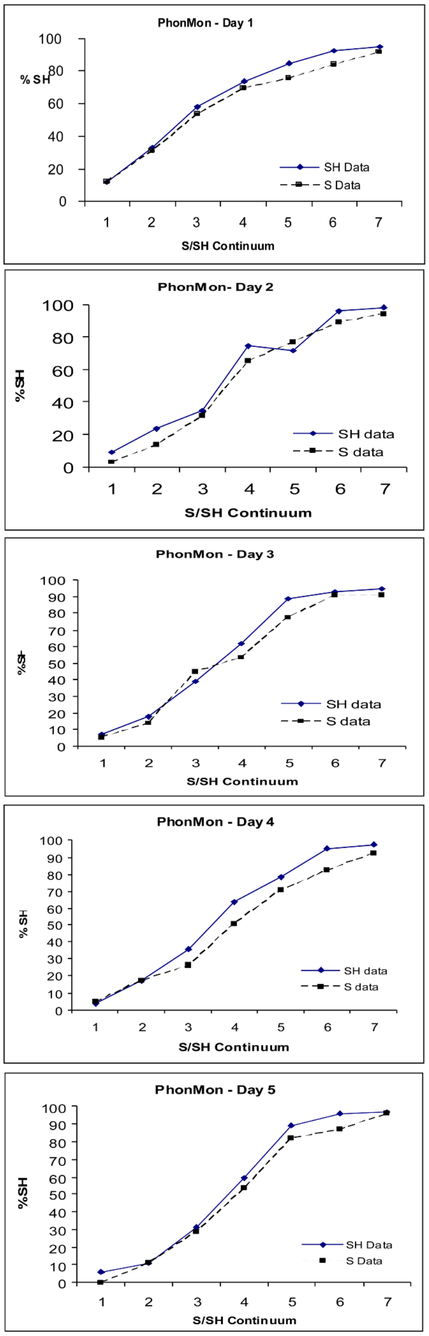


Figure 4.

Experiment 1: Perceptual learning task, averaging labeling as "sh", for groups who would have enlarged the "sh" category (solid line) or the "s" category (dashed line)

Figure 5.

Experiment 2: Pictures that were assigned to the 12 new words

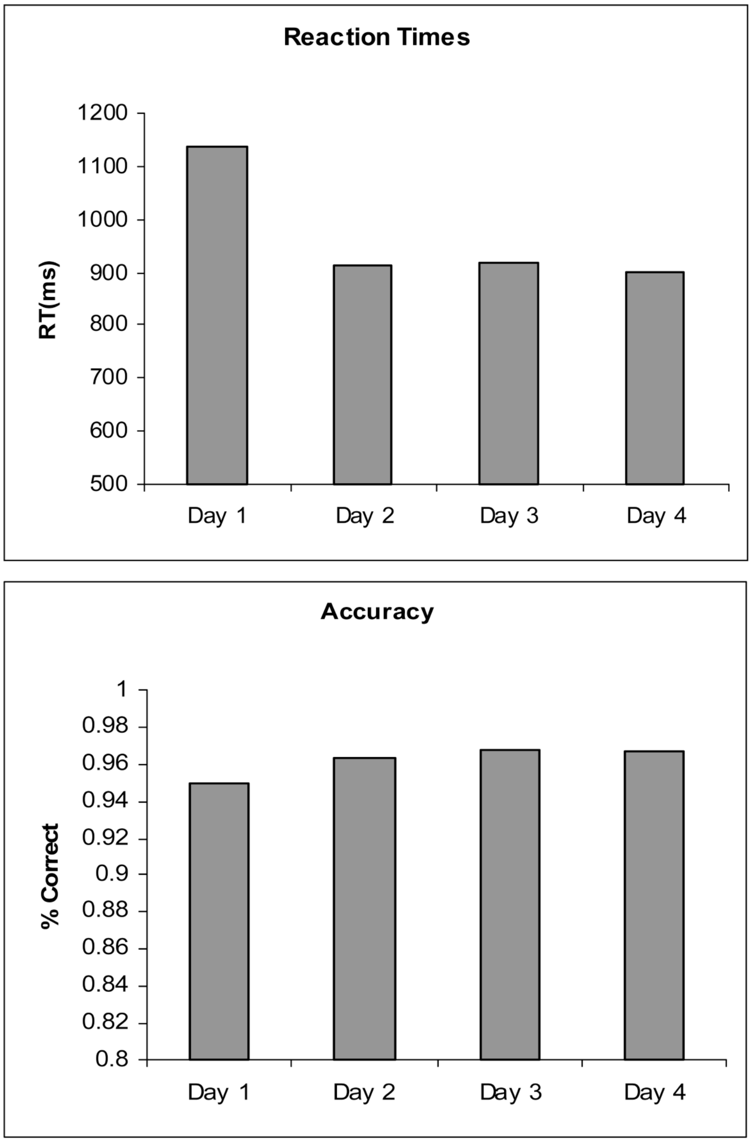


Figure 6.
Experiment 2: Response times and accuracy for selecting pictures

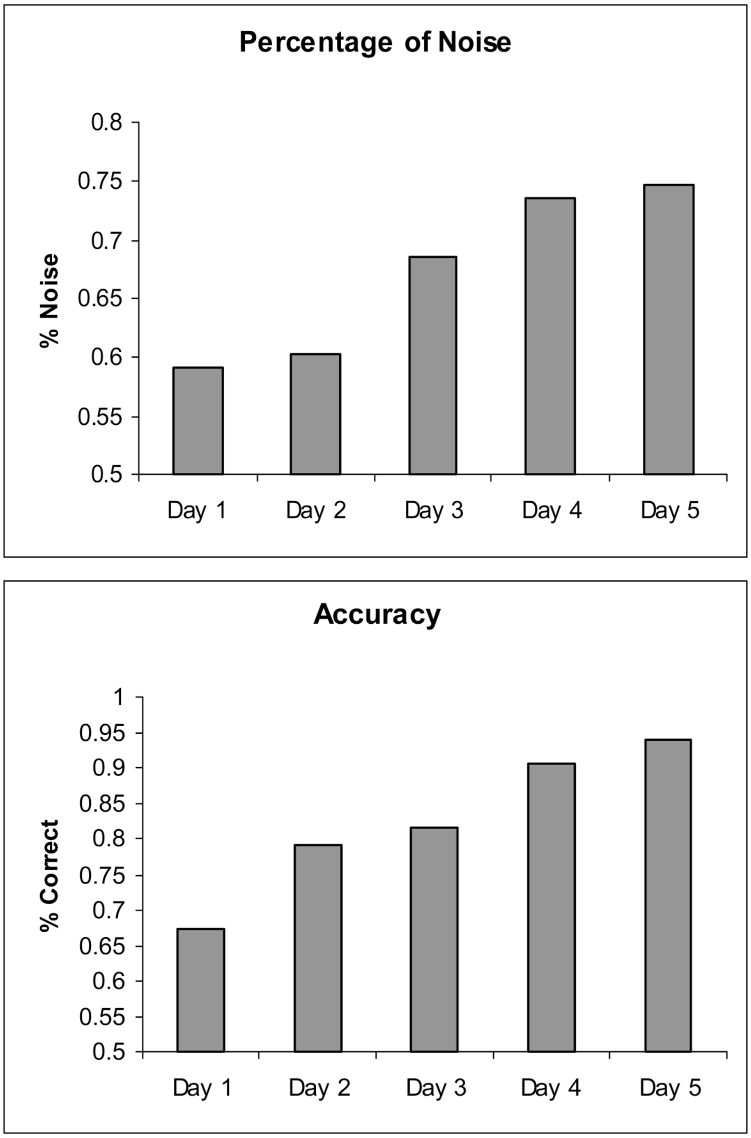
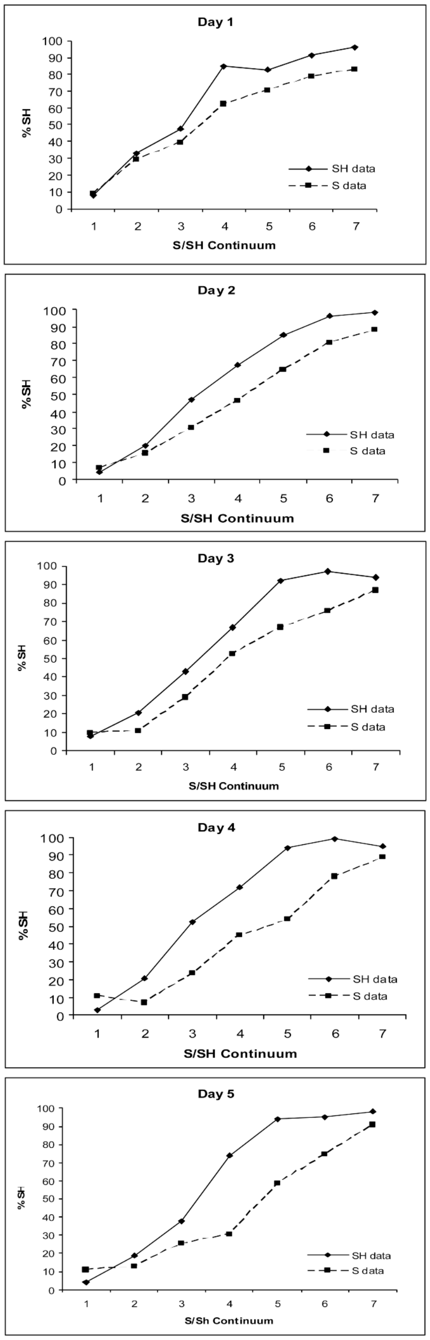


Figure 7.
Experiment 2: Threshold task, average noise level and average accuracy for recognizing items in noise



Day 1

Day 2

Day 3

Day 4

Day 5

Figure 8.

Experiment 2: Perceptual learning task, averaging labeling as “sh”, for groups who would have enlarged the “sh” category (solid line) or the “s” category (dashed line)

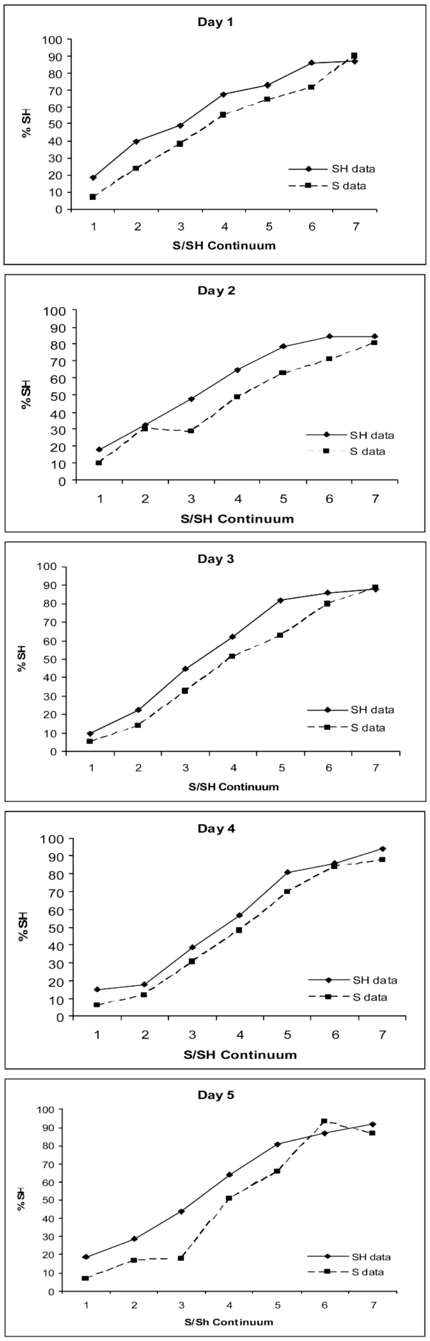


Figure 9.

Experiment 3: Perceptual learning task, averaging labeling as “sh”, for groups who would have enlarged the “sh” category (solid line) or the “s” category (dashed line)

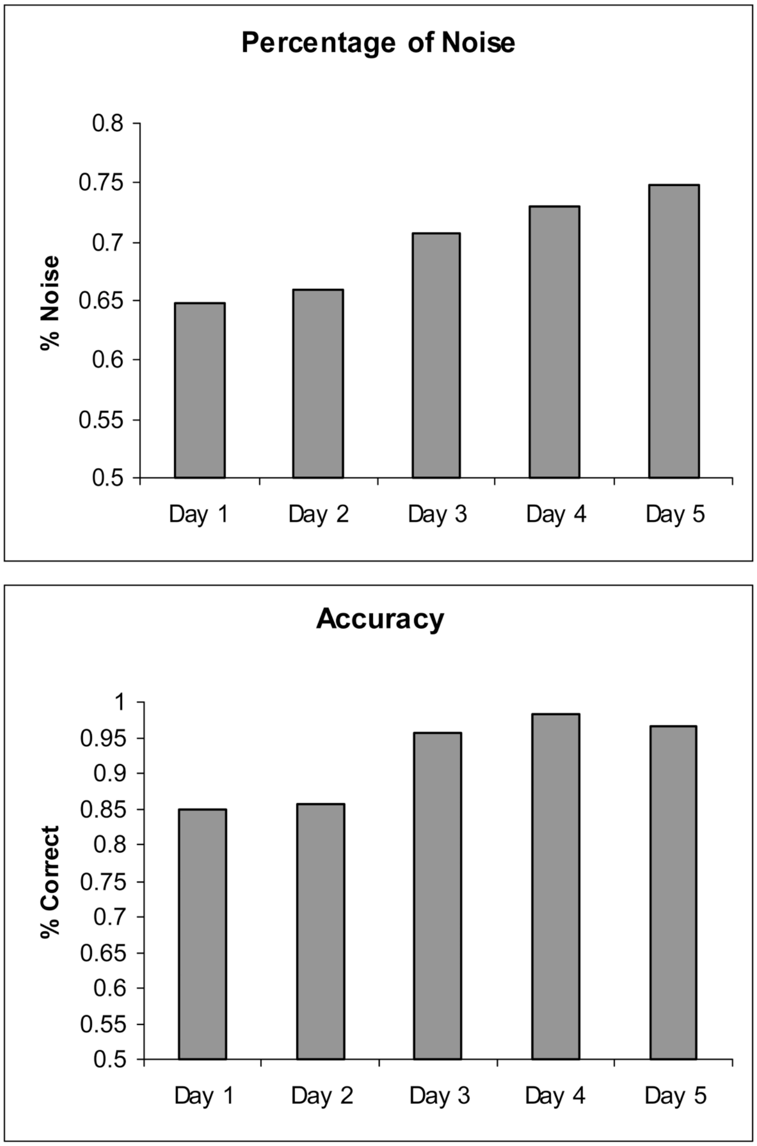


Figure 10.

Experiment 4: Threshold task, average noise level and average accuracy for recognizing items in noise

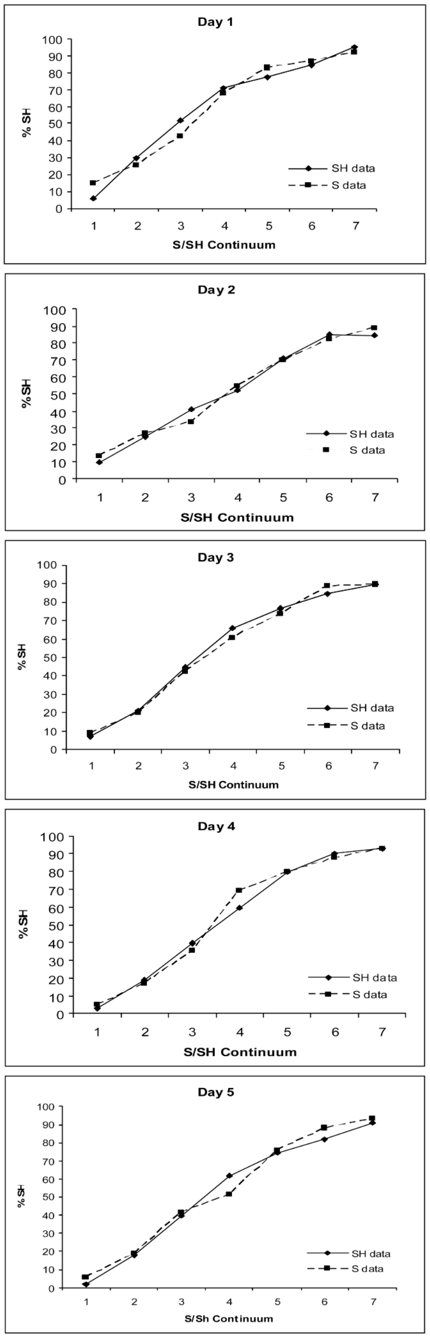


Figure 11.

Experiment 4: Perceptual learning task, averaging labeling as “sh”, for groups who would have enlarged the “sh” category (solid line) or the “s” category (dashed line)

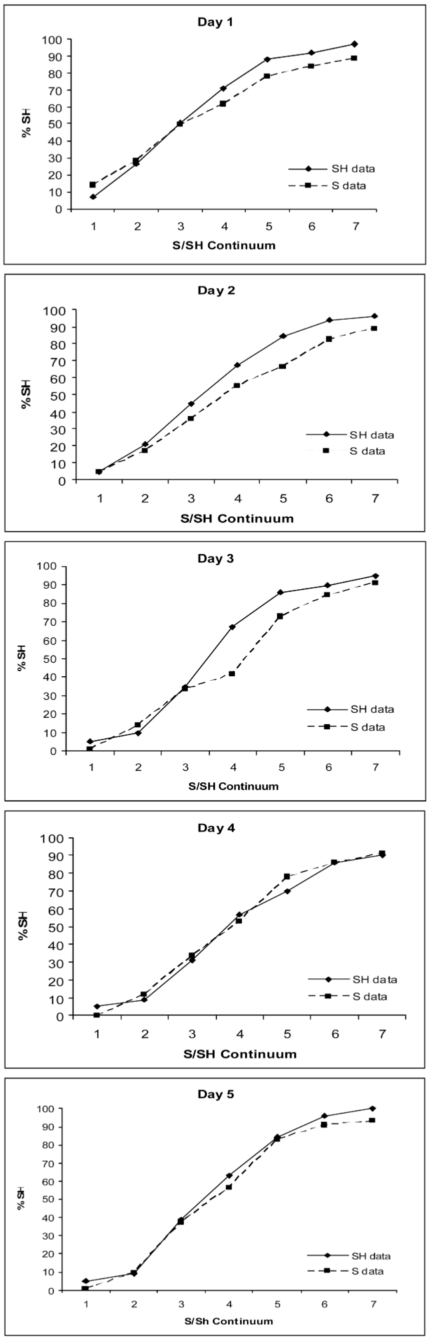


Figure 12.

Experiment 5: Perceptual learning task, averaging labeling as “sh”, for groups who would have enlarged the “sh” category (solid line) or the “s” category (dashed line)

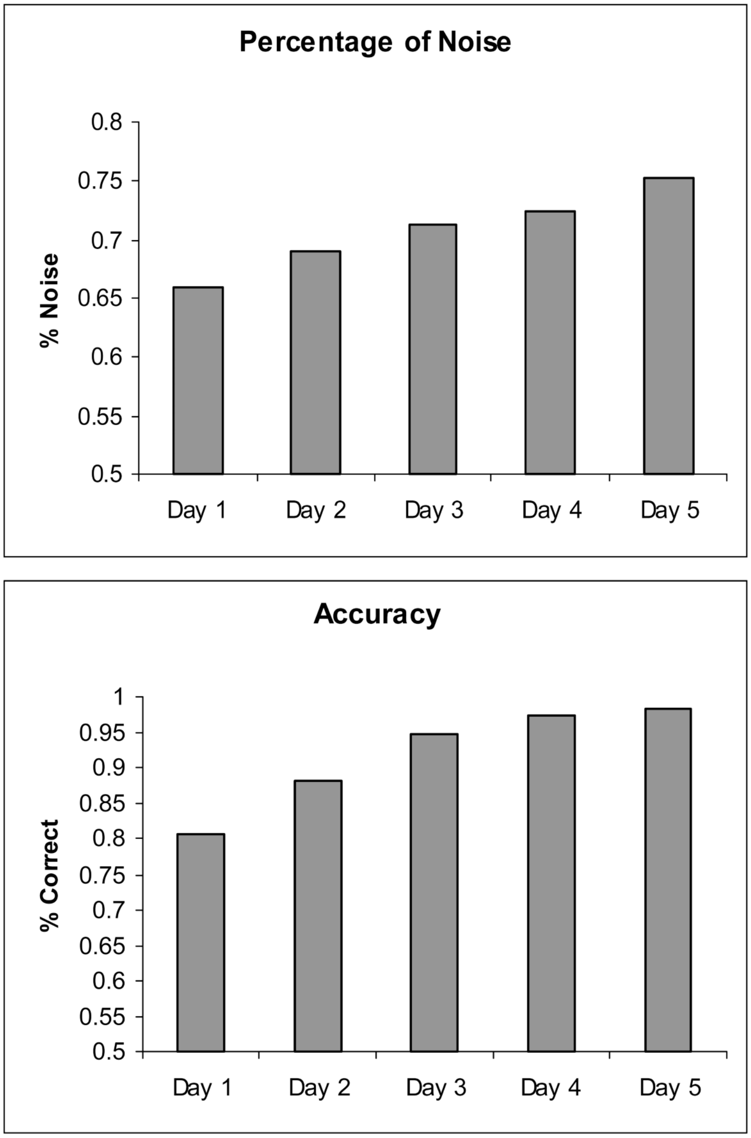


Figure 13.
Experiment 5: Threshold task, average noise level and average accuracy for recognizing items in noise

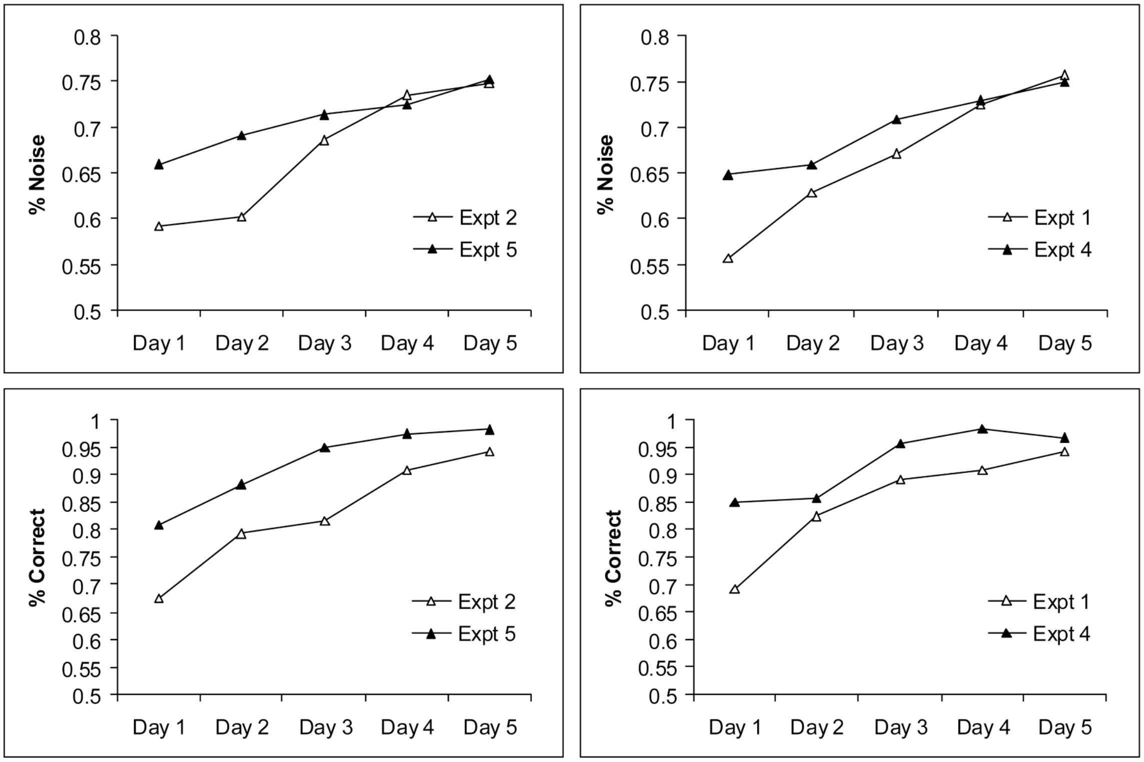


Figure 14.
Between experiment comparison of threshold task performance

Table 1

The twelve critical new words

Words with /s/	**Words with /ʃ/**
lifrisen	wickoshah
gatersy	bibershack
galasod	marfashick
figondalis	benemshalow
naronesay	penivasher
nomemsoly	komalsheum

Table 2

The targets and foils used in the three-alternative forced choice task

Critical Item	Foil #1	Foil #2
marfashick	marfashil	marpashick
bibershack	bizershack	bibershav
gatersy	gabersy	gaterfee
galasod	gapasod	galasot
lifrisen	lifriken	lifrisib
wikoshah	wifoshah	wikogah
nomemsoly	nomeksoly	nomempoly
figondalis	figonbalis	figondalit
benemshalow	benezshelow	benemvelow
komalsheum	komarsheum	komalsheul
penivasher	penikasher	penivawer
naronesay	narobesay	naronetay

Table 3

Picture choice accuracy, with and without a production requirement

Experiment	Production?	Day 1	Day 2	Day 3	Day 4
2	No	95.0	96.3	96.8	96.6
5	Yes	85.7	96.3	97.4	97.5