Published in final edited form as:
Cogn Psychol. 2014 May ; 70: 1–30. doi:10.1016/j.cogpsych.2014.01.001.

How Lexical is the Lexicon? Evidence for Integrated Auditory Memory Representations

April Pufahl[a] and **Arthur G. Samuel**[a,b,c]

April Pufahl: april.pufahl@stonybrook.edu; Arthur G. Samuel: arthur.samuel@stonybrook.edu

[a]Department of Psychology, 100 Nicholls Road, Stony Brook University, Stony Brook, NY 11794-2500, USA, Tel: 1 631 632 7792, Fax: 1 631 632 7876

[b]IKERBASQUE Basque Foundation for Science, Alameda Urquijo, 36-5, Plaza Bizkaia, 48011 Bilbao, Bizkaia, Spain, Tel: +34 944 05 26 60

[c]Basque Center on Cognition Brain and Language, Paseo Mikeletegi 69, 2nd Floor, 20009 Donostia-San Sebastián, Gipuzkoa, Spain, Tel: +34 943 309 300, Fax: +34 943 309 052

Abstract

Previous research has shown that lexical representations must include not only linguistic information (what word was said), but also indexical information (how it was said, and by whom). The present work demonstrates that even this expansion is not sufficient. Seemingly irrelevant information, such as an unattended background sound, is retained in memory and can facilitate subsequent speech perception. We presented participants with spoken words paired with environmental sounds (e.g., a phone ringing), and had them make an “animate/inanimate” decision for each word. Later performance identifying filtered versions of the words was impaired to a similar degree if the voice changed or if the environmental sound changed. Moreover, when quite dissimilar words were used at exposure and test, we observed the same result when we reversed the roles of the words and the environmental sounds. The experiments also demonstrated limits to these effects, with no benefit from repetition. Theoretically, our results support two alternative possibilities: 1) Lexical representations are memory representations, and are not walled off from those for other sounds. Indexical effects reflect simply one type of co-occurrence that is incorporated into such representations. 2) The existing literature on indexical effects does not actually bear on lexical representations – voice changes, like environmental sounds heard with a word, produce implicit memory effects that are not tied to the lexicon. We discuss the evidence and implications of these two theoretical alternatives.

Address Correspondence to: Arthur G. Samuel, Department of Psychology, 100 Nicholls Road, Stony Brook University, Stony Brook, NY 11794-2500, USA, Tel: 1 631 632 7792, Fax: 1 631 632 7876, arthur.samuel@stonybrook.edu.

Keywords

implicit memory; indexical effects; mental lexicon; priming; specificity effects; spoken word recognition

Fundamentally, the mental lexicon is a memory system: It is the place where language and memory meet. Most models of spoken word recognition (e.g., TRACE: McClelland & Elman, 1986; Shortlist: Norris, 1994; PARSYN: Luce, Goldinger, Auer, & Vitevitch, 2000; Distributed Cohort Model: Gaskell & Marslen-Wilson, 1997; 1999; 2002) assume the incoming speech signal is mapped onto abstract linguistic representations. As such, in these models, the input codes for lexical representations include only abstract phonological features that differentiate between words. One major challenge to this assumption comes from empirical evidence that speech recognition is sensitive to changes in surface characteristics such as the voice of the speaker – a set of properties that collectively constitute "indexical" information. These specificity effects have led to an expansion of the mental lexicon to include episodic features reflecting this indexical variation (Goldinger, 1996, 1998, 2007; Johnson, 1997, 2005, 2006; Palmeri, Goldinger, & Pisoni, 1993; Pierrehumbert, 2001; Sheffert, 1998). Other models have retained abstract linguistic representations but also included probabilistic information about their occurrence that can be altered based on input by a given speaker (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Norris & McQueen, 2008).

In the present study, we ask whether the expansion of the mental lexicon to include both linguistic and indexical information is sufficient. Arguably, voices are an important source of variation in comprehending spoken language, so the inclusion of indexical information in the mental lexicon may serve a pragmatic purpose. On the other hand, it is impossible to hear a word without also hearing the voice speaking it, so the inclusion of indexical information in the mental lexicon could simply be due to its co-occurrence with linguistic information. From this perspective, the indexical properties added to some speech recognition theories are not necessarily indexical *per se*, but simply properties that happen to be co-present with the linguistic information. To test this possibility, we compared the co-occurrence of words and voices to the co-occurrence of words and irrelevant environmental sounds. Given that speech and non-speech sounds are frequently encountered simultaneously, how does the system treat additional variation from this co-occurring non-speech? Do listeners discard variability in the incoming auditory signal that comes from non-human sources when attending to speech, or does this variability, like that from voices, persist in memory?

0.1 Talker Variability in Speech Perception

Previous research has shown that listeners retain speaker-specific auditory details in memory, and that these memories help facilitate future understanding of previously encountered speakers (for a review, see Luce & McLennan, 2005). These indexical effects refer to any performance advantage (e.g., improved accuracy or response time) for tokens repeated in the same voice (or with similar properties) over a different voice.

In a typical indexical study, participants first perform a task to encode the stimuli into memory. After some delay, they then complete a memory test with stimuli repeated in the same voice (or with similar properties) or in a different voice. Encoding tasks have varied in terms of depth of processing, such as classifying words according to the speaker's gender (shallow), initial phoneme (moderate), and syntactic class (deep; Goldinger, 1996). Other encoding tasks have drawn attention to the voice by requiring participants to identify the speaker (Allen & Miller, 2004; Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994) or to rate the pitch or clarity of pronunciation (Church & Schacter, 1994; Schacter & Church, 1992). These tasks contrast with others that do not require processing of voice characteristics, such as making a "word/non-word" lexical decision (Luce & Lyons, 1998), identifying the category to which a word belongs (Schacter & Church, 1992), or counting the number of meanings for a word (Church & Schacter, 1994; Schacter & Church, 1992). Overall, indexical effects appear to be relatively insensitive to task changes at encoding, as evidenced by a performance advantage for same-voice over different-voice items across levels of processing (Goldinger, 1996) and across tasks that do and do not draw attention to voice characteristics (Schacter & Church, 1992).

0.2 Explicit and Implicit Tests

In addition to task changes at encoding, memory tests have varied in the extent to which they overtly referred to the initial encoding; tests have been more explicit or more implicit in nature. Explicit memory tests assessed participants' conscious memory for the original stimuli, with a recognition test of "old/new" items (Church & Schacter, 1994; Goldinger, 1996; Luce & Lyons, 1998; Schacter & Church, 1992), a continuous recognition test (Bradlow, Nygaard, & Pisoni, 1999; Palmeri et al., 1993), or a cued recall test (Church & Schacter, 1994). Other recognition tests have required participants to identify both whether the item is old or new, and whether old items are repeated in the same voice or a different voice, by choosing either "new", "old – same", or "old – different" (Bradlow et al., 1999; Palmeri et al., 1993). Above-chance performance on these "old-same" and "old-different" judgments has demonstrated that participants can consciously access memories of words that include voice information, and can use this information to make judgments when asked to do so. However, when the judgment is simply "new" versus "old", changes in surface characteristics of words between exposure and test do not reliably lead to measurable differences in performance.

Studies using explicit tests to measure indexical effects have produced inconsistent results. Some researchers have found significant differences (Bradlow et al., 1999; Goldinger, 1996; Luce & Lyons, 1998; Palmeri et al., 1993) while others have not (Church & Schacter, 1994; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Schacter & Church, 1992). Explanations for these inconsistent results are complicated by the number of methodological differences between studies (see Goh, 2005 for a review). One explanation favors the transfer-appropriate processing approach (Roediger, 1990). According to this account, a change in surface characteristics will influence memory performance when perceptual processing of stimuli is encouraged during both encoding and test, so that the type of processing used is the same at encoding and test. Most explicit tests encourage conceptual or semantic processing over perceptual processing and are therefore less sensitive to changes in

surface characteristics. Consistent with this view, as processing during encoding becomes deeper (or more conceptual), explicit tests produce smaller indexical effects (Goldinger, 1996).

Implicit memory tests, specifically those that encourage perceptual processing, have proved more reliable for detecting indexical effects. Participants have performed better on stimuli repeated in the same voice, as compared to a different voice, on a variety of implicit tests. These include word identification tasks for filtered words or words presented in noise (Church & Schacter, 1994; Goldinger, 1996; González & McLennan, 2007; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Schacter & Church, 1992), stem-completion tasks (Church & Schacter, 1994; González & McLennan, 2007; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Schacter & Church, 1992), speeded classification tasks (Goldinger, 1996), and lexical decision tasks (Luce & Lyons, 1998). These findings are consistent with the transfer-appropriate processing approach (Roediger, 1990) because these implicit tests are sensitive to the kind of surface information (e.g., a speaker's voice) that characterizes indexical information. Given that implicit tests have proven to produce more reliable indexical effects, in the current study we used an implicit test in order to provide the most sensitive measurement.

0.3 Types of Changes between Exposure and Test

As we have noted, the most common manipulation of indexical properties has been a change of voice between exposure and test. However, the performance advantage for same-token exemplars extends to a variety of voice characteristics, including gender, emotional intonation (happy/sad), phrasal intonation (statement/question), fundamental frequency (Church & Schacter, 1994); speaking rate (Bradlow et al., 1999); and voice-onset-time (Allen & Miller, 2004). However, there was no advantage for tokens repeated at the same volume (Bradlow et al., 1999; Church & Schacter, 1994). The effect has also proven robust over time, with a significant advantage for same-voice tokens after a week delay (Goldinger, 1996). When listening to speech, people appear to have encoded not only what was said, but also how it was said and by whom, and these memories help facilitate future understanding of previously encountered speakers.

0.4 Expanding the Mental Lexicon

Indexical effects are problematic for traditional models of a lexicon composed of abstract linguistic representations. In order to accommodate variability due to speakers, such models have assumed a speaker normalization process, in which variability is stripped away and the input is "normalized" to map onto abstract representations. A strong view of speaker normalization would predict that voice details are not maintained in memory, whereas a weaker view would predict that voice details are maintained in a separate system and would not influence speech processing on-line. Contrary to both the strong and weak versions, indexical effects provide evidence that voice details are maintained in memory and can produce on-line effects during speech processing. This suggests that lexical representations may not be as abstract as previously thought.

Some theorists have looked at the variability due to speakers as a source of information, rather than a source of noise. This perspective is consistent with a lexicon composed of detailed episodic traces rather than abstract units (Goldinger, 1996, 1998; Johnson, 1997, 2005, 2006; Palmeri et al., 1993; Pierrehumbert, 2001; Sheffert, 1998). In Goldinger's (1998) model, these episodic traces have feature slots both for linguistic information and for indexical information. During speech perception, auditory input activates previous traces with matching features, producing an advantage for items repeated in the same voice over those in a different voice, because of the features that represent indexical properties. Episodic models like this have expanded the mental lexicon by including both linguistic and indexical information in the underlying representations. A related set of models retains abstract linguistic representations but also includes probabilistic information about their occurrence (Clayards et al., 2008; Norris & McQueen, 2008). For example, not only is a given phoneme stored, but also probabilistic information about the perceptual context in which occurs. In this way, these models are incorporating a kind of episodic information that could be extended to account for indexical effects.

0.5 Walling Off the Mental Lexicon

Alternatively, it could be argued that indexical effects have little to do with the lexicon. Instead of extending lexical representations to include episodic information, this alternative view would place any effects due to surface changes outside the lexicon, thereby retaining abstract linguistic representations. This approach keeps the lexicon purely lexical, but it then requires some additional theoretical apparatus to address how co-occurring surface information, such as that from voices, is stored in memory and can affect speech perception on-line. If the implicit (and sometimes explicit) memory for changes in indexical information demonstrated in the literature reviewed above stems from a detailed memory trace of the original episode at encoding, then the challenge for this approach is to place these representations somewhere outside the lexicon while still allowing them to influence speech perception online.

A specific and somewhat daunting challenge for this option is the need for an indexical advantage (better word recognition when an indexical property matches across initial exposure and later test) to be word-specific: Hearing a particular word in a particular voice, with a particular tone of voice, leads to improved recognition of *that same word* later if those properties recur. If the indexical advantage does not reside in the lexicon, then one must specify how a word-specific advantage could be represented elsewhere. We will consider this issue, in the context of the results of our six experiments, in the General Discussion.

0.6 Source Variability in Environmental Sound Perception

While the expansion of the mental lexicon to include both linguistic and indexical information (e.g., Goldinger, 1996, 1998) has provided an explanation of indexical effects, it remains a language-centric kind of idea, focused on only those aspects of auditory input integral to language. However, there are other sources of variability, besides those due to speakers, which can co-vary with speech. When perceiving speech in the real world, there

are usually co-occurring background sounds from a variety of environmental sources such as footsteps in the hall or traffic in the street. What effect, if any, does this additional variability have on the way spoken words are encoded? Most views of language processing call for linguistic information to be streamed to brain regions that are specialized for language processing, with all variability peripheral to speech (like a telephone ringing or a dog barking) filtered out and processed elsewhere, if at all. However, if the episodic view is followed to its logical outcome, then "tainting" of lexical representations by such extraneous sounds is a natural consequence of co-occurrence. If this is the case, then there may be nothing special about voices to warrant their inclusion in the mental lexicon. Perhaps the lexicon is not a unique form of storage specifically for words, but rather more like storage for auditory memories in general.

If lexical representations are in fact like other auditory memories, then there should be evidence for "indexical" effects for other types of sounds. Thus, the question is, are specificity effects unique to words, or do other sounds share the same properties? If sounds more generally share these properties, it would provide further support for the view that the mental lexicon is much like auditory memory generally. In fact, similar to the indexical effect found with words, Chiu (2000) found evidence for an exemplar specificity effect using environmental sounds (such as a doorbell, a helicopter, and a ticking clock). At encoding, participants rated five-second recordings of these sounds on familiarity or pitch. After a distractor task, participants identified one-second sound stems (i.e., the first second of the five-second sound) by writing down the name of the sound source. Critically, the test items were either the same exemplars presented during encoding, or different exemplars (e.g., a different doorbell). In addition, instructions given to participants were either implicit (write the first sound that comes to mind) or explicit (write the sound only if it was a previously heard sound, either an identical instance or another instance). Performance was better on both the implicit and explicit test when the same exemplar was repeated, rather than a new exemplar. This same-exemplar advantage for perception of environmental sounds parallels the same-voice advantage for perception of speech shown in the indexical literature reviewed above. González and McLennan (2009) have replicated Chiu's findings - participants were more accurate identifying sound stems when the exemplar was the same as during encoding rather than a new exemplar. Thus, it appears that indexical effects are not unique to words and voices, as similar effects have been found with environmental sounds.

0.7 The Present Research

The goal of the current series of experiments is to clarify the nature of the representations that underlie spoken word recognition, addressing several fundamental questions: Are the contents of lexical representations limited to abstract linguistic codes, or do they include information from the full auditory context in which a word was encountered? Is the expansion of the mental lexicon under episodic models, accommodating both linguistic and indexical information, sufficient? Are spoken words stored in a fundamentally different way than other sounds, or do words and sounds share similar representations and processing properties?

We first explore the effects of surface changes in co-occurring auditory information on the ability to recognize spoken words (Experiment 1) or environmental sounds (Experiment 2) under difficult listening conditions. We then examine the effects of repeated co-occurrence (words: Experiment 3; sounds: Experiment 4). Finally, we focus on perception of environmental sounds, testing a population that should have particular expertise with sound (Experiment 5) and testing whether increasing the perceptual distance of changes in co-occurring words produces effects more like those found in the indexical literature (Experiment 6).

Part I: What Co-occurs Together, Stays Together

In Experiment 1, we begin by testing whether indexical effects are limited to variability from voices, or if they can result from any change in the auditory input. As we have noted, the robust indexical effects in the literature have been taken as evidence that lexical representations must include voice-based information in addition to the more traditional sets of vowels and consonants. The goal of the first experiment was to determine whether variability from voices is included in the lexicon because of the integral relationship between speech and speaker, or if instead it is a consequence of the fact that speech and speaker necessarily co-vary. Outside of the laboratory, there are all manner of sounds that co-occur with words and voices in the auditory stream, such as birds chirping and cars honking. What does the auditory system do with this additional variability? In particular, when processing spoken language, is the system sensitive only to phonetic variation and indexical variation? Or is it sensitive to a much wider range of auditory input, even from sources that are irrelevant to language?

The experimental approach was simple: We extended previous indexical studies by having participants listen to spoken words that were each accompanied by a simultaneous unrelated environmental sound. Thus, we increased the variability in the auditory stream to include not only the phonetic and indexical variation from words and voices, but also variation from a specific exemplar of an environmental sound. Then, at test, we examined how changes to the surface characteristics of different aspects of the auditory stream affected spoken word identification. When participants heard the (degraded) word-sound pair at test, the test word was either identical (no change), spoken by a different voice (voice changed), mixed with a different exemplar of the same environmental sound (sound changed), or both spoken by a different voice and mixed with a different exemplar of the same environmental sound (voice and sound changed). Our central question is: Will a change in the co-occurring sound affect word identification performance in the same way that a change in voice affects performance?

In Experiment 2, participants were exposed to the same stimuli (co-occurring spoken words and environmental sounds) that were used in Experiment 1. However, in Experiment 2 our focus was on perception of environmental sounds, and how their recognition is affected by variations in prior exposure. Across the two experiments, we had subjects make the same judgment about each word-sound pair, and only varied whether we had them make this judgment about the word (Experiment 1) or the sound (Experiment 2). To keep things matched this way, we chose a semantic judgment: Was the item referred to animate or

inanimate? For example, if an exposure trial was the word “sparrow” combined with the sound of a doorbell, in Experiment 1 a subject should respond “animate” (a sparrow is animate), and a subject in Experiment 2 should respond “inanimate” (a doorbell is inanimate). As we discussed above, a wide range of exposure tasks have been used successfully in the indexical literature, including both semantic and nonsemantic tasks (Schacter & Church, 1992).

In both Experiments 1 and 2, the exposure task was followed by a short filler task, and then a perceptual identification task. Like the exposure task, the final identification task was identical across the two experiments, differing only in whether subjects were instructed to deal with the words (Experiment 1) or the sounds (Experiment 2). In both cases, we presented heavily filtered versions of the pairs that had been presented during the exposure phase, and asked subjects to identify either the words (Experiment 1) or the sounds (Experiment 2). Based on the research reviewed in the introduction, we expect to observe a performance cost during the perceptual identification test for surface changes in the attended information (i.e., a different-voice cost for words and a different-exemplar cost for sounds). The open question is whether we will also observe a performance cost for surface changes in the unattended information, and if this cost is of a similar magnitude. It is possible that since features of unattended information are suppressed (Mesgarani & Chang, 2012) the performance cost for surface changes of unattended information may be smaller in magnitude than changes of attended information.

Experiment 1

1.1 Method

1.1.1 Participants—Seventy-three undergraduates from Stony Brook University participated in exchange for course credit or $10 payment. Eight participants were excluded for responding to the environmental sounds instead of the spoken words during the “animate/inanimate” decision task. One participant was excluded because she indicated she was not a native speaker of English. All remaining participants identified themselves as native speakers of English.

1.1.2 Materials—Stimuli were constructed from audio recordings of spoken words and environmental sounds. Half of the words referred to animate things (e.g., “butterfly”, “rabbit”) and half referred to inanimate ones (e.g., “microwave”, “hammer”). Similarly, half of the sounds were from animate sources (e.g., a cow mooing), and half were from inanimate sources (e.g., a ping-pong ball). See the Appendix for the complete list of stimuli.

All words were recorded by both a male and a female speaker in a sound attenuated chamber. Thus, there were two versions of each word (male or female voice). The environmental sounds were adapted from Gregg and Samuel (2008) or gathered from various online sources. Similar to the words, there were two versions of each sound (exemplar 1 and exemplar 2), for example, a small dog and a large dog barking, or two different melodies played on a piano. Words and sound names were matched by animacy for average number of syllables (animate words $M = 1.66$, sounds $M = 1.53$; inanimate words $M = 1.97$, sounds $M = 1.91$). All words and sounds were filtered to remove background noise,

edited to a maximum of one second in length, and converted to a sampling rate of 16 kHz (Goldwave, version 5.55). Additionally, amplitude ramps were imposed on the first and last 10ms of the sounds to avoid abrupt onsets and offsets.

A list of 64 experimental pairs, preceded by eight practice pairs, was created in which pairs of words and sounds were equally distributed across the possible combinations of congruency, voice, and exemplar. For congruency, words and sounds were paired such that there were equal numbers of animate words with animate sounds, animate words with inanimate sounds, inanimate words with animate sounds, and inanimate words with inanimate sounds. For voice, words were spoken equally by the male and female voices, and, for exemplar, sounds were distributed equally between exemplars 1 and 2. Items were randomly assigned to pairs according to these criteria. A second exposure list was then created by switching the voices and/or exemplars used in the pairs (e.g. switching the voice from male to female).

1.1.3 Exposure Phase—Participants sat in a sound-attenuated chamber and listened to the paired word and sound stimuli. On each trial, the word and sound were digitally mixed and the result was played binaurally over headphones at a comfortable listening level. The experimenter instructed participants to pay attention to the spoken words and to ignore the background sounds. For each word, they were told to make an “animate/inanimate” decision. The experimenter defined animate and inanimate and gave category examples of each (e.g., birds, instruments). Participants responded on a button board with two buttons labeled “animate” and “inanimate.” Participants were given a maximum of three seconds to respond, with a one-second inter-trial interval that was timed from the participant’s response, or from the end of the three-second response window if no response was made. Eight practice trials preceded 64 experimental trials. Order of trials was randomly determined for each participant.

Although the simultaneous presentation necessarily meant that the environmental sound partially obscured the spoken word, participants scores were near ceiling (93%) on the animacy task, indicating they had little difficulty understanding the spoken word. We predetermined a cutoff of 80% accuracy on the animacy task to ensure that participants paid attention and completed the task as instructed. All eight participants who were eliminated had an accuracy score around 50% because they responded based on the animacy of the environmental sound instead of the spoken word.

It appears that the remaining participants had little difficulty attending to the spoken words and ignoring the background sounds, as instructed. If participants experienced interference from the background sounds, we would expect to observe more errors on the animacy judgment when the spoken word and environmental sound were incongruent (e.g., an animate word paired with an inanimate sound). However, this was not the case. Accuracy on congruent (M = 93.9%, SE = 0.5%) and incongruent (M = 92.9%, SE = 0.6%) pairs was comparable, $t(4050) = 1.25$, $p = 0.21$. Furthermore, because congruency was counterbalanced across the key manipulation of exposure-test match (described in the Test Phase section below), any differences in error rate would not affect our measurement of indexical effects or their extension to co-occurring sounds. Finally, as we noted previously,

effects in the indexical literature have generally been robust across a range of implicit memory tests; in experiments in our lab (Pufahl & Samuel, 2013) that used a quite different exposure task (a size judgment), we have found similar results to those we report here.

1.1.4 Delay Phase—In order to ensure that performance during the test phase was not based on short-term or working memory, participants spent five to ten minutes on an unrelated distractor task prior to the memory test. Participants were given a sheet with 24 semantic illusions, like the Moses Illusion. In this illusion, when people are asked the question "How many animals of each kind did Moses take on the ark?" they generally respond "two" even though they know it was Noah, not Moses, who built the ark (Erickson & Mattson, 1981). Participants wrote their answers on the sheet below each question and circled "yes" or "no" to indicate if they had ever heard the question before.

1.1.5 Test Phase—To assess the consequences of the exposure phase, participants completed a word identification test on filtered versions of the original (unfiltered) word-sound pairs heard during exposure. We varied how well each test pair matched the corresponding exposure episode, with some pairs presented in different voices or with different instances of the same environmental sound. For example, if the sound was a dog barking, it changed to a different dog; similarly if the sound was a piano, it changed to a different melody. As illustrated in Figure 1, there were four possible relationships between an exposure pair and a test pair. For example, if during exposure participants heard the word "termite" spoken in the male voice paired with the large dog barking, the corresponding test pair could be one of four possibilities: 1) the word "termite" spoken in the male voice paired with the large dog barking (no change); 2) the word "termite" spoken in the female voice paired with the large dog barking (voice changed); 3) the word "termite" spoken in the male voice paired with the small dog barking (sound changed), or 4) the word "termite" spoken in the female voice paired with the small dog barking (voice and sound changed). Eight test lists (four test lists for each of the two exposure lists) were created in which word-sound pairs were counterbalanced for exposure-test match across participants.

Additionally, all words and sounds were filtered to make perceptual identification challenging. We chose to use filtering rather than embedding the paired words and sounds in white noise, because previous work has demonstrated that noise embedding can reduce indexical effects (Schacter & Church, 1992). Previous indexical studies have used low pass or high pass filtering for this purpose (Church & Schacter, 1994; González & McLennan, 2007; Luce & Lyons, 1998; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Schacter & Church, 1992; Sheffert, 1998), but a slightly different approach was needed here for the broad range of frequencies of the environmental sounds. Based on pilot testing of the words and sounds, we applied multiple band-pass filtering (Goldwave, version 5.55) using the following frequency bands: 200–250Hz, 400–450Hz, 600–650Hz, 800–850Hz, 2000–2500Hz, 4000–4500Hz, 6000–6500Hz, and 8000–8500Hz. The pilot testing was used to find a set of filtering bands that would lower recognition of the speech and of the wide range of environmental sounds roughly equally. This property was desirable because it ensured that during the recognition test the words and the sounds would be similarly affected, rather than having one or the other remain audible while the other one largely disappeared.

For the word identification test, the experimenter told participants that they would hear the same words and sounds they had heard during the “animate/inanimate” decision task, but that the pairs would now be muffled and difficult to hear. Participants were instructed to write each word they heard, guessing if necessary, on a numbered answer sheet. The word identification test was self-paced. After writing a response, participants pressed a button to move on to the next trial. As in the exposure phase, there were eight practice trials followed by 64 experimental trials. Order of trials was randomized for each participant. Responses to the word identification test were scored generously, allowing for minor deviations or spelling errors (e.g., accepting “staple” for “stapler”). During scoring, the experimenter was blind to which words appeared in each condition for each participant.

1.1.6 Posttest Phase—Our goal in the perceptual identification test was to measure how well people could perceive degraded words as a function of their prior exposure to the words. However, it is at least conceivable that participants might use explicit strategies during this test, especially since the test was self-paced and allowed participants time to think before writing down the word they heard. Perhaps, instead of retrieving the word based on the auditory cues alone, they could use the paired environmental sound as a cue to the word’s identity. For example, if participants heard the dog barking, they might explicitly remember that the bark was previously paired with the word “termite” and therefore guess “termite.” To assess this possibility, we added a post-test phase for the final 32 participants in which we measured how well they knew particular pairings of words and sounds.

Words and sounds were filtered just as they had been in the test phase. However, unlike the test phase, changes made to the pairings were always identity changes of the environmental sound. Half the words remained paired with the same identity and instance of the environmental sound (old combination), e.g., the word “termite” remained paired with the big dog barking. The other half of the words switched pairs (new combination). For example, if the word “rabbit” was originally paired with a chainsaw and the word “vulture” was originally paired with a drum roll, in the posttest, the pairings were changed such that “rabbit” was paired with the drum roll and “vulture” was paired with the chainsaw. These changes were always within-animacy; in this example, inanimate sounds switched with inanimate sounds.

The experimenter instructed participants that they would hear the same muffled words and sounds once more. The task was to remember which sound the word was combined with. Participants pressed a button labeled “old combination” if the same word and sound were heard together previously, and pressed a button labeled “new combination” if the word was previously combined with a different sound. They were instructed to guess if unsure. As in previous phases, there were eight practice trials and 64 experimental trials, after which the experimenter debriefed the participants. Order of trials was randomized for each participant.

1.2 Results and Discussion

Previous research (Church & Schacter, 1994; Goldinger, 1996; Nygaard & Pisoni, 1998; Nygaard et al., 1994; Schacter & Church, 1992) has shown that word identification under challenging conditions is better if the word is presented in the same voice as in a previous

exposure. The central theoretical question in Experiment 1 is whether a change in an unrelated accompanying sound also affects spoken word identification. We modeled our data using a generalized linear mixed model fit by the Laplace approximation. Analyses were conducted in R (R Core Team, 2013), using the *lmer* function within the *lme4* package (Bates, Maechler, & Bolker, 2012). Since our dependent variable was binomial (correct or incorrect, scored as 1 or 0), we used the binomial linking function by specifying family=binomial in the model. We modeled participants' word identification accuracy by including random intercepts for both subjects and items and the fixed effect of exposure-test match, which had four levels (no change, voice changed, sound changed, voice and sound changed) dummy coded such that the no change condition was the intercept. We chose this dummy coding because the critical comparisons are between each of the conditions in which we changed the surface characteristics and the baseline no-change condition; this method can and will also be applied in the subsequent experiments. We report the means and standard errors, as well as the model estimates (*b*), standard errors (SEs), *z* values, and *p* values.

We tested whether the inclusion of the fixed effect exposure-test match significantly improved the overall fit of the model, as compared to a model with only random effects, using a log-likelihood ratio test (Baayen, 2008). As predicted, the degree to which the surface characteristics of word-sound pairs matched between exposure and test accounted for a small but significant 1.3% of the variance in participants' word identification accuracy at test, χ^2 (3) = 18.31, $p < .001$. Participants were most accurate identifying words when surface characteristics did not change between exposure and test (M = 64.7%, SE = 1.5%). Table 1 presents participants' average accuracy identifying the words in the four conditions (along with the results for Experiment 3, to be discussed below).

Our central question is what happens when the voice remains the same, but there is a change in the accompanying environmental sound. The answer is clear: There was a significant 8% drop in word identification accuracy when the background sound changed (M = 57.1%, SE = 1.5%, b = −0.43, SE = 0.11, z = −4.02, $p < .001$), relative to when both the voice and sound exemplar remained the same. There was also a significant 5% drop in word identification accuracy when both the voice and the background sound changed (M = 59.3%, SE = 1.5%, b = −0.30, SE = 0.11, z = −2.82, p = .005). Furthermore, there was no difference between the sound changed and the voice and sound changed conditions, as evident in a model in which the levels of exposure-test match were recoded such that sound changed became the intercept, b = 0.13, SE = 0.11, z = 1.21, p = .23).

Curiously, although performance was worse in the voice change condition than in the no change case, the 2% drop in word identification accuracy did not reach significance (M = 62.7%, SE = 1.5%, b = −0.13, SE = 0.11, z = −1.22, p = .22). This is clearly a Type 2 error, as this classic indexical effect is reliable in Experiment 3 (see below), and in several additional experiments we have run using similar materials and procedures (Pufahl & Samuel, in preparation). Across the various experiments we have run, the traditional indexical effect averages about 6%.

We designed the sound-change condition to be formally identical to the voice-change manipulation that has been the hallmark of the indexical literature, a literature that has implicated richer lexical representations. The results of Experiment 1 implicate lexical representations that are even more episodic than episodic models have suggested. Those models expanded lexical representations to include aspects of a speaker's voice along with traditional linguistic information. This expansion was motivated by the observed drop in word recognition performance when voice properties changed between initial exposure and later test. The fact that, like changes in voice, a similar drop occurs when a completely separate environmental sound changes between exposure and test indicates that what is critical is simple co-occurrence, rather than any properties integral to the spoken word. As we noted in the Introduction, indexical effects (and now, the environmental sound effect) require either surprisingly detailed/episodic lexical representations, or a locus for the effects outside the lexicon that still allows word-specific representation of the co-occurring nonlinguistic information.

Could this pattern of results be due to subjects reconstructing what a test word had been by remembering the sound that it was paired with? Recall that we included a posttest to measure how well subjects knew these pairings. The results of this old/new combination posttest indicated that participants were not able to explicitly remember the word-sound pairs. A one-sample *t*-test confirmed their accuracy was no different from chance (50%): Participants correctly identified the old and new combinations as such 50.9% (*SE* = 0.8%) of the time, $t(2036) = 0.82$, $p = 0.41$. These results suggest that it is very unlikely participants could have successfully used the paired environmental sounds as an associated cue to the word's identity in any strategic way.

Another possible issue concerns masking effects during the exposure phase. Due to the simultaneous presentation of the word-sound pairs, each potentially obscured the other to some extent. We believe that it is very unlikely that such masking can account for our results. First, any masking was relatively weak, as it did not appear to diminish performance on the unfiltered pairs heard during exposure; participants performed near ceiling (93.4%) on the animacy task. Performance was low at test due to the filtering (by design), not to a failure to hear the words in the exposure phase. Second, items were counterbalanced for exposure-test match across participants, reducing the likelihood that the changes in pairings led to an advantage in the segmental integrity of the words in one condition over another.

To provide a more definitive and empirical assessment of potential masking effects during exposure, we conducted a follow-up experiment ($n = 36$) in which we manipulated how much time the word and sound overlapped, splitting the items into high- and low-overlap. High-overlap pairs included sounds like a chainsaw, which filled the entire one second stimulus duration. Low-overlap pairs included sounds like a duck quacking, in which the one second included two quacks with pauses before, between, and after. During these pauses, portions of the co-occurring word could be heard unobscured. As would be expected, words in the low-overlap pairs were easier to recognize than words in the high-overlap pairs, $F(1, 35) = 14.00$, $p = .001$, partial $\eta^2 = .29$. Critically, however, the pattern of results due to exposure-test match was unaffected by overlap, as evidenced by the absence of an interaction, $F(2, 70) = 0.08$, $p = .93$; the cost of changing the voice and/or the

accompanying environmental sound was essentially the same for the low-overlap pairs and the high-overlap pairs. For low-overlap words, there was a 6–7% cost for surface changes (no change: 74.0%, voice changed: 67.4%, sound changed: 67.7%). For high-overlap words, there was a 4–5% cost for surface changes (no change: 65.6%, voice changed 60.8%, sound changed 61.8%). This resulted in a marginally significant trend for exposure-test match, $F(2, 70) = 2.91$, $p = .06$, partial $\eta^2 = .08$, which we analyzed with pairwise comparisons (adjusted using Fisher's Least Significant Difference). Replicating the results of previous indexical studies, and our new finding in Experiment 1, we observed performance costs due to surface changes in both attended (voice changed) and unattended information (sound changed) present in the auditory stream, $p = .04$ and .06 respectively.

Experiment 2

The purpose of Experiment 2 was to see if the pattern of results with words found in Experiment 1 generalized to environmental sounds. A new set of participants listened to the stimuli from Experiment 1 and performed the same "animate/inanimate" decision task during encoding. However, they now judged the animacy of the sound exemplar rather than the spoken word. At test, participants identified the sound exemplar by writing the name of the thing making the sound, rather than the spoken word. Thus, Experiment 2 parallels the design of Experiment 1, but with participants attending to the sound portion of the paired word-sound stimuli.

Previous research has demonstrated an exemplar effect with sounds similar to the indexical effect found with words. Chiu (2000) reported better performance identifying environmental sounds when the sound presented at test was the same exemplar that was presented at exposure. The facilitated performance occurred both when the test sounds were presented in noise, and when they were shortened from the five-second exposure version to a one-second test version. In Experiment 2, we aimed to replicate this exemplar effect. In addition, we tested whether a change in an unrelated co-occurring spoken word would affect sound source identification, in the same way that a change in an unrelated co-occurring background sound affected spoken word identification in Experiment 1. If sounds behave like words, then these effects are not a unique property of the mental lexicon, but instead a general property of auditory memory.

2.1 Method

2.1.1 Participants—Seventy-two undergraduates from Stony Brook University participated in exchange for course credit or $10 payment. All participants identified themselves as native speakers of English. Five participants were excluded for responding to the spoken words instead of the environmental sounds during the "animate/inanimate" decision task. Three additional participants were excluded for writing the spoken words during the sound source identification task.

2.1.2 Materials—Stimuli were the same as in Experiment 1.

2.1.3 Exposure Phase—The procedure was identical to Experiment 1, except that participants now attended to the sounds, ignoring the spoken words. During the "animate/

inanimate" decision task, participants judged the animacy of the thing making the sound. There were two exposure lists that each included 64 experimental pairs preceded by eight practice pairs.

2.1.4 Delay Phase—Participants completed the same distractor task used in Experiment 1.

2.1.5 Test Phase—As in the word identification task participants performed in Experiment 1, participants in Experiment 2 performed a sound source identification task on filtered versions of the sound-word pairs. The experimenter instructed participants to write the name of the sound source (i.e., the thing making the noise), and to try to be as specific as they could. They were reminded that they heard several birds and instruments during the "animate/inanimate" decision task, and that therefore they should try to identify the type of bird or instrument if possible.

As in Experiment 1, we varied how well a test pair matched the corresponding pair heard during exposure, with some pairs presented with different exemplars of the environmental sound or with a different voice speaking the word. There were four possible relationships between an exposure pair and a test pair, as illustrated in Figure 1. For example, if at exposure participants heard the large dog barking paired with the word "termite" spoken in the male voice, at test this pair could appear in one of four ways: 1) the large dog barking paired with the word "termite" spoken in the male voice (no change); 2) the small dog barking paired with the word "termite" spoken in the male voice (sound changed); 3) the large dog barking paired with the word "termite" spoken in the female voice (voice changed); or 4) the small dog barking paired with the word "termite" spoken in the female voice (sound and voice changed). Sound-word pairs were counterbalanced by exposure-test match across participants, such that each pair appeared in all four conditions.

As in Experiment 1, responses to the sound source identification task were scored generously (e.g., accepting variations like "car horn", "car", "horn", "traffic", or "car honk") if they seemed to uniquely identify that sound. Minor spelling deviations were also scored as correct (e.g., "symbols" for "cymbals"). Non-specific responses such as "animal", "instrument", "bird", "bugs", or "machine" and descriptions of the sound such as "beeping", "bark", "breathing", "buzzing", "growl", "music" or "popping" were scored as incorrect. Additionally, all responses using the word stimuli were scored as incorrect (e.g., "wolf" was not accepted for "coyote", or for any other item, since "wolf" was one of the spoken words). During scoring, the experimenter was blind to which sounds appeared in each condition for each participant.

2.1.6 Posttest Phase—As in Experiment 1, we added a posttest phase for the final 32 participants to assess whether participants could explicitly remember which word was paired with each sound. As in Experiment 1, half of the environmental sounds remained paired with the same word, spoken in the same voice ("old combination") and half the sounds switched which word they were paired with ("new combination"). Eight practice trials preceded 64 experimental trials. Order of trials was randomized by participant.

2.2 Results and Discussion

Based on the results of Experiment 1, we know that word identification performance is sensitive not only to changes in voice between exposure and test, but also to changes in unrelated co-occurring background sounds. In Experiment 2, we reversed the roles of words and sounds to see if sound source identification performance is sensitive to both changes in sound exemplars and changes in the voice speaking an unrelated word heard simultaneously. To assess these issues, we modeled the accuracy of participants' sound source identification responses in the same manner as Experiment 1, including the fixed effect of exposure-test match (no change, sound changed, voice changed, sound and voice changed) with random intercepts for both subjects and items.

Table 2 presents the average accuracy in identifying the sounds in the four conditions (along with the results for Experiments 4–6, to be discussed below). As shown in the table, overall accuracy at recognizing sounds was, as one would expect, lower than accuracy for recognizing words. Average sound identification performance when there was no change in surface characteristics between exposure and test was 38.2% (*SE* = 1.5%). As in Experiment 1, the degree to which sound-word pairs matched between exposure and test accounted for a small but significant 1.1% of the variance in participants' identification accuracy at test as compared to a model with only random effects, χ^2 (3) = 23.18, $p < .001$. We observed a robust exemplar effect, i.e. a 6% drop in sound source identification accuracy when the sound presented at test was a different exemplar than participants heard at exposure (*M* = 32.4%, *SE* = 1.5%, *b* = −0.40, *SE* = 0.12, *z* = −3.34, $p < .001$). In fact, how well the sound exemplar matched the sound presented during the exposure phase seems to be the sole predictor of participants' performance, with no observed effect of changing the voice speaking the co-occurring word. Participants showed comparable identification accuracy when there was no change and when the voice changed between exposure and test (*M* = 39.8%, *SE* = 1.5%, *b* = 0.12, *SE* = 0.12, *z* = 1.04, *p* = .30). Finally, participants showed a significant 4% drop in identification accuracy when both the sound and voice changed (*M* = 34.5%, *SE* = 1.5%, *b* = −0.25, *SE* = 0.12, *z* = −2.13, *p* = .03) relative to no change, similar to the effect found when only the sound changed. In fact, there was no difference between the sound changed and the sound-and-voice changed conditions, as evident in a model in which the levels of exposure-test match were recoded such that sound changed became the intercept, *b* = 0.15, *SE* = 0.12, *z* = 1.23, *p* = .22).

Results of the old/new combination posttest indicated that participants were not able to explicitly remember the sound-word pairs. A one-sample *t*-test confirmed their accuracy was no different from chance (50%). Participants correctly identified old and new combinations as such 51.3% (*SE* = 1.1%) of the time, *t*(1972) = 1.15, *p* = 0.25.

Paralleling Experiment 1, performance was best when identifying stimuli that were presented in pairs that matched the exposure phase. Changing the exemplar of the environmental sound produced a drop in sound identification performance (an exemplar effect), much as changing the voice speaking the word produced a drop in word identification performance (an indexical effect). Previous studies (Chiu, 2000; González & McLennan, 2009) showed the cost of a change of exemplar, and the results of Experiment 2

confirm that finding. Collectively, the data show that surface effects are not unique to words and voices in the lexicon, but rather are a general property of auditory memories.

Despite this similarity, the results of the first two experiments suggest a difference in the way words and sounds are stored. In Experiment 1, we found that a change in a co-occurring background sound impaired word identification performance. In Experiment 2, we found no such effect for sound source identification, with no performance cost for changing the voice speaking a co-occurring word. However, we need to consider two differences between the words and sounds used in our experiments. First, listeners generally have much more experience hearing and identifying words than they have for the environmental sounds. Second, the two sound exemplars paired with a word in Experiment 1 were generally more acoustically dissimilar than the two tokens of the spoken words paired with a sound in Experiment 2. In Part II, we test whether the frequency of recent exposure to particular sounds affects the influence of changing a co-occurring word or environmental sound. In Part III, we test a blind population that arguably relies more heavily on auditory input, both from speech and non-speech, when perceiving the world around them compared to sighted populations. In that sense, these subjects are more experienced listeners than those tested in the first two experiments. In addition, in Part III we test whether varying the perceptual distance between a co-occurring word at exposure and a co-occurring word at test affects its influence on sound recognition. The results of these experiments will guide our inferences regarding the representation and processing of words and environmental sounds.

Part II: Repetition and Co-occurrence: Experiment 3

Experiment 1 provided evidence that lexical representations may be more episodic than previously theorized, including not only variability from voices, but also from co-occurring sounds. In Experiment 3, we explore a property that might be associated with episodic representations of words: the effect of repetition of similar episodes. The notion is that repetition could enhance indexical and exemplar effects by increasing the influence of recent episodes among a lifetime's experience with these common words. For example, participants are likely to have heard a word like "termite" spoken by many speakers across a variety of contexts, but would have only heard it spoken by our male voice and paired with a large dog barking in the context of the experiment. In an episodic model, this would mean there is one episode of "termite" with the surface characteristics we presented during exposure along with multiple episodes of "termite" in other contexts. If all these episodes of "termite" are activated at test, then presumably the surface characteristics of the one episode heard during exposure will have relatively little weight among all episodes. But, if there were more episodes of "termite" with the surface characteristics used in our experiment through repetition, then perhaps there will be a greater cost in word identification accuracy when the pair heard at test is not an exact match to those surface characteristics.

In Experiment 3, we aim to replicate the drop in word identification performance when the voice changes from exposure to test (the indexical effect found in previous studies) and when the background sound changes (found in Experiment 1) relative to when both the voice and background sound remain the same. In addition, we test if repeated exposure, i.e. hearing the word-sound pair twice, four times, or eight times during exposure, enhances the

magnitude of this drop in word identification performance when surface characteristics change from exposure to test. Experiment 3 also includes an "unpaired word" condition that provides an upper bound on word recognition performance, when any masking by an accompanying sound is removed. Since our design requires that the words be partly obscured by the co-occurring sound, this unpaired condition will index participants' highest performance and provide an indication of how much of a benefit can be expected as a result of increased repetition.

3.1 Method

3.1.1 Participants—Sixty-five undergraduates from Stony Brook University participated in exchange for course credit or $10 payment. All participants identified themselves as native speakers of English. One participant was excluded for responding to the environmental sounds instead of the spoken words during the "animate/inanimate" decision task.

3.1.2 Materials—The words and sounds in the Exposure Phase, Delay Phase, and Test Phase were the same as those used in Experiment 1. However, we created new word-sound pairs (i.e., different pairings than those used in Experiments 1 and 2), randomly assigning items such that they were equally distributed across the possible combinations of congruency (animate-animate, animate-inanimate, inanimate-animate, inanimate-inanimate), exemplar (exemplars 1 and 2), and voice (male and female speakers). Furthermore, stimuli were counterbalanced across participants such that each item appeared in each of the 16 combinations of exposure-test match and levels of repetition. Thus, there were 16 sets of lists used during exposure and test.

3.1.3 Procedure—The procedure followed that of Experiment 1 with two changes. First, we eliminated the exposure-test match condition where both the voice and the background sound changed. We used the stimuli freed up by eliminating this condition to include test trials in which a word was presented alone, with no environmental sound. Second, we added repetition as a variable, so that participants now heard each pair either once, twice, four times, or eight times (1x, 2x, 4x, 8x) during the exposure phase, resulting in a total of 240 presentations across the 64 experimental pairs. Order of items was randomly determined for each participant with repeated presentations randomly spaced throughout. As in Experiment 1, each of the 64 experimental items was heard only once during the final word identification test, with the same filtering as before, again with order of items randomly determined for each participant.

3.2 Results and Discussion

In Experiment 1, we found that changes in irrelevant background sounds between exposure and test decreased word identification performance just as changes in the voice of the speaker have in previous experiments. In our analyses of Experiment 3, we tested whether the same pattern occurred, and whether multiple exposures to the pairs affected the results. We modeled the accuracy of participants' word identification responses in the same manner as Experiment 1, including the fixed effects of exposure-test match (dummy coded: no change, voice changed, sound changed, unpaired word), repetition (numerically coded and

such that 1, 2, 4, and 8 repetitions became 0, 1, 3, 7 so the intercept reflected the baseline of 1 exposure) and their interaction. We included random intercepts for both subjects and items. We used forward selection to determine the best fitting model, beginning with a null model including only the random intercepts, then adding the fixed effects and interaction until the fit did not improve. To assess the full set of simple comparisons among the four levels of exposure-test match, we reordered the dummy variables in the model, so that each condition served as the intercept.

The two models including the fixed effects of either exposure-test match or repetition both improved the fit as compared to a model with only random effects, χ^2 (3) = 313.86, $p < .001$ and χ^2 (1) = 13.04, $p < .001$ respectively. Furthermore, each fixed effect accounts for a unique portion of the variance, as adding the fixed effect of repetition to exposure-test match (and vice versa) improved the fit of the model, χ^2 (1) = 14.35, $p < .001$ and χ^2 (3) = 315.17, $p < .001$ respectively. Finally, the inclusion of the interaction term did not improve the fit beyond that of the model with the two fixed effects, χ^2 (3) = 4.53, $p = .21$. Therefore, the best fit for the data included the fixed effect of exposure-test match and the fixed effect of repetition but not the interaction. This model accounted for 24.7% of the variance in participants' word identification accuracy.

As shown in Table 1, results for changes in exposure-test match replicated those found in Experiment 1. Collapsing across the repetition factor, participants accurately identified 72% of the words when there was no change in surface characteristics from exposure to test (M = 71.6%, SE = 1.4%). Compared to this baseline, there was an 8% drop in word identification accuracy when the voice changed between exposure and test (M = 63.5%, SE = 1.5%, b = −0.50, SE = 0.11, z = −4.57, $p < .001$). There was a 4% drop in word identification accuracy when the background sound changed (M = 67.5%, SE = 1.5%, b = −0.25, SE = 0.11, z = −2.23, p = .03). Performance identifying words when the sound changed was impaired to a lesser degree than when the voice changed (z = 2.35, p = .02). Finally, performance was highest, and near ceiling, for the newly added condition in which the word was presented alone (M = 90.1%, SE = 0.9%, b = 1.64, SE = 0.14, z = 11.54, $p < .001$ as compared to no change).

For repetition, the pattern of results showed that word identification accuracy improved modestly with repeated presentation of the word-sound pairs at exposure, b = 0.06, SE = 0.02, z = 3.72, $p < .001$. As reported above, the interaction between repetition and exposure-test match was not significant, indicating that presenting a particular episodic pairing up to eight times does not produce a substantial change in the way that a word is retrieved from the lexicon beyond the change that a single pairing produces. However, numerically the largest difference from the no change condition was found for the 8× condition, for both the voice change, and for the sound change cases. Thus, we examined the effect of change for each of these cases individually. The analyses performed on these subsets of the data yielded a marginally significant interaction when the voice changes, χ^2 (1) = 3.21, $p < .07$, and no interaction when the sound changes, χ^2 (1) = 0.55, p = .46. For the voice change case, much of that marginal interaction is presumably driven by the oddly small effect for the 4× case, rather than by a systematic increase in the effect as a function of repetition.

If episodic memory representations are responsible for the specificity effects we observed in Experiment 1, then all other things being equal, exposing participants to additional episodes should enhance the effect. It is of course possible that eight repetitions were insufficient to enhance the influence of recent episodes compared to years of episodes of the common words chosen as stimuli. The repetition manipulation relies on there being a much heavier weighting of recent episodes over more remote ones. The effect of a single instance supports the importance of recent episodes. However, it is possible that only the most recent episode is highly influential. Alternatively, it is possible that only the first unique episode is highly influential, e.g., only the first time "termite" is heard in the male voice with the big dog barking, and not the subsequent repetitions of this pairing.

Experiment 4

In Experiment 3, we tested the effect of episodic repetition on the perception of spoken words. Now, in Experiment 4, we test the effect of episodic repetition on the perception of environmental sounds. As in Experiment 2, we ask whether the pattern of effects found with words is specific to the lexicon or is instead a property of auditory memory more generally.

There were three main objectives. First, Experiment 4 tests whether the effects found in Experiment 2 will appear with a new set of listeners with a different mapping of stimuli to conditions. Second, Experiment 4 tests the effect of eight episodes of pairing a sound with a particular word, versus a single episode. Although we did not find enhanced specificity effects for words repeated during exposure in Experiment 3, sounds may be more sensitive to repetition because in general they should be of lower frequency than words. Although both the word and sound stimuli we selected were fairly common items, presumably participants have less experience with the environmental sound stimuli than the word stimuli. As a result, recent episodes may exert a greater influence on the sound stimuli than the word stimuli. Third, as in Experiment 3, Experiment 4 has an "unpaired sound" condition that provides a best-case measure of sound identification performance. Since our observed sound identification performance was overall much lower than word identification performance, this unpaired condition will clarify participants' ability to identify filtered sounds, independent of additional noise in the signal from the paired word.

4.1 Method

4.1.1 Participants—Fifty -two undergraduates from Stony Brook University participated in exchange for course credit or $10 payment. Four participants were excluded as they indicated they were not native speakers of English. All other participants identified themselves as native speakers of English.

4.1.2 Materials—The words and sounds in the Exposure Phase, Delay Phase, and Test Phase were the same as those used in Experiment 2. Stimuli were assigned to new pairs and counterbalanced across participants such that each item appeared in each of the eight combinations of exposure test match and repetition. Thus, there were eight sets of exposure and test lists. Sound and word stimuli were once again randomly assigned to pairs such that they were equally distributed across the possible combinations of congruency (animate-

animate, animate-inanimate, inanimate-animate, inanimate-inanimate), exemplar (exemplars 1 and 2), and voice (male and female speakers).

4.1.3 Procedure—The procedure followed that of Experiment 2, with the same two changes we made to the parallel set of experiments (1 and 3) conducted with spoken words. First, as in Experiment 3, we eliminated the exposure-test match condition where both the environmental sound and the voice speaking the paired word changed and included test trials in which the environmental sound was presented alone, with no paired spoken word. Second, as in Experiment 3, we added repetition as a variable, so that participants now heard each pair either once or eight times (1x, 8x) during the exposure phase, resulting in a total of 288 presentations across all 64 experimental pairs (based on the modest effect of repetition in Experiment 3, we focused on the most extreme cases). As in Experiment 2, each of the 64 experimental pairs was heard only once during the sound source identification task. Order of items was randomly determined for each participant with repeated presentations randomly spaced throughout.

4.2 Results and Discussion

Table 2 reports the average sound recognition rates. We begin our analyses by looking at whether the results of the current experiment replicate the specificity effects reported in Part I. In particular, we examine whether the exemplar effect found for environmental sounds in Experiment 2 was also obtained in Experiment 4. Recall that in Experiment 2, participants attended to and identified environmental sounds, which were paired with co-occurring spoken words, and we found that changing the environmental sound (e.g., big dog to little dog barking) reduced the ability to recognize the sound when tested under heavy filtering. However, performance was not affected by changing the co-occurring word.

We modeled the accuracy of participants' sound source identification responses, including the fixed effects of exposure-test match (dummy coded: no change, sound changed, voice changed, unpaired sound) and repetition (numerically coded and such that 1 and 8 repetitions became 0 and 7 so the intercept reflected the baseline of 1 exposure) and their interaction. We also included random intercepts for both subjects and items. As in Experiment 3, we used forward selection to determine the best fitting model. To assess the full set of simple comparisons among the four levels of exposure-test match, we reordered the dummy variables within the model, so that each condition served as the intercept.

Including the fixed effect of exposure-test match improved the fit as compared to a model with only random effects, χ^2 (3) = 32.72, $p < .001$. However, including the fixed effect of repetition did not, χ^2 (1) = 2.41, $p = .12$. Furthermore, adding the interaction to the fixed effect of exposure-test match did not improve the fit, χ^2 (4) = 4.81, $p = .31$. Therefore, the best fit for the data included only the fixed effect of exposure-test match and accounted for a small but significant 2.1% of the variance in participants' sound source identification accuracy. Therefore, replicating Experiment 2, the match between surface characteristics of sound-word pairs heard during exposure and test significantly affected participants' sound source identification accuracy at test. However, repetition had no effect.

As shown in Table 2, results for changes in exposure-test match replicated those found in Experiment 2. Collapsed across repetition, participants once again showed an exemplar effect, with an 8% drop in sound identification performance when the sound changed (M = 35.7%, SE = 1.7%, b = −0.50, SE = 0.13, z = −3.83, $p < .001$) relative to when there was no change between exposure and test (M = 43.4%, SE = 1.8%). As we found before, compared to the no change baseline, there was no observed performance cost when the voice of the co-occurring word changed between exposure and test (M = 44.3%, SE = 1.8%, b = 0.05, SE = 0.13, z = 0.41, p = .68). Finally, sound source identification performance was highest in the newly added unpaired sound condition, with participants recognizing about 47% of the sounds (M = 46.6%, SE = 1.8%). However, this did not significantly exceed performance when there was either no change (b = 0.21, SE = 0.13, z = 1.69, p = .09) or when the voice changed (b = 0.16, SE = 0.13, z = 1.28, p = .20).

The present experiment replicated the core findings from Experiment 2, showing an exemplar effect together with no drop in sound identification performance when the voice of a co-occurring word changed. Furthermore, as in Experiment 3, repetition at exposure did not enhance the specificity effects we observed with respect to changes in surface characteristics of auditory stimuli between exposure and test. In fact, there was no significant benefit of repetition for the sound-word pairs. As we suggested in Experiment 3, it seems that only one episode is effective, but additional ones do not produce much additional change.

An intriguing difference between Experiment 3 and Experiment 4 is the much bigger advantage for the unpaired word over the paired cases than the advantage for an unpaired sound over the paired cases. In fact, in the no change 8× condition in Experiment 4, performance was actually slightly better than in the unpaired 8× case. These results, together with the more robust effect of change found for the word items (both voice and sound changes mattered, versus only sound changes here), suggest that memory for spoken words may be more episodic than memory for environmental sounds. Both show significant effects of change between exposure and test, but the effects for words are more pervasive. This pattern is coupled with a much higher overall level of recognition for words, indicating that the more detailed episodic representations support better recognition.

Part III: Experience and Perceptual Distance

As we just noted, the results of the first four experiments suggest that the representations of environmental sounds may not retain as much episodic detail as those for words. However, as we noted in Experiment 2, listeners generally have much more experience hearing and identifying words than they have for environmental sounds. Perhaps the level of episodic detail in an auditory representation depends on the amount of experience the person has with the stimulus. If so, our tests may be underestimating the possible episodic nature of sound representations due to the relatively low level of experience most people have with these environmental sounds, relative to their experience with particular words. It is also possible that the experiments so far have underestimated the episodic nature of sound representations if the “changed word” condition is not a substantial enough change to produce a reliable difference.

With these possibilities in mind, we conducted two additional experiments to look for episodic effects for the environmental sounds. In Experiment 5, we test Blind participants who presumably rely more heavily on environmental sounds to perceive the world around them than sighted individuals do. If experience with sounds affects the level of episodic detail that is retained, Blind listeners may show more episodic results. Experiment 6 is motivated by the possibility that there were systematic differences in the perceptual distance between the two spoken word tokens used in the other experiments and the two environmental sound exemplars in those experiments. In particular, it is possible that the male and female versions of a spoken word are more perceptually similar than the two exemplars of each environmental sound (e.g., male and female tokens of “termite” may be acoustically more similar to each other than the sound of a large dog barking is to the sound of a small dog barking). In Experiment 6, we increased the perceptual distance between the paired spoken words heard at exposure versus test in order to see whether this will lead to an observable drop in sound identification performance when the (now quite different) co-occurring word changes.

Experiment 5

The results of the previous studies have demonstrated a dissociation between how changes in surface information affect the perception of words and sounds. In Experiments 1 and 3, word recognition was influenced by changes in voice as well as changes in co-occurring environmental sounds. In Experiments 2 and 4, sound recognition was only influenced by changes in the sound itself, and not by changes in co-occurring spoken words. This pattern is consistent with the view that word representations contain variability from co-occurring events, while sound representations lack as much episodic detail.

However, although we used common words and sounds, the words were more recognizable. Presumably, our participants had more experience identifying the spoken words than the environmental sounds. In Experiment 5, we replicate Experiment 2 using a population that presumably relies more heavily on, and therefore has more experience with, identifying environmental sounds: a Blind population. This experience may allow the perceptual system to optimize its use of the full variability in the incoming auditory stream.

5.1 Method

5.1.1 Participants—Twenty-three blind adults were recruited from the community served by Arizona State University. The age at which participants lost their eyesight ranged from birth to 52 years, with 12 who lost their eyesight early (before age 2) and 11 who lost their eyesight late (after age 7). Four participants (3 late and 1 early onset) had difficulty completing the tasks, as evidenced by low scores (63–77% accuracy) on the animacy task, and were dropped from the analysis. The resulting sample of only 19 subjects is smaller than the sample in our other experiments, an unavoidable consequence of the difficulty of finding Blind subjects to participate in the study. As will become clear, the sample was nevertheless sufficient.

5.1.2 Materials—Materials in the Exposure Phase, Delay Phase, and Test Phase were those used in Experiment 2.

5.1.3 Procedure—The procedure followed that of Experiment 2 with minor changes to accommodate the Blind participants. During the delay phase, an experimenter read the questions to the participant (e.g., the Noah's Ark question) and wrote their verbal responses. During the test phase, the participants responded verbally to identify the environmental sound and an experimenter wrote their response on the answer sheet; in Experiment 2, the listeners wrote their answers down themselves. Finally, there was no post-test phase.

5.2 Results and Discussion

As in previous experiments, we modeled the accuracy of participants' sound source identification responses by including the fixed effect exposure-test match (no change, sound changed, voice changed, sound and voice changed) and random intercepts for both subjects and items.

The fixed effect of exposure-test match improved the fit of the model as compared to a model with only random effects, χ^2 (3) = 28.07, $p < .001$, and accounted for a small but significant 4.2% of the variance in participants' sound source identification accuracy. We observed a robust exemplar effect, replicating Experiment 2. Participants showed a 14% drop in sound source identification accuracy when the sound presented at test was a different exemplar than participants heard at exposure (M = 28.0%, SE = 2.6%, b = −0.82, SE = 0.21, z = −3.86, $p < .001$), relative to when there was no change in surface characteristics between exposure and test (M = 42.1%, SE = 2.8%). Once again, how well the sound exemplar matched seems to be the sole predictor of participants' performance, with no observed effect of changing the voice speaking the co-occurring word. Participants showed comparable identification accuracy when there was no change and when the voice changed between exposure and test (M = 41.1%, SE = 2.8%, b = −0.22, SE = 0.20, z = −0.11, p = .92). Finally, participants showed a significant 14% drop in identification accuracy when both the sound and voice changed (M = 28.6%, SE = 2.6%, b = −0.83, SE = 0.21, z = −3.93, $p < .001$) relative to no change.

The 14% exemplar effect for the Blind participants is numerically much larger than the 6% drop we observed with undergraduate participants in Experiment 2. We conducted a post hoc comparison by modeling the data from Experiments 2 and 5 including the fixed effects of exposure-test match (no change, sound changed, voice changed, sound and voice changed) and population (sighted, blind) as well as their interaction. We included random intercepts for both subjects and items and used forward selection to determine the best fitting model. The fixed effect of exposure-test match improved the fit of the model as compared to a model with only random effects, χ^2 (3) = 44.99, $p < .001$, but population did not, χ^2 (1) = 0.46, p = .50. Critically, adding the interaction to the model including the fixed effect of exposure-test match provided the best fit for the data, χ^2 (4) = 9.49, p = .05, accounting for a small but significant 2.2% of the variance in participants' sound source identification accuracy. The model estimates indicate that the cost of changing the exemplar was greater for the blind population than the sighted population. In addition to the pattern of the fixed effect of exposure-test match described above and in Experiment 2, the Blind participants showed a greater drop in performance when the sound changed, b = −0.51, SE =

0.25, $z = -2.05$, $p = .04$, and when the sound and voice changed, $b = -0.66$, $SE = 0.25$, $z = -2.67$, $p = .008$.

The results of Experiment 5 suggest that experience plays a role in the representation of episodic detail. Practice may enhance perception by allowing the system to process and store a finer grain of episodic detail present in the incoming auditory stream. However, we should add two caveats. First, overall sound recognition performance was not significantly better for the Blind subjects than for the sighted ones. Second, and of more interest, the enhanced representation of episodic detail for the Blind subjects was only seen for the non-speech sounds – even for subjects who rely heavily on auditory perception, we saw no effect of changing the voice speaking the accompanying word. Whereas we have observed repeatedly that memory for spoken words is contaminated by co-occurring variability, so far we have not observed that memory for non-speech sounds shares this property.

Experiment 6

Our experiments on spoken word perception (1 and 3) show that word recognition is influenced by the sound that accompanies it. So far, our experiments on environmental sound perception (2, 4, and 5) show that environmental sound recognition is not influenced by the word that accompanies it. Thus, there appears to be an asymmetry in the way speech and non-speech sounds are stored.

An alternative explanation for this dissociation is that the change in a background sound was on average greater than the change of a background word. The sound exemplars were chosen to be highly dissimilar, so it would be clear to participants that these were different instances of the sound. For example, instead of using two exemplars of barking from the same dog, one exemplar was from a large dog and another from a small dog. Thus, the acoustic characteristics of the two exemplars for each sound were quite different. This was not the case for the spoken words because when the same word is produced, even by different speakers, the two tokens will necessarily share many acoustic characteristics. Thus, on average, it is plausible that the perceptual distance between the two sound exemplars was greater than that between the male and female tokens of the spoken words. Previous research (Goldinger, 1996) has shown that indexical effects are sensitive to the perceptual distance between speakers, with greater distances being correlated with lower word recognition rates, lower identification accuracy, and slower reaction times. It is possible that sound recognition, like word recognition, is influenced by co-occurring variability, but that in our experiments, the perceptual distance between the tokens of the spoken words was too small to show a measurable effect on performance.

In Experiment 6, we tested this possibility. To do this we created a condition in which the sound exemplar remained the same between exposure and test but the paired word changed both in identity and voice (e.g. from “peacock” spoken in the female voice to “moose” spoken in the male voice). The question in Experiment 6 is whether increasing the perceptual distance of the change in the co-occurring spoken word affects sound recognition in a manner similar to changing the sound exemplar itself.

6.1 Method

6.1.1 Participants—Fifty-one undergraduates from Stony Brook University participated in exchange for course credit or $10 payment. All participants identified themselves as native speakers of English.

6.1.2 Materials—Materials in the Exposure Phase, Delay Phase, and Test Phase were similar to those used in Experiment 2, with additional sounds and words to allow for 10 practice and 72 experimental pairs of sound and word stimuli. We created new sound-word pairs (i.e., different from the pairings used in all previous experiments) by randomly assigning items such that they were equally distributed across the possible combinations of congruency (animate-animate, animate-inanimate, inanimate-animate, inanimate-inanimate), exemplar (exemplars 1 and 2), and voice (male and female speakers). Furthermore, stimuli were counterbalanced across participants such that each item appeared in each of the three combinations of exposure-test match. Thus, there were three exposure lists and one test list. Again, order of items was randomly determined for each participant.

6.1.3 Procedure—The procedure followed that of Experiment 2, with a change in the way sound-word pairs matched between exposure and test. For one-third of the sound-word pairs there was no change to the sound exemplar or the spoken word between exposure and test, and for one-third of the sound-word pairs, the exemplar of the sound changed between exposure and test (e.g., from a large dog barking to a small dog barking). In the third condition, the sound exemplar remained the same but both the identity of the spoken word, as well as the voice speaking the word, changed between exposure and test. For example, if at exposure a participant heard a melody played on a piano paired with the female voice saying the word “peacock”, then the test pair could be the same melody on a piano paired with the male voice saying the word “moose”. By changing the word identity in addition to the voice, we increased the dissimilarity or perceptual distance between the two instances of the sound-word pairs.

6.2 Results and Discussion

We modeled the accuracy of participants’ sound source identification responses by including the fixed effect of exposure-test match (no change, sound changed, word and voice changed) and random intercepts for both subjects and items. To assess the full set of simple comparisons among the three levels of exposure-test match, we reordered the dummy variables within the model, so that each condition served as the intercept.

The best fit for the data included the fixed effect of exposure-test match as compared to a model with only random effects, χ^2 (2) = 7.30, p = .03, and accounted for a small but significant 0.3% of the variance in participants’ sound source identification accuracy. As shown in Table 2, the results for sound recognition now paralleled those found with word recognition. Participants showed an exemplar effect, with a significant 4% drop in sound identification performance when the sound changed (M = 35.9%, SE = 1.4%, b = −0.29, SE = 0.11, z = −2.61, p = .009) relative to when there was no change between exposure and test (M = 39.7%, SE = 1.4%). Critically, there was a similar and significant performance cost when both the identity and the voice speaking the co-occurring word changed between

exposure and test (M = 36.7%, SE = 1.4%, b = −0.22, SE = 0.11, z = −2.03, p = .04). Performance identifying sounds when the word and voice changed was no different from that when the sound exemplar changed (b = −0.06, SE = 0.11, z = −0.58, p = .56).

The results from Experiment 6 highlight the similarity between word recognition and sound recognition. Both are impaired when there are relatively large changes in surface characteristics of co-occurring information between exposure and test. These results provide support for the hypothesis that the lack of a performance cost in sound recognition due to changes in the voice of the co-occurring word reported in Experiments 2, 4 and 5 was due to these surface changes being too small to show a measurable effect on performance. When we changed both the identity of the word as well as the voice, adding more acoustic variability between the two instances heard at exposure and test, the effects were similar to those we have observed for words. This supports idea that indexical (words) and exemplar (sounds) effects, as well as the effect of changing co-occurring information, result from a general mechanism that applies to all auditory input: Words and environmental sounds are processed and represented in similar ways.

General Discussion

The extensive literature on indexical effects has repeatedly demonstrated that when listeners hear spoken words they encode more than just the succession of vowels and consonants – details of the speaker's voice and tone of voice are represented, as shown by impaired word recognition when those properties change from the initial exposure. In the current study, we have employed exactly those procedures that have been used in the indexical literature. With these procedures, we have two critical new findings: First, changing an accompanying environmental sound from exposure to test produces a cost in word recognition, just as with classic indexical changes. Second, when the change in an accompanying word is sizable, there is a comparable cost in environmental sound recognition.

The second finding bears on recognition of environmental sounds as a function of how those sounds differ from a previous instantiation. By running complementary experiments in which participants either focused on the spoken words or on the environmental sounds, we tested whether specificity effects are unique to the mental lexicon, or are a property of auditory perception more generally. Overall, the indexical effects we observed for spoken words were similar to the effects we observed for environmental sounds, with the caveat that participants were less accurate overall identifying environmental sounds than spoken words. Cohen, Evans, Horowitz and Wolfe (2011) also found that participants were less accurate at recognizing environmental sounds than speech clips. Clearly, spoken words are optimized for recognition in ways that environmental sounds are not; after all, language evolved as the primary method of human communication, and that communication depends on very high word recognition rates in ways that do not hold for most environmental sounds. Our results nonetheless indicate that there are very strong commonalities in the way that words and other sounds are encoded.

We believe that the most important result of the current study is our observation that changes in a co-occurring environmental sound affect how well listeners can recognize a

spoken word. In the Introduction, we outlined two very different accounts that might apply to such a result. In the following discussion, we will consider these two alternatives, and the issues that they raise. The first possibility is that the mental lexicon includes detailed episodic information that goes well beyond prior suggestions. The second alternative is that classic indexical effects do not actually inform us about lexical representations. In either case, we believe that a substantial theoretical re-evaluation is needed.

A Non-Lexical Lexicon?

A parsimonious interpretation of our results is that the mental lexicon may not be as lexical as originally conceived, at least to the extent that "lexical" implies word-related information and nothing else. The ability to identify previously heard words under difficult listening conditions is facilitated not only by keeping the voice consistent, but also by keeping irrelevant background sounds consistent. It is worth emphasizing here that classic indexical effects (and the effects in Experiments 1 and 3) *are effects on spoken word recognition*, and that the core function of the mental lexicon is to support word recognition (and production). Given this, the natural locus for these effects is the memory structure that holds the representations used to recognize spoken words – the lexicon. A particular indexical effect reported by Creel, Aslin, & Tanenhaus (2008) reinforces the connection between indexical effects and word recognition, and thus, the lexicon. Creel et al. found that lexical competition between words like "sheep" and "sheet" is reduced when they are consistently produced by different speakers, e.g., "sheet" is only produced by a male speaker and "sheep" is only produced by a female speaker. Lexical competition is a fundamental property of the lexicon, and finding that this competition is subject to indexical experience is strong evidence for a lexical locus of the indexical effect.

As we have noted, the impact of indexical variation on lexical access has led a number of researchers to argue for a lexicon composed of detailed episodic traces rather than abstract units (Goldinger, 1996, 1998; Johnson, 1997, 2005, 2006; Palmeri et al., 1993; Pierrehumbert, 2001; Sheffert, 1998). These previous expansions of the mental lexicon to include both lexical and voice information (e.g., Goldinger, 1998) are appropriate but are apparently not sufficient, as they cannot account for the drop in performance we observed by changing the surface characteristics of co-occurring sounds.

Goldinger's (1998) model could be expanded to include feature slots for not only voice information but also any co-occurring variability. With this modification, it would make sense to view the classic indexical effects as being a consequence of voices co-occurring with words, rather than something particular about voices being related to words *per se*. Taking this approach leads to a somewhat different conception of the lexical representations, moving away from the idea of having "slots" for particular types of information. Instead, the lexicon is seen as being composed of integrated memory representations, containing the acoustic information and variation received during speech perception. As Goldinger has shown, a system with these kinds of rich representations can be coupled with access processes that are sensitive to central tendencies to produce many of the behaviors that are associated with a more traditional lexicon. The access process relies on the idea that as lexical episodes accumulate, the features that remain constant across the episodes will form

a central tendency, so that a new input will tend to resonate with the more central part of the lexical collection. A useful analogy is the additive process used in ERP studies in which the sum of many waveforms, each of which has considerable noise, produces an emergent pattern because all of the between-item variation that is not relevant tends to sum to zero, leaving the relevant peaks and valleys intact.

There are alternative models that take a different approach to deriving the central tendency of a set of inputs that could produce a similar pattern of results. For example, in some models the system uses each successive input to update the central tendency, without necessarily storing the new episode itself (e.g., an extension of Clayards et al., 2008; Norris & McQueen, 2008). These models nicely capture the central tendencies across the episodes, but it is less clear whether they can produce the full range of episodically-driven effects in the indexical literature, and in the current experiments.

Although auditory memories appear episodic, they showed little or no benefit from repetition in Experiments 3 and 4. This is potentially problematic for the episodic view, as repetition provides multiple episodes, and in an episodic model, these additional tokens should strengthen the observed specificity effects. Instead, our results suggest that any such effect is primarily driven by the most recently experienced relevant episode (or, perhaps, the first experience of a particular instantiation – our data do not discriminate between these two possibilities). Given the episodic results here and elsewhere, and given the clear evidence for various types of abstraction (e.g., the word “table” is the same lexical event, regardless of who produces it), there is a growing consensus that the lexicon must incorporate both abstract and episodic information (Cutler & Weber, 2007; Goldinger, 2007).

We have described our results in terms of the representation of words in memory, as representations that are at least partially episodic can account for our observed specificity effects. From this perspective, the priming advantage at test arises from the activation of a detailed memory trace, with greater priming when the trace and input match across the set of lexical, indexical, and other auditory properties that comprise the episodic representation. A related but conceptually different explanation is that the observed specificity effects are a result of previous experience selecting the word-unique variability from the overall signal and completing the processing steps involved (see Kirsner, Dunn & Standen, 1987; Kolers, 1976; Kolers & Ostry, 1974; Kolers & Roediger, 1984 as cited in Goldinger, 1998; Nygaard, Sommers, & Pisoni, 1994). This is also an episodic approach, but it is not one that posits episodic representations. Rather, the episode is the set of processes applied to the stimulus. For example, if “termite” had been presented during the exposure phase, a participant will have had practice correctly mapping the word “termite” onto an abstract representation, in a particular voice (e.g., male) with a particular co-occurring sound (e.g., a particular dog barking in the background). At test, if these surface characteristics are maintained, there is an advantage in processing fluency. Both explanations assume that some sort of detailed memory is maintained – either a memory for the specific instance or a memory for the processing of that specific instance. While both the episodic representational view and the episodic processing view are viable explanations, they are difficult to distinguish experimentally since they support identical hypotheses.

Is the Indexical Literature about Something Other than the Lexicon?

In the preceding discussion, our analysis was based on a simple observation: Changes in co-occurring sounds produced the same type of decrement in word recognition that has been found for voice changes in the indexical literature. Following the logic used in the indexical literature, we argued that lexical representations must include episodic detail because a change in such details hurts word recognition, just as a change of voice does. Of course, it is conceivable that finding the same pattern of results with environmental sounds as with voice changes is just a coincidence, but invoking "coincidence" does not get one anywhere if one believes in empirical tests: If something quacks, waddles, and tastes awfully good with *l'orange* sauce, the best working hypothesis is that that thing is a duck. In this case, changes in accompanying sounds produce the same decrement in performance as changes in voice characteristics.

There is, however, another possible alternative that seems more interesting. Our inference that the environmental sounds are impacting lexical representations is valid if and only if the logic of the studies in the indexical literature is sound. It is possible that indexical effects, and our newly observed effect for surface changes in co-occurring information, have little to do with the lexicon. On this account, the lexicon contains traditional abstract linguistic representations, and any effects of surface changes take place outside the lexicon.

The challenge for a proponent of this approach is to provide an account of the indexical effects that does not introduce episodic details into the lexical representations. As we pointed out, indexical effects have been taken to be within the lexical sphere because they affect word recognition – in most such studies, including ours, the task is to recognize a word, and that is precisely what lexical access is about. Presumably, if the observed effects are to be separated from lexical representations, they must instead be attributed to some non-lexical memory process. It does seem plausible that recognition (of any stimulus) would be better if the recognition probe is more similar to a preceding event than if it differs. However, we see two challenges to this approach. First, there still must be some specification of what representations are being used in this memory process; if they turn out to be isomorphic to lexical representations, then this approach is not actually different than the standard indexical interpretation. Second, to the extent that this approach relies heavily on the results being a function of memory processes (separate from lexical activation), it is potentially problematic that indexical effects are very fragile using explicit memory tasks, and are much more robust with implicit memory tests. If the effects are memory based, then asking subjects to use what they remember should increase the effects, not decrease them. Thus, if a non-lexical and non-explicit-memory explanation is to have any substance, it must provide an account of how the properties of implicit memory would produce the observed effects, without introducing lexical representations as the site of the memory effect.

This is challenging because it is the lexicon that represents each word separately, leaving it unclear what other memory structure would be that could keep such word-specific information available. To understand the problem, it is useful to contrast this pattern with recent research (Norris, McQueen, & Cutler, 2003) which demonstrates that perceptual retuning of phonetic category boundaries can be guided by lexical context. For example,

when listeners hear a speech segment that is ambiguous between /s/ and /f/, it will be heard as /s/ in “witne?”, but as /f/ in “gira?”. Listeners exposed to a dozen or so words with such ambiguities learn to interpret the ambiguous sound, based on the lexical contexts it occurs in. Critically, they generalize this learning to new tokens of the ambiguous segment, demonstrating a prelexical locus to the effect, probably at the level of phonetic feature analysis (Kraljic & Samuel, 2006). The kind of prelexical representation implicated in perceptual retuning applies to all words and nonwords (in the learned voice). As such, this type of representation is fundamentally incapable of producing the indexical effect because the indexical effect is word-specific. This example is useful in illustrating the word-specific nature of indexical effects, and how they implicate a type of representation that is therefore also word-specific. Any non-lexicon-based account of these effects must offer a form of memory in which the indexical property is associated with an individual word, yet remains outside of the lexicon.

With this constraint in mind, we can consider some examples of non-lexical explanations for our effects and previous indexical findings. For example, we can assume that in our experiments, the pairing of each word with some environmental sound provides the listener with a degraded input of the word due to a certain amount of masking the sound produces. At test, if the same pairing is presented (with the additional difficulty introduced by filtering), the listener could do better because of the prior exposure to the same pattern of residual information left in the word. This could be viewed as comparable to applying a type of filter to the word at exposure, with the plausible expectation that later recognition would be better if the same pattern of filtering were applied at test as had been applied at exposure.

This approach does not assume any representation of the environmental sound in the lexicon; in fact, it does not assume any representation of the environmental sound at all, just the consequences the sound (or the filter) had on the word. This possibility is consistent with our data, but as we noted, any explanation outside the lexicon must assume a memory for item-specific information, where the items are words. This approach predicts that indexical effects would be found for nonwords – items that are not represented in the lexicon. We are aware of one (unpublished) study consistent with this prediction (Azuma & Hickox, 2010) in which indexical effects were found with words from a language that listeners did not know, meaning these items were essentially nonwords for these listeners. Furthermore, our own study provides evidence that nonlexical stimuli produce the indexical pattern of results: In Experiment 6, we found that environmental sounds themselves were better recognized if they were heard with the same word, in the same voice, at exposure and test than with a different word in a different voice. A nonlexical locus may also be favored by earlier demonstrations of the generality of the phenomenon, including Church and Schacter’s (1994) observation of comparable effects for manipulation of emotional tone (happy versus sad) and prosodic pattern (question versus statement). Given the impact of all of these factors, it seems likely that most noticeable differences in the realization of a word (and, we now know, other familiar sounds) will impact the later recognizability of the test stimulus.

Models that Seem Consistent with our Findings

We thus have an apparent conundrum: The association of indexical effects with later recognition of a specific word precludes a prelexical explanation of the type that seems appropriate for phenomena such as perceptual recalibration (Norris, McQueen, & Cutler, 2003). At the same time, the sheer generality of indexical effects seems inconsistent with a lexical account, at least one that is grounded in the traditional view of the lexicon as a kind of list of abstract word representations.

Fortunately, in recent years, there have been theoretical developments in how the lexicon can be conceptualized that seem potentially congenial to our results. For example, Gaskell and Marslen-Wilson's (1997, 2002) distributed cohort model (DCM), and Elman's (2004, 2009) simple recurrent network (SRN) both conceptualize lexical representations in a way that may allow them to incorporate the kind of episodic information that our results call for. For example, Elman's model assumes a distributed representation of word knowledge in which categories emerge over time based on the distributional properties of the input that the system receives. According to this perspective, words are cues that activate the co-occurring information with which they have appeared, based on the frequency of their co-occurrence. Within this type of framework, Creel et al. (2008) cited evidence that words coactivate phonological representations, motor codes, and visual speech information and extended this by adding their own evidence that words coactivate talker information.

Our results are consistent with a distributed view of the mental lexicon, extending coactivation to the full co-occurring variation available in the auditory stream. In Gaskell and Marslen-Wilson's DCM, and in Goldinger's (1998) "Echo" model, when a word is presented, its mapping to the lexicon is conceived as a vector in a potentially high-dimensional space. If we assume that the entries for this vector are not limited to dimensions that index consonants, vowels, and voices, but can instead reflect many kinds of acoustic variation (e.g., the variation introduced by an accompanying environmental sound), the effects that we have observed here can be accommodated. Note that this approach does not treat the lexicon as inherently separate from nonlinguistic information, but at the same time, one would expect words to form clusters in the multi-dimensional space that are largely in different regions than clusters for things like environmental sounds.

This approach at the lexical level is conceptually very similar to the approach that Kat and Samuel (1984) suggested for speech processing at the acoustic-phonetic level. Kat and Samuel tested selective adaptation effects (a contrastive shift in phonetic identification that occurs when a particular sound is presented repeatedly) when the adaptors were nonspeech sounds, while the test items were speech segments. They found that particular combinations of acoustic properties yielded adaptation across the speech-nonspeech divide. In particular, white noise segments that differed in abruptness of onset produced differential adaptation shifts for a speech continuum of aperiodic sounds – a test continuum varying between "ch" (abrupt onset) and "sh" (gradual onset); periodic tone complexes that differed in abruptness of onset (the gradually-onsetting sound was similar to a note played by bowing a violin string, whereas the abrupt onset sound was like a violin string being plucked) shifted identification of a /b/ (abrupt) -- /w/ (gradual onset) speech continuum (/b/ and /w/ are periodic sounds). There was no adaptation effect of the aperiodic adaptors on the periodic

speech sounds, nor of the periodic adaptors on the aperiodic speech sounds. This pattern is consistent with representations that are initially coded in terms of combinations of acoustic features (such as abruptness of onset, and periodicity), that can map onto both speech and nonspeech sounds.

At both the acoustic-phonetic level, and at the lexical level, the set of features that define the whole recognition space will generally be more similar within speech sounds than between speech and nonspeech sounds. Because of this, a model with these properties will simultaneously achieve effective specialization for phonetic segments or words (because these representations are close to each other, and items that are close to each other would generally be expected to interact more with each other than ones that are more distant). Critically, this specialization is achieved entirely within a more general memory structure.

At least at the lexical level, one might ask why a system would develop that stores so much information that is arguably useless – in general, recognizing the word “termite” will not be helped by including information about the sound of a dog that happened to be barking at the time the word was heard. We suggest that although it seems inefficient to operate this way, the alternative may be much more difficult: In order to avoid this type of storage, the listener must continuously evaluate the input to decide what to filter out of the word’s representation, and what to include. Under the real-time conditions that the listener faces, such decisions may well exceed the available processing capacity. And, if the system operates as we have suggested, any information that is truly extraneous will eventually not have an impact on the lexical item’s central tendency, just as the noise in each waveform of an ERP experiment ultimately has no impact. Thus, the system can simply map each input onto a point in this high-dimensional space, and clusters will form in this space for episodes that share the most relevant information.

Conclusion

We began this paper with what we believe was a rather uncontroversial claim: “Fundamentally, the mental lexicon is a memory system: It is the place where language and memory meet.” The results of our six experiments have shown that this memory system does not have a wall between linguistic and nonlinguistic information – word recognition is affected by the environmental sounds that a word has previously occurred with, and recognition of a sound is similarly affected by the sound’s recent history with co-occurring words. The most parsimonious account is that similar perceptual processes support not only indexical effects for words and exemplar effects for sounds, but that these same processes also lead to impaired performance as a result of changes in co-occurring information. The symmetric effects for words and environmental sounds indicate that the incoming auditory stream is not automatically divided into linguistic and non-linguistic streams, as even when participants are actively trying to attend to one or the other, variability from the unattended stream permeates and affects later performance.

Thus, we conclude that representations stored in the mental lexicon are not limited to linguistic information, nor are they limited to the addition of information from highly related sources like voices. Instead, these representations appear to reflect more episodic traces of

words and co-occurring auditory events, even from unrelated sources like background sounds. Overall, the results of the current study suggest that models of the lexicon should treat lexical representations as a proper subset of auditory memory representations. The lexicon is best viewed as a relatively clustered region within a multidimensional space that includes both words and other sounds. Each word within the lexicon is represented as its own tighter cluster, with a central tendency (that can either be induced, or represented explicitly) that minimizes the effects of tokens that contain less typical properties. Recent models of the lexicon have this structure, and offer a plausible platform to account for the pervasive effects of co-occurring information.

Acknowledgments

This material is based upon work supported by NIH grants R01-51663 and R01-059787. We would like to thank Donna Kat and Elizabeth Cohen for technical assistance, Gregory Perryman and Megan Tudor for lending us their voices, and Yevgen Borodin for providing data from the Blind participants.

References

Allen JS, Miller JL. Listener sensitivity to individual talker differences in voice-onset-time. The Journal of the Acoustical Society of America. 2004; 115(6):3171–3183.10.1121/1.1701898 [PubMed: 15237841]

Azuma, T.; Hickox, A. Auditory Cognitive Neuroscience Society. Tucson, AZ: 2010. The effect of voice consistency on memory for English and French words.

Baayen, RH. Analyzing linguistic data: A practical introduction to statistics using R. Vol. 2. Cambridge University Press; 2008.

Bradlow AR, Nygaard LC, Pisoni DB. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. Perception & Psychophysics. 1999; 61(2):206–219.10.3758/BF03206883 [PubMed: 10089756]

Chiu CYP. Specificity of auditory implicit and explicit memory: Is perceptual priming for environmental sounds exemplar specific? Memory & Cognition. 2000; 28(7):1126–1139.10.3758/BF03211814 [PubMed: 11126936]

Church BA, Schacter DL. Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1994; 20(3):521–533.10.1037/0278-7393.20.3.521

Clayards M, Tanenhaus MK, Aslin RN, Jacobs RA. Perception of speech reflects optimal use of probabilistic speech cues. Cognition. 2008; 108(3):804–809.10.1016/j.cognition.2008.04.004 [PubMed: 18582855]

Cohen MA, Evans KK, Horowitz TS, Wolfe JM. Auditory and visual memory in musicians and nonmusicians. Psychonomic Bulletin & Review. 2011; 18(3):586–591.10.3758/s13423-011-0074-0 [PubMed: 21374094]

Creel SC, Aslin RN, Tanenhaus MK. Heeding the voice of experience: The role of talker variation in lexical access. Cognition. 2008; 106(2):633–664.10.1016/j.cognition.2007.03.013 [PubMed: 17507006]

Cutler, A.; Weber, A. International Congress of Phonetics Sciences. Saarbrücken; Germany: 2007. Listening experience and phonetic-to-lexical mapping in L2; p. 6-10.

Elman JL. An alternative view of the mental lexicon. Trends in Cognitive Sciences. 2004; 8(7):301–306.10.1016/j.tics.2004.05.003 [PubMed: 15242689]

Elman JL. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. Cognitive Science. 2009; 33(4):547–582.10.1111/j.1551-6709.2009.01023.x [PubMed: 19662108]

Erickson TD, Mattson ME. From words to meaning: A semantic illusion. Journal of Verbal Learning and Verbal Behavior. 1981; 20(5):540–551.10.1016/S0022-5371(81)90165-1

Gaskell MG, Marslen-Wilson WD. Integrating form and meaning: A distributed model of speech perception. Language and Cognitive Processes. 1997; 12(5):613–656.10.1080/016909697386646

Gaskell MG, Marslen-Wilson WD. Ambiguity, competition, and blending in spoken word recognition. Cognitive Science. 1999; 23(4):439–462.10.1016/s0364-0213(99)00011-7

Gaskell MG, Marslen-Wilson WD. Representation and competition in the perception of spoken words. Cognitive Psychology. 2002; 45(2):220–266.10.1016/s0010-0285(02)00003-8 [PubMed: 12528902]

Goh WD. Talker variability and recognition memory: Instance-specific and voice-specific effects. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31(1):40–53.10.1037/0278-7393.31.1.40

Goldinger SD. Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996; 22(5):1166–1183.10.1037/0278-7393.22.5.1166

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105(2):251–279.10.1037/0033-295x.105.2.251 [PubMed: 9577239]

Goldinger, SD. International Congress of Phonetics Sciences. Saarbrücken; Germany: 2007. A complementary-systems approach to abstract and episodic speech perception; p. 6-10.

González J, McLennan CT. Hemispheric differences in indexical specificity effects in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33(2):410–424.10.1037/0096-1523.33.2.410 [PubMed: 17469976]

González J, McLennan CT. Hemispheric differences in the recognition of environmental sounds. Psychological Science. 2009; 20(7):887–894.10.1111/j.1467-9280.2009.02379.x [PubMed: 19515117]

Gregg MK, Samuel AG. Change deafness and the organizational properties of sounds. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34(4):974–991.10.1037/0096-1523.34.4.974 [PubMed: 18665739]

Johnson K. The auditory/perceptual basis for speech segmentation. OSU Working Papers in Linguistics. 1997; 50:101–113.

Johnson, K. Speaker normalization in speech perception. In: Pisoni, DB.; Remez, RE., editors. The Handbook of Speech Perception. Malden, MA: Blackwell Publishing Ltd; 2005. p. 363-389.

Johnson K. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. Journal of Phonetics. 2006; 34(4):485–499.10.1016/j.wocn.2005.08.004

Kat D, Samuel AG. More adaptation of speech by nonspeech. Journal of Experimental Psychology: Human Perception and Performance. 1984; 10(4):512–525.10.1037/0096-1523.10.4.512 [PubMed: 6235316]

Kraljic T, Samuel AG. Generalization in perceptual learning for speech. Psychonomic Bulletin & Review. 2006; 13(2):262–268. [PubMed: 16892992]

Luce PA, Goldinger SD, Auer ET, Vitevitch MS. Phonetic priming, neighborhood activation, and PARSYN. Perception & Psychophysics. 2000; 62(3):615–25.10.3758/BF03212113 [PubMed: 10909252]

Luce PA, Lyons EA. Specificity of memory representations for spoken words. Memory & Cognition. 1998; 26(4):708–715.10.3758/BF03211391 [PubMed: 9701963]

Luce, PA.; McLennan, CT. Spoken word recognition: The challenge of variation. In: Pisoni, DB.; Remez, RE., editors. The Handbook of Speech Perception. Malden, MA: Blackwell Publishing Ltd; 2005. p. 591-609.

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology. 1986; 18(1):1–86.10.1016/0010-0285(86)90015-0 [PubMed: 3753912]

Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012; 485(7397):233–6.10.1038/nature11020 [PubMed: 22522927]

Norris D. Shortlist: A connectionist model of continuous speech recognition. Cognition. 1994; 52(3):189–234.10.1016/0010-0277(94)90043-4

Norris D, McQueen JM. Shortlist B: A Bayesian model of continuous speech recognition. Psychological Review. 2008; 115(2):357–95.10.1037/0033-295X.115.2.357 [PubMed: 18426294]

Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. Perception & Psychophysics. 1998; 60(3):355–376.10.3758/BF03206860 [PubMed: 9599989]

Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. Psychological Science. 1994; 5(1):42–46.10.1111/j.1467-9280.1994.tb00612.x [PubMed: 21526138]

Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1993; 19(2):309–328.10.1037/0278-7393.19.2.309

Pierrehumbert, JB. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J.; Hopper, P., editors. Frequency Effects and the Emergence of Linguistic Structure. Amsterdam: John Benjamins; 2001. p. 137-158.

Pilotti M, Bergman ET, Gallo DA, Sommers MS, Roediger HL III. Direct comparison of auditory implicit memory tests. Psychonomic Bulletin & Review. 2000; 7(2):347–353.10.3758/BF03212992 [PubMed: 10909144]

Pufahl, A.; Samuel, AG. Let sleeping dogs lie: The persistence of co-occurring variability in memories supporting speech perception. 54th Annual Meeting of The Psychonomic Society; Toronto, Ontario, Canada. 2013.

Roediger HL III. Implicit memory: Retention without remembering. American Psychologist. 1990; 45(9):1043–1056.10.1037/0003-066x.45.9.1043 [PubMed: 2221571]

Schacter DL, Church BA. Auditory priming: Implicit and explicit memory for words and voices. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1992; 18(5):915–930.10.1037/0278-7393.18.5.915

Sheffert SM. Voice-specificity effects on auditory word priming. Memory & Cognition. 1998; 26(3): 591–598.10.3758/BF03201165 [PubMed: 9610127]

Appendix

List of unpaired stimuli

Animate	**Inanimate**		
Words	**Sounds**	**Words**	**Sounds**
ant	bear	airplane	accordion
beaver	bee	axe	alarm clock
beetle	big cat (e.g. tiger)	bagpipe	bell
buffalo	canary	bassoon	bike bell
butterfly	cat	blender	boiling water
camel	chick	broom	camera
cheetah	chicken/rooster	bugle	can (opening)
chipmunk	chimp/monkey	bus	car horn
cobra	cicada	cello	cash register
crab	cow	clarinet	chainsaw
crocodile	coyote	drill	chimes (wind)
deer	cricket	fan	coins/change
eel	crow	french horn	cowbell
flamingo	dog	guitar	cuckoo clock
fox	dolphin	hammer	cymbal
giraffe	donkey	keys	doorbell

Animate	**Inanimate**		
Words	**Sounds**	**Words**	**Sounds**
gopher	dove	lawnmower	drum roll
hyena	duck	matches	flute
lark	eagle	microwave	glass (breaking)
llama	elephant	modem	harmonica
lobster	fly	motorcycle	harp
mole	frog	oboe	helicopter
moose	goat	organ	jackhammer
ostrich	goose	piccolo	music box
otter	gorilla/ape	pinball	page (turn)
parakeet	horse	printer	party favor
peacock	lamb	radio	phone (ring)
pelican	lion	saxophone	piano
penguin	loon	scissors	ping pong
pheasant	mosquito	shower	saw
rabbit	mouse	shredder	ship
robin	owl	sprinkler	shuffle cards
shark	parrot	stapler	siren
skunk	pig	stopwatch	steel drum
snail	raccoon	subway	tambourine
sparrow	rat	teapot	train (whistle)
spider	rattlesnake	thunder	trumpet
squirrel	seagull	toaster	tuba
swan	seal	toilet	typewriter
termite	turkey	toothbrush	violin
turtle	woodpecker	triangle	zipper
vulture		trombone	
whale		vacuum	
worm			

Highlights

- We presented participants with spoken words paired with environmental sounds.
- Subsequent word recognition of filtered stimuli was impaired if the voice changed.
- Word recognition was similarly impaired if the paired environmental sound changed.
- We observed the same result when we reversed the roles of the words and the sounds.
- Models should treat lexical representations as a subset of memory representations.

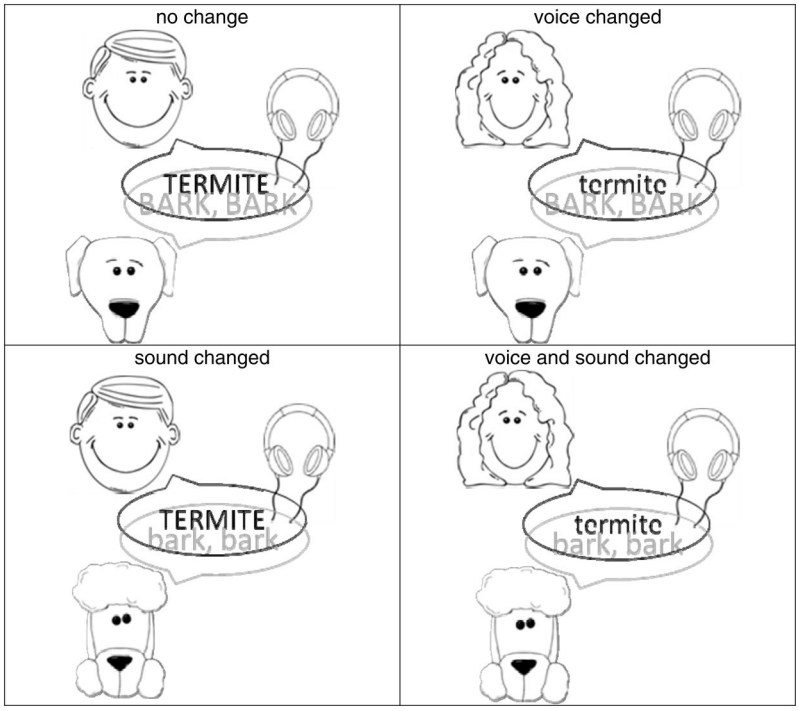

Figure 1.
Illustration of the four experimental conditions in Experiments 1 and 2.

Table 1

Mean percent accuracy on the word (Exp 1 & 3) identification task by exposure-test match. For Experiment 3, mean percent accuracies are by exposure-test match and repetition. Standard errors are in parenthesis. Difference scores compared to the "no change" condition are also shown.

Exposure-test match	no change	voice changed	sound changed	voice and sound changed	unpaired word
Exp 1 (n=64)	64.7% (1.5%)	62.7% (1.5%)	57.1% (1.5%)	59.3% (1.5%)	-
	difference from no change	***−2.0%***	***−7.6%***	***−5.4%***	-
Exp 3 (n=64) 1x	69.1% (2.9%)	58.2% (3.1%)	64.5% (3.0%)	-	86.7% (2.1%)
	difference from no change	***−10.9%***	***−4.6%***	-	***17.6%***
2x	68.6% (2.9%)	62.9% (3.0%)	63.3% (3.0%)	-	93.4% (1.6%)
	difference from no change	***−5.7%***	***−5.3%***	-	***24.8%***
4x	69.5% (2.9%)	69.9% (2.9%)	70.7% (2.9%)	-	90.6% (1.8%)
	difference from no change	***0.4%***	***1.2%***	-	***21.1%***
8x	79.3% (2.5%)	62.9% (3.0%)	71.5% (2.8%)	-	89.8% (1.9%)
	difference from no change	***−16.4%***	***−7.8%***	-	***10.5%***

Table 2

Mean percent accuracy on the sound (Exp 2, 4, 5 & 6) identification task by exposure-test match. For Experiment 4, mean percent accuracies are by exposure-test match and repetition. Standard errors are in parenthesis. Difference scores compared to the "no change" condition are also shown.

Exposure-test match	no change	sound changed	voice changed	sound and voice changed	unpaired sound	word and voice changed
Exp 2 (n=64)	38.2% (1.5%)	32.4% (1.5%)	39.8% (1.5%)	34.5% (1.5%)	-	-
	difference from no change	***−5.8%***	***1.6%***	***−3.7%***	-	-
Exp 4 (n=48) 1x	40.6% (2.5%)	34.4% (2.4%)	43.2% (2.5%)	-	47.1% (2.6%)	-
	difference from no change	***−6.2%***	***2.6%***	-	***6.5%***	-
8x	46.1% (2.5%)	37.0% (2.5%)	45.3% (2.5%)	-	46.1% (2.5%)	-
	difference from no change	***−9.1%***	***−0.8%***	-	***0.0%***	-
Exp 5 (n=19)	42.1% (2.8%)	28.0% (2.6%)	41.1% (2.8%)	28.6% (2.6%)	-	-
	difference from no change	***−14.1%***	***−1.0%***	***−13.5%***	-	-
Exp 6 (n=51)	39.7% (1.4%)	35.9% (1.4%)	-	-	-	36.7% (1.4%)
	difference from no change	***−3.8%***	-	-	-	***−3.0%***