

# The importance of semantics in auditory representations

MELISSA K. GREGG AND ARTHUR G. SAMUEL  
Stony Brook University, Stony Brook, New York

The purpose of the present study was to examine the nature of auditory representations by manipulating the semantic and physical relationships between auditory objects. On each trial, listeners heard a group of four simultaneous sounds for 1 sec, followed by 350 msec of noise, and then either the same sounds or three of the same plus a new one. Listeners completed a change-detection task and an object-encoding task. For change detection, listeners made a same-different judgment for the two groups of sounds. Object encoding was measured by presenting probe sounds that either were or were not present in the two groups. In Experiments 1 and 3, changing the target to an object that was acoustically different from but semantically the same as the original target resulted in more errors on both tasks than when the target changed to an acoustically and semantically different object. In Experiment 2, comparison of semantic and acoustic effects demonstrated that acoustics provide a weaker cue than semantics for both change detection and object encoding. The results suggest that listeners rely more on semantic information than on physical detail.)

Would you notice if the person sitting next to you switched heads? How about if the person you were listening to switched voices? Recent research in change detection suggests that most people do not notice substantial changes such as these (see Grimes, 1996; Vitevitch, 2003), even though the changes seem blatantly obvious after the fact. These failures to detect large changes to visual and auditory scenes—called *change blindness* and *change deafness*—call into question the extent of detail with which we perceive and represent our environment.

Change blindness has been demonstrated with a variety of stimuli in many different scenarios (see Simons & Rensink, 2005, for a review). Change blindness can occur during a visual interruption, such as a flicker (e.g., Rensink, O'Regan, & Clark, 1997), a blink (e.g., O'Regan, Deubel, Clark, & Rensink, 2000), or a mudsplash (O'Regan, Rensink, & Clark, 1999). For example, people fail to notice a huge airplane engine disappearing in a flickering picture display (Rensink et al., 1997). Around two thirds of participants fail to notice a change in the identity of a real-life conversational partner when the conversation is interrupted by a brief visual occlusion (Simons & Levin, 1998). Change blindness can also occur when attention is focused away from the changing object: When instructed to attend to a group of basketball players, viewers rarely notice a person dressed up in a gorilla suit walking behind the players (Simons & Chabris, 1999).

The change-deafness literature is much smaller than the change-blindness literature, but the little work that has been done in the auditory domain indicates that listeners miss substantial changes to auditory scenes at rates

comparable to or worse than those in the corresponding visual literature. Failures to hear auditory changes were demonstrated as far back as the early shadowing work of Cherry (1953) and Treisman (1960), which showed that changes in an unattended channel, such as a reversal of speech or change of language, are often missed while shadowing a spoken message in the other channel. More recent investigations have shown that listeners can even be unaware of changes to the shadowed information. For example, Vitevitch (2003) found that listeners fail to hear a voice change in the middle of a shadowed message and was the first to call such failures *change deafness*. Other recent auditory paradigms more closely resemble the change-blindness work. For example, Eramudugolla, Irvine, McAnally, Martin, and Mattingley (2005) presented interrupted scenes of auditory objects, such as instruments and animal sounds, and found substantial change deafness that increased with increasingly large numbers of sounds. Using a similar paradigm, Gregg and Samuel (2008) found that change deafness follows an orderly pattern as a function of the acoustic distribution of stimuli on pitch and harmonicity dimensions. They demonstrated that auditory change detection strongly depends on how far away the sounds are on both dimensions, with change detection being enhanced when the object that changes is relatively far on both dimensions from the sound it replaced and quite poor when the objects are close together acoustically on both dimensions.

Although change deafness has only recently begun to receive attention, initial results indicate that perceptual representations may not be as rich and detailed as subject-

---

A. G. Samuel, [asamuel@ms.cc.sunysb.edu](mailto:asamuel@ms.cc.sunysb.edu)

tive experience would suggest. If perceptual representations of the world do not capture intricate detail, how much and what kind of information do they contain? In the better developed change blindness literature, various claims regarding the nature of perceptual representations have been offered. Converging tests from the auditory modality have the potential to inform this debate, if perceptual representations have similar properties across modalities.

Perhaps the most extreme position in the visual literature is that people do not represent visual information in any enduring format (see O'Regan, 1992; O'Regan & Noë, 2001). According to this theory, we perceive objects by treating the world as a memory store, rather than by comparing them with preexisting representations in memory (presumably because there is no such memory). Thus, poor change detection in visual scenes results from having no internal representations of our visual environment (see O'Regan, 1992; O'Regan & Noë, 2001). This theory is not likely to be entirely correct, however, because, as Simons and Ambinder (2005) noted, change blindness does not occur on 100% of trials; the fact that observers can sometimes detect change indicates that there must be some sort of representation.

In more mainstream theories, it is assumed that we do have visual representations. According to these theories, change-detection failures occur because of a failure to compare current perceptual information with recently encoded information. Although there is general agreement among such theories that visual perceptual representations do exist, there has been conflicting evidence regarding the kind of information they contain. In most theories, it is assumed that visual representations contain information about an object's physical features, but theories vary with respect to how much physical information is included and how detailed it may be. The most prominent view of the nature of visual representations is probably object-file theory (Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992; Treisman, 1988, 1992). In this theory, temporary perceptual representations—called *object files*—are formed when attention is directed toward an object; the object files are updated with new physical feature information as necessary when an object in the visual field changes (e.g., by being occluded or by changing position). These visual representations are rough and abstract, consisting of physical information that is descriptive of and useful for object identification and continuity. One of the key empirical findings in support of object-file theory is a larger benefit in naming time from previewing a letter when it later appears as part of the same object rather than as part of a different object (see Kahneman et al., 1992). Using slightly modified versions of the priming paradigm, other researchers have found support for the claim that visual representations encode some abstract physical information, with this information primarily serving the purpose of object identification (e.g., Gordon & Irwin, 1996, 2000).

In contrast with the view that visual representations are abstracted forms of actual objects, some researchers have claimed that they consist mostly of physical feature information (e.g., Pollatsek, Rayner, & Collins, 1984; Pollatsek,

Rayner, & Henderson, 1990) and that this information can be quite detailed (e.g., Hollingworth & Henderson, 2002; Hollingworth, Williams, & Henderson, 2001). These investigators compared detection-of-change trials involving different types of objects (e.g., a straight-edged notebook changes to a calculator) with trials involving changes to a different token from the same basic-level category (e.g., a straight-edged notebook changes to a spiral-edged notebook); they also included trials in which there were small physical changes, such as a rotation change (e.g., Hollingworth, 2003; Hollingworth & Henderson, 2002). These studies indicated that participants often do notice subtle token and orientation changes. Detection of changes to objects that are so visually similar has led to the claim that visual representations must be quite detailed.

Although it is fairly well established that transient representations contain some level of physical feature information, there have been mixed claims regarding whether or not they contain semantic information. Some theorists have suggested that visual representations do contain semantic information, at least about the identity of the object (e.g., Gordon & Irwin, 1996; Kahneman et al., 1992). According to object-file theory (e.g., Kahneman et al., 1992) and related models (e.g., Hollingworth & Henderson, 2002), perceptual representations interact with long-term memory. As soon as the physical features of an object provide enough information for identification, the representation is updated to include semantic information from long-term memory.

Empirical support for this position has been provided in several studies (e.g., Gordon & Irwin, 1996). The finding that object identity information seems to persist despite changes to an object's physical appearance has led to the claim that identity may take precedence in representations and overshadow any physical information. For example, manipulation of the font of the preview and of the target letter in a priming paradigm has shown that the object-specific benefit is just as strong when the letters are presented in different cases (one lower and one upper) as it is when the two are presented in the same case (Gordon & Irwin, 1996; Henderson, 1994). In addition, object-specific effects have been found to persist even when gross physical changes are involved in a priming paradigm (at least from words to pictures; see Gordon & Irwin, 2000). These findings suggest that representations do encode semantic identity information.

On the other hand, there is some evidence that perceptual representations may not contain semantic information. For example, two identical drawings of an object in a priming paradigm were found to result in more naming facilitation than did two different drawings from the same basic-level category (e.g., Pollatsek et al., 1984; Pollatsek et al., 1990). Also, it has been demonstrated that two conceptually related but physically dissimilar objects (e.g., a baseball bat and a baseball) do not result in naming facilitation, whereas two conceptually unrelated, but physically similar objects (e.g., a tomato and a baseball) do lead to facilitation (see Pollatsek et al., 1984). These findings suggest that semantic information may not be included in perceptual representations.

In summary, work in the visual domain has not yet offered a clear answer as to whether transient representations contain semantic information. Taking the issue into the auditory domain may offer another way to explore this representational issue. Auditory and visual objects are quite different; auditory objects are sources of vibrations, and visual objects are surfaces that reflect light. The differences between the two types of objects and between the ways in which the two perceptual systems are organized raise the question of whether auditory and visual object representations are different or whether there are domain-general properties of perceptual representations that apply to both modalities. The purpose of the present study was to explore this issue by investigating the nature of auditory representations and the extent to which both semantic and acoustic information can influence these representations.

Exploring representational issues in the auditory domain may also provide a cleaner dissociation of physical identity from basic-category semantic identity than would such exploration in the visual domain. Physical information and semantic information are tightly coupled in vision, and disentangling the contributions from each is problematic, because most semantically similar visual objects are also physically similar. The nature of auditory stimuli offers an advantage in comparing objects that are physically and semantically different with objects that are physically different and semantically the same (e.g., a small bird chirping and a seagull squawk are the same basic objects—bird sounds—but they have quite different spectral patterns).

To examine the nature of auditory representations, we added a semantic manipulation to our previous change-deafness paradigm (see Gregg & Samuel, 2008). Our paradigm was modeled after a visual change-detection study (Mitroff, Simons, & Levin, 2004) in which both change detection and object encoding were measured. In that study, change detection was measured by presenting a visual scene, followed by an interruption and then a second visual scene. The participants' task was to determine whether the two visual scenes were the same or different. Object encoding was measured by presenting the same sequential events followed by two visual objects; one of these had been present in the visual scenes, and one was a foil. The participants' task was to determine which one of the two objects had been present in the visual scenes (see Mitroff et al., 2004). This study provided a method for distinguishing between two possible causes of change-detection failure: a failure to encode the objects in the first place or a failure to detect a change to an object that has been encoded. The encoding/detection distinction is critical in an examination of why failures to perceive changes occur. Thus, we adopted it in our previous change-deafness work (Gregg & Samuel, 2008) and in the present study. The change-detection and object-encoding paradigms used in the present study were used to test whether semantic and acoustic information are encoded in perceptual representations and, if so, whether this information can be utilized in a change-detection task.

The purpose of the present study was to examine whether perceptual representations of acoustic stimuli contain semantic information. We used an acoustic space (in Experi-

ment 1) and a perceptual space of sounds (in Experiments 3A and 3B) to examine a potential effect of semantics in an auditory change-detection task and in an object-encoding task by comparing trials in which a target changed to a different *type* of object (e.g., a dog bark changed to a bell) with trials in which a target changed to a different *token* of the same object (e.g., a small dog's bark changed to a large dog's bark). Hollingworth and Henderson (2002) used a similar approach in a visual change-detection paradigm.

We also set out to compare semantic and acoustic influences on both change detection and object encoding. We know from our previous study (Gregg & Samuel, 2008) that change deafness depends on the physical acoustic information, but it remains to be seen whether this information is more important than, less important than, or equally as important as any potential semantic information. To explore this issue, in Experiment 2, we compared trials in which there was a token change (e.g., one dog barking to a different dog barking) with trials in which there was a change to a different object type that was either acoustically smaller than (and thus harder to detect) or acoustically larger than (and easier to detect) the acoustic distance between the two tokens.

In all three experiments, in principle, one would want to have a fully specified map of psychoacoustic space that reflects all and only psychoacoustics. In such a hypothetical space, all sounds could be located, and then semantic information would be added in order to see whether it affects change deafness. Of course, such a map cannot be created; there can always be some additional acoustic property that one did not consider, and there will always be influences on perceptual structure that go beyond acoustics. Our approach, therefore, was to use two very different ways to approximate psychoacoustic distances between objects, each of which had strong internal validity, and to then see whether semantic influences would operate similarly in measures of change deafness across the two methods of constructing psychoacoustic distances. The goal was to obtain converging evidence.

The approach used in Experiments 1 and 2 was to rely on a large body of previous research that has demonstrated the fundamental psychoacoustic importance of two factors: pitch (based on a sound's fundamental frequency) and harmonicity (a dimension that distinguishes noisy sounds from more periodic ones). There is a substantial literature that supports the importance of these two cues in auditory scene organization (e.g., Bregman, 1990; Dannenbring & Bregman, 1976; Kat & Samuel, 1984; Van Noorden, 1975; Warren, Obusek, & Ackroff, 1972). In addition to this evidence for the general importance of these two factors, our previous change-deafness work demonstrated that  $f_0$  and harmonicity produce additive and large effects on change detection (Gregg & Samuel, 2008). In our previous study, we showed that the ability to detect changes to a set of sounds strongly depends on how far away the sounds are in the two-dimensional space that these factors provide: Changes to sounds from the same region of the acoustic space were difficult to detect, whereas spacing stimuli out along both the  $f_0$  and harmonicity dimensions made it easier to detect a change in an acoustic scene.

In Experiments 3A and 3B, we used a quite different approach to estimating the psychoacoustic distances between our stimulus sounds. First, listeners made similarity judgments for pairs of the sounds, and these similarity ratings were used as the input for a multidimensional scaling (MDS) analysis. The output of such an analysis is a perceptual space of  $N$  dimensions, with  $N$  determined by the number of dimensions that are needed to account for most of the variance in the similarity judgments (see Gygi, Kidd, & Watson, 2007, for an example using this approach). We then used this derived space to test for semantic versus acoustic effects with a new group of listeners.

The two approaches produced different psychoacoustic spacings for our sounds, reflecting the different strengths and weaknesses of each approach in approximating the true psychoacoustic reality. As noted above, our overall strategy was to examine the influence of semantic identity information using these two different spaces, to see whether there would be a consistent effect of such information across the different psychoacoustic spaces. Such converging evidence would provide strong support for a significant role of semantic identity in the representations that listeners use.

## EXPERIMENT 1

The purpose of Experiment 1 was to determine whether auditory representations contain semantic identity information. This was assessed using two different tasks: a change-detection task and an object-encoding task. In both tasks, the semantic relationship of the targets was manipulated so that they were either acoustically different but semantically from the same basic-level category, (e.g., two different birds singing) or acoustically different and semantically unrelated (e.g., a bird singing and a phone ringing).

### Method

#### Participants

Twenty-five undergraduate listeners from the State University of New York, Stony Brook, with no reported hearing deficiencies participated in this experiment. Listeners received credit toward a course requirement for their participation.

An a priori criterion was set, such that listeners had to perform above chance on the object-encoding question to be included in the data analyses. One listener did not meet this criterion and was thus excluded from all reported analyses. The mean age of the remaining 24 listeners was 19 years.

#### Stimuli

Sounds drawn from a set of 22 common sound categories (e.g., bells, phones, and dogs), with four variants of each category (e.g., four dog barks), were initially rated by five members of our lab for similarity. Most of the sounds were gathered from various online sources, and only highly identifiable stimuli were selected (Shafiro & Gygi, 2004). All stimuli were truncated to 1 sec, and a linear off-ramp to zero amplitude was imposed over the final 10 msec to avoid abrupt offsets. Stimuli were filtered for any background noise using a noise reduction filter specific to the spectral envelope of each stimulus (Goldwave, ver. 5.12). All stimuli were then digitized at or converted to a sampling rate of 44.1 kHz and were matched for RMS amplitude to roughly equate for loudness. In order to ensure that the variants of each category were acoustically distinct, only

those sounds within a category that were consistently rated as most dissimilar were selected for use in the experiment. This process yielded 12 sound pairs, with each pair consisting of two sounds from the same basic-level category. These sounds can be found at [www.psychology.stonybrook.edu/asamuel/ldemo.html](http://www.psychology.stonybrook.edu/asamuel/ldemo.html) by clicking on *Change Deafness*.

Each stimulus was submitted to Praat (Boersma & Weenink, 1992) for analysis of harmonicity (i.e., mean amount of acoustic periodicity in the signal) and pitch ( $f_0$ ). The  $f_0$  measurement was a spectral mean and, for sounds that were more aperiodic (e.g., a clap), was computed by averaging the  $f_0$  measurements throughout the duration of the signal. Figure 1 depicts the distribution of the 24 sounds on both dimensions. As noted above, the two instances of each sound (e.g., dog 1 and dog 2) were ones that had been selected to be relatively dissimilar, on the basis of an initial screening of tokens. The purpose of that selection was to try to generate stimuli in which two sounds sharing semantic identity were nonetheless not particularly close acoustically, since this property would be needed in the experimental design (see below). The placement of the stimuli shown in Figure 1 indicates that we were successful in this regard. To quantify this, using the coordinates of each sound in the two-dimensional space shown in the figure, we compared the average distance between each pair sharing semantic identity (e.g., dog 1–dog 2; bell 1–bell 2) with the average distance between all other pairs (e.g., dog 1–bell 1; dog 1–piano 2). The average distances for the two cases were extremely similar (6.8 vs. 7.2 units in this space) [ $t(274) = 0.27, p = .79, d = 0.510$ ].

#### Procedure

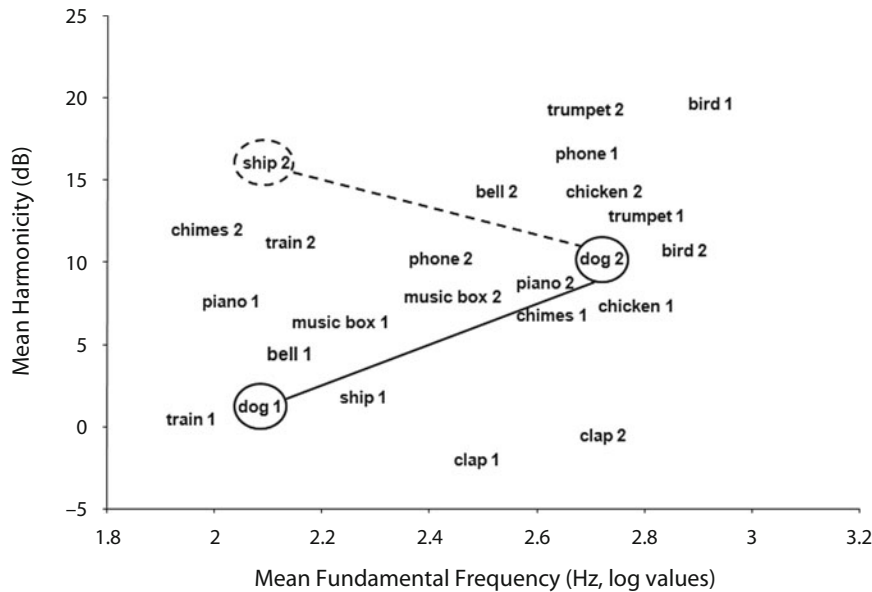
All stimuli were presented at a comfortable listening level over headphones in a sound-attenuated chamber. Each listener did two tasks: one assessing change detection and the other assessing object encoding. The change-detection and object-encoding tasks were run in two separate blocks, with order counterbalanced across listeners.

**Change detection.** On all trials, four of the sounds were first presented over headphones simultaneously for 1 sec (Group 1). Group 1 was immediately followed by 350 msec of white noise. Then, on *same* trials, the same four sounds were played again in Group 2. On *different* trials, Group 1 was presented, followed by the white noise, and then all but one of the sounds from Group 1 was presented in Group 2, with the missing sound replaced by a sound that had not been used in Group 1. The listeners' task on the change-detection trials was to indicate whether the two groups of sounds were the same or different by pressing the corresponding key. *Same* and *Different* were displayed on a screen in front of the participants when the second group of sounds had been presented. Within the change-detection block, half of the trials were *same* trials and half were *different*.

There were two kinds of *different* trials: type change and token change. In the type-change condition, Group 1 was presented, followed by the white noise, and then all but one of the sounds from Group 1 were presented, with the missing sound replaced by an acoustically and semantically different object (i.e., a sound that did not share the same identity) from one of the 24 sounds that had not been used in Group 1. For example, *dog 2*, *phone 1*, *train 1*, and *bird 1* could constitute Group 1, and *ship 2*, *phone 1*, *train 1*, and *bird 1* could make up Group 2. Trial structure in the token-change condition was the same, except that the missing sound was replaced by a sound that was acoustically different from but semantically the same as one of the sounds from Group 1 (i.e., the other member of a pair of sounds sharing the same identity). For example, *dog 2*, *phone 1*, *train 1*, and *bird 1* could constitute Group 1, and *dog 1*, *phone 1*, *train 1*, and *bird 1* could make up Group 2.

Critically, the acoustic distance<sup>1</sup> separating the changed items was matched between the token- and type-change conditions. That is, the distance between the sound that changed from Group 1 to Group 2 in the type-change condition (e.g., the distance between *dog 2* and *ship 2*, represented by the dotted line in Figure 1) was matched for the distance between the sound that changed from Group 1 to Group 2





**Figure 1.** Acoustic distribution of stimuli used in Experiment 1 and—except for clap 1, clap 2, music box 1, and music box 2—in Experiment 2. The distance between targets in token-change trials (in this example, the distance between dog 1 and dog 2, represented by the solid line) was matched for the distance between targets in the type-change trials (in this example, the distance between dog 2 and ship 2, represented by the dotted line).

in the token-change condition (e.g., the distance between dog 2 and dog 1, represented by the solid line in Figure 1). There were 72 *same* trials and 72 *different* trials (36 type change and 36 token change) in the change-detection block. The 24 sounds were balanced across the *different* trials, such that each sound served as the changed item three times. All other sounds presented on a trial were randomly selected, with the exception that no two sounds sharing semantic identity (e.g., dog 1 and dog 2) were presented together on a trial, aside from where intended in the token-change condition.

**Object encoding.** Object-encoding trials began identically to change-detection trials, in that each trial began with four simultaneous sounds, followed by 350 msec of white noise and then by four more simultaneous sounds. However, on object-encoding trials, Group 2 was followed by two sequential sounds (Sound 1 and Sound 2), presented 250 msec apart (beginning 500 msec after Group 2). The listeners' task was to determine whether it was Sound 1, Sound 2, or neither sound that had been presented in the two groups of sounds. The three choices (Sound 1, Sound 2, Neither) were displayed on the screen.

Within the object-encoding block, the same 144 trials (72 *same* and 72 *different*) that occurred in the change-detection block were presented. There were three kinds of trials for both *same* and *different* trials: valid, type change, and token change. In all three conditions, one of the two sequential sounds was a foil that had not occurred in either group of the trial; the other sound (the target) varied by condition. On a valid trial, the target had occurred in both Group 1 and Group 2, and, thus, the correct answer would be selection of this target. In the other two conditions, the correct response was *neither*. In the token-change condition, the target sound was acoustically different from but semantically matched to one of the sounds that had occurred in both groups on the trial (e.g., dog 2, when dog 1 had been present on the trial). In the type-change condition, the target sound was a second foil and was acoustically and semantically different from all of the sounds in Group 1 and Group 2.

As in the change-detection trials, the acoustic distances between token-change trials and type-change trials were matched; the distance between the two semantic tokens (e.g., the distance between piano 1 and piano 2) was equal to the distance between the variant

that was presented within both groups of sounds and one of the two foils (e.g., the distance between piano 1, presented in both Group 1 and Group 2, and clap 1, which would be presented as one of the two sound choices after the Group 1–noise–Group 2 sequence). There were 24 trials in each condition for both *same* and *different* trials (with target and foil selection balanced so that all 24 sounds served as targets equally often).

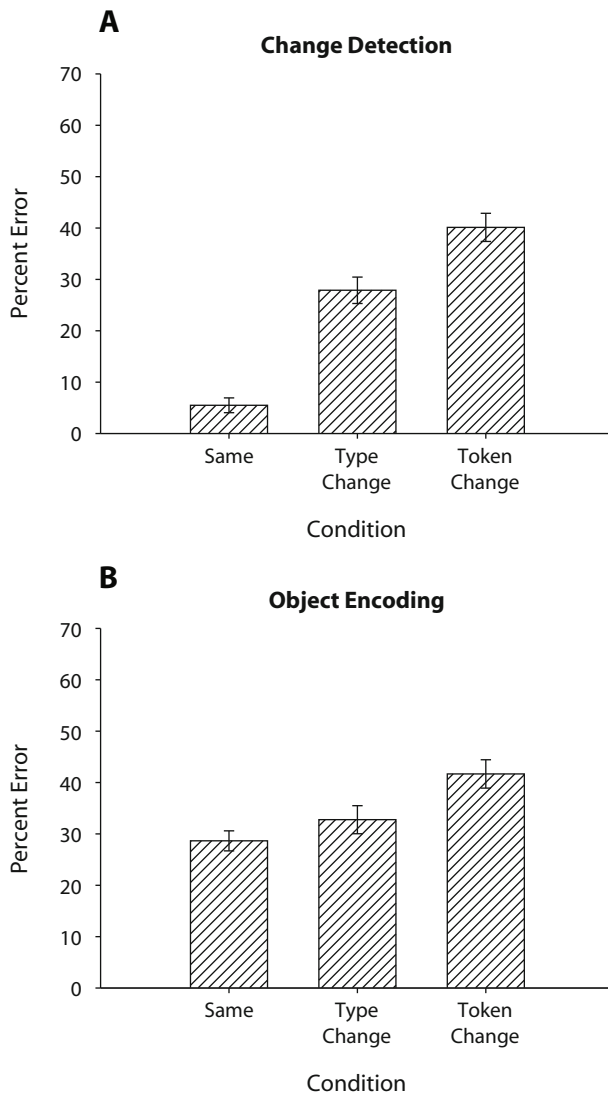
There was a brief rest break in between the change-detection and the object-encoding blocks. The session lasted approximately 1 h. Order of trials within each block was randomized. Each trial was launched 500 msec after the listener's response, timing out after 5 sec. The between-subjects variable of block order (change detection first vs. object encoding first) did not result in a main effect, nor did it interact with any other variable in Experiment 1 or 2, so data were collapsed across this factor in all reported analyses.

## Results and Discussion

The central goal of Experiment 1 was to determine whether auditory representations contain semantic—that is, basic-level identity—information. We approached this issue by examining a potential effect of semantics on auditory object encoding and change detection. Figure 2 presents the pattern of error responses for both tasks as a function of condition.

### Change Detection

Change deafness is a failure to detect a change in the set of sounds being presented. In our paradigm, a direct measure of change deafness is the error rate on *different* trials;<sup>2</sup> these are trials in which the sound array changes, yet listeners choose *same*. As Figure 2A shows, these errors were very common, occurring on over a third of the *different* trials, whereas errors on *same* trials were very low. This pattern contributed to a significant effect of condi-



**Figure 2.** Mean percent error on the change-detection (A) and object-encoding (B) tasks as a function of condition.

tion [ $F(2,46) = 99.95$ ,  $MS_e = 74.01$ ,  $p < .01$ ,  $\omega^2 = .614$ ]. Change deafness is demonstrated by the high error rate on the two types of *different* trials (mean error = 34.0%). In contrast, on *same* trials, the mean error rate was 5.5%. Planned comparisons revealed that this difference was significant ( $p < .01$ ).

If semantics plays a role in change detection, not only should we observe change deafness, we should also find significantly more change deafness on *different* trials in which the target changed to a sound from the same semantic category. As can be seen in Figure 2, this is exactly the pattern of data we obtained: Change deafness was substantially increased on *different* trials that involved a change within a semantic category. Planned comparisons demonstrated that listeners made significantly more errors on token-change trials (40.1%) than on type-change trials (27.9%) ( $p < .01$ ). These findings suggest that the auditory objects were semantically encoded on change-

detection trials, and that object identity information was used in making a decision of whether there was a change or not.

### Object Encoding

For the object-encoding trials, an effect of semantics would be indicated by higher error rates on trials in which listeners were presented with an acoustically different but semantically identical, yet incorrect choice and a foil (token-change condition) than on trials in which listeners were presented with two invalid choices, both acoustically and semantically different from any sound on the trial (type-change condition). This pattern can be seen in Figure 2B; the difference yielded a significant effect of condition [ $F(2,46) = 5.42$ ,  $MS_e = 195.64$ ,  $p < .01$ ,  $\omega^2 = .08$ ], with listeners making more errors on the token-change trials than on the type-change trials. Planned comparisons revealed that the error rate in the token-change condition (41.7%) was significantly higher than in either the type-change (32.8%) or the valid (28.6%) condition ( $p < .01$ ). There was no significant difference between the type-change and valid conditions.

The finding of higher error rates in the token-change condition suggests that listeners were encoding abstract identity information about the objects. If this is the case, then not only should error rates be higher in the token-change condition, listeners should also have actually chosen the semantic variant, rather than the foil, on trials in which they did not choose the correct response (*neither*). In fact, listeners chose the semantic variant on 69% of the incorrect trials, significantly greater than chance (50%) performance [ $t(23) = 6.92$ ,  $p < .01$ ,  $d = 1.41$ ]. This finding provides further evidence that listeners encode and rely on semantic category information when they perceive auditory scenes.

## EXPERIMENT 2

In Experiment 1, the careful matching of physical similarity of token changes and type changes demonstrated that identity information is encoded in auditory representations. The purpose of Experiment 2 was to determine the importance of semantic information when we manipulate physical cues to make changes either acoustically easier to detect or acoustically harder to detect than in the matched conditions. In addition to targets changing to a different type of object or to a different token of the same object type, there were also type changes with targets that were either systematically psychoacoustically closer to or systematically farther from the matched type (and token) changes.

### Method

#### Participants

Twenty-nine undergraduate listeners from the State University of New York, Stony Brook, with no reported hearing deficiencies participated in this experiment. Listeners received credit toward a course requirement or payment for their participation.

As in Experiment 1, an a priori criterion was set such that listeners had to perform above chance on the object-encoding block to

be included in the data analyses. Three listeners failed to meet this criterion and, thus, were excluded from all analyses. Two additional listeners were excluded from the data analyses because they admittedly did not understand the object-encoding task and so pushed one button for the entire block. The mean age of the remaining 24 listeners was 21 years.

### Stimuli

Twenty of the stimuli—10 semantically related pairs—from Experiment 1 were used. In order to balance the design of the experiment, 2 sound pairs from Experiment 1 were eliminated. The 2 sound pairs that were the closest together acoustically—the claps and the music boxes—were not used in this experiment.

### Procedure

Two conditions were added to the change-detection and object-encoding blocks. All other procedures were the same as those in Experiment 1. In both blocks, a short-type-change and a long-type-change condition were added.

**Change detection.** Within the change-detection block, the short-change condition involved a change from a sound in Group 1 (i.e., the target) to a sound in Group 2 that was the shortest acoustic distance away from the target. In the change-detection long-change condition, the target from Group 1 changed to a sound in Group 2 that was the longest acoustic distance away. For example (see Figure 1), for the dog 2 target, the piano 2 sound would serve in the short-change condition, and the train 1 sound in the long-change condition. Within the short- and long-change trials, only two targets were permitted to share a sound that was the shortest or longest distance away. In cases in which more than two targets shared the same shortest or longest distance sound, the Group 2 sound was chosen from the second shortest or longest sound away.

There were five conditions in the change-detection block: *same*, token change, type change (with the acoustic distance between targets matched for the acoustic distance between targets in the token-change condition), short type change, and long type change. There were 100 *same* trials and 100 *different* trials, with 25 trials of each *different* condition. All 20 targets were balanced so that they occurred once in each of the *different* conditions. Targets for the remaining five trials within each condition, as well as the order of trials within the change-detection block, were randomly selected.

**Object encoding.** Within the object-encoding short-change condition, one of the two sequential sounds presented after the Group 1–noise–Group 2 sequence was a sound that was the shortest (or second shortest) acoustic distance away from a sound that had been presented in both Group 1 and Group 2, and the other was a randomly selected foil that was not present in either of the groups. In the long-change condition, one of the two sequential sounds was a sound that was the longest (or second longest) acoustic distance away from a sound that was presented in both Group 1 and Group 2, and the other sound was a foil. The correct answer in both of these conditions was *neither*.

The five conditions within the object-encoding block included these two new conditions and the three conditions tested in Experiment 1 (valid, token change, and type change matched for acoustic distance with the token-change condition). There were 20 trials within each condition for both *same* and *different* trials (200 trials total). Trials were balanced so that each sound served as the object-encoding target once in each condition. The order of trials within the object-encoding block was randomized.

## Results and Discussion

The central goal of this experiment was to determine whether auditory representations contain both acoustic and semantic information and, if so, to determine which kind of information is more important in change detection and in object encoding.

### Change Detection

The mean error rates for each condition in the change-detection block are depicted in Figure 3A. The differences in error rates across all change-detection conditions were significant [ $F(4,92) = 66.79$ ,  $MS_e = 89.80$ ,  $p < .01$ ,  $\omega^2 = .532$ ]. If acoustics provide useful information in change detection, errors in the acoustically more difficult short-change condition should be higher than in the acoustically easier long-change condition. Error rates in the type-change condition should also be between those in the short- and long-change conditions, given the intermediate acoustic distance between targets in this condition. Planned comparisons demonstrated that errors in the long-change condition (23.5%) were significantly lower than those in the type-change (33.7%) and short-change (34.8%) conditions ( $p < .01$ ); the latter two did not differ. The reduced error rate in the long-change condition shows

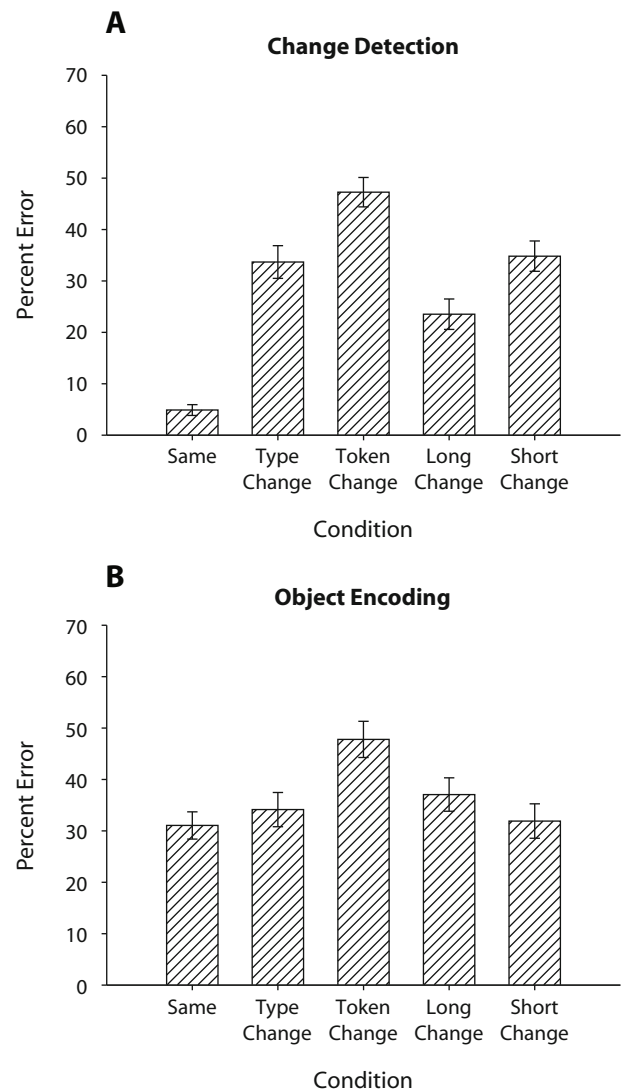


Figure 3. Mean percent error on the change-detection (A) and object-encoding (B) tasks as a function of condition.

both that acoustic information about the auditory objects was useful in the change-detection task, and that the auditory representations were not reencoded to the point of losing such detail.

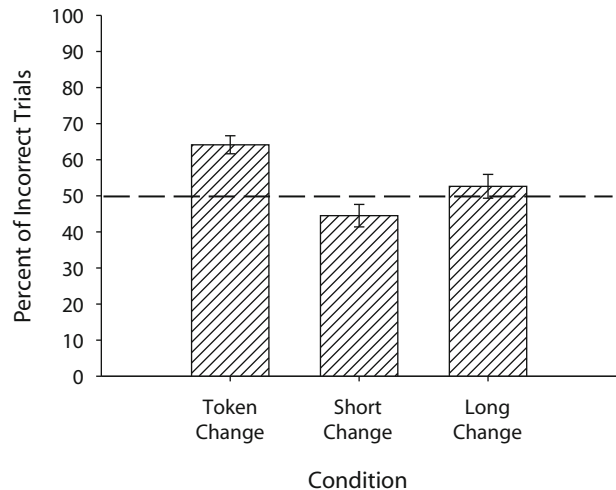
In addition to this acoustic effect on change detection, we replicated the semantic effect of Experiment 1. Recall that an effect of semantics should produce more change deafness on token-change trials than on acoustically matched type-change trials. As can be seen in Figure 3, this is the pattern of data we obtained. Planned comparisons confirmed that the token-change condition produced a higher error rate (47.2%) than did the type-change condition that was matched for acoustic distance (33.7%) ( $p < .01$ ).

Having obtained an acoustic and a semantic effect in the change-detection task, we considered which one was more important to the task. To do so, we compared the token-change and short-change conditions. If semantic information is more important, the error rate in the short-change condition should be lower than that in the acoustically advantaged token-change condition; the reverse pattern would indicate that acoustics are more important. As can be seen in Figure 3, semantic matching had a stronger impact. Planned comparisons indicated that errors were lower in the short-change condition (34.8%) than in the token-change condition (47.2%) ( $p < .01$ ). These results provide further evidence that semantic identity plays an important role in change detection. The greater change deafness for trials with semantic matching (token change) than for those with more similar acoustics (short change) suggests that the perceptual representations in audition (at least in these conditions) are already based on relatively abstract codes.

### Object Encoding

An acoustic effect on object encoding would be indicated by higher error rates on short-change trials than on long-change trials, with error rates on type-change trials somewhere in between. The mean error rates for each condition in the object-encoding task are presented in Figure 3B. Although there was a significant difference across the conditions [ $F(4,92) = 7.95$ ,  $MS_e = 138.57$ ,  $p < .01$ ,  $\omega^2 = .115$ ], the pattern was not driven by an acoustic effect: The mean error rate in the long-change condition (37.0%) was numerically—but not significantly—higher (rather than lower) than that in the short-change condition (31.9%). Acoustics did not play much of a role in the object-encoding task. This result is consistent with Gregg and Samuel's (2008) finding that acoustic properties affected change detection but not object encoding.

Although there was no acoustic effect in the object-encoding task, there was a substantial semantic effect. Planned comparisons demonstrated that the error rate was higher in the token-change condition (47.8%) than in any other condition ( $p < .01$ ). To further explore the semantic effect, the token-change trials were analyzed to see how often listeners were choosing the semantic variant. The resulting data are presented in Figure 4. As can be seen in the figure, listeners did pick the semantic variant on most of the incorrect trials in the token-change condition.



**Figure 4.** Mean percentage of incorrect object-encoding trials (i.e., trials on which the listeners failed to choose *neither*) in which the semantic or acoustic variant was selected.

An ANOVA comparing the three conditions in which *neither* was the correct response (token change, long change, and short change) yielded a significant effect of condition [ $F(2,46) = 9.21$ ,  $MS_e = 253.89$ ,  $p < .01$ ,  $\omega^2 = .211$ ]. Planned comparisons revealed that the effect was primarily driven by the significantly higher percentage of semantic variant (token) responses (64.1%) than long (52.6%) or short (44.5%) variant responses ( $p < .01$ ). Further analysis indicated that only the percentage of semantic variant responses in the token-change condition was significantly different from chance [ $t(23) = 5.68$ ,  $p < .01$ ,  $d = 1.158$ ]. This finding and the substantially higher error rates in the token-change condition than in any other condition indicate that semantic information is more likely to be encoded and utilized than is fine physical detail.

## EXPERIMENTS 3A AND 3B

Experiments 1 and 2 demonstrated that semantic information plays a surprisingly strong role in auditory object representations. In our comparison of semantic effects with acoustic influences, Experiments 1 and 2 share the assumptions that mean  $f_0$  and harmonicity are critical acoustic dimensions of auditory objects and that acoustic similarity can be estimated using Euclidean distance in a two-dimensional space defined by these properties. There is a good deal of previous work to support this assumption. The importance of frequency information has been widely established, beginning with the tonotopic organization of the auditory system (e.g., Clarey, Barone, & Imig, 1992) and including evidence that frequency spacing is critical in sound segregation (e.g., Carlyon, 2004). Periodicity has also been shown to play an important role in sound organization. For example, sound similarity judgments depend heavily on periodicity (Kat & Samuel, 1984), and patterns of selective adaptation also depend on this dimension (Samuel & Newport, 1979). Our previous work (Gregg & Samuel, 2008) showed that both dimensions



influence change-deafness levels, and the results of the short- versus long-change manipulation in Experiment 2 replicate this result.

Nevertheless, it is of course possible that other acoustic properties of sounds contribute to acoustic quality, such as temporal structure and timbre. If some such additional acoustic property is correlated with semantic similarity, our comparison of token changes with matched distance type changes could conceivably be confounded with acoustic similarity. The possibility that acoustic similarity of object tokens explains the semantic effects observed seems small, given that the error rates on the object-encoding task were affected by semantic similarity but not by frequency or harmonicity. Regardless, we designed Experiments 3A and 3B to address this possibility. To account for any acoustic characteristics of auditory objects that may be perceptually important in our tasks, we first conducted an MDS analysis on the 24 sounds used in Experiment 1 (Experiment 3A). We then used the resulting perceptual space of the MDS to replicate Experiment 1 (Experiment 3B).

### Experiment 3A

#### Method

**Participants.** Seventeen undergraduate listeners from the State University of New York, Stony Brook, with no reported hearing deficiencies participated in this experiment for course credit. Two listeners were excluded from all data analyses because they failed to respond to more than 20 trials in each session. The mean age of the remaining 15 listeners was 19 years.

**Stimuli.** The same 24 sounds that were used in Experiment 1 were used in this experiment.

**Procedure.** In most respects, our procedures were similar to those used by Gygi et al. (2007) in their MDS analysis of environmental sounds. With one exception (see below), all of the possible combinations of the 24 sounds were presented to listeners in sequential pairs, including each sound paired with itself and both orders of every pair. Sounds in a pair were separated by 250 msec. For each pair of sounds, the listener's task was to rate their similarity on a scale of 1 (*not similar at all*) to 7 (*as similar as they can possibly be*). Listeners were instructed to use the full scale.

Because the goal of this study was to determine whether semantic information affects change-detection and object-encoding performance (implying a role for such information in auditory perceptual representations), we sought to minimize listeners' use of semantic information in our measurement of psychoacoustic space. In examining the MDS results of Gygi et al. (2007), we saw clear evidence for some semantic clustering, indicating that the MDS procedure does not necessarily provide a pure assessment of psychoacoustic space. As we discuss below, semantic influences on the MDS solution can potentially make it more difficult for us to observe semantic effects on change deafness (i.e., they make our change-deafness experiment more conservative).

For this reason, we took two steps to try to minimize semantic influences on judgments of sound similarity. The first was to explicitly instruct listeners to base their judgments on what the sounds sounded like, rather than on what the objects were. We anticipated that listeners might nonetheless have difficulty ignoring the semantic relationships between the objects. Therefore, our second control to reduce semantic influences and to make the space reflect psychoacoustics as much as possible was to eliminate pairs that might promote listeners' attention to semantic similarity. Trials were constructed so that no two object tokens (i.e., sounds sharing the same identity, such as dog 1 and dog 2) were paired together; this eliminated 24 trials. Similarly, no two objects from the same general

category (i.e., no two animal sounds, such as a dog and a chicken; no two musical instruments, such as a trumpet and a piano; and no two modes of transportation, such as a ship and a train) were paired; this eliminated 14 pairs. This precaution eliminated 4% of the pairs, leaving 992 sound-pair trials. We did not expect these two mild steps to eliminate semantic influences on the similarity judgments, but we did hope to moderate such influences to some extent.

The 992 trials were split into two sessions of 496 trials each. Listeners completed each session on consecutive days. One order of the two sounds within each pair was presented in the first session, and the reverse order of the items within each pair was played in the second session (session order was counterbalanced across listeners). Trials within each session were randomized and split into two blocks of 248 trials to allow for a brief rest break halfway through the session. Each trial was launched 500 msec after a listener's response, timing out after 5 sec. Each session lasted approximately 40 min.

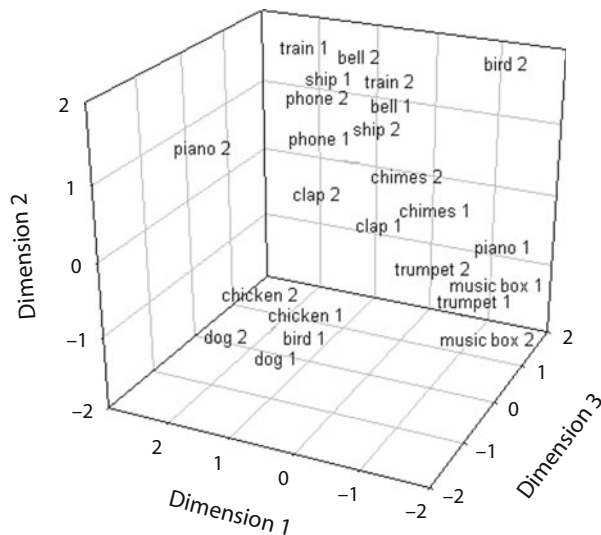
#### Results and Discussion

Similarity judgments of the sound pairs were highly consistent across listeners; the interlistener correlations were high (all above .6,  $ps < .01$ ), and the interlistener reliability was significant [ $W = .299$ ,  $\chi^2(14) = 1,087.21$ ,  $p < .01$ ]. Same-sound pairs were consistently rated with a 7; any same-sound pair given a different rating (fewer than 1% of the total responses) was changed to a 7. Trials in which no response was recorded (fewer than 1%) were eliminated, and, given that there was no effect of order of sound pairs ( $F < 1$ ), data were collapsed across the two orders before being submitted to the MDS solution.

The resulting similarity matrix of sound-pair ratings was submitted to an MDS analysis using the alternating least squares scaling method and a standard Euclidean model. The three-dimension solution accounted for more variance ( $R^2 = .703$ ) and had a lower Kruskal's stress score (i.e., a measure of the badness of fit of the MDS model to the original data; stress = .223) than did the two-dimension solution ( $R^2 = .520$ , stress = .329), so the three-dimension solution was used. Figure 5 depicts the resulting distribution of the 24 sounds in perceptual space.

As can be seen in Figure 5, the MDS perceptual space was quite different from the space used in Experiments 1 and 2. This difference was expected; although  $f_0$  and harmonicity are important properties for auditory change detection (as the data in Gregg & Samuel, 2008, and as the results of Experiments 1 and 2 in the present study demonstrate), we assumed that the two dimensions could only capture a subset of the relevant variance. Any MDS analysis is limited to a representation of the perceptual distances between items; it does not provide any interpretation of the dimensions, and those dimensions may not have any obvious mapping. As in the MDS space of environmental sounds in Gygi et al. (2007), it does not appear that there is any obvious single factor associated with each dimension in our space. What is apparent in both spaces is a clustering of semantically related items.

The semantic clustering in our MDS space resulted in object tokens being generally closer together than they were in the  $f_0 \times$  harmonicity space used in previous experiments. This semantic clustering indicates that our MDS space likely reflects acoustic properties of sounds, along with some semantic properties. If there is a semantic influence on our MDS space, it is important to note that such a



**Figure 5.** Perceptual distribution of stimuli used in Experiments 3A and 3B. The distance between targets in token-change trials (e.g., the distance between ship 1 and ship 2) was matched for three-dimensional Euclidean distance between targets in the type-change trials (e.g., the distance between ship 1 and phone 1).

confound would work *against* our hypothesis that change detection is harder for token than for type changes: A semantic component to our MDS space would mean that two items that should be a certain distance apart on the basis of psychoacoustics will be moved closer or farther apart by some bit of semantic similarity or dissimilarity. Recall that matched pairs of token and type changes are used in our tasks (i.e., two items equidistant in our space, one sharing semantics and the other not, were used). If the distance estimates we use from our MDS space have some semantic influence, then any pairs we create will have some bit of semantics added. Either this should have no effect on our results (because both pairs have the same average extra boost from semantics), or it should dilute the semantic contrast we are aiming for, because even the purely acoustic pair has some semantic component. The former case would have no effect on our test, and the latter would make it harder for us to find the semantic effect we are looking for.

### Experiment 3B

#### Method

**Participants.** Twenty-seven undergraduate listeners with no reported hearing deficits from the State University of New York, Stony Brook, participated in this experiment for course credit.

As in Experiments 1 and 2, listeners were required to meet a criterion of above-chance performance on the object-encoding task. Three listeners were excluded from the data analyses because they failed to meet this criterion. Two additional listeners were excluded because they admittedly did not understand the object-encoding task. The mean age of the remaining 22 listeners was 21 years.

**Stimuli.** The 24 sounds from Experiment 1 were used, and the same kinds of object-encoding and change-detection *same* and *different* (type and token) trials were constructed. In this experiment, the perceptual space obtained from the MDS solution in Experiment 3A

was used to construct trials (see Figure 5), rather than the  $f_0 \times$  harmonicity space used in Experiment 1. That is, in matching acoustic distance for our type versus token trials, distance was based on three-dimensional Euclidean distance. As in Experiment 1, preliminary analyses showed that interobject distances based on a city-block metric were essentially perfectly correlated with Euclidean distance.

**Procedure.** The procedure was the same as that in Experiment 1. Listeners completed both a change-detection and an object-encoding block of trials. On all trials, two groups of sounds were presented over headphones. The two groups were sequential and were separated by 350 msec of white noise. Half of the trials were *same* trials, in which the same sounds were played in both groups, and half were *different* trials, in which one of the sounds from Group 1 was replaced with a different sound in Group 2. There were 144 trials within each block, and block order—change detection first or object encoding first—was counterbalanced across listeners.

**Change detection.** Within the change-detection block, the listeners' task was to indicate whether the two groups of sounds were the same or different by pressing the corresponding key. There were two kinds of *different* trials: token change and type change. In the token-change condition, a sound from Group 1 (i.e., the target) was replaced by a sound in Group 2 that was a different token of the same object (e.g., ship 1 would change to ship 2). In the type-change condition, the target was replaced by a sound that was a different type of object and that was approximately the same distance away from the target in the three-dimensional Euclidean space as were the object tokens (e.g., ship 1 would change to phone 1, because, as can be seen in Figure 5, the distance between the two objects is approximately equal to the distance between the two ship tokens). There were 36 type-change and 36 token-change trials in the change-detection block.

**Object encoding.** Trials within the object-encoding block were similar, except that rather than the trial ending after the second group of sounds, two sequential sounds (*Sound 1* and *Sound 2*) were presented 250 msec apart and 500 msec after the end of Group 2. The task was to determine whether Sound 1, Sound 2, or neither sound had been presented in the trial (i.e., in both Groups 1 and 2) by pressing the corresponding key. There were three conditions: valid, token change, and type change. In all three conditions, one of the two sequential sounds was a foil that had not occurred in either group of the trial; the other sound (the target) varied by condition. In the valid condition, the target sound had occurred in both groups, and thus the correct answer would be selection of this target. In the other two conditions, the correct response was *neither*. In the token-change condition, the target sound was a different token of one of the sounds that had occurred in both groups (e.g., ship 2 could be one of the two sounds when ship 1 had been present on the trial). In the type-change condition, the target sound was a second foil and was a different type of object than any of the sounds in either group. The Euclidean distances on token- and type-change trials were matched; the distance between the two tokens (e.g., the distance between ship 1 and ship 2) was equal to the distance between the token that was presented within both groups and one of the two foils (e.g., the distance between ship 1, which would be presented in both Group 1 and Group 2, and phone 1, which would be presented as one of the two sequential sound choices). There were 48 trials in each condition.

The order of trials within each block was randomized. Each trial was launched 500 msec after the listener's response, timing out after 5 sec. As in Experiment 1, there was no effect of block order ( $F < 1$ ); therefore, data were collapsed across order in all reported analyses.

### Results and Discussion

The central goal of this experiment was to determine whether semantic identity affects change deafness, using the new perceptual space obtained from an MDS solution (which reflects psychoacoustic properties of the sounds that were potentially not represented in Experiments 1

and 2). Figure 6 presents the pattern of error responses for both the change-detection and the object-encoding tasks as a function of condition. As can be seen by comparing Figures 6 and 2, the pattern of error responses obtained in this experiment is nearly an exact replica of what was found in Experiment 1.

**Change detection.** As in Experiment 1, the pattern of error responses indicated substantial change deafness. As Figure 6A shows, errors on *different* trials were very common, whereas errors on *same* trials were rare. This pattern produced a significant effect of condition [ $F(2,42) = 159.66$ ,  $MS_e = 76.99$ ,  $p < .01$ ,  $\omega^2 = .757$ ]. Planned comparisons revealed that error responses were higher on the two types of *different* trials (mean error rate = 42.8%) than on *same* trials (mean error rate = 2.5%) ( $p < .01$ ).

The change-detection data also replicated the semantic effect found in Experiment 1. As can be seen in Figure 6, change deafness was substantially different on *different* trials that involved a semantic change relative to *different* trials that did not involve a semantic change. Planned comparisons demonstrated that this difference was significant: Listeners made significantly more errors on token-change trials than on type-change trials ( $p < .01$ ). This finding provides further evidence that the auditory objects were semantically encoded on change-detection trials.

**Object encoding.** The data from this experiment also replicated the semantic effect found for the object-encoding task of Experiment 1. Error rates were higher on trials in which listeners were presented with an acoustically different but semantically identical (yet incorrect) choice and a foil (token-change condition) than they were on trials in which listeners were presented with two invalid choices, both acoustically and semantically different from any sound on the trial (type-change condition) [ $F(2,42) = 12.10$ ,  $MS_e = 110.09$ ,  $p < .01$ ,  $\omega^2 = .215$ ]. Planned comparisons revealed that the error rate in the token-change condition (42.3%) was significantly higher than that in either the type-change (31.0%) or the valid (27.3%) condition ( $p < .01$ ). There was no significant difference between the type-change and valid conditions.

As in Experiment 1, the data from the present experiment revealed that listeners actually chose the semantic variant, rather than the foil, on trials in which they did not choose the correct (*neither*) response. In fact, listeners chose the semantic variant on 72% of the incorrect trials, significantly greater than chance (50%) performance [ $t(21) = 7.40$ ,  $p < .01$ ,  $d = 1.58$ ]. This finding provides further evidence that listeners encode and rely on semantic category information in auditory scenes.

In summary, we obtained a robust effect of semantics. Recall that in Experiment 3B, we provided a conservative test, because any residual semantic influences on our MDS-generated space would work against finding this semantic effect on change deafness. Despite the very different spaces used in Experiments 1 and 3, we got exactly the same semantic influence on change detection and object encoding.

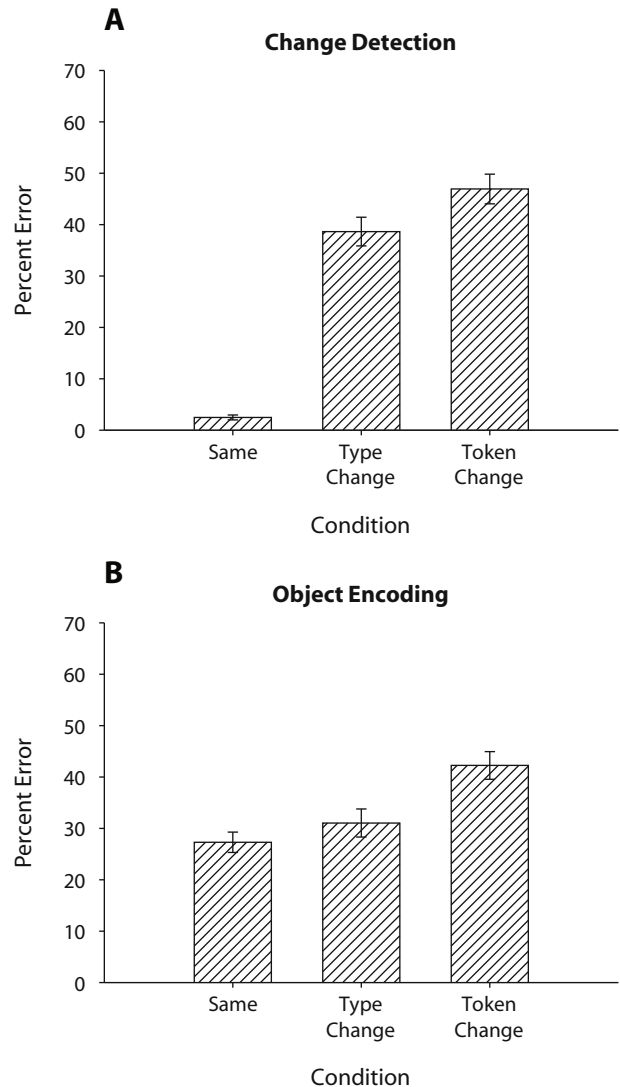


Figure 6. Mean percent error on the change-detection (A) and object-encoding (B) tasks as a function of condition.

## GENERAL DISCUSSION

In this study, we explored the nature of auditory perceptual representations. Our goal was to determine whether auditory representations contain semantic information. The results of all three experiments provide a strong affirmative answer to this question. Auditory representations do contain semantic information, at least about the basic identity of an object. When the acoustics of the sounds were controlled for in the present set of experiments, semantics affected both change detection and object encoding; change-detection and object-encoding performance were more accurate on trials in which an auditory object changed to a different type of object (e.g., a dog bark changed to a ship horn) than on trials in which an object changed to a different token of the same object (e.g., a small dog barking changed to a large dog barking). This



finding is consistent with previous visual perception work (e.g., Gordon & Irwin, 1996) and with recent statistical learning work (see Brady & Oliva, 2008) showing that semantic categories play a role in the representation of visual information.

In Experiment 2, we tested the importance of semantic information relative to acoustic detail in the auditory representations. When semantic information was controlled, acoustic information (at least in terms of pitch and harmonicity) was useful for the change-detection task, but not for the object-encoding task. More change deafness occurred on trials in which there was a short acoustic distance change (e.g., dog 2 changed to piano 2; see Figure 1) than on trials in which there was a long acoustic distance change (e.g., dog 2 and train 1 in Figure 1). These results replicate a pattern in our earlier work (Gregg & Samuel, 2008). Although acoustics were useful for change detection, a comparison of short-change trials with the acoustically advantaged token-change trials showed that more errors were made in the latter condition. Change detection based primarily on acoustic information would have yielded the opposite pattern of results. This finding, along with the lack of an effect of acoustics in the object-encoding task, suggests that auditory representations encode both semantic and acoustic information, but that listeners often encode abstract identity information more than much physical detail. Our findings are consistent with object-file theory claims that transient representations contain identity information plus some physical information (e.g., Gordon & Irwin, 1996; Kahneman et al., 1992). The present results extend object-file theory to the auditory domain and suggest that the transient representations in perception may have consistent properties across modalities.

The results of the present study appear to be inconsistent with the view (e.g., Hollingworth & Henderson, 2002) that transient perceptual representations are richly detailed. In the visual work, one of the key findings in support of the claim that scenes are represented in rich detail is accurate token- and rotation-change detection (see Hollingworth & Henderson, 2002). In contrast, we found high error rates on trials in which there was a token change. If listeners encode perceptual objects in much detail, our listeners should have been able to detect these physical changes. How can this apparent inconsistency be reconciled? The difference may be more apparent than real. Detection rates are actually rather poor in much of the work that claims accurate change detection performance to token changes (over 70% error in Hollingworth et al., 2001, and in Hollingworth & Henderson, 2002). As such, the claim that change-detection performance was accurate may be a bit overstated. Although Hollingworth and Henderson's model may be able to explain good long-term memory for scenes (even when the memory test is a surprise; see Castelhamo & Henderson, 2005), it does not appear to give a complete picture of the nature of perceptual representations. The high error rates in Hollingworth and Henderson's study and in the present study suggest that visual and auditory representations often do not contain the fine details of a scene.

There is an interesting issue raised by one of the main findings of the present study that should be considered. Although our results suggest that perceptual representations contain identity information, the nature of this identity information is a matter for further speculation. There are two possibilities: The identity information may be a rough concept of the object's basic identity, or it may be a verbal code. In practice, it is very difficult to tease the two possibilities apart. It is not clear that there is a method that could eliminate verbal encoding without simultaneously masking the contents of auditory working memory, making the task impossible. A few visual studies have eliminated verbal encoding by using homonyms (see Pollatsek et al., 1984) or by having participants rehearse random numbers aloud (e.g., Hollingworth, 2003) or silently (e.g., Vogel, Woodman, & Luck, 2001). Our study was very similar to the Hollingworth (2003) visual change-detection study, and, in that study, verbal encoding did not have any effect on detection performance. Although visual and auditory working memory may differ, these results suggest that verbal encoding is not solely responsible for our results. Similarly, the procedures that minimized this concern in the present study (using auditory objects rather than words) make it less likely that verbal labels are the primary basis of the observed semantic effect.

In summary, our study provides evidence that semantic information plays a significant role in representations used in perceiving environmental sounds, whereas detailed acoustic information appears to be less influential in these representations. As was reviewed in the introduction, the role of semantic identity information in visual representations is an important and ongoing theoretical issue. The data presented in this article make an important contribution to this issue and to our understanding of how the world is initially represented. Our data indicate that both semantic and acoustic information are likely to be included in auditory representations and that abstract identity information may be especially salient in these representations. As Hollingworth and Henderson (2002) have shown, physical feature information can be accessed when it is useful for the task at hand.

#### AUTHOR NOTE

This work was supported by NIMH Grant R0151663. This project was presented at the 2007 meeting of the Cognitive Science Association for Interdisciplinary Learning and at the 2007 meeting of the Psychonomic Society. We thank Donna Kat for programming the experiments and for the tireless technical assistance. We also thank Carolyn Creary and Kavita Patel for their assistance with data collection. Correspondence concerning this article should be addressed to A. G. Samuel, Department of Psychology, State University of New York, Stony Brook, NY 11704-2500 (e-mail: asamuel@ms.cc.sunysb.edu).

#### REFERENCES

- BOERSMA, P., & WEENINK, D. (1992). Praat: Doing phonetics by computer (Version 4.3.2) [Computer software and manual]. Retrieved October 2005 from www.praat.org.
- BRADY, T. F., & OLIVA, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*, 678-685.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.



- CARLYON, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences*, **8**, 465-471.
- CASTELHANO, M. S., & HENDERSON, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, **12**, 1017-1040.
- CHERRY, C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.
- CLAREY, J. C., BARONE, P., & IMIG, T. J. (1992). Physiology of thalamus and cortex. In A. N. Popper & R. R. Fay (Eds.), *The mammalian auditory pathway: Neurophysiology* (pp. 232-334). New York: Springer.
- DANNENBRING, G. L., & BREGMAN, A. S. (1976). Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 544-555.
- ERAMUDUGOLLA, R., IRVINE, D. R. F., MCANALLY, K. I., MARTIN, R. L., & MATTINGLEY, J. B. (2005). Directed attention eliminates "change deafness" in complex auditory scenes. *Current Biology*, **15**, 1108-1113.
- GORDON, R. D., & IRWIN, D. E. (1996). What's in an object file? Evidence from priming studies. *Perception & Psychophysics*, **58**, 1260-1277.
- GORDON, R. D., & IRWIN, D. E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 136-150.
- GREGG, M. K., & SAMUEL, A. G. (2008). Change deafness and the organizational properties of sounds. *Journal of Experimental Psychology: Human Perception & Performance*, **34**, 974-991.
- GRIMES, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception* (Vancouver Studies in Cognitive Science, Vol. 5, pp. 89-110). Oxford: Oxford University Press.
- GYGI, B., KIDD, G. R., & WATSON, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics*, **69**, 839-855.
- HENDERSON, J. M. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General*, **123**, 410-426.
- HOLLINGWORTH, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 388-403.
- HOLLINGWORTH, A., & HENDERSON, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 113-136.
- HOLLINGWORTH, A., WILLIAMS, C. C., & HENDERSON, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, **8**, 761-768.
- KAHNEMAN, D., & TREISMAN, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29-61). New York: Academic Press.
- KAHNEMAN, D., TREISMAN, A., & GIBBS, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, **24**, 175-219.
- KAT, D., & SAMUEL, A. G. (1984). More adaptation of speech by non-speech. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 512-525.
- MITROFF, S. R., SIMONS, D. J., & LEVIN, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics*, **66**, 1268-1281.
- O'REGAN, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, **46**, 461-488.
- O'REGAN, J. K., DEUBEL, H., CLARK, J. J., & RENSINK, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, **7**, 191-211.
- O'REGAN, J. K., & NOË, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral & Brain Sciences*, **24**, 939-1031.
- O'REGAN, J. K., RENSINK, R. A., & CLARK, J. L. (1999). Change-blindness as a result of "mudsplashes." *Nature*, **398**, 34.
- POLLATSEK, A., RAYNER, K., & COLLINS, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, **113**, 426-442.
- POLLATSEK, A., RAYNER, K., & HENDERSON, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 199-210.
- RENSINK, R. A., O'REGAN, J. K., & CLARK, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, **8**, 368-373.
- SAMUEL, A. G., & NEWPORT, E. L. (1979). Adaptation of speech by nonspeech: Evidence for complex acoustic cue detectors. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 563-578.
- SHAFIRO, V., & GYGI, B. (2004). How to select stimuli for environmental sound research and where to find them. *Behavior Research Methods, Instruments, & Computers*, **36**, 590-598.
- SIMONS, D. J., & AMBINDER, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science*, **14**, 44-48.
- SIMONS, D. J., & CHABRIS, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, **28**, 1059-1074.
- SIMONS, D. J., & LEVIN, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, **5**, 644-649.
- SIMONS, D. J., & RENSINK, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, **9**, 16-20.
- TREISMAN, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, **12**, 242-248.
- TREISMAN, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology*, **40A**, 201-237.
- TREISMAN, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, **47**, 862-875.
- VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology.
- VITEVITCH, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 333-342.
- VOGEL, E. K., WOODMAN, G. F., & LUCK, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 92-114.
- WARREN, R. M., OBUSEK, C. J., & ACKROFF, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, **176**, 1149-1151.

## NOTES

1. We chose to use standard Euclidean distance, rather than city-block distance. The two methods of computing distance yielded similar results and were highly correlated ( $r = .98$ ).

2. We made an explicit decision not to use  $d'$  scores in our analyses of change detection (and, therefore, in that of object encoding). Change deafness is specifically indexed by the likelihood of listeners reporting *same* when, in fact, there had been a change (*different*). It is important to know how often the other type of error occurs (responding *different* to a *same* trial), but combining the two error types would obscure, rather than clarify, change deafness. For example, if a particular manipulation increased the likelihood of responding *different* on a *same* trial, this would lower  $d'$ , but would not indicate that there was greater change deafness. The actual data pattern for this error type in our experiments (low and essentially flat across the experimental manipulations) assures us that there is nothing odd driving the critical error type, but combining the error types would only tend to obscure what we are trying to observe.

(Manuscript received August 27, 2008;  
revision accepted for publication October 23, 2008.)