Effects of Context on the Classification of Everyday Sounds

James A. Ballas and R. Timothy Mullins

Center for Behavioral and Cognitive Studies
George Mason University
Fairfax, VA 22030

This research was supported by the Perceptual Science Program
Office of Naval Research.

Technical Report #ONR-89-1
October, 1989

Approved for public release; distribution unlimited. Reproduction in whole or part is permitted for any purpose of the United States Government.
The effects of context on the classification of everyday sounds was examined in five experiments. Context was produced by meaningful sounds and by phrases describing an environmental scene. All experiments presented listeners with pairs of test sounds that are confused in identification, but which are discriminable. These test sounds were presented for classification in isolation, and embedded in sequences of other everyday sounds. Three types of embedding sequences were used: 1) sequences consistent with the correct response; 2) sequences biased toward an incorrect choice; and 3) neutral sequences composed of randomly arranged sounds. Two paradigms, binary-choice and free classification were used. The results indicated that context could bias the response against the correct response, but did not raise performance above isolated classification performance. Performance was consistently poorest in biased context and best in both isolated and consistent context. Performance in random context depended upon the paradigm and the performance measure. In the free response paradigm, biased sequences produced responses that were appropriate for the context but incorrect as classifications of the sound. A signal detection analysis indicated that sensitivity in detecting a sound that is out-of-context remains constant for different paradigms, and that response bias is conservative, especially with a free response paradigm. Labels added to enhance context generally did not change the effects of context, suggesting that sounds alone are usually sufficient to generate these contextual effects.
Abstract

The effects of context on the classification of everyday sounds was examined in five experiments. Context was produced by meaningful sounds and by phrases describing an environmental scene. All experiments presented listeners with pairs of test sounds that are confused in identification, but which are discriminable. These test sounds were presented for classification in isolation, and embedded in sequences of other everyday sounds. Three types of embedding sequences were used: 1) sequences consistent with the correct response; 2) sequences biased toward an incorrect choice; and 3) neutral sequences composed of randomly arranged sounds. Two paradigms, binary-choice and free classification were used. The results indicated that context could bias the response against the correct response, but did not raise performance above isolated classification performance. Performance was consistently poorest in biased context and best in both isolated and consistent context. Performance in random context depended upon the paradigm and the performance measure. In the free response paradigm, biased sequences produced responses that were appropriate for the context but incorrect as classifications of the sound. A signal detection analysis indicated that sensitivity in detecting a sound that is out-of-context remains constant for different paradigms, and that response bias is conservative, especially with a free response paradigm. Labels added to enhance context generally did not change the effects of context, suggesting that sounds alone are usually sufficient to generate these contextual effects.
Effects of Context on the Classification of Everyday Sounds

There is ample evidence that context influences the perception of sounds. Southworth (1969) found that evaluative judgments of sounds in a city depended in part upon the site where the sound occurred. Anderson, Mulligan, Goodman, & Regen (1983) found that sounds influenced perceptual appreciation of an environmental site. The experiments in this report examine the effects of contextual stimuli, including sounds and descriptive phrases, on the classification of everyday sounds. Classification is defined as the labelling of a sound.

The effect of context has a long history in language research, and certain findings are well established. Ogden and Richards (1923), in their triangular model of signification, claimed that the meaning of a word can only be grasped in the context of its usage. Olsen (1970) argued that words or utterances are used to specify a particular referent among the alternatives that are possible in the existing context. In his model, the context prescribes a set of alternatives, and the words actually used specify the intended alternative among this set. More recent research has defined the different types of context that operate in language, and established the effects and the conditions in which the effects are obtained. Massaro and Oden (1980) distinguish between auditory and linguistic context. Auditory context is provided by the other sounds in a sequence. The literature on auditory context is extensive both for speech and non-speech sounds. Auditory context influences the perception of individual sounds (Bennett, Parasuraman, Howard, & O'Toole, 1984; 1973; Howard, O'Toole, Parasuraman, & Bennett, 1984; Watson, 1987), the perception of spectral components within a sound (Green, 1983), as well as the perception of patterns (Bregman, 1978; Warren, Obusek, Farmer, & Warren, 1969; Warren, 1974; Vicario, 1982).

Linguistic context can influence perception through both lexical and sentential effects (Warren, 1970), especially when ambiguities are present or when the stimulus is impoverished (Ganong, 1980). By manipulating signal to noise ratios (S/N), Miller, Heise, and Lichten (1951) found that speech intelligibility was related to linguistic context. At high S/N levels, sentence presentation of words provided a 20% or more advantage in intelligibility over isolated presentation. However, the effects are found only under certain conditions. Forster (1976) argues that sentential effects of context will not be found when the stimulus information is sufficient and the observer is given sufficient time. Marslen-Wilson and Welsh
(1978) concluded that the effects of sentential context on restoration of mispronounced syllables occurred with the 2nd and 3rd syllables in a word, but not with first, and only when the mispronounced word was syntactically and semantically appropriate. Thus, the timing of the contextual information is important. Miller and Glucksberg (1988) reviewed studies on polysemy and ambiguity and concluded that context resolves lexical ambiguity in a variety of language tasks including phoneme monitoring, lexical decision and word naming. The effects of context were not always immediate but rather took several hundred milliseconds to become fully effective. They proposed that upon presentation of an ambiguous word, all of its potential meanings may be momentarily activated; however, within a brief interval, only the meaning appropriate to the sentence context is still activated.

In the studies which follow, the effect of context on the classification of an everyday sound was assessed. Conditions were manipulated to examine some of the effects that have been found in language research. Stimulus ambiguity was introduced by using items that can be confused, and contextual conditions were manipulated to influence the classification of the ambiguous items. Temporal position of the ambiguous item was varied. This research was designed to pursue the implications of recent studies which have suggested that the classification of everyday sounds involves a consideration of the alternative causes of the sound when the sounds are presented in isolation, and that the number of and probability of alternative causes is related to the speed of classification (Ballas, 1989; Ballas, Dick, & Groshek, 1987). These results are similar to well-established relationships in cognitive research between the probability of class membership and category verification time. Current research suggests that at least two factors govern the consideration of alternative causes: the stereotypy of the sound and the probability of the cause. In experiments where a possible cause was presented prior to the sound, and the listener asked to verify or reject the cause, confirmation response time was faster for stereotypical sounds and high probability causes, and increased as sound stereotypy and the probability of the cause were reduced (Ballas & Sliwinski, 1986). These effects can be thought of as a kind of stimulus-response compatibility. Although generally there is good compatibility between an everyday sound and its primary cause, this compatibility decreases as the sound becomes less typical of its type and as the cause becomes a less likely alternative.

The alternative causes for a sound are not necessarily of a similar type. Rather, alternative causes define a causal equivalence class because the acoustic
effects of different causes are confused, even though the sounds of these causes are discriminable. A quantification of alternative causation for an equivalence class of sounds was developed using the uncertainty metric from Information theory. H values for sounds can be calculated which quantify the uncertainty about the actual cause of a sound, or more specifically, the response equivalence in identifying the sound. These values have been found to be stable for different exemplars of a sound. The H values correlate with time to identify a sound, with subject ratings of uncertainty, and with ratings of the identifiability of a sound (Ballas, 1989). Finally, they are consistent for different groups of listeners, including high school students, college students, and older listeners (Ballas & Barnes, 1988; Ballas, Dick, & Groshek, 1987). The values offer a practical quantification of the identifiability of everyday sounds, and as such, can be used to select a set of identifiable sounds.

If a sound has reasonable, alternative causes which denote very different events, this constitutes an equivalence class of causes for the sound. The choice of a cause within this class would be determined by contextual factors, if indeed the alternative causes cannot be distinguished on the basis of stimulus properties. This possibility guided the design of the studies in this report. Test sounds were chosen which had, in previous studies, been confused when presented in isolation for classification. The test sounds were selected in pairs to represent similar acoustics, but very different causes. It should be emphasized that although the sounds were confused in classification tasks, they were discriminable. It was expected that context would affect classification of these sounds. Context was generated by embedding the test sound in a sequence of sounds consistent with the true cause, and alternatively, in a sequence which was biased toward the incorrect cause. In two control conditions, the test sounds were embedded in random sequences and presented for classification in isolation. Other effects examined included the position of the test sound in the sequence, the type of classification paradigm (binary-choice and free response), and labeling of the contextual scene.

Experiment 1

Experiment 1 was the baseline study for the series of experiments. In it, the effect of context on classification was examined by comparing performance on test sounds presented in isolation, and embedded in three contextual conditions, consistent, biased, and random. A binary-choice paradigm was used to control
task demands which otherwise would vary between isolated and contextual sound presentation. The order of isolated and contextual trials was counter balanced to assess transfer effects. The position of the test sound in the sequence of contextual sounds was varied to assess the effects of short-term memory interference.

Method

Listeners. Forty undergraduate students participated, twenty in each of two orders of task completion.

Stimuli. Eighty-two sounds were obtained from a compact disc sound effects library of 2000 sounds. These sounds were digitized at 20 kHz and presented using a Macintosh SE with a GW Instruments MacADIOS digitizing system. The sounds were low-pass filtered at 9 kHz and presented at a comfortable sensation level. The sounds were combined into forty-four sequences of three to six sounds. Each sequence was designed to reproduce a scene such as a basketball game, using a phone, or taking a walk in the park. Within each of the sequences one sound was used as the test item. This item was placed in the last, second to last, third to last, and fourth to last position in the different sequences. There were 23 test sounds altogether. Using confusion data from previous studies, the test sounds were chosen in 12 pairs of confusable sounds (one sound was used in two pairs). For example, one test pair consisted of the sound of a fuse burning and the sound of food frying. Previous research had shown that these two causes had been attributed to the same sound and therefore could be confused. The discrimination of these pairs was tested in an ABX paradigm. Three listeners made no errors in discriminating the pairs of sounds.

The test sounds were embedded in three types of sound sequences. Twenty of the 24 test sounds were embedded in both consistent and biased sequences. The remaining four test sounds were used only in one of these conditions. Twenty four sequences with a random arrangement of sounds were also constructed to test the effect that embedding itself would have on classification.

Procedure. The listeners participated in two binary-choice tasks, classification of the test sounds in isolation, and classification of the test sounds in context. The order of the two tasks was reversed for half of the listeners. In the isolated classification task, the listener heard the pair of test sounds and made a choice of the sequence in which the pair occurred. The options in the choice were labels for the sounds, so the task required a classification of the sounds, as well as a
discrimination. In the context classification task, the listener heard a sequence of sounds, including the test sound (but only one from each pair) and made a choice of which sound in the pair had occurred. Each test sound was presented in a consistent sequence half of the time, and in a biased sequence the other half. For example, the two test sounds in one pair were food frying and a fuse burning. The sound of food frying was inserted into a food preparation sequence (i.e., slicing-chopping-frying) which would be consistent with the correct classification of the sound. It was also inserted into a blasting sequence (i.e., match lit-frying-explosion) which would be biased toward the other member of the test pair (fuse). The test sounds were also presented in randomly-designed sequences. The three types of context were presented in a block of 68 trials, randomly ordered.

Results

Average performance by listener was calculated for each combination of context and order. The four context levels included no-context (isolated binary-choice classification), and classification in each of the three types of context (consistent, biased, and random). Note that the order variable modified only the order of the no-context condition. This condition was presented either before (Order A) or after (Order B) the other three context conditions. These data were analyzed in a 4 (context) by 2 (order) mixed ANOVA. All effects were significant. The interaction of task by order was significant, $F(3,114) = 5.09, p = .0024$. The means are shown in Figure 1. The simple effect of context was significant for both Order A, $F(3,114) = 6.88, p < .001$, and Order B, $F(3,114) = 62.53, p < .0001$. Newman-Keuls tests indicated that classification in Order A with biased sequences was significantly less than classification with consistent sequences and less than classification of isolated test sounds. In Order B, these same comparisons were significant and in addition, all other comparisons were significant. The performance pattern was similar in both orders, supported by a significant effect for context, $F(3,114) = 18.15, p = .0001$. Specifically, performance was best on isolated classification and classification with consistent sequences, poorest on classification with biased sequences, and in between on classification with random sequences. This pattern held for test sounds that could only be poorly discriminated. Two pairs of test sounds that were poorly discriminated were the stapler-light switch combination and the hammer-basketball combination. The mean performances in identifying these four sounds in the isolated, consistent, biased and random conditions were .58, .57, .52, and .56 (for the sounds as a
whole, a difference of .06 was significant). A similar pattern was found with the remaining sounds, which were well classified in isolation (.92, .89, .81, .85). These patterns are similar to the effect of context on performance for all the sounds (Figure 1).

Performance in classifying sounds was significantly better within the three contexts when these conditions followed isolated classification, $F(1,38) = 12.16, p = .0013$. The isolated classification context gave the listeners exposure to the pair of sounds prior to hearing them in a sequence of sounds.

The relationship between performance in the contextual conditions and properties of the sequences was analyzed with correlations. The properties that were examined included the number of sounds, the position of the test sound, the number of sounds after the test sound, and the duration of the test sound. Three correlations were significant (Table 1). In consistent context, for both orders, there was an inverse relationship between performance and the position of the test sound, such that better performance occurred with test sounds earlier in the sequence. In the biased context, but only for Order A, there was an inverse relationship between performance and the number of sounds in the sequence, such that better performance occurred with shorter sequences.

Discussion

One of the interesting findings was that context had a negative effect reducing performance below isolated classification levels but not a positive effect raising performance above isolated classification levels. Identification with consistent sequences was not significantly better than classification of isolated sounds when isolated classification preceded contextual classification, and was significantly less when isolated classification followed contextual classification. It is possible that isolated classification performance defines a performance ceiling, which is not exceeded when the sounds are later presented in a sequence that supports the correct response. It should be emphasized that the isolated classification task was different from the contextual classification task in two ways. First, the sounds were not embedded in other sounds. Second, both alternatives were presented and the listener had to choose the presentation order. This means that both members of the equivalence class were presented to the listener. Performance would probably be reduced if only one member of the class had been presented, but studies have not been conducted to test this prediction.
It is interesting that the effects of context were found with sounds that could be well identified in isolation. That is, even though listeners could identify the two sounds at better than 90% correct when the two alternatives were presented alone, their performance dropped to 81% when these sounds were placed in a biased context. Furthermore, prior exposure to the sounds in context did not improve performance on isolated classification when this condition followed contextual classification, even though the listeners had heard the test sounds two times before they started isolated classification. The relative pattern of the effects of context persisted even after prior exposure to the test sounds in the isolated classification condition. However, not all of the differences were significant. Finally, in Order B, without prior exposure to the sounds in isolation, classification of sounds in context was poorer than isolated classification performance, even when the test sounds were presented in consistent context.

Relationships between sequence properties and performance were generally minor, but those that were significant were consistent with context effects. Improved performance occurred in the consistent context as the number of sounds following the test sound increased. This is consistent with findings in language research that context effects become stronger with an increased number of contextual elements between the test item and test response. In the biased context, better performance was related to fewer sounds in the sequence, which suggests a weaker context.

Experiment 2

In order to establish the reliability of some of the results in the first experiment, a second experiment was conducted using a different group of listeners. Conditions were similar to the first study, except that only one order of trial presentation was used.

Method

Listeners. Thirteen graduate students participated for class credit.

Stimuli. Same as Experiment 1.

Procedure. Same as Experiment 1 except that all the listeners first performed isolated classification and then classification in context.
Results

The average performance was calculated for each subject in each context condition and is shown in Figure 2. Note that only data for one order of context (Order A in Experiment 1) are shown because the listeners were only tested in this order. The pattern of these means is identical to the pattern found in Experiment 1. A planned comparison of the difference between the consistent and biased condition was significant, $F(1,36) = 6.22, p < .05$. Comparisons between the other conditions were not significant. A similar result was also found in Experiment 1. Thus the significant effect of consistent and biased context was replicated, and the same pattern of contextual effects were found, with a different group of listeners. Relationships between performance and sequence properties (Table 2) were similar in this experiment to those found for Order A in Experiment 1. However, the only significant correlation was an inverse relationship between performance and the position of the test sound in consistent context, supporting the finding in Experiment 1 that the context effect in consistent context became stronger as more sounds intervened between the test sounds and the response.

Experiment 3

The results of Experiment 1 and 2 supported the prediction that context would influence the interpretation of ambiguous sounds. As shown in Figure 1, accuracy was highest when the context was consistent with the correct response, lowest when the context was biased for the alternative test sound, and intermediate when the context was a random set of sounds. However the effects of consistent context were not sufficient to raise performance above isolated classification levels. One could conclude from these results that: 1) isolated classification performance is the limiting factor in classification of sounds in context; and 2) the bias effect of consistent contextual can only offset the negative effects of embedding the test sound within a sequence of other sounds. In support of the first conclusion, the similar performance level between consistent context and isolated classification suggests that classification of a test sound is a limiting factor. It should be emphasized that the isolated classification task is not a measure of discrimination ability. In fact, the test sounds used were discriminable. Furthermore, if contextual information is available, the listeners need only give the response that is appropriate to the context to be completely correct in the consistent context
condition. This is probably what occurs in natural listening, and is what would be expected based on the ability to easily interpret polysemous words in sentences. Listeners did not do this in either Experiment 1 or 2.

In support of the second conclusion, embedding would produce interference in short term memory as the sounds are produced. If the random contextual condition is used as a measure of this interference effect, and set as a baseline of performance in a contextual classification task, then consistent context does improve performance above this baseline, and significantly so in Order B in Experiment 1. This assumes that the sounds are not being identified immediately. Otherwise, the classification of each sound in a consistent contextual condition would be handily chunked. Ballas and Sliwinski (1986) showed that the classification of isolated sounds took from .6 to 6.2 s for the 41 sounds they used. Many of those sounds were similar to the ones used in this study. Thus the effects of consistent context may be seen as offsetting the negative effects of embedding the test sound in other sounds. The inverse relationship between performance in consistent context and position of the test sound is consistent with this interpretation. As the test sound occurs earlier in the series, or as more sounds follow the test sound, then responses tend to be more accurate for the consistent contextual condition. The basis for this relationship could be that listeners rely less on the stimulus information, and more on the contextual cues. However, the opposite relationship was not observed in the biased condition, casting doubt on a simple interpretation.

Although the two conclusions are not mutually exclusive, it would be helpful to know the relative importance of each. Distinguishing between these alternatives requires manipulating response bias through context (response bias could be manipulated directly through a payoff matrix, but the interest in these studies is contextual effects on bias). If isolated classification is the limiting factor, then contextual changes to increase response bias should not raise contextual classification performance above isolated classification levels.

Experiments 3 and 4 were designed to manipulate contextual effects on response bias. Because the consistent context was generated by other sounds, each of which needs interpretation, the intended context may not have been produced. This possibility was investigated in Experiment 3 by giving the listeners a phrase describing the scene represented by the series of sounds. This phrase would explicitly define the context. Howard and Ballas (1980) found that a brief description of the class of sounds to be heard enhanced classification of sound...
patterns that were semantically consistent with the description but disrupted classification of sound patterns that were semantically inconsistent with the description. A similar effect was expected in Experiment 3. A descriptive phrase should suggest classification responses consistent with the phrase and cast doubt about classification responses inconsistent with the phrase. It was expected that performance would increase with consistent context, and decrease in biased context, relative to isolated contextual performance.

**Method**

*Listeners.* The listeners were twenty undergraduate students who participated for class credit.

*Stimuli.* The sounds and sequences were identical to those in Experiment 1 except that 1) the random sequences of sounds were not used because a legitimate description of the meaning of these sequences could not be given; and 2) the four test sounds that had been used only in consistent or biased context were not used, nor were they used in subsequent experiments. Thus 20 sequences were used in these contextual conditions, and stimuli were identical for consistent and biased context.

Written phrases describing each of the meaningful sequences were generated and provided to the listener prior to presenting the sounds themselves. These phrases described the environmental scenes represented by the sounds.

*Procedure.* The procedure was similar to Experiment 1, except that a phrase describing the sound sequence was presented prior to the sounds. After reading the phrase, the listener initiated the sequence of sounds. Furthermore, after making a choice, the listeners indicated their confidence in the choice on a four-point scale with 1 being "very unsure" and 4 being "very sure". All listeners were tested in isolated classification and in the two context conditions in a repeated measures design.

**Results**

The percentage of correct responses by context condition was obtained for each subject. The three context conditions were no-context, consistent context, and biased context. In a one-way, repeated-measures ANOVA, contextual condition was significant, $F(2,38) = 5.26, p < .02$. Mean percent correct for the three context conditions is shown in Figure 3. Newman-Keuls tests revealed that the biased context was significantly less than both the no context and the consistent
context conditions. The difference between the no context and the consistent context conditions was not significant. Correlations between performance and properties of the sequences (Table 3) were consistent with the results of Experiments 1 and 2; the only significant outcome was an inverse relationship between performance and the position of the test sound in the consistent context.

Confidence ratings on correct and incorrect responses were significantly different, $t(662) = 10.90$, $p < .001$, with greater confidence for correct responses. Confidence ratings for the consistent context were not significantly different from the rating for the biased context.

Discussion

Comparing the results of this experiment to Experiments 1 and 2, the addition of a label had little effect on performance, and it would seem that the effects of context are limited to offsetting the negative effects of embedding a sound in a series of other sounds. However, there are other factors involved which need examination. The experiments were repeated-measures studies, and prior experience to the limited set of sounds might reduce the potential effects of context. Furthermore, the binary-choice paradigm is less susceptible to bias effects than other paradigms because it limits the response alternatives. The binary-choice nature of the classification task in Experiments 1, 2 and 3 may have helped the listeners ignore contextual cues. Experiment 4 was designed to pursue these possibilities.

The results of the confidence ratings showed that listeners had some knowledge of when they were inaccurate, but this knowledge was not related to the changes in contextual conditions. This apparent contradiction can be explained by the relative magnitude of differences in performance with changes in context versus changes in sounds. Changes in context produced average differences of less than 10%. Changes in sounds produced average differences as large as 50% in consistent context and 70% in biased context.

Experiment 4

In this experiment the same stimuli were used in an unconstrained classification task, and the listeners asked to identify all of the sounds heard. The test sounds were presented both in isolation and embedded in three types of context. If changing the task produces better classification of the test sounds when
presented in consistent context compared to isolated presentation, then binary-choice task demands would account for the similar performance in isolated and consistent context classifications. However, this task-demand explanation would not account for the poorer classification in biased context compared to random context, which was significant in Order B of Experiment 1. Nevertheless, in order to determine if the binary-choice nature of the task was mitigating the effects of context, listeners were presented the same sounds as those used in Experiment 3 and asked to identify sounds by writing descriptions. We were particularly interested in the classification of the test sounds, but did not convey this to the listeners. No information about the sounds was presented to the listeners. In order to get a baseline on classification performance, the test sounds were presented singly in a no-context condition. This procedure was a departure from the first three experiments which had paired presentation of the test sounds in the no-context condition. It was felt that paired presentation might provide contextual information about the causes of the sounds, and/or would bias the unconstrained classification of the sounds toward one or another member of the pair.

An important advantage of this design is that it would allow us to measure the denotative effects of context when the listener is uninformed about which sounds are the test sounds, and uninformed that the test sounds were selected in pairs to represent equivalence classes. To insure this, a completely randomized design was used in this study to eliminate the transfer effects between isolated and contextual classification. The advantage of this procedure occurs especially in the analysis of biased-context performance. By examining the free classifications of test sounds in biased-context sequences, we could determine if the biased context would suggest the other member of the pair and lead to more classifications that were appropriate for the context, but incorrect as a classification of the sound. The binary-choice task did not permit this determination because the listener was always informed about the alternatives we were testing.

Method

Listeners. The participants were twenty undergraduate students who participated for class credit. Five listeners were used in each of four conditions: no context, consistent context, biased context, and random context.

Stimuli. The test sounds and sound sequences were the same as those used in Experiment 3 with the addition of the 20 random sequences that contained the same test sounds used in Experiment 3, and with the elimination of one of the
consistent sequences because it contained a test sound that was presented as a contextual sound in another sequence in the consistent condition. With these changes, there were 20 random and biased sequences, and 19 consistent sequences.

Procedure. Listeners in the no context condition were presented with one of the test sounds and asked to identify it by writing a phrase. A single presentation was given for each sound. Listeners in the three context conditions were told they would hear a series of sounds and were to write down as many as they could identify. Each sequence was presented once. After all the sequences were heard and identified, they were presented again (and repeated if the listener wished) to provide an opportunity to identify the sounds missed in the first hearing, or an opportunity to change an answer. Initial and second responses were kept separate. The sequences included the test sounds, chosen in pairs as in the previous experiments. The test sound was not highlighted in any manner, nor was the possibility or identity of alternative versions and meanings of the test sounds suggested.

Results
The responses were scored on the number of sounds identified on the first round and the second round, whether the classification of the test sound was correct, and whether the classification response identified the sound as the other sound in the test pair (equivalence class). These responses will be called Test Alternatives (TAs).

A one-way ANOVA on the percentage of correct responses for the different conditions was significant, $F(3,16) = 7.43, p < .003$. The means are shown in Figure 4. A Newman-Keuls analysis showed that the consistent and no context conditions had a greater percentage of correct classifications than the other two conditions, but there were no significant differences within these groupings.

A one-way ANOVA on the percentage of Test Alternatives (TAs) was significant, $F(3,16) = 4.47, p < .02$. The means are shown in Figure 5. A Newman-Keuls analysis revealed that the biased context produced a greater percentage of TA responses than the other conditions, which did not differ from each other. Thus, the biased context produced a greater number of responses which would be consistent with the biased context of this condition, but which were actually incorrect.
Analyses of variance were carried out on the number of sounds identified on the first and second rounds and were not significant for effects of context. Correlations between performance and properties of the sequences were not significant (Table 4) with the exception of a positive correlation between TAs and the position of the test sound in consistent context. Generally, there were very few TAs with consistent context, but those that were given were produced when the test sound occurred towards the end of the sequence.

Discussion

The results of Experiment 4 confirm the results of Experiments 1 to 3 with a different paradigm. The pattern of effects due to context are similar, with performance poor in biased context, and best in isolated and consistent context. Although performance with consistent sequences was greater than performance on isolated sounds, the difference was not significant. Most importantly, biased context was shown to produce significantly more TA classifications. This finding supports the selection of the test sounds in pairs representing a class of equivalent acoustic stimuli.

Overall, the results of Experiments 1 to 4 support the general prediction that context influences the interpretation of ambiguous sounds. Two paradigms were used to study the effects of context. The experiments involved the presentation of a test sound in two contextual conditions, consistent and biased. The studies showed that a set of contextual sounds biased the listener away from the correct cause of the sound and offset the effects of embedding a sound in a sequence, but did not raise classification performance above isolated classification performance. This outcome occurred in both a binary-choice and a free classification task, and when verbal labels were used to enhance the generation of a specific contextual meaning.

Experiment 5

One further explanation for the limited positive effects of context was tested in Experiment 5. It could be that the biasing effect of a label was not strong enough to offset the response restriction of the binary-choice paradigm. An experiment that combined a label intended to produce a stronger context effect than sounds alone, and used a free classification paradigm to eliminate response restriction was
conducted. The labels were expected to enhance both the negative and positive effects of context. A further advantage of this experiment is that it permitted us to examine the effects of a label, and thus context, on the unconstrained classification of isolated sounds. Both consistent and biased effects could be examined.

Method

Listeners. Twenty students participated for class credit. Five listeners were used in each of four conditions.

Stimuli. The sound sequences were the same consistent and biased sequences used in Experiment 4. Random sequences were not used because it would be inappropriate to label them. Phrases describing the scene were the same as those used in Experiment 3. Labeling was used for both isolated and embedded stimuli.

Procedure. The listeners were randomly assigned to one of four conditions and asked to identify the sounds they heard. Listeners in the first and second conditions heard the sounds in isolation. Listeners in the first condition saw a phrase which was consistent with the sound they heard. Listeners in the second saw a phrase which was biased against the sound they heard, but consistent with the TA in the test pair. Listeners in the third and fourth conditions heard the same sounds embedded in sequences, which were consistent with the test sound in the third condition and biased in the fourth condition. They also saw labels which were consistent in the third and biased in the fourth condition. Labels and the sequences were always matched for meaning. A single presentation was given for each sound or sound sequence. Listeners were told they would hear a sound (sequence of sounds in the third and fourth conditions) and were to write it down (write down as many as they could identify). After all the trials had been completed, the sounds were presented again to give the listeners an opportunity to identify the sounds they had missed round, or to change their answer. Initial and second responses were kept separate in the analysis. The test sound was not highlighted in any manner, nor was the possibility or identity of alternative versions and meanings for the test sounds suggested.

Results

The responses were scored on whether the classification of the test sound was correct, whether the classification response was TA response, and on the number of sounds identified on the first round and on the second round.
A one-way ANOVA on the percentage of correct responses for the four conditions was significant, $F(3,16) = 12.61, p < .0002$. Newman-Keuls tests showed that the means for the two conditions with consistent context were significantly greater than the two conditions with biased context (Figure 6), but there were no significant differences within these two groupings.

The TA responses were analyzed with a one-way ANOVA for the four conditions and a significant effect was found, $F(3,16) = 39.05, p < .0001$. Newman-Keuls tests showed that the means for the two conditions with biased context generated significantly greater percentages of the TA responses than the two conditions with consistent context (Figure 7). This occurred not only for sounds embedded in other sounds, but also for sounds presented in isolation. None of the correlations of performance with properties of the sequences were significant (Table 5).

Four sequences generated responses for a third alternative to one pair of test sounds. The two test sounds in the pair were the pull-chain light switch and the stapler. These were placed in either a laundry room sequence or a copying machine sequence, depending upon the context condition. For example, the light switch would be consistent with a laundry room, but inconsistent with a copying machine. The opposite relations would hold for the stapler sound. However, in this free classification experiment, a majority of listeners thought that the test sounds were coins being placed in machines, in both the laundry room and the copy machine sequence. This serendipitous effect of context would affect the results of the experiment, so the analyses were done without these four sequences. The ANOVA and multiple comparison results were not changed with the exclusion of these sequences. The means were similar, except for the average correct responses in the consistent labels and sounds condition which increased from 69% to 74%, reflecting the elimination of these four sequences in which correct responses were low because of the unforeseen third alternative.

General Discussion

Experiment 5 showed that the effect of both consistent and biased context was similar for isolated sounds and embedded sounds. Together with the previous studies, it would appear that sounds alone are sufficient to generate the positive and negative context effects, and that labeling and embedding have equivalent but
not additive effects, unless the labeling serves to identify and restrict the alternatives, as in a binary-choice paradigm.

The exception to this conclusion is the effect of random context, which is based upon embedding alone, not the semantic content of the embedding. The effect of random context is complex, and changed depending on the paradigm. Random context had an effect similar to consistent context both in Order A of Experiments 1 and 2 and in the generation of TA responses in Experiment 4. The similar effect of consistent and random context in Experiment 4 is reasonable because neither would suggest the TA response. The similar effect of these contexts in Order A means that prior exposure to the sounds can mitigate the usual negative effects of random context.

Random context had an effect similar to biased context in the generation of accurate responses in Experiment 4. This is reasonable because neither context would suggest the accurate response in a free classification paradigm. Finally, in Order B of Experiment 1, random context had an intermediate effect between consistent and biased context. In this condition, without prior exposure to the sounds, the negative effect of embedding the test sound in other sounds is most clearly seen.

**Signal Detection Analysis**

The results have implications for contextual effects in a task in which the alternative classifications of a sound are known. To draw out these implications, a signal detection analysis was performed. This analysis was made possible because the test sounds were chosen to represent two alternative classifications of sounds that can be confused, and thus there was a basis for defining the signal and noise conditions. All the experiments involved the presentation of a test sound in two contextual conditions, consistent and biased. From a signal detection perspective, the context provides noise and a signal would be a sound semantically unrelated to the noise, or in the present experiments, a sound which has a cause inconsistent with the meaning of the context. The listener's signal detection task is to identify occasions when the test sound is inconsistent with the context. These would be signal trials. A noise trial would be an occasion when the sound was consistent with the context because in effect, the contextual sounds and the test sound belong to the same auditory scene.
A hit is a correct response on a biased sequence. A false alarm (FA) is a TA response on a consistent sequence, when in fact the test sound was consistent with the context. The percentage of hits would reflect ability to correctly ignore context, while percentage of FAs would reflect incorrect rejection of the contextually correct response. With these definitions for the binary-choice experiments, \( p(H) = \frac{\text{% correct in biased context}}{1} \), and \( p(FA) = 1.00 - \frac{\text{correct performance in consistent context}}{1} \).

In the free classification studies, the definition of a hit is also a correct response on a biased sequence, and \( p(H) = \frac{\text{% correct in biased context}}{1} \). However, a FA could be defined in two ways, and both of these are included in the analysis. First it could be a TA response on a consistent sequence, and thus \( p(FA) = \frac{\text{% TA responses in consistent context}}{1} \). This focuses on the distribution of responses between the two members of the test pair. Alternatively, a FA could be any response other than the correct one on a consistent sequence, which means that \( p(FA) = 1.00 - \frac{\text{correct performance in consistent context}}{1} \).

Using these definitions, non-parametric measures of sensitivity and response bias (\( A' \) and \( B'' \)) were calculated from Hits and FAs. The results of the six experiments are shown in Table 6. Isolated performance is included for reference. The first experiment is referred to as Binary-Choice 1. Recall that in that study, two orders of context were used. Half of the listeners received no-context, isolated-classification trials before contextual classification trials (Order A), and half received the trials in reversed order (Order B). All listeners in the second and third binary-choice experiments received Order A. Note that the hits and false alarms are consistent, and sensitivity and bias similar in the three experiments with Order A. With Order B, hits were lower and false alarms were greater, producing a lower measure of sensitivity. \( B'' \), a measure of response bias is similar in all binary-choice experiments, and the value represents conservative responding in detecting sounds that are inconsistent with the context.

In the free classification studies in which the test sounds were embedded in other sounds and the listener was uninformed about the alternatives, the results differ depending on whether FAs are defined as a TA response or as any alternative to the correct response. When FAs are TA responses as in the binary-choice studies, then sensitivity remains about the same as it is in the binary-choice studies. Thus, in both binary-choice and free classification paradigms, sensitivity in classifying a sound as one of two sounds in an equivalence class remains about the same, even when labeling is introduced. Bias values reflect changes in the
paradigm, and indicate that responding is more conservative in the free classification task, compared to a binary choice task. The bias becomes even more conservative when labels are added to the free classification task. These results confirm the expectation that the free classification studies were expected to enhance the biasing effects of context.

When FAs are defined as any alternative to the correct response in consistent sequences, not just the alternative that was within the test pair, then sensitivity is reduced in the free classification task, especially when labels are added. This is expected because consistent context is not designed to suggest the specific TA response, and equivalence classes were not limited to the two sounds in each pair. As was discovered in Experiment 5, other classifications are possible for the sounds besides the correct response and the TA response.

In general, when the listeners are tested in different paradigms with the same set of sounds, sensitivity for a specific pair of test sounds remains constant, but bias becomes more conservative as the paradigm shifts to produce greater context effects. Overall, these results show that the effects of context are reliable with changes in the experimental paradigm, and are clearly detrimental when the task is to detect a meaningful sound embedded in a context that is biased against the meaning of the sound.
References


Table 1

Correlations between Performance and Sequence Properties in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Order A</th>
<th>Order B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consistent</td>
<td>Biased</td>
</tr>
<tr>
<td># Sounds</td>
<td>-.07</td>
<td>-.43*</td>
</tr>
<tr>
<td>Test sound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>position</td>
<td>-.46*</td>
<td>-.07</td>
</tr>
<tr>
<td># sounds</td>
<td>.29</td>
<td>-.20</td>
</tr>
<tr>
<td>after test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test sound</td>
<td>.18</td>
<td>.32</td>
</tr>
<tr>
<td>duration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05

FIGURE 1. Performance on discriminating pairs of sounds (□), and on recognizing one of the sounds in three contextual conditions: semantically consistent (■), semantically inconsistent (□), and random (■).
Figure 2. Performance on discriminating pairs of sounds (□ □), and recognizing one of the sounds in three contextual conditions: semantically consistent (■), semantically inconsistent (□), and random (□).

Table 2
Correlations between Performance and Sequence Properties in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Consistent</th>
<th>Biased</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sounds</td>
<td>-.25</td>
<td>-.18</td>
<td>.18</td>
</tr>
<tr>
<td>Test sound position</td>
<td>-.41*</td>
<td>-.11</td>
<td>.06</td>
</tr>
<tr>
<td>Number of sounds after test</td>
<td>.15</td>
<td>-.03</td>
<td>.07</td>
</tr>
<tr>
<td>Test sound duration</td>
<td>.23</td>
<td>.12</td>
<td>.31</td>
</tr>
</tbody>
</table>

* $p < .05$
Figure 3. Performance on discriminating pairs of sounds ( ), and recognizing one of the sounds in two contextual conditions with verbal labels: semantically consistent ( ), semantically inconsistent ( ).

Table 3
Correlations between Performance and Sequence Properties in Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Consistent</th>
<th>Biased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sounds</td>
<td>-.09</td>
<td>-.31</td>
</tr>
<tr>
<td>Test sound position</td>
<td>-.64*</td>
<td>-.15</td>
</tr>
<tr>
<td>Number of sounds after test</td>
<td>.39</td>
<td>-.07</td>
</tr>
<tr>
<td>Test sound duration</td>
<td>.33</td>
<td>.30</td>
</tr>
</tbody>
</table>

* p < .05
Figure 4. Percentage of responses listing the correct experimental sound to sounds presented alone (□), and in three contextual conditions: semantically consistent (■), semantically inconsistent (▲), and random ( □ ).

Figure 5. Percentage of responses listing the alternative experimental sound to sounds presented alone (□), and in three contextual conditions: semantically consistent (■), semantically inconsistent (▲), and random (□).
Table 4
Correlations between Performance and Sequence Properties in Experiment 4

<table>
<thead>
<tr>
<th></th>
<th>% Correct</th>
<th></th>
<th>% Test Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consis.</td>
<td>Biased Random</td>
<td>Consis. Biased Random</td>
</tr>
<tr>
<td># Sounds</td>
<td>-.14</td>
<td>-.05</td>
<td>.01</td>
</tr>
<tr>
<td>Test sound</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>position</td>
<td>.04</td>
<td>-.11</td>
<td>-.04</td>
</tr>
<tr>
<td># sounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>after test</td>
<td>-.11</td>
<td>.04</td>
<td>.04</td>
</tr>
</tbody>
</table>

* p < .05
Figure 6. Performance on identifying isolated test sounds with consistent labels (□), and with inconsistent labels (☒), and on identifying test sounds with both labels and contextual sounds that are consistent (□□) and inconsistent(☒☒).

Figure 7. Performance on identifying alternatives to the correct cause of isolated test sounds with consistent labels (□), and with inconsistent labels (☒), and on identifying alternatives to the correct cause of test sounds with both contextual sounds or labels that are consistent (□□) and inconsistent(☒☒). Note the change in scale from Figure 6.
Table 5
Correlations between Performance and Sequence Properties in Experiment 5

<table>
<thead>
<tr>
<th></th>
<th>Correct Identification</th>
<th>Alternate Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Consistent.</td>
<td>Biased</td>
</tr>
<tr>
<td># Sounds</td>
<td>-.12</td>
<td>.00</td>
</tr>
<tr>
<td>Test sound position</td>
<td>.07</td>
<td>.18</td>
</tr>
<tr>
<td># sounds after test</td>
<td>-.12</td>
<td>-.13</td>
</tr>
</tbody>
</table>
Table 6
Detection of Sounds in Biased Context

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Hits</th>
<th>FAs</th>
<th>A'</th>
<th>B'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary-choice 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order A</td>
<td>.80</td>
<td>.14</td>
<td>.90</td>
<td>.14</td>
</tr>
<tr>
<td>Order B</td>
<td>.65</td>
<td>.20</td>
<td>.81</td>
<td>.17</td>
</tr>
<tr>
<td>Binary-choice 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order A</td>
<td>.78</td>
<td>.14</td>
<td>.89</td>
<td>.18</td>
</tr>
<tr>
<td>Binary-choice 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistent label</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order A</td>
<td>.79</td>
<td>.14</td>
<td>.89</td>
<td>.16</td>
</tr>
<tr>
<td>Free classification 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test alternative</td>
<td>.49</td>
<td>.07</td>
<td>.83</td>
<td>.57</td>
</tr>
<tr>
<td>Any alternative</td>
<td>.49</td>
<td>.26</td>
<td>.69</td>
<td>.13</td>
</tr>
<tr>
<td>Free classification 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test alternative</td>
<td>.28</td>
<td>.01</td>
<td>.81</td>
<td>.90</td>
</tr>
<tr>
<td>Any alternative</td>
<td>.28</td>
<td>.31</td>
<td>.46</td>
<td>-.03</td>
</tr>
</tbody>
</table>