Syntactic and semantic factors in the classification of nonspeech transient patterns

JAMES H. HOWARD, JR., and JAMES A. BALLAS
The Catholic University of America, Washington, D.C. 20064

Three experiments were conducted to investigate the role of both syntactic (i.e., temporal structure) and semantic (i.e., knowledge of the source events) factors in a two-alternative (target/ nontarget) categorization task involving patterns of nonspeech acoustic transients. The results demonstrated that both factors can play an important role in the classification of such patterns. Although pattern syntax influenced performance in all three experiments, the effects of syntactic structure were clearest in Experiment 1, in which listeners categorized meaningless tonal patterns. Listeners who categorized a syntactically structured target set performed better than did those with an unstructured set. Experiments 2 and 3 were similar to Experiment 1, but listeners classified patterns of familiar, brief-duration, complex sounds rather than tones. When listeners in Experiment 3 were given explicit descriptive information about the pattern components in their instructions, performance actually improved for interpretable, but not for uninterpretable, patterns. This suggests that syntactic and semantic factors interact in an important way to influence performance. It was argued that many complex nonspeech patterns have both syntactic and semantic structure, which is determined by the sequence of source events that produce them. In classifying such patterns, as in the case of speech, listeners rely on their knowledge of these factors as well as on the perceptual information in the sound itself.

Many current theorists have argued that the recognition of fluent speech involves both top-down, or knowledge-driven, and bottom-up, or data-driven, processes (Cole & Jakimik, 1978, 1980; Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978). In other words, when human listeners perceive speech, they appear to use their general knowledge of linguistic structure (both syntactic and semantic) as well as the specific perceptual information in the signal. In contrast, relatively little research has investigated the role of syntactic and semantic factors in the perception of complex nonspeech patterns. Although less obvious than the case of speech, many of the complex nonspeech sounds that we recognize in everyday life have a specifiable sequential structure (syntax) as well as semantic content. The importance of sound effects in radio drama illustrates this point. One sequence of acoustic transients is heard as someone opening a door to enter a room, whereas another can depict the escape of bank robbers with the police in hot pursuit. Ordered sequences of nonspeech transients of this sort will be referred to as transient patterns. These patterns have temporal structure since their individual components occur in

an order and with durations determined by the source events.[1] For many patterns, an experienced listener can identify the source events when presented with only the sound pattern. The present paper investigates the role of syntactic (i.e., temporal structure) and semantic (i.e., knowledge of the source events) factors in transient pattern recognition.

The most convincing evidence that auditory perception involves both bottom-up and top-down processing is found in the speech perception literature. Logically, the "raw data" or specific sounds in continuous speech cannot be sufficient to account for language understanding since the "raw" input is neither complete nor unambiguous (Cole & Jakimik, 1978). Rather, the listener must rely on his or her knowledge of the syntax and semantics of language and the constraints introduced by those elements that can be interpreted unambiguously. It is not uncommon for us to "hear" missing words or to correct mispronounced words when they occur in fluent speech. For example, Warren (1970) has demonstrated a "phoneme restoration" effect in the perception of spoken text. His listeners consistently reported hearing phonemes that had actually been replaced by a buzz or other nonspeech sound. Warren argued that this reflects the operation of a higher level process that perceptually produces the missing phoneme.

Similarly, Marslen-Wilson and Welsh (1978) investigated the tendency for listeners to correct mispronounced words while shadowing continuous text. They concluded that although speech perception in

0031-5117/80/110431-09$01.15/0

this context is primarily data driven, top-down processes serve to make the system more resistant to input noise and to enhance overall recognition efficiency. In other papers, Cole and Jakimik (1978, 1980) described experiments that investigated a variety of factors in word recognition. They concluded that listeners use a number of knowledge sources in speech perception ranging from fairly specific item-to-item syntactic constraints to global semantic considerations such as the theme or title of a story. In conclusion they argued that "it is not only what we hear that tells us what we know; what we know tells us what we hear" (1978, p. 113).

The evidence from the speech literature that top-down processes play a major role in perception is not particularly surprising. Less obvious, however, is the evidence reported in recent years by Bregman (1978) and his associates, which demonstrates a parallel role of knowledge-driven processes in the perception of relatively simple and semantically impoverished tonal sequences.

Bregman's basic assumption is that listeners are "built to pay attention to acoustic sources, not to acoustic components" (p. 74). At any point in time, the single waveform we hear is likely to represent information combined from several sources, and yet we perceive sounds from each of the separate sources individually (e.g., the dog barking, a Bach cantata, the telephone ringing) rather than as a nonsensical hodgepodge. Bregman has referred to this perceptual phenomenon—the act of sorting our perceived acoustic world into separate sources—as auditory streaming. From a theoretical vantage, he has argued that streaming occurs as the result of a perceptual parsing of the complex acoustic input (Bregman, 1978), and his research has focused on identifying the rules involved in the formation of auditory streams.

To date, auditory streaming research has been restricted largely to relatively simple stimulus contexts. For example, a listener may be asked to judge whether one or two melodic passages are heard when six low-frequency pure tones are played alternately with six high-frequency pure tones. Two separate streams or patterns are readily heard in this context, and the influence of factors such as tonal frequency separation, duration, and alternation rate on stream formation has been investigated. The results of these experiments have been consistent with similar research on simple visual patterns in revealing general heuristics related to the classical Gestalt principles. Such factors as similarity, good continuation, simplicity, common fate, and closure all operate to influence the auditory streams that will be formed in a specific context.

In general, Bregman (1978) has concluded that two kinds of factors are important in auditory streaming. First, there are general—and probably innate—rules or heuristics that can be applied to parse all signals.

Second, there are more specific rules related to a listener's skills, intentions, and knowledge of the stimuli that apply to only selected patterns. By restricting his research to relatively simple acoustic patterns, Bregman has necessarily concentrated on the former of these two kinds of factors. In the present study, we investigate more complex transient patterns and consequently extend Bregman's analysis to heuristics of the second type.

Our approach to this problem involves an extension of Reber's "implicit learning" procedure. In his research, Reber (Reber, 1969, 1976; Reber & Allen, 1978; Reber & Lewis, 1977) examined the role of syntactic structure in the classification of visually presented letter patterns. Subjects classified either grammatical or nongrammatical patterns. The grammatical patterns were generated by a simple finite-state grammar similar to the one shown in the state-transition diagram of Figure 1. An additional letter in the pattern string is produced with every legal state-transition made between the initial and terminal states. For example, the letter pattern "AAACDD" could be produced by the grammar and is therefore grammatical, whereas the pattern "AADDCC" would
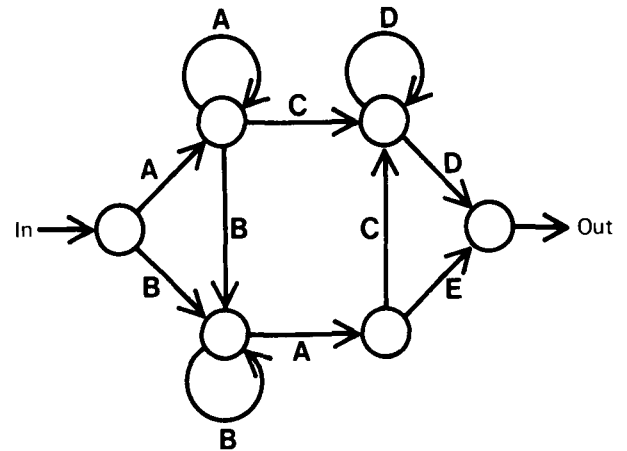
Grammatical	Experimental Stimuli		
	Exp 1	Exp 2	Exp 3
Codes			
A	1157 Hz	Drill	Valve
B	1250 Hz	Clap	Drop
C	1345 Hz	Steam	Steam
D	1442 Hz	Clank	Clang
E	1542 Hz	Wood	Flush

Figure 1. State-transition diagram for the finite-state grammar used to generate structured target patterns in all three experiments.

be ungrammatical since it could not be produced by the grammar.

Reber's extensive research has demonstrated consistently superior classification performance with the grammatical, as opposed to nongrammatical, arbitrarily grouped patterns. This has been shown using a variety of tasks and dependent variables. In the present study, three experiments were conducted to investigate the role of syntactic structure in a two-alternative transient pattern classification task. The simple finite-state grammar of Figure 1 was used to generate syntactically structured patterns for all three experiments. In the first experiment, listeners classified meaningless patterns of brief-duration pure tones. The second experiment was similar to the first, but the pattern components consisted of complex familiar sounds rather than simple tones. Although the individual components were familiar in this experiment, the transient *patterns* were not interpretable. Finally, the role of both syntactic and semantic factors was investigated in the third experiment, in which some listeners classified semantically interpretable patterns of complex sounds.

EXPERIMENT 1

The first experiment was designed to demonstrate that listeners can use syntactic information to facilitate the classification of nonspeech transient patterns. In Experiment 1, listeners were required to classify sequences of brief-duration pure tones as either "target" or "noise" patterns. For the grammatical group, the target patterns were generated by the finite-state grammar of Figure 1, whereas for the other, nongrammatical, group the targets were randomly determined but matched to the grammatical targets in length. By comparing performance with structured (grammatical group) and unstructured (nongrammatical group) target patterns, we can assess the importance of syntactic structure in the classification of simple unfamiliar tonal patterns.

Method

Participants. Ten student volunteers served as listeners in the experiment. Five were assigned to each of the two groups.

Stimuli. Individual transient events consisted of five pure tones selected to be approximately equally spaced in pitch (1,157, 1,250, 1,345, 1,442, and 1,542 Hz). The grammatical patterns were produced by assigning one of the five tones to each of the output letters shown in Figure 1 in corresponding ascending order. Twelve grammatical patterns ranging in length from four to six events (three, four, and five patterns of each length, respectively) were selected to make up the "grammatical target" category. A corresponding "nongrammatical target" set was produced by randomly sampling patterns from the total set of possible patterns with the restriction that sampled targets match those of the grammatical target patterns in length. Similarly, 48 randomly constructed "noise" patterns were selected to be nonoverlapping with the target sets but to match them in length. Within the patterns, each tone was presented for 80 msec at a comfortable listening level (87 dB SPL). Successive tones were separated by 20 msec of silence.

Apparatus. All experimental events were controlled by a general-purpose laboratory computer. The tones were synthesized with the computer using standard digital techniques. They were output on a 12-bit digital-to-analog converter at a sampling rate of 12.5 kHz, low-pass filtered at 5 kHz (Khron-Hite Model 3550), attenuated, and presented binaurally over matched Telephonics TDH-49 headphones with MX-41/AR cushions. Verbal prompts were presented on a video monitor in the testing booth, and listeners indicated their responses by pressing buttons on a solid-state keyboard.

Procedure. Listeners were tested individually in a sound-attenuated booth. The experiment began when the listeners were instructed that they would be hearing patterns made up of several notes played very quickly. They were told that some of the patterns were designated as targets and that their task would be to pick out the targets. Although listeners were told that targets and nontargets could occur equally often, no information was provided regarding the composition of the target set. However, they were told that the pattern categories were determined by the order of components and that loudness and duration were not relevant to the classification. The grammatical and nongrammatical groups received identical instructions.

Each trial began when the word "LISTEN" appeared on the listener's screen. A second prompt—"TARGET (Y OR N)?"—followed the pattern presentation. The listener then responded by pressing "Y" or "N" on the keypad, and visual feedback was provided immediately after the response. After a brief intertrial interval (1.5 sec), the screen was erased and the next trial began. Each listener received 96 trials (4 presentations of each of the 12 targets and 48 presentations of nontargets) in each of 12 blocks. Pattern presentation order was randomized within blocks, and listeners completed four blocks on each of the 3 days. Overall, there were 1,152 trials per individual.

Immediately after the last block, listeners in both groups were told that we had used a set of rules—like the rules of language—to construct the target patterns. We explained that they would be hearing a new set of patterns and that their task would be to classify each pattern as target or nontarget: "Just as you can tell if a sentence is grammatically correct without knowing all the rules for sentences, so should you be able to tell whether any sound is consistent with the rules we used by remembering how the targets sounded." They then completed an additional block of 96 trials, responding as before but without feedback. The target sounds in this test block were selected from the grammatical patterns produced by the grammar in Figure 1 that were not used as targets in the experiment. This test condition was included to determine whether the listeners in the grammatical group could use their syntactic knowledge to classify novel, but grammatical, patterns. Each listener filled out a short, open-ended questionnaire and was interviewed before leaving.

Results and Discussion

The hit (responding "yes" to a target) and false-alarm (responding "yes" to a nontarget) rates were used to compute a response-bias-free (d') index of performance for each individual on each block. These data were then averaged across subjects within each of the two groups to assess group performance. These results are displayed in Figure 2.

Although the data are not strictly monotonic over blocks, it is clear that performance improved with practice for both groups. This effect was revealed statistically by a significant Block effect in a two-way (Group by Block) mixed-design analysis of variance with repeated measures on the Block factor [$F(11,88) = 14.91, p < .001$].

It is also clear from Figure 2 that the grammatical group reached a substantially higher overall perfor-

mance level than did the nongrammatical group. This finding was demonstrated by a significant Block by Group interaction in the analysis of variance [$F(11,88) = 2.23$, $p < .025$]. The main effect of Group was only marginally significant [$F(1,8) = 4.27$, $p < .10$], indicating that the effect of grammaticality developed primarily with practice. Overall, these findings are consistent with Reber's earlier results with letter strings and with our hypothesis that listeners can use syntactic structure to help them classify complex nonspeech patterns.

Of further interest is the performance of listeners in the grammatical group with unfamiliar patterns in the final posttest block. If subjects in the grammatical group had internalized the syntactic or grammatical structure of the target patterns during the experiment, then their performance should have been substantially better than chance on the test block. On the other hand, listeners in the nongrammatical group would have no opportunity to learn about the pattern grammar, and consequently, their performance should be considerably worse than that of the grammatical group. A d' index of performance was computed for each of the 10 listeners on the final test block. These data are presented in Table 1. It is evident from these data that listeners in the grammatical group performed substantially better than the near-chance levels (i.e., d' = .0) of listeners in the nongrammatical group. This difference in group mean performance was found to be statistically reliable [$t(8) = 1.98$, $p < .05$, one-tailed]. This difference occurred despite a relatively large negative d' value observed for Listener 1 in the grammatical group. The large absolute magnitude of this value suggests that this individual was able to distinguish the grammatical and nongrammatical patterns to some extent, but had simply confused the response keys designating the two categories. Since no feedback was provided during the test block, a response reversal would
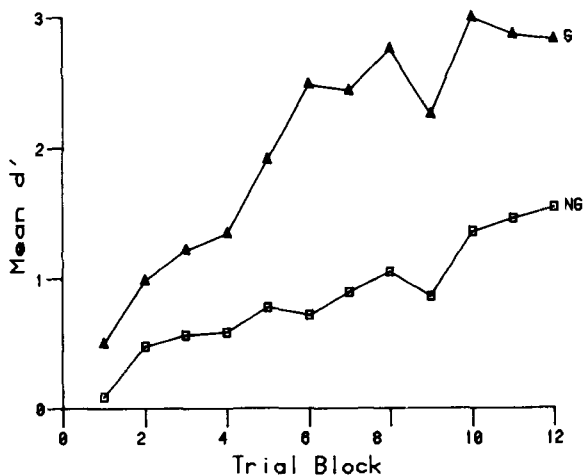
Table 1
Final Test Block Performance (d') for All Listeners in Each of the Three Experiments

| | Individual Listener | | | | | |
Group	1	2	3	4	5	Mean
	Experiment 1					
G	−1.48	1.48	1.84	2.58	3.31	1.55
NG	−.66	−.43	−.23	.10	.56	−.13
	Experiment 2					
G	−.34	.10	.32	1.76	3.79	1.13
NG	−.40	−.12	−.10	.10	.20	−.06
	Experiment 3					
G/S	.95	1.26	1.48	2.35		1.51
G/NS	.86	1.23	1.95	2.55		1.65
NG/S	−1.20	.00	.00	.34		−.21
NG/NS	−.88	−.10	.30	1.12		.11

not be detected easily by the subject. Overall, performance on the final block supports our position that listeners in the grammatical group had actually learned something about the syntactic rules used to generate the target patterns they had classified previously.

EXPERIMENT 2

Experiment 1 demonstrated that listeners can use syntactic pattern structure to their advantage in classifying simple tonal transient patterns. The question arises as to whether a similar result would occur for sound patterns made up of complex, realistic transients. To investigate this, a second experiment was conducted in which listeners classified patterns as either targets or nontargets under conditions similar to those of Experiment 1. One group of listeners had targets produced by the grammar of Figure 1, whereas the other group had a randomly constructed target set. However, unlike the previous experiment, the individual transient events used here were familiar, but unrelated, real-world sounds recorded in the laboratory.

Method

Participants. Ten student volunteers served in the experiment, five in the grammatical group and five in the nongrammatical group. None had served in the previous experiment.

Stimuli. Five individual acoustic transients were selected from a larger set of common "real-world" sounds collected in our laboratory. The larger set was produced by recording a variety of events such as a "clank" (hammer striking a heavy metal object), a "thump" (a hollow, resonant sound from striking a metal drum), and other similar sounds. These samples were then digitized using standard signal processing techniques with a 10-bit analog-to-digital converter at a 12.5-kHz sampling rate.[2] A name for and brief description of each transient are presented in Table 2.

The 12 grammatical targets, 12 nongrammatical targets, and 48 noise patterns were generated as in Experiment 1. Each sound was presented for 82 msec at a comfortable listening level that differed for each sound. Successive transients were separated by 510 msec of silence within a pattern.

Figure 2. Mean performance on classification of brief-duration pure tones for structured (G) and unstructured (NG) patterns.

Table 2
Transient Sounds Used in Experiments 2 and 3

	Experiment 2
Drill	82-msec recording of a high-speed drill being turned on
Clap	82-msec recording of a hand clap
Steam	82-msec recording of white noise bandpass filtered between 4.6 kHz and 5.4 kHz
Clank	82-msec recording of a hammer striking a C-clamp
Wood	82-msec recording of two pieces of wood being struck together
	Experiment 3
Open Valve	320-msec recording of a radiator valve being turned
Water Drop	38-msec recording of a drop of water
Steam	320-msec recording of white noise bandpass filtered between 4.6 kHz and 5.4 kHz
Clang	320-msec recording of a metal object striking a radiator pipe
Water Flush	320-msec recording of water flushing down a drain

Apparatus. The apparatus was the same as in Experiment 1.

Procedure. The procedure was identical to that of Experiment 1.

Results and Discussion

Hit and false-alarm rates were used to compute a response-bias-free (d') index of performance for each individual on each block. These data were then averaged across subjects within the two groups to assess group performance. These data are presented in Figure 3.

Overall, the results are similar to those of Experiment 1. Although it appears that performance for the nongrammatical group may have reached an asymptote by Block 9, performance generally improved with practice for both groups as indicated by a significant Block effect in a two-way (Group by Block) mixed-design analysis of variance [F(11,88) = 6.95, p < .001]. Furthermore, the grammatical group did better than the nongrammatical group on all but the first block. As in Experiment 1, however, the performance difference between the two groups

Figure 3. Mean performance on classification of "real-world" sounds for structured (G) and unstructured (NG) patterns (Experiment 2).

developed with practice since the Group factor interacted reliably with Block [F(11,88) = 2.82, p < .005], but did not produce a significant main effect [F(1,8) = 2.79, p > .10]. On the basis of these findings, we can generalize the conclusion of Experiment 1 to include patterns of familiar transients as well as patterns of simple tones. It appears that listeners are able to use syntactic or grammatical structure to facilitate classification in both cases.

However, it is interesting to note that the overall performance level reached by the grammatical group in the present experiment (mean d' = 1.51) was considerably lower than that reached by the corresponding grammatical group in the earlier, tonal pattern experiment (mean d' = 2.06). Furthermore, this difference cannot be attributed to the different pattern components or different intercomponent intervals used since the mean performance levels observed for the nongrammatical groups were virtually identical across the two experiments (mean d' = .87, Experiment 1; mean d' = .88, Experiment 2). This suggests that structured patterns of unrelated, familiar sounds may be more difficult for listeners to classify than structured patterns of simple tones.[3]

Although the present experiment was designed to investigate the role of syntactic processes in pattern classification, the overall difference observed between Experiments 1 and 2 can be explained most easily be referring to semantic processing. In the present experiment, some listeners recognized the individual transients as familiar everyday sounds. In the words of one participant, "These are the kinds of sounds we walk around all day trying to ignore." Since the sounds were familiar, the listener could not avoid using a parsing strategy that tried to make "sense" out of the patterns. Since the finite-state grammar we used was semantically arbitrary, this could prove to be an impossible task. On the other hand, grammatical listeners in Experiment 1 had little difficulty in using purely syntactic parsing rules since they did not expect the tonal transients to form semantically sensible patterns. This point is drama-
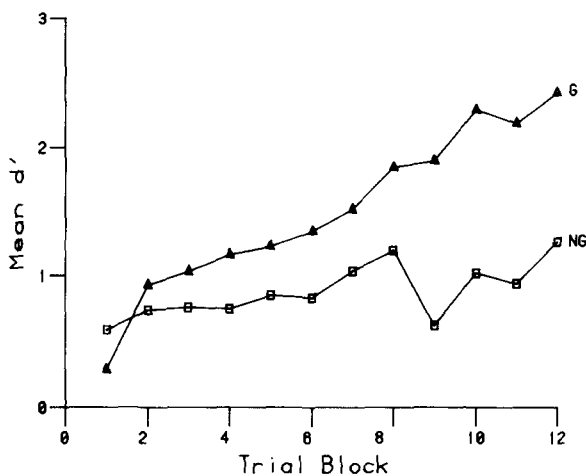
tized by one outspoken listener in Experiment 2 who offered the unsolicited advice that we should have used tones instead of sounds to make the task easier!

Finally, the results obtained for the final test block with novel grammatical patterns revealed generally poor performance for all listeners, with only two individuals in the grammatical group (Listeners 4 and 5) performing better than chance. The performance level observed for each individual is shown in Table 1. Although the group mean performance was somewhat greater for the grammatical group than the nongrammatical group, this difference did not approach significance [$t(8) = .99$]. This finding is consistent with the above discussion in suggesting that the listeners have a great deal of difficulty in abstracting syntactic structure from patterns of familiar sounds that do not relate to any interpretable sequence of source events. The two listeners in the grammatical group who seemed able to do this may have successfully ignored the distracting semantic contents of the individual pattern components.

EXPERIMENT 3

The results of the two experiments reported above suggest that a more elaborate investigation of both syntactic and semantic factors in transient classification would be appropriate. In Experiments 1 and 2, we found that syntactically structured patterns were easier to classify than unstructured patterns. Listeners in the present experiment were also required to classify either grammatical or nongrammatical target patterns, but the procedure was extended so that we could evaluate explicitly the semantic component in auditory classification.

A comparison of our findings from Experiments 1 and 2 suggested that the semantic cues provided by individual transient components may be distracting when the overall pattern does not lend itself to an interpretable semantic analysis. In the present experiment, some listeners were required to classify transient *patterns* that were semantically sensible. To include this condition, it is essential to have patterns that are both syntactically *and* semantically reasonable. In other words, the grammar cannot be arbitrary but must reflect the temporal structure of possible real-world events. Similarly, the selection of individual transient events must be consistent with the grammar.

The simple finite-state grammar of Figure 1 had been developed with these criteria in mind. The grammar can represent possible temporal relations among a series of water- and steam-related events when appropriate complex transients are substituted for the tones and sounds used in the preceding experiments. The five sounds employed in Experiment 3 are described briefly in Table 2. To illustrate how semantically interpretable patterns can be produced,

consider an output string A-A-A-C-D-D from the grammar in Figure 1. This corresponds to a pattern that could represent someone taking three turns to open a valve that releases steam, which, in turn, causes pipes to clang twice. Similar source scenarios can be provided for other grammatical patterns. To evaluate the possible role of semantic information in auditory transient classification, one-half of the listeners in the present study were read a brief paragraph that suggested a schema or theme for the patterns they would hear. The paragraph was suggestive but did not identify any specific patterns explicitly:

> All of the individual sounds relate to water and steam. You will hear such things as drips, water flushing down a drain, a valve being turned on, steam escaping, and radiator pipes clanging.

The remaining subjects received no semantic information about the patterns. The role of gross semantic factors in transient classification can be assessed by comparing performances across the two instructional conditions.

To summarize, four groups were tested in the present experiment. The groups were determined by factorially combining the two syntactic (grammatical and nongrammatical) and two semantic (semantic instructions and no-semantic instructions) variables. The grammatical/semantic group classified structured target patterns and received the semantic information described above, whereas the grammatical/no-semantic group categorized the same structured target patterns without any explicit semantic instructions. Two corresponding nongrammatical target groups were tested (nongrammatical/semantic and nongrammatical/no-semantic). The possible interaction of the syntactic and semantic factors was of particular interest here. Specifically, the findings discussed above suggest that explicit semantic instructions may induce a semantic parsing strategy that would facilitate classification for interpretable patterns (grammatical group) and possibly interfere with performance when the patterns were not interpretable (nongrammatical group).

Method

Participants. Sixteen student volunteers served as subjects in the experiment, four in each group. No listener had served in either of the previous experiments.

Stimuli. The five thematically related transient sounds described in Table 2 were recorded and digitized as in Experiment 2. These were then combined to form sequential transient patterns as in the earlier experiments. Twelve grammatical target patterns were produced using the grammar of Figure 1. The unstructured target and noise patterns were constructed randomly as in the earlier experiments. Within a pattern, each transient was presented for a brief duration (38 msec for the drip and 320 msec for all others) at a comfortable listening level that differed slightly for the various sounds to enhance realism. Successive sounds were separated by 510 msec within the patterns.

Apparatus. The apparatus used was the same as in Experiments 1 and 2.

Procedure. The procedure was identical to that of Experiments 1 and 2. All four groups were tested for 12 blocks on the target patterns with feedback and for one block on the test patterns without feedback.

Results and Discussion

The hit and false-alarm rates were used to compute a response-bias-free index of performance for each individual on each block. These data were then collapsed across individuals within each group to determine group performance levels. These mean data are plotted across blocks for each of the four groups in Figure 4.

A number of findings are apparent on initial inspection of these data. First, overall performance improved with practice as in the earlier experiments. This finding was confirmed statistically by a significant main effect of Block in a three-way (Pattern Structure by Semantic Instructions by Block), mixed-design analysis of variance [$F(11,132) = 25.33$, $p < .001$]. It is clear from Figure 4, however, that large differences exist in the effects of practice across the four groups. In particular, the two grammatical groups showed considerable improvement with practice, whereas the two nongrammatical groups showed relatively little improvement. This was revealed by a statistically significant Pattern Structure by Block interaction [$F(11,132) = 9.96$, $p < .001$] and is consistent with the results of Experiments 1 and 2. No interaction was observed between the Semantic Instruction and Block factors [$F(11,132) < 1.0$].

Second, listeners in the two grammatical groups also performed at a significantly higher level than did listeners in the two nongrammatical groups. This was supported by a significant main effect of Pattern Structure [$F(1,12) = 59.11$, $p < .001$].

Third, there appears to be no overall difference between listeners who received semantic instructions and those who did not. The main effect of Semantic Instruction did not approach statistical significance [$F(1,12) < 1.0$]. Nevertheless, it is obvious that the semantic instructions influenced classification performance. Specifically, semantic instructions appeared to enhance performance only for those listeners who also received syntactically structured target patterns (grammatical/semantic vs. grammatical/no semantic). This result was seen in a statistically reliable Pattern Syntax by Semantic Instruction interaction [$F(1,12) = 4.73$; 4.75 required for $p = .05$]. This effect was examined further in a post hoc analysis. Lindquist's test of critical differences revealed that the grammatical/semantic group performed significantly better than the grammatical/no-semantic group (observed difference of .51; critical difference of .51 for $p = .05$). However, the apparent difference between the nongrammatical/semantic and nongrammatical/no-semantic groups did not reach statistical significance (observed difference of .27; critical difference of .51 for $p = .05$).

The above finding clearly demonstrates the importance of semantic factors in auditory transient classification. In addition, it underscores our earlier conclusion that these effects are not always facilitative. In particular, semantic cues—in our case, the explicit semantic description of the sounds—appear to induce semantic parsing strategies that can enhance performance only for semantically interpretable patterns (grammatical groups). Listeners in the grammatical/semantic group were able to use to their advantage the semantic information we provided. Here, a semantic parsing strategy was appropriate in that it could lead to sensible interpretations for the target patterns. These listeners performed reliably better than their counterparts who had to depend on the syntactic structure of the patterns and the implicit semantic cues in the isolated transients alone.

On the other hand, the semantic instructions did not lead to improved performance when semantically anomalous patterns were used (nongrammatical groups). For these individuals, the specific thematic instructions inappropriately led them to search for sensible interpretations of the patterns when none existed.

Although the explicit thematic instructions were an obvious source of semantic information in the present study, it is also obvious that the familiar sounds themselves provided an additional source of semantic cues. Since these cues were available for listeners in all four groups, it was not possible to assess their effects explicitly in the present study.

Finally, the results of the no-feedback test block were examined to determine whether any listeners were able to generalize their knowledge of the target set to new, grammatical test patterns. The d' performance levels on the final test block are shown for each individual in Table 1. As expected, listeners in the two nongrammatical groups responded at
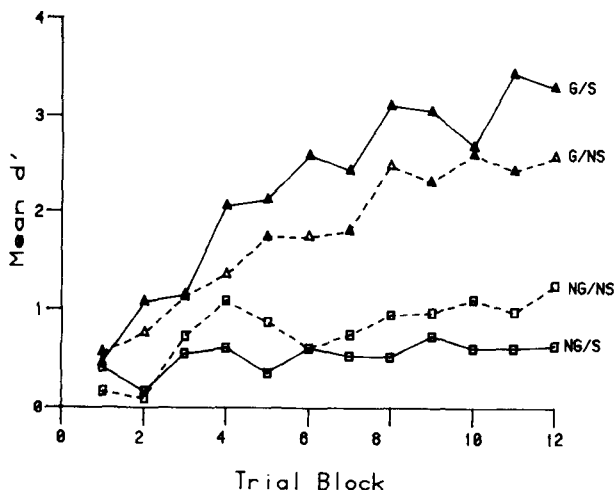

Figure 4. Mean performance on classification of water-related sounds for structured (G) and unstructured (NG) patterns, with (S) and without (NS) thematic information about the sounds.

approximately chance levels, whereas listeners in both grammatical groups responded at above-chance levels. Overall, listeners in the two grammatical groups (mean d' = 1.58) performed reliably better on the test patterns than did listeners in the two non-grammatical groups (mean d' = − .05) [t(14) = 2.40, p < .025, one-tailed]. Those who received semantic instructions performed slightly worse than did those with no explicit semantic instructions, but this difference was not statistically reliable [t(6) = .21]. These findings indicate that listeners in the grammatical groups were able to internalize aspects of the pattern grammar regardless of whether explicit semantic instructions were provided.

GENERAL DISCUSSION

Overall, the results presented above have demonstrated that both syntactic and semantic factors can play an important role in the classification of acoustic transient patterns. Although pattern syntax influenced performance in all three experiments, the effects of syntactic structure were most clearly seen in Experiment 1, in which listeners categorized meaningless tonal patterns. Here, listeners who categorized a grammatically structured target set performed substantially better than those with an unstructured set. Listeners in the former group were also able to generalize their knowledge of the grammar to a novel set of grammatical test patterns. These results are consistent with Reber's earlier findings with visually presented, meaningless letter strings.

Reber has argued that subjects exposed to structured stimuli internalize a "conceptual structure" that represents the underlying grammar or rules, and that this abstraction process occurs implicitly rather than explicitly. In this sense, he has argued that the learning of synthetic grammars of the sort employed in the present study is similar to the acquisition of natural language grammars (Reber & Allen, 1978). Our findings are generally consistent with this interpretation since, in the postexperimental questionnaire and interview, listeners found it impossible to articulate the rules they used to identify the target patterns. Although a few listeners were able to indicate some obvious properties of the grammatical patterns (e.g., the fact that they began with one of two sounds), most specified such vague classification rules as: the targets were "more coherent" or "flowed better" and the nontargets were "unexplainably different" or "not harmonious." Regardless of how the pattern structure is internalized, however, the present results make it clear that syntactic structure does influence the processing of complex transient patterns.

Furthermore, it is obvious from Experiments 2 and 3 that the effects of pattern syntax cannot be considered in isolation. Rather, syntactic and semantic factors interact in an important way to determine categorization performance. For example, in Experi-

ment 2, listeners categorized uninterpretable patterns of familiar sounds. Although clear syntactic effects were observed in this experiment, the effect of pattern structure was considerably smaller than the corresponding effect in Experiment 1. This suggests that the listeners' semantic knowledge (i.e., their familiarity with the pattern components) may have interfered with their ability to abstract the pattern structure.

The importance of semantic factors in auditory classification was even more obvious in the third experiment. When listeners were given explicit descriptive information about the pattern components in their instructions, performance improved for grammatical patterns but did not improve for nongrammatical patterns. Although Cole and Jakimik (1978) have demonstrated that the theme or title of a story influences the linguistic processing of specific words, the strength of the present effect with nonspeech patterns is somewhat surprising.

One explanation of the effect is based on a relatively simple labeling strategy. It is clear that the descriptive instructions we provided equipped the listeners in the semantic groups with a consistent set of labels for the pattern components. In contrast, listeners in the no-semantic groups would have to provide their own labels for the components through absolute recognition. As a result, the semantic listeners could employ more effective encoding and chunking strategies than the no-semantic listeners to facilitate the learning of pattern/category pairs. While it is likely that labeling differences of this sort occurred in Experiment 3, it is apparent that the semantic instruction effect cannot be attributed exclusively to labeling. In particular, this explanation cannot account for the absence of labeling facilitation for the uninterpretable patterns (the nongrammatical/ semantic condition). Since these listeners presumably shared the same labels as listeners in the grammatical/semantic group, they too should have performed at a higher level than the corresponding no-semantic group (the nongrammatical/no-semantic condition).

A more compelling explanation proposes that listeners are influenced by existing semantic structures when perceiving patterns of familiar complex sounds. These existing structures have been referred to as *frames* (Minsky, 1975) or *scripts* (Schank & Abelson, 1977). In Minsky's view, a frame is simply a "data-structure for representing a stereotyped situation" (p. 212). We propose that most individuals have frames for a wide range of possible acoustic transient patterns. Each frame represents the source events for a particular pattern. When an initial sound occurs, the listener refers to the likely source scenarios—the frames—that contain the sound as a beginning component. In other words, the listener constructs hypotheses about what the entire pattern will be, based on partial perceptual information and his or her existing knowledge. As successive transients are heard and interpreted, inappropriate frames can

be eliminated until, ultimately, enough information is accumulated for the pattern to be associated with an appropriate source scenario. In this view, the interpretation of complex transient patterns results from an interplay of bottom-up and top-down processes.

In such a system, explicit semantic instructions would provide the listener not only with a set of component labels, but also with a set of possible frames as well. In Experiment 3, the instructed subjects would attempt to relate the patterns they heard to familiar scenarios involving steam and water flow. These frames or scenarios would be appropriate in the case of interpretable (i.e., grammatical) patterns, but inappropriate for the uninterpretable patterns. In the latter case, the listeners' inability to interpret the patterns using the suggested frames would prove distracting. In other words, these listeners would be unable to make semantic sense out of the patterns despite the fact that the individual pattern components were consistent with the labels we provided. On the other hand, when no semantic information was provided explicitly, the listeners may have simply constructed their own labels and frames for the patterns. Since the listener's spontaneous interpretations of the sound components would be less coherent than those we provided, it would be difficult for them to develop frames that were consistent with all of the target patterns without additional semantic information.

In conclusion, we have argued that many complex sound patterns have both syntactic and semantic structure, which is determined by the sequence of source events that produce them. In interpreting such patterns, human listeners rely on their knowledge of these factors as well as on the perceptual information available in the sound itself. Most theorists agree that this occurs in the processing of linguistic information, and current research is underscoring the importance of syntactic and semantic factors in the perception of complex visual scenes (Biederman, Note 1). Despite this, however, the role of these factors in the classification of nonlinguistic acoustic patterns has not been demonstrated previously. In the present study, we have shown that these factors can play a significant role in even relatively simple classification tasks. Additional work is needed to elaborate their effects and to determine the influence of pattern structure on more traditional psychoacoustic measures such as the listener's ability to resolve individual pattern components (Watson & Kelly, in press).

REFERENCE NOTE

1. Biederman, I. *Human information processing of real-world scenes.* Paper presented at the U.S. Army Research Institute Colloquium on Selected Topics in Behavioral Science Research, Alexandria, Virginia, April 1980.

REFERENCES

BALLAS, J. A., & HOWARD, J. H., JR. Preliminary research on perceiving patterns of underwater acoustic transients. *Proceedings of the Human Factors Society,* 1980, 292-296.

BREGMAN, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII.* Hillsdale, N.J: Erlbaum, 1978.

COLE, R. A., & JAKIMIK, J. Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing.* New York: Academic Press, 1978.

COLE, R. A., & JAKIMIK, J. A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, N.J: Erlbaum, 1980.

MARSLEN-WILSON, W. D., & TYLER, L. K. The temporal structure of spoken language understanding. *Cognition,* 1980, **8**, 1-71.

MARSLEN-WILSON, W. D., & WELSH, A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology,* 1978, **10**, 29-63.

MINSKY, M. A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975.

REBER, A. S. Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology,* 1969, **81**, 115-119.

REBER, A. S. Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, **2**, 88-94.

REBER, A. S., & ALLEN, R. Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition,* 1978, **6**, 189-221.

REBER, A. S., & LEWIS, S. Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition,* 1977, **5**, 331-361.

SCHANK, R. C., & ABELSON, R. P. *Scripts, plans, goals and understanding.* Hillsdale, N.J: Erlbaum, 1977.

WARREN, R. M. Perceptual restoration of missing speech sounds. *Science,* 1970, **167**, 392-393.

WATSON, C. S., & KELLY, W. J. The role of stimulus uncertainty in the discrimination of auditory patterns. In D. J. Getty & J. H. Howard, Jr. (Eds.), *Auditory and visual pattern recognition.* Hillsdale, N.J: Erlbaum, in press.

NOTES

1. It is clear that the syntactic constraints that exist in nonspeech patterns are more flexible and less well-defined than those for speech. Nevertheless, such constraints do exist. In the sounds produced by "an automobile leaving garage" scenario, for instance, one would expect to hear the garage door open before the car pulls out! On the other hand, the engine may be started either before or after the garage door opens.

2. Although the same 12-bit digital-to-analog converter was used to present signals in all three experiments, the sounds used in Experiments 2 and 3 had only 10-bit precision since the analog-to-digital converter used to digitize them had only 10-bit accuracy.

3. Additional support for this interpretation was found in some earlier data collected in our laboratory. Two experiments were conducted that paralleled Experiments 1 and 2 in the present study, but with different listeners and a different finite state grammar. As in the present study, listeners who classified grammatical targets of pure tones (as in Experiment 1) performed at a higher level than those who classified grammatical targets of unrelated sounds (as in Experiment 2) (Ballas & Howard, 1980). This indicates that the between-experiments difference reported here is a reliable finding.