

Feature and conjunction processing in the auditory modality

BENJAMIN J. DYSON and PHILIP T. QUINLAN
University of York, York, England

In five experiments, participants made speeded target/nontarget classification responses to singly presented auditory stimuli. Stimuli were defined via vocal identity and location in Experiments 1 and 2 and frequency and location in the remaining experiments. Performance was examined in two conditions inspired by visual search: In the feature condition, responses were based on the detection of unique stimulus features; in the conjunction condition, unique combinations of features were critical. Experiment 1 showed a conjunction benefit, since classifications were faster in the conjunction condition than in the feature condition. Potential confounds were eliminated in Experiments 2 and 3, which resulted in the observation of conjunction costs. In Experiments 4 and 5, we examined, respectively, whether the cost could be explained in terms of differences in interstimulus similarity and target template complexity across the main conditions. Both accounts were refuted. It seems that when the identification of particular feature combinations is necessary, conjunction processing in audition becomes an effortful process.

Over the past 20 years, Treisman's feature integration theory (hereafter, FIT; Treisman & Gelade, 1980) has played a pivotal role in many investigations of human visual information processing. The empirical assessment of FIT has largely centered around performance in various kinds of visual search tasks, in which participants are presented with a number of spatially distributed stimuli and are asked to indicate whether or not certain predefined targets are present within the display. A basic distinction has been made between cases in which targets are defined relative to a unique value of a stimulus dimension (i.e., feature conditions) and cases in which targets are defined relative to a unique combination of dimensional values (i.e., conjunction conditions). A basic finding is that search for a conjunction target takes longer than search for a feature target.¹ This cost in processing conjunctions has been variously attributed to (1) the need to call attention to the target location in order to accurately conjoin stimulus features (Treisman & Gelade, 1980), (2) increased target/nontarget similarity in the conjunction condition (Duncan & Humphreys, 1989), or (3) increased target feature activation at the nontarget locations in the conjunction condition (Chun & Wolfe, 1996). Much effort is still devoted to

trying to tease these different ideas apart (Duncan & Humphreys, 1992; Treisman, 1991, 1992).

Although the visual search literature is extensive (see Wolfe, 1998, for a review), attempts to generalize the findings to other modalities are relatively sparse. This is not entirely surprising, for whereas the simultaneous presentation of multiple, spatially distributed stimuli is unproblematic in the visual domain, this is not so in audition. The simultaneous presentation of several sounds is likely to produce interference effects, such as peripheral masking at the cochlea (Woods, Alain, & Ogawa, 1998), and this may militate against the development of auditory search paradigms similar in nature to their visual counterparts (although see Hall, Pastore, Acker, & Huang, 2000). Perhaps as a consequence of this, the study of auditory features and conjunctions has traditionally taken place under conditions of sequential presentation. A recent behavioral study of this nature is that reported by Woods et al. (1998, Experiment 2). Here, stimuli were tones of 50-msec duration, each defined relative to one of three values of frequency (250, 612, and 1500 Hz) and one of three values of location (left, binaural, and right headphone presentation). The task was to press a key when a target stimulus was detected within a rapid sequence of tones. Within a block of trials, the target could be defined either with respect to a single frequency or location (i.e., in feature conditions) or with respect to a specific combination of frequency and location (i.e., in conjunction conditions). For example, in a feature block defined by location, participants were required to ignore any stimuli presented binaurally or only in the right headphone and respond to any stimuli presented only in the left headphone, regardless of frequency. Similarly, for a feature block defined by frequency, participants ignored 250- and 612-Hz tones and responded to any 1500-Hz

The research reported here was supported in part by a Ph.D. Committee Award Studentship funded by the Biotechnical and Biological Sciences Research Council, U.K., awarded to the first author. Part of this work formed the basis of an oral presentation at the Annual Meeting of the Cognitive Section of the BPS, Essex, September 6-8, 2000. We thank Douglas Creelman, Tom Sanoki, Charles Spence, David Woods, and a number of anonymous reviewers for helpful comments made regarding earlier versions of this manuscript. Correspondence should be addressed to P. T. Quinlan, Department of Psychology, University of York, Heslington, York YO10 5DD, England (e-mail: ptq1@york.ac.uk).

tone, regardless of location. In the conjunction blocks, participants responded only when the specific combination of frequency and location appropriate to that block (e.g., left, 1500 Hz) was presented, ignoring all other stimuli.

Woods et al. (1998) found that although feature target responses made on the basis of frequency were significantly faster than conjunction target responses, conjunction target responses were significantly faster than feature target responses made on the basis of location. The former finding will be referred to henceforth as a *conjunction cost*, whereas the latter finding will be referred to as a *conjunction benefit*. The observation of a conjunction benefit in audition is striking, since it stands in contrast to a traditional visual FIT framework, in which feature information is available at an early, preattentive stage of analysis and conjunction information is available at a later, capacity-limited stage of analysis (Treisman & Gelade, 1980). In order to account for the conjunction benefit in audition, Woods et al. (1998) put forward a parallel interactive model of stimulus processing, in which the representation of an auditory conjunction develops concurrently with that of individual features.

As a result, substantially different accounts of stimulus processing in the respective auditory and visual modalities suggest themselves, and consequently, the present experiments were carried out to explore the nature of auditory feature integration in more detail. The experiments were

designed to examine choice reaction time (RT) performance in a number of simple speeded classification tasks involving auditory stimuli. In all of the experiments reported, the participants were initially provided with the definitions of various target stimuli. On a given trial, a single stimulus was presented, and the participants were told to decide whether the stimulus was either a target or a non-target as quickly and as accurately as possible. Each stimulus was presented well above threshold, and the participants had an unlimited amount of time to respond. Of critical importance was the fact that in some conditions, target/nontarget (hereafter, T/NT) discrimination was dependent on featural information, whereas in other conditions, T/NT discrimination was dependent on particular conjunctions of features. In order to clarify the structure of the paper, Table 1 gives an overview of the experiments in terms of the nature of the auditory stimuli employed, the aim of the investigation, and the outcome of the investigation defined as either a conjunction benefit or a cost.

EXPERIMENT 1

As a first exploratory step, the stimulus sets in Experiment 1 were based on those used previously in a standard visual search paradigm. Although there is no sense in which the present tasks can be construed as search tasks, because only a single stimulus appeared on each trial, the structure

Table 1
Overview of Experiments 1–5

Experiment 1	
Mode of presentation	headphones
Stimuli	speech sounds
Design	direct mapping of stimulus contingencies from visual search
Conjunction cost or benefit	benefit
Experiment 2	
Mode of presentation	headphones
Stimuli	speech sounds
Design	control of quantity and exposure to target tokens across feature and conjunction conditions
Conjunction cost or benefit	cost
Experiment 3	
Mode of presentation	headphones
Stimuli	tones
Design	replication of Experiment 2 utilizing Woods et al.'s (1998) stimuli
Conjunction cost or benefit	cost
Experiment 4	
Mode of presentation	headphones
Stimuli	tones
Design	control of NT–NT similarity across feature and conjunction conditions
Conjunction cost or benefit	cost
Experiment 5	
Mode of presentation	free-field
Stimuli	tones
Design	test of simple feature-checking account of performance via additional control condition
Conjunction cost or benefit	cost

of the stimulus sets in the feature and conjunction classification conditions in Experiment 1 were identical to those used previously in research on feature binding in vision. Specifically, stimulus sets based on those defined by Quinlan and Humphreys (1987, Experiment 1) were used.² Figure 1 provides a schematic representation of the stimulus sets used in that experiment. Specifically, the feature condition contained six stimuli: four targets and two nontargets, defined relative to the visual dimensions of color and shape. The two nontargets shared no common features between themselves, yet shared one of their features with one of the targets. Therefore, two of the feature targets were discriminable from the nontargets on the basis of a unique color (in this case, blue), and two were discriminable on the basis of a unique shape (in this case, C). In contrast, the conjunction condition contained only three stimuli: one target and two nontargets, again defined by color and shape. The two nontargets were identical to those used in the feature condition, and also like the former case, each nontarget shared one of its features with the target. However, the conjunction target was now simply a recombination of nontarget color and shape (in this case, green H); the target could be discriminated from the nontargets only on the basis of its unique *combination* of features.

In generating the auditory stimuli for the present study, location was substituted for the color dimension used by Quinlan and Humphreys (1987, Experiment 1). Although the location of auditory stimuli can be conveyed in a number of different ways—for example, through interaural intensity differences (Hansen & Hillyard, 1983), interaural time differences (Schröger, 1995), or free-field speaker position (Rhodes, 1987)—in the present experiment, headphone channel presentation was chosen so as to maintain parity with Woods et al.'s (1998) study. Here, the three

location values of left, right, and central were used, with left and right presentation requiring output from the appropriate lateral headphone speaker and central presentation involving simultaneous output from both speakers.

For the second auditory dimension, speech sound was substituted for the shape dimension used by Quinlan and Humphreys (1987, Experiment 1). In previous research, the assignment of values along a so-called linguistic dimension has varied considerably, from the use of whole words (e.g., high and low, up and down; Melara & Marks, 1990a, 1990b) to the use of specific types of consonant–vowel syllables (e.g., /ba/ and /da/; Day, 1972; Springer, 1973). Here, /i/, /di/, and /bi/ were chosen as the three values along which auditory stimuli would be allowed to vary in terms of identity. These stimuli were selected in order to achieve a compromise between the use of two consonants and one vowel in Quinlan and Humphreys's study and the more restrictive use of stop consonant–vowel syllables in Day and in Springer. In summary, in Experiment 1, we examined cases in which, on any given trial, participants listened to a single speech sound delivered over headphones and categorized the stimulus according to a predefined T/NT classification scheme.

Method

Participants

An initial sample of 12 participants was taken from the student population at York University. One participant was removed owing to an average error rate of 20.00%, relative to a corresponding average error rate of 3.51% for the remaining participants. As a result, a 13th individual was run in Experiment 1. Of the final sample, 10 were female, 2 were male, and all were right-handed. The mean age within the sample was 19.58 ($SD = 3.06$). All the participants completed the investigation either for course credit or for the receipt of £2. All reported normal hearing.

A FEATURE CONDITION				
Q&H		Green	Blue	Orange
	Ex. 1	Left	Center	Right
A	E	NT1	T1	
C	D	T2		T3
H	B		T4	NT2

B CONJUNCTION CONDITION				
Q&H		Green	Blue	Orange
	Ex. 1	Left	Center	Right
A	E	NT1		
C	D			
H	B	T1		NT2

Figure 1. Schematic representation of the stimulus sets used in Quinlan and Humphreys (1987, Experiment 1; Q&H in the figure) and the present Experiment 1. For the feature condition, T1, T2, T3, and T4 all refer to the target stimuli, and NT1 and NT2 refer to the nontargets. For the conjunction condition, T1 refers to the single target stimulus, and NT1 and NT2 refer to the nontargets. Each stimulus set was defined on two dimensions, and each dimension comprised three feature values. Columns and rows represent, respectively, the dimensional values of color and letter identity (Q&H) or location and vocal identity (Experiment 1).

Stimuli and Apparatus

The spoken stimuli—/i/ (hereafter, *E*), /di/ (hereafter, *D*) and /bi/ (hereafter, *B*)—were recorded via SoundEdit 16 (MicroMedia). Initially, these sounds were edited so that (1) each lasted 700 msec, (2) the onset of speech occurred after, approximately, 5 msec, and (3) the wave forms between 450 and 700 msec were identical. The sounds were then compressed into a 500-msec time window. During the experiment, these stimuli were presented over headphones (Sennheiser HD414), with the ordering and presentation of stimuli controlled via a PsyScope script (Cohen, MacWhinney, Flatt, & Provost, 1993) running on a Macintosh computer. The speech sounds were presented either monaurally via one headphone speaker or binaurally through both speakers. An attempt was made to equate loudness between binaural and monaural presentation (cf. Woods et al., 1998) by changing the sound parameters in the PsyScope script. This procedure yielded the following decibel values for each stimulus as measured by an artificial ear (Brüel & Kjøer Type 4153) and sound level meter: left *E* = 78 dB, central *E* = 73 dB, left *D* = 76 dB, right *D* = 75 dB, left *B* = 77 dB, central *B* = 72 dB, and right *B* = 75 dB.

The participants were tested in an isolated testing cubicle. Responses were made via a PsyScope button box. Other instructions and visual feedback were presented on a computer monitor in the cubicle. The participants viewed the monitor through an adjustable head- and chinrest placed at a distance of approximately 57 cm from the screen of the monitor.

Design

The trial order within all practice and experimental blocks was randomized, and the completion order of the two conditions (feature and conjunction) was counterbalanced between participants. In addition, handedness was also controlled for, in that one half of the participants were asked to make target-present responses with their preferred hand and target-absent responses with their nonpreferred hand, and the remaining participants were asked to do the reverse. Figure 1 provides a schematic representation of the stimulus contingencies employed in both conditions.

Feature condition. The feature condition consisted of one block of 48 practice trials and one block of 160 experimental trials. For the 48 practice trials, half required target-absent responses, and half required target-present responses. Of the 24 target-absent trials, 12 contained the nontarget left *E*, and 12 contained the nontarget right *B*. For the 24 target-present trials, each of the four targets (central *E*, central *B*, left *D*, and right *D*) were presented six times. For the 160 experimental trials, half required target-absent responses, and half required target-present responses. The 80 target-absent trials comprised 40 left *E* and 40 right *B* nontarget presentations. The 80 target-present trials comprised 20 presentations of each of the four targets. Before the initiation of the feature condition, the participants were informed that the target stimuli were “any central or any *D* stimuli,” thereby requiring them to hold in memory one location value and one identity value throughout the condition.

Conjunction condition. The conjunction condition had the same number of practice and experimental trials and also used the same nontargets as the feature condition. However, the four targets in the feature condition were now replaced by a single conjunction target: left *B*. As a result of this change, the target-absent trials were the same as those in the feature condition, but the 24 target-present trials in the practice block and the 80 target-present trials in the experimental block now contained left *B* presentations only. Before the initiation of the conjunction condition, the participants were informed that the target was “any left *B* stimuli.” Despite the change in target, the participants were still required to hold in memory only one location value and one identity value during the condition.

Procedure

At the beginning of a testing session, the participants were provided with a brief overview of the experiment so that informed con-

sent could be given, following which the sequence of events on a trial was explained. Each trial began with the presentation of a central fixation cross for 500 msec. The offset of the fixation cross was immediately followed by the presentation of an auditory stimulus through the headphones. The participants were asked to respond to the auditory stimulus by pressing one key for target-present responses and a second key for target-absent responses. They were also asked to respond as quickly as possible to the stimuli, but not so quickly as to sacrifice accuracy for speed. At the end of each trial, visual feedback (“Correct” or “Error”) was presented at the center of the screen for 500 msec. The next trial began after an intertrial interval of 60 msec.

At the end of the practice trials, the participants were encouraged to address any problems they had had with the experimenter before continuing with the experimental trials. After all issues had been resolved, the participants initiated the experimental trials by clicking on a mouse button. At the end of the experimental trials, the participants were allowed to take a short break while the experimenter prepared the next condition. After completing both conditions, the participants were thanked for their time and were debriefed as to the nature of the investigation.

Results

In all of the experiments, RTs and errors were analyzed separately using identical analysis of variance (ANOVA) techniques. Only correct responses from experimental blocks were considered in the analysis of the RT data. Error rates were transformed prior to analysis according to the standard arcsin procedure as detailed in Winer (1962, p. 400).³ To aid exposition, the complete results of these statistical analyses are presented in the Appendix. In all cases, statistical significance was assessed at the $\alpha = .05$ level. For Experiment 1, the removal of all incorrect responses from the experimental trials resulted in the elimination of 4.97% of the scores. Outliers were also removed via the procedure described by Jolicoeur (1999),⁴ which resulted in a further 2.33% of the scores being removed. Summary statistics of the data for the conditions of interest are shown in Table 2.

The data were analyzed using two-way repeated measures design ANOVAs. Condition (feature and conjunction) and response (target present and target absent) were fixed factors, and participants was entered as a random factor. The RT data were clear in showing that the participants responded faster to targets than to nontargets and, more important, that they also responded faster in the conjunction condition than in the feature condition. The error rate data showed a similar pattern of significance, al-

Table 2
Summary Statistics for the Four Stimulus Types
in Experiment 1

Stimulus Type	RT	SE	%Error	%Outlier
Feature				
Targets	744	32	6.77	2.61
Nontargets	780	36	3.54	1.72
Conjunction				
Targets	637	33	2.81	2.89
Nontargets	687	30	3.85	2.10

Note—Reaction times (RTs) and standard errors (SEs) are given in milliseconds.

though the condition \times response interaction was also found to be statistically reliable. The interaction was decomposed with Tukey's HSD test. This revealed that error rates for feature targets were significantly larger ($p < .05$) than those for the other three stimulus types. No other comparisons were statistically significant ($p > .05$). The error rate data are clear in showing that the participants were particularly inaccurate on target-present trials in the feature condition. It is important to note that this relatively large miss rate does not reveal any form of speed/accuracy tradeoff that might compromise the critical main effect of condition found in the RT analysis.

Additional Analyses

In a manner similar to Woods et al. (1998), a consideration of the RTs for individual stimuli allowed for the more detailed comparison of performance with feature targets defined relative to a distinctive location (i.e., central B and central E; henceforth, distinctive location targets), feature targets defined by a distinctive letter identity (i.e., left D and right D; henceforth, distinctive identity targets), and the conjunction target (i.e., left B). Mean RTs for these targets were 749, 738, and 637 msec, respectively. The corresponding RTs were entered into a one-way repeated measures ANOVA. The ANOVA revealed a statistically significant main effect of target type [$F(2,22) = 11.99, p < .001$]. When analyzed further using Tukey's HSD test, the main effect revealed that RTs to the conjunction target were faster than RTs to either of the feature targets ($p < .05$). In addition, there was no difference in RTs between the two types of feature targets ($p > .05$).

Similar analyses were conducted on the corresponding arcsin transformed error rates. The respective mean error rates were 7.50%, 6.04%, and 2.81%. The ANOVA revealed a statistically significant main effect of target type [$F(2,22) = 8.92, p = .001$]. Additional exploration using Tukey's HSD test indicated that error rates for the distinctive location targets and distinctive identity targets were equivalent ($p > .05$), yet significantly more errors were made in responding to both of these feature targets than in responding to conjunction targets ($p < .05$). The error rates are consistent with the RT data and clearly demonstrate that the participants were slower and less accurate in responding to targets defined by a single feature than to targets defined by a conjunction of features.

Overall, these analyses have shown that there was no difference in performance with the distinctive location and distinctive identity targets. However, since different targets were used in the feature and the conjunction conditions, it is possible that the present conjunction benefit may simply have arisen because there was some quality inherent within a left B that facilitated responding relative to any central or any D stimulus. However, if there was some benefit in a left B presentation, one would anticipate that responses made to feature targets sharing one of these values (i.e., left D and central B) would be quicker than responses made to feature targets that did not share any of these features (i.e., right D and central E). A related t test

was carried out comparing RTs between these two kinds of feature targets (763 and 725 msec, respectively). Although this test revealed a statistically significant, one-tailed difference between the two sets of stimuli [$t(11) = 1.95, p = .039$], this difference was in the opposite direction to that predicted by the notion that the conjunction benefit was due to the specific combination of salient features, left and B. Nevertheless, comparing different targets across different conditions remains problematic and will be considered in more detail later.

Discussion

The central result in Experiment 1 is that participants were overall faster in responding in the conjunction condition than in the feature condition. The basic effect is of a conjunction benefit and thus may be compared with the similar effect reported by Woods et al. (1998). Whereas the conjunction benefit discussed by Woods et al. (1998) was revealed as a benefit on location/frequency conjunction trials relative to location feature trials only, the present benefit for responses to location/identity conjunction trials was robust with respect to responses on both location and identity feature trials. Woods et al. (1998) argued that their auditory conjunction benefit fitted comfortably with claims about the parallel and interactive processing of individual features. By this account, auditory features and conjunctions can be processed simultaneously so that their operations are mutually dependent. This is an entirely plausible explanation of their data and the present conjunction benefit in Experiment 1. However, before such a functional account of processing can be accepted, it is important to consider certain methodological considerations that may stand in the way of such an interpretation.

One notable aspect of Experiment 1 was the different number of targets used in the feature and the conjunction conditions (i.e., four and one, respectively). This difference was justified on the grounds that the experiment simply replicated the T/NT definitions used extensively in the visual search literature. These definitions hinge on a type/token distinction in which the number of target *types* are equated across the feature and the conjunction conditions but the number of target *tokens* differ. In both Quinlan and Humphreys (1987, Experiment 1) and the present Experiment 1, the participants were required to search for two target types defined relative to four target tokens, in the feature condition, and two target types defined relative to one target token, in the conjunction condition. Although this confound works against the standard experimental hypothesis of improved performance in the feature condition, relative to the conjunction condition, it also provides a potential explanation of the data when the opposite result obtains. In this respect, the interpretation of the conjunction benefit in Experiment 1 is not straightforward, and consequently, several possibilities now present themselves.

If the number of target tokens influences overall condition performance, the control of target token differences between the feature and the conjunction conditions clearly becomes an issue.⁵ A second, concomitant design concern

relates to the amount of exposure that the participants received per target token. By adopting the stimulus contingencies in standard visual search, the single conjunction target token in Experiment 1 received much more exposure than any one of the four feature target tokens (i.e., 80 vs. 20 trials, respectively).

In order to address these methodological issues, a follow-on experiment to Experiment 1 was carried out. However, since the details of this study are tangential to the main focus of the experimental series, only a brief summary will be included here. Again, the idea was to compare performance between a feature condition in which a target response could be based on a simple *or* rule (i.e., any central *or* any D stimuli) and a conjunction condition in which target responses could be based on an *and* rule (i.e., any left *and* B stimuli). Here, the conjunction condition in Experiment 1 was reused, but for the feature classification task only two targets and two nontargets were presented in a block of trials, instead of the previous four targets and two nontargets. The number of target tokens was again different across the feature and conjunction conditions (i.e., two vs. one, respectively), but this difference was simply unavoidable if the intention was to compare a feature *or* decision with a conjunction *and* decision, since both decisions must be defined relative to two target features. The feature conditions were also designed so that the participants had equal amounts of exposure to each feature target token relative to the single conjunction target token.

Average RTs in the feature and conjunction conditions were now 671 and 607 msec, respectively, and this conjunction benefit was again found to be statistically significant ($p < .05$). In comparison with Experiment 1, these data revealed a diminution in the size of the conjunction benefit, in that the effect was now 64 msec, relative to Experiment 1, in which it was 100 msec. Nevertheless, in controlling the amount of exposure to targets tokens across the feature and the conjunction conditions and reducing the number of target stimuli defined for any given block of trials in the feature condition, the previous conjunction benefit still remained. This experiment demonstrates the importance of these confounding factors, and as a result, both the number of target tokens and their relative amount of exposure were explicitly controlled in the remaining experiments. However, it appears that neither of these factors critically determined the presence of the conjunction benefit.

EXPERIMENT 2

Perhaps the critical issue is whether performance in the conjunction conditions of Experiment 1 and Woods et al. (1998) depends on the successful conjoining of features at all. For the conjunction condition in Experiment 1 and for every conjunction block within the Woods et al. (1998) study, there was only one predefined conjunction target. In Experiment 1, the conjunction target was presented among nontargets that always shared one feature with the target. In Woods et al.'s (1998) study, conjunction nontar-

gets possessed either one target feature or no target features.⁶ The central point is that although feature binding is required for the correct identification of a conjunction target in visual search, alternative mechanisms may explain the identification of a single conjunction target in the auditory sequential presentation tasks used here and by Woods et al. (1998). One possible explanation is that performance in both conjunction tasks could simply have been based on the operation of a mechanism that was sensitive to the number of target features present on a given trial, rather than on feature binding per se. Since it was always the case that the conjunction target possessed two critical features, whereas nontargets possessed either one or no critical feature, the detection of a "conjunction" target in these studies may simply rest on the ability to detect two target features, and a nontarget response would be based on the detection of fewer than two target features.

Models such as these have been discussed in the visual information processing literature under the description of coactivation processes (Miller, 1982; Mordkoff & Yantis, 1991; Mordkoff, Yantis, & Egeth, 1990) and are based on the assumption that there are independent feature analyzers for each of the constituent features of an input stimulus. In contrast to independent-race models, coactivation models argue that a decision about the presence of a stimulus is made on the basis of a single mechanism that pools activation from the individual feature analyzers. Such models account for a conjunction benefit on the grounds that total activation builds up faster when there are two target features present than when there is just one (see Mordkoff et al., 1990). Such a feature activation account is different from a corresponding feature integration account in which the act of conjoining basic features is not simply a matter of summing target feature activation but, rather, is based on the active process of binding particular features together.

Given such concerns, the aims of Experiment 2 were threefold. First, an attempt was made to generate an auditory conjunction task in which T/NT discriminations could be achieved only by some form of feature binding. To satisfy this primary goal, a condensation task (cf. Fitts & Biederman, 1965; Smith, 1968) was adopted so as to ensure that the participants were basing their responses on specific feature combinations. In a standard condensation task, two stimuli (x_1y_1 and x_2y_2) are assigned to one response and two stimuli (x_1y_2 and x_2y_1) are assigned to another. Here, let x and y denote different stimulus dimensions, 1 and 2 denote different values on these dimensions, and the two responses represent target present and target absent, respectively. Under such conditions, T/NT discriminations can now no longer be made at the level of target feature activation, since both targets and nontargets possess two target features. Rather, the participant must determine the separate values of the stimulus and respond on the basis of the particular combination of these two values (see Lockhead, 1972, p. 416).

Second, the same two targets were employed in all the conditions in Experiment 2. Although there was no evi-

dence to suggest that the different target stimuli used in the feature and conjunction conditions produced the pattern of results observed in Experiment 1, use of the same targets throughout prevented being attributed any contingent performance differences across conditions to stimulus differences. Third, the same number of stimuli were presented for the same number of trials in every condition in Experiment 2. This also ensured that performance differences observed between the feature and the conjunction classification tasks could not be attributable to relative differences in stimulus uncertainty or stimulus exposure. Figure 2 provides a schematic representation of the stimulus sets used in Experiments 2 and 3.

Method

Participants

A sample of 16 participants was taken from the student population at York University. Of the 16, 14 were female and 2 were male, 13 were right-handed and 3 were left-handed. The mean age within the sample was 19.19 ($SD = 0.98$). All the individuals completed the

investigation for course credit or for the receipt of £2. All reported normal hearing.

Stimuli and Apparatus

The stimuli and apparatus were the same as before, apart from the addition of one new stimuli: right E, measured as 75 dB by an artificial ear and sound level meter.

Design

Experiment 2 consisted of four conditions. Specifically, two different versions of the feature condition and two identical versions of the conjunction condition were completed by participants. The same two targets were used in all four conditions (left E and right B), and consequently, the only variation across the four conditions was the presentation of different nontargets. T/NT discrimination in the feature conditions could be completed on the basis of unique featural properties of each stimulus. In contrast, T/NT discrimination in the conjunction conditions could be achieved only by the consideration of specific combinations of features. Each condition consisted of 24 practice trials and 80 experimental trials. In all cases, 12 of the practice trials and 40 of the experimental trials required target-present responses, and the other half of the practice and experimental trials re-

A FIRST FEATURE CONDITION				
Ex. 2		Left	Center	Right
	Ex. 3	Left	Center	Right
E	High	T1	NT1	
D	Mid			NT2
B	Low			T2

B SECOND FEATURE CONDITION				
Ex. 2		Left	Center	Right
	Ex. 3	Left	Center	Right
E	High	T1		
D	Mid	NT1		
B	Low		NT2	T2

C FIRST CONJUNCTION CONDITION				
Ex. 2		Left	Center	Right
	Ex. 3	Left	Center	Right
E	High	T1		NT1
D	Mid			
B	Low	NT2		T2

D SECOND CONJUNCTION CONDITION				
Ex. 2		Left	Center	Right
	Ex. 3	Left	Center	Right
E	High	T1		NT1
D	Mid			
B	Low	NT2		T2

Figure 2. Schematic representation of the stimulus sets used in Experiments 2 and 3. T1 and T2 refer to the target stimuli, and NT1 and NT2 refer to the nontargets. Each stimulus set was defined on two dimensions, and each dimension comprised three feature values. The columns represent the values of the location dimension, and the rows represent the values of the vocal identity (Experiment 2) or the frequency (Experiment 3) dimensions.

Table 3
Summary Statistics for the Four Stimulus Types
in Experiment 2

Stimulus Type	RT	SE	%Error	%Outlier
Feature				
Targets	735	33	3.44	1.97
Nontargets	766	31	2.42	2.46
Conjunction				
Targets	853	42	5.86	1.62
Nontargets	948	57	6.88	2.07

Note—Reaction times (RTs) and standard errors (SEs) are given in milliseconds.

quired target-absent responses. Since four stimuli (two targets and two nontargets) appeared in all the conditions, each stimulus was assigned to 6 presentations in the practice trials and 20 presentations in the experimental trials. Trial order within each of the blocks was randomized, and condition order and handedness were counterbalanced across participants.

Procedure

The procedure was essentially the same as before: The participants made speeded classification responses to spoken letter names presented over different headphone channels. For each condition, the participants were informed that target stimuli were “any left E or any right B stimuli.”

Results

Real errors accounted for 4.65% of the scores, and a subsequent 2.03% of the scores were classified as outliers. Only correct responses from experimental blocks were considered in subsequent RT analyses. Summary statistics of the data for the conditions of interest are shown in Table 3.

The RT data were analyzed in a manner identical to that in Experiment 1, but now the critical finding was of a conjunction cost, in that the participants were generally slower to respond in the conjunction condition than in the feature condition. In addition, the participants were slower in responding to nontargets than to targets, although this pattern reached statistical significance only in the conjunction condition (Tukey's HSD test, $p < .05$). The error data show consistency with the RT data, in that the participants were also less accurate in the conjunction condition than in the feature condition. Consequently, there was no evidence of a speed/error tradeoff within the data.

Additional Analyses

Although the same two targets were used in the feature and the conjunction conditions, it was still possible to examine the data by classifying the trials according to distinctive location, distinctive identity, and conjunction targets. The respective mean RTs were 727, 744, and 853 msec. Analysis of the corresponding RT data showed a statistically significant main effect of target type [$F(2,30) = 23.42$, $p < .001$]. Further analysis using a Tukey's HSD test revealed that responses to distinctive location and distinctive identity targets did not differ ($p > .05$) but that both were significantly faster than responses to the conjunction target ($p < .05$).

Mean error rates for the distinctive location, distinctive identity, and conjunction targets were 3.91%, 2.97%, and 5.86%, respectively. Analysis of the corresponding transformed scores revealed that the main effect of target type failed to reach statistical significance [$F(2,30) = 2.70$, $p = .083$], thereby indicating the absence of a speed/error tradeoff.

Discussion

The results of Experiment 2 are quite clear in showing that the participants were slower and less accurate in categorizing conjunction stimuli than they were in categorizing feature stimuli. Importantly, these effects arose when other potential confounds across the feature and the conjunction conditions had been removed. Now, the number of different target stimuli for each task was equated, as was the number of presentations for any particular target stimulus. Under these circumstances, both the RT and the error rate data revealed a conjunction cost, as opposed to a conjunction benefit (cf. Experiment 1 and Woods et al., 1998). Moreover, additional analyses revealed that the cost did not occur because performance with just one of the feature targets was particularly efficient relative to the conjunction targets.

Before attempting to examine the underlying reasons for the present conjunction cost in more theoretical detail, it seemed appropriate to generalize the finding to a different set of stimuli. Therefore, to forestall any concerns that the conjunction cost was limited to the particular stimuli used in Experiment 2, a third experiment was carried out. The same sorts of speeded classification tasks were used, but now the stimuli were tones presented over headphones. The specific values of location and frequency selected for Experiment 3 were identical to those used by Woods et al. (1998).

EXPERIMENT 3

Method

Experiment 3 was modeled on Experiment 2 in all respects, apart from the fact that tones were substituted for the previously used speech sounds. All other differences will now be detailed.

Participants

An initial sample of 16 participants was taken from the student population at York University. One individual had to be removed from the study owing to a large error rate (24.25%), as compared with the error rate of the remaining participants, which averaged 6.58%. As a result, a 17th participant was substituted. Of the final sample, 9 were female and 7 were male, 15 were right-handed and 1 was left-handed. The mean age within the sample was 23.75 ($SD = 7.00$). All individuals completed Experiment 3 for course credit or for the receipt of £2, and all reported normal hearing.

Stimuli and Apparatus

Three separate tones equally spaced on a log scale—that is, 250 Hz (or *low*), 612 Hz (or *middle*), and 1500 Hz (or *high*)—were created using SoundEdit 16. Each tone was 250 msec in length with linear 5-msec rise/fall times. Stimuli were equated for subjective loudness between tone frequencies and between monaural and binaural pre-

Table 4
Summary Statistics for the Four Stimulus Types
in Experiment 3

Stimulus Type	RT	SE	%Error	%Outlier
Feature				
Targets	652	44	6.31	1.40
Nontargets	700	45	5.88	1.74
Conjunction				
Targets	720	44	7.75	2.30
Nontargets	769	54	6.44	2.28

Note—Reaction times (RTs) and standard errors (SEs) are given in milliseconds.

sentation, according to the guidelines set out by Woods et al. (1998). The 1500-Hz tones were attenuated by 2 dB relative to the 250- and the 612-Hz tones, and binaural (i.e., *central*) presentation was attenuated by 6 dB relative to monaural presentation. This yielded the following decibel levels for the stimuli, as measured by an artificial ear and sound level meter: left, 1500 Hz = 85 dB; central, 1500 Hz = 79 dB; right, 1500 Hz = 85 dB; left, 612 Hz = 87 dB; right, 612 Hz = 87 dB; left, 250 Hz = 87 dB; central, 250 Hz = 81 dB; and right, 250 Hz = 87 dB. All of the apparatus used was the same as in Experiment 1.

Design and Procedure

The design and procedure of Experiment 3 replicated those of Experiment 2, except that there were now 32 trials in each practice block (8 presentations of the four stimuli within each condition) and 100 trials in each experimental block (25 presentations of the four stimuli within each condition). Although the values of location were consistent with Experiments 1 and 2, the values of identity were changed so that each vocal stimulus was substituted for a tonal stimulus (E = 1500 Hz, D = 612 Hz, and B = 250 Hz). Target stimuli in all conditions were defined as “any left high or any right low stimuli.” Figure 2 provides a detailed schematic representation of the stimulus set used in Experiment 3.

Results

Real errors accounted for 6.59% of the scores, and an additional 1.93% of the scores were identified as outliers. Only correct responses from experimental blocks were considered in subsequent RT analyses. Table 4 provides summary statistics of the data for the conditions of interest.

Analysis of the RT data again revealed the presence of a statistically significant conjunction cost, in that the participants were overall slower to respond in the conjunction condition than in the feature condition. In addition, the participants were slower in making target-absent responses than target-present responses, an effect that was present in both the feature and the conjunction trial data. Corresponding analysis of the error data failed to produce any statistically significant results. Error rates were generally low, all being less than 7%, and consequently there was no evidence of a speed/error tradeoff in the data.

Additional Analyses

Mean RTs for distinctive location, distinctive frequency, and conjunction targets were 642, 662, and 720 msec, respectively. Analysis of the corresponding RT data revealed a statistically significant main effect of target type [$F(2,30) = 3.95, p = .030$]. A Tukey's HSD test showed that the only pairwise comparison to reach statistical significance was between the data for the distinctive location and

the conjunction targets ($p < .05$), with responses being quicker for the former type of target than for the latter.

Identical analyses were conducted for the transformed error rates for the three target types (mean error rates were 6.88%, 5.75%, and 7.75%, respectively). The main effect of target type failed to reach statistical significance [$F(2,30) = 1.29, p = .290$], indicating the absence of any speed/error tradeoff.

Discussion

Overall, the data in Experiment 3 are clear in showing a conjunction cost similar to that reported in Experiment 2. The participants were slower in making classifications in the conjunction condition than they were in the feature condition, and this cost was shown when the stimuli were defined relative to location and vocal identity, and to location and frequency. Consequently, the conjunction cost cannot be attributed to some special strategy associated with a particular kind of auditory stimulus.

Although it can be argued that the conjunction costs found in Experiments 2 and 3 reflect the operation of processes that take place after the detection of simple features—in line with a basic tenet of FIT—it is important to consider whether the present data necessitate such a view. In this regard, an alternative account of the data derived from the attentional engagement theory (AET) of Duncan and Humphreys (1989) was explored.

Duncan and Humphreys (1989) questioned several basic claims of FIT and, in particular, argued that the differences between feature and conjunction search performance could be accounted for in terms of the similarity relations that existed between the respective displays. By their account, two factors interact to make an inefficient search: (1) increased target–nontarget (T–NT) similarity and (2) decreased nontarget–nontarget (NT–NT) similarity. T–NT similarity provides an index of the degree to which the target groups with the nontargets, whereas NT–NT similarity provides an index of how well the nontargets group together. Target detection will be efficient when the target fails to group strongly with the nontargets in the display (as in feature search), and target detection will be relatively inefficient when the target and the nontargets group together (as in conjunction search).

Although there is no sense in which similar processes of grouping can explain performance in the present sequential presentation tasks, general notions of interstimulus similarity can be used to generate an account that does not invoke any appeal to feature binding. Indeed, such an exercise seems reasonable given the recent findings reported by Mondor, Zatorre, and Terrio (1998, Experiment 3). Here, participants were required to detect a target tone defined by a combination of location and frequency within streams of distractor tones. These distractor tones were either similar to or different from the target with respect to location and/or frequency, thereby yielding four factorial T–NT combinations. In brief, the results showed that target detection was quick when T–NT similarity was low, and slow when T–NT similarity was high. Hence, Mondor et al.'s study provides evidence to suggest that T–NT sim-

ilarity can affect auditory classification performance in a way that is consistent with AET.

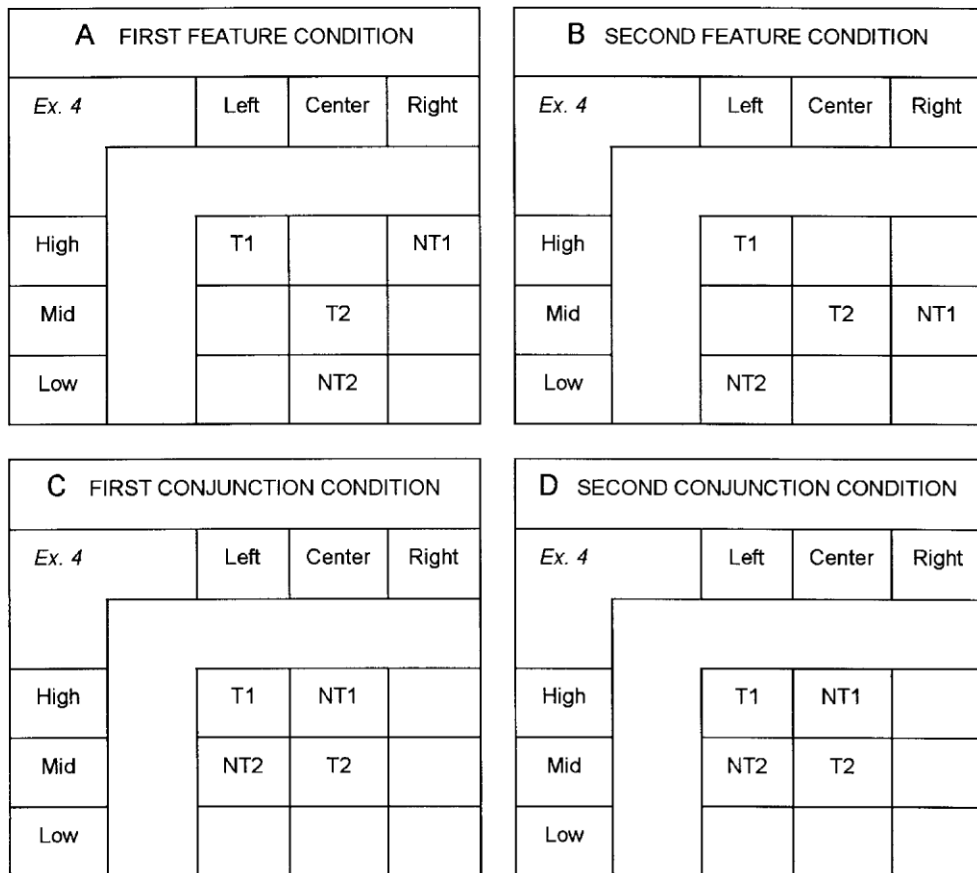
The issue now is whether notions of interstimulus similarity can be used to explain performance in the present tasks. One obvious way to operationalize interstimulus similarity is with reference to the number of common features shared between stimuli. However, with reference to both the feature and the conjunction conditions in Experiments 2 and 3, T–NT similarity was the same, since any given target stimulus shared one common feature with any given nontarget stimulus. Likewise, NT–NT similarity between feature and conjunction tasks was identical, since under all conditions, nontargets failed to share any common features. Since these measures of interstimulus similarity were equivalent in both types of conditions, the observed difference in performance cannot be due to differences in T–NT or NT–NT similarity as defined by shared features.

A different way to conceptualize T–NT and NT–NT similarity is with respect to dimensional proximity (as discussed by Treisman, 1991). Here, notions of interstimulus similarity are captured by the distance between points in

a multidimensional space. In this regard, it is possible to take Figure 2 as a starting point and simply assume that the underlying space can be conceptualized as the particular two-dimensional plane shown.⁷ In reviewing an earlier version of the paper, Tom Sanocki pointed out that in Figure 2, NT–NT similarity is higher in the feature condition than in the conjunction condition, because the nontargets are closer together in the former case relative to the latter case. It follows, therefore, that performance in the feature condition may have benefited from increased NT–NT grouping that was not present in the conjunction condition. This account is clearly independent of any differences between the feature and the conjunction conditions in terms of the presence or absence of additional feature-binding mechanisms.

EXPERIMENT 4

Experiment 4 retained the design and procedure of Experiment 3, but now an attempt was made to decrease NT–NT similarity in the feature condition and to increase NT–NT similarity in the conjunction condition. The hypoth-



esis under investigation was that the critical factor in facilitating responding was the ease with which nontargets *grouped together*. Figure 3 provides a schematic representation of the stimulus sets used in Experiment 4. As can be seen in this figure, nontargets are now more distant in the feature case than in the conjunction case. If NT–NT similarity is indeed a critical factor, the previous conjunction cost observed in Experiments 2 and 3 should be either eliminated or reversed, since the nontargets are now more proximal in the conjunction condition than in the feature condition.

Method

Participants

Sixteen participants (10 female and 6 male) were taken from the student population at York University. Fifteen were right-handed, and 1 was left-handed. The mean age within the sample was 23.75 ($SD = 4.46$). All the individuals completed the investigation for course credit or for the receipt of £2, and all reported normal hearing.

Stimuli and Apparatus

The stimuli and apparatus used in Experiment 4 were identical to those used in Experiment 3, except for the addition of a central 612-Hz stimulus, calibrated at 81 dB by an artificial ear (Brüel & Kjøer Type 4153) and sound level meter.

Design and Procedure

The participants completed four conditions as before (two different-feature conditions and two identical-conjunction conditions), with four stimuli (two targets and two nontargets) present within each condition. Each condition involved one practice block of 32 trials (four stimuli presented for 8 trials each), followed by one experimental block of 100 trials (four stimuli presented for 25 trials each). Half the trials within each block were allocated to target-present responses, and the other half were allocated to target-absent responses. Target stimuli were consistently defined as left high and central middle throughout all four conditions.

Results

Real errors accounted for 7.28% of the scores, and an additional 2.02% of the scores were identified as outliers. Only correct responses from experimental blocks were considered in subsequent RT analyses. Table 5 provides summary statistics of the data for the conditions of interest.

The central result from the analysis of the RT data was again of a conjunction cost, in that the participants were slower in making conjunction than feature classifications. The participants were also generally quicker in making target than nontarget responses, but this effect was statistically significant only in the conjunction condition (Tukey's HSD test, $p < .05$). The error data also indicated the presence of a conjunction cost, in that significantly more errors were made in the conjunction condition than in the feature condition. This is consistent with the RT data and, therefore, provides no evidence of a speed/error tradeoff in Experiment 4.

Additional Analyses

Mean RTs for the distinctive location, the distinctive frequency, and the conjunction targets were 558, 570, and 639 msec, respectively. The corresponding analysis of the RT data revealed a statistically significant main effect of target type [$F(2,30) = 9.94$, $p < .001$]. A Tukey's HSD

test, in turn, revealed that RTs for both types of feature target were equivalent ($p > .05$) and that both were significantly faster than RTs to conjunction targets ($p < .05$). Mean error rates for the distinctive location, distinctive frequency, and conjunction targets were 6.13%, 5.62%, and 8.44%, respectively. Analysis of the corresponding transformed scores showed that the main effect of target type was significant [$F(2,30) = 4.00$, $p = .029$]. Further analysis using a Tukey's HSD test revealed that error rates for conjunction targets were significantly greater than error rates for distinctive frequency targets ($p < .05$). All other pairwise comparisons failed to reach significance ($p > .05$). Therefore, the error rate data show broad consistency with the RT data and do not suggest the presence of a speed/error tradeoff.

Discussion

In Experiment 4, the hypothesis that differences in NT–NT similarity across the feature and the conjunction conditions in Experiments 2 and 3 gave rise to the conjunction cost was examined. In this experiment, NT–NT similarity was decreased in the feature condition and increased in the conjunction condition, and it was predicted that this manipulation ought to have eliminated or even reversed the conjunction cost. In contrast to both of these predicted outcomes, the data in Experiment 4 were clear in showing that the participants were slower and less accurate in making classifications in the conjunction condition than in the feature condition. Initially, it was predicted that the reason for the better performance in the feature condition was due to the fact that nontargets were easier to *group together* than they were in the conjunction condition. In contrast, it actually appears as though increasing the distance between the nontargets in Experiment 4 facilitated overall responding in the feature condition, relative to performance in Experiment 3. Further discussion of stimulus similarity is perhaps unwarranted at this point. Rather, it is sufficient to state that no simple account of T–NT or NT–NT similarity can explain the differences observed between feature and conjunction conditions.

EXPERIMENT 5

Taken collectively, the data from Experiments 2–4 accord well with predictions made on the basis of taking a FIT framework and applying it to auditory processing.

Table 5
Summary Statistics for the Four Stimulus Types
in Experiment 4

Stimulus Type	RT	SE	%Error	%Outlier
Feature				
Targets	572	33	4.50	1.43
Nontargets	583	42	5.75	1.47
Conjunction				
Targets	666	40	6.75	2.43
Nontargets	739	43	7.75	1.90

Note—Reaction times (RTs) and standard errors (SEs) are given in milliseconds.

According to a simple reading of FIT, features are registered preattentively and in parallel, which makes feature detection an effortless operation. In contrast, conjunctions can be registered only after the initial stage of feature processing is complete and additional feature-binding operations have taken place. This idea is supported by the observation in audition that when the appropriate methodological controls are put in place so that the conjunction condition depends on the registration of specific combinations of features, T/NT discriminations completed on the basis of auditory conjunctions are slower than similar discriminations made on the basis of auditory features. Yet, despite this potential correspondence between auditory and visual stimulus processing, alternative accounts of performance continue to suggest themselves.

One such explanation relates to the different levels of complexity implicated by respective feature and conjunction classification rules.⁸ Consider the processing required to identify targets in the feature condition in Experiment 3. From Figure 2A, it can be seen that the first target (left high) was discriminable from nontargets on the basis of the unique location feature *left*. The second target (right low) was discriminable from nontargets on the basis of the unique frequency feature *low*. Therefore, rather than considering both location and frequency on every trial, the participants could simply respond *target*, given the presence of the unique features described above. If the participants also checked for the presence of these target features in a sequential manner, the average number of feature checks needed for target-present responses was 1.5. That is, a response to the first target could be made after checking for the first unique feature, *left*, and a response to the second target could be made after checking for the second unique feature, *low*.

However, in the conjunction condition, the participants had to register the presence of both *left* and *high* features in order to respond to the first conjunction target, and had to register the presence of both *right* and *low* features in order to respond to the second conjunction target. Therefore, the sequential feature-checking account predicts that target-present responses in the conjunction condition demand, on average, three feature checks, since a response to the first conjunction target can be made after two feature checks, whereas a response to the second conjunction target can be made after four feature checks. Therefore, the conjunction cost may simply have reflected the difference in the number of feature checks that were required in the feature and the conjunction conditions. Other feature-checking accounts also suggest themselves, but they all assert that the difference in performance between the feature and the conjunction conditions was simply due to the fact that fewer feature checks were needed in the former case.

To test this basic claim, it was important to design a new control condition in which the same number of feature checks were required as in the conjunction condition. However, it was also important that T/NT discriminations in this new control condition need not be predicated on additional feature-binding operations. In this respect, the

feature-checking account would predict no difference in performance across the new control and the conjunction conditions. In contrast, the FIT account would predict that a cost would still arise in the conjunction condition relative to the new control condition.

Experiment 5 therefore reported performance across three conditions: a standard feature condition, a standard conjunction condition, and a novel control condition. In order to develop the new control condition, the space of possible stimuli had to be increased, and consequently, a speaker array for free-field presentation was constructed so that one of four possible tone locations could be used to present one of four possible frequencies. Whereas the design of the feature conditions and the conjunction conditions follows on from Experiments 2–4, the control condition requires additional explanation.

As before, two target stimuli were defined as different combinations of location (L) and frequency (F) features. Let L1F1 represent the first target stimulus and L4F4 represent the second target stimulus, with the numbers referring to different values of location and frequency (L1 representing the leftmost speaker, F1 representing the highest frequency, and so on). Four nontargets may also be defined—namely, L2F1, L1F2, L4F3, and L3F4. As in the previous experiments, each of the nontargets shared one feature with any given target, and no common features were shared between nontargets. However, owing to the design of the new control condition, targets could no longer be discriminated from nontargets on the basis of single unique features. That is, according to the sequential feature-checking model, a response to the first target could be made only after two feature checks, and a response to the second target could be made only after four feature checks. On average, therefore, three feature checks would be needed to make target-present responses, which is the same as in the standard conjunction condition. So if the average number of feature checks is critical to task performance, target detection in the feature condition (which depends on an average of 1.5 checks) should be significantly faster than target detection in both the control and the conjunction conditions (which, in both cases, depends on an average of 3 checks). Moreover, if feature checking is *sufficient* for the completion of all three tasks, there should be no observable difference between the control and the conjunction conditions, given the equivalent number of feature checks in both of these conditions.

Not only does this new control condition provide a test of the feature-checking account, but it also makes predictions regarding putative processes of coactivation (see Mordkoff et al., 1990). According to these accounts, target evidence builds up significantly faster when two target features are present (i.e., conjunction targets) than when just a single target feature is present (i.e., feature targets). Therefore, the coactivation hypothesis predicts that RTs to target responses in the control condition should be significantly faster than target responses in the feature condition. However, in contrast to both the feature-checking and the coactivation hypotheses, the feature-binding idea

predicts a different pattern of results. Namely, target RTs in the conjunction condition should be longer than those in both the control and the feature conditions, since it is only in the conjunction tasks that the critical feature-binding operations are assumed to operate. Experiment 5 was therefore carried out in order to resolve these competing accounts of the data.

Method

Participants

An initial sample of 32 participants was taken from the student population at York University. However, 4 individuals had to be removed, since they produced large error rates (all greater than 20%) relative to the sample average of 8.40%. As a consequence, 4 new participants were run as substitutes. Of the final sample of 32, 17 were female and 15 were male, 27 were right-handed and 5 were left-handed. The mean age within the sample was 22.59 ($SD = 5.33$). All the individuals received course credit or £4 for the completion of Experiment 5, and all reported normal hearing.

Stimuli and Apparatus

Four tone frequencies were selected, with the criteria that adjacent frequencies were separated by an equivalent log-related distance. The four tone frequencies were 300, 420, 588, and 823 Hz, created in SoundEdit 16. In order to maintain parity with previous studies using this particular experimental setup (see Dyson & Quinlan, 2002), each tone was 25 msec in length, with 5-msec linear rise/fall times, and was presented over one of four free-field speakers (Roll Master LS-500) located at a radius distance of 70 cm, at intervals of -50° , -25° , $+25^\circ$, and $+50^\circ$ from the central 0° head position of the listener. The negative sign indicates the pair of speakers to the left of center, and the positive sign indicates the pair of speakers to the right of center.

All four tones were calibrated to 70 dB (B) at each of the four speakers in the array, using an artificial ear and omnidirectional microphone (Brüel & Kjør Type 4145) and sound level meter. Of the 16 possible combinations of frequency and location, 8 were selected for use in Experiment 5. These were -50° , 823 Hz; -50° , 588 Hz; -50° , 300 Hz; -25° , 300 Hz; $+25^\circ$, 823 Hz; $+50^\circ$, 823 Hz; $+50^\circ$, 420 Hz; and $+50^\circ$, 300 Hz. A PsyScope script (Cohen et al., 1993) was again used to order and present the stimuli. In addition to listening to auditory stimuli over the speaker array, the participants also attended to a computer monitor situated at the straight-ahead position (i.e., 0°) placed 70 cm from a chin- and headrest. The participants responded via a PsyScope button box placed between the headrest and the monitor.

Design

The three conditions in the experiment were known as the feature condition, the control condition, and the conjunction condition. As in the previous experiments, the participants were instructed to make a speeded response to singly presented stimuli on the basis of a pre-defined T/NT classification scheme. Within each block, a random half of trials contained a target stimulus. In all the conditions, the same two targets were used (-50° , 823 Hz, and $+50^\circ$, 300 Hz), and so the only systematic difference between the conditions was the allocation of stimuli to the nontarget category. Figure 4 provides a schematic representation of the stimulus sets used.

Feature condition. The feature condition essentially replicated the design of the standard feature task used in Experiments 2–4. For each condition, two nontargets were defined so that each shared a common feature with one of the targets (see Figures 4A and 4B, respectively). Two versions of the feature condition were run in a bid to balance the overall design across the feature, control, and conjunction conditions. These two versions were always completed one after the other, with the order of completion being counterbalanced between participants. For each of these versions, the participants

completed a practice block of 24 trials (6 presentations of each stimulus), followed by an experimental block of 100 trials (25 presentations of each stimulus). In total, the feature condition comprised 48 practice trials and 200 experimental trials.

Control condition. In the control condition, the participants also took part in a total of 48 practice trials and 200 experimental trials, but in contrast to the feature condition, these trials were now completed in a single condition (see Figure 4C). The same two target tones were used as before, but now four nontargets were defined by combining the two sets of two nontargets used in the separate versions of the feature condition. Each of the targets was presented 12 times in the practice block and 50 times in the experimental block. Each of the nontargets was presented 6 times in the practice block and 25 times in the experimental block.

Conjunction condition. The conjunction conditions in Experiment 5 mirrored the other conjunction conditions tested previously in Experiments 2–4 (see Figure 4D). The targets were the same as those used in the feature and control conditions, but now two different nontargets were defined. The participants completed the conjunction condition twice. In each case, a practice block of 48 trials (12 presentations of each stimulus) was followed by an experimental block of 200 trials (50 presentations of each stimulus). It should be noted that across the feature and the control conditions, the same two targets were presented an equal number of times (100 target-present trials). However, the participants had twice as much exposure to these targets in the conjunction condition (200 target-present trials). Therefore, the participants had more practice at processing the targets in the conjunction condition than in the other two conditions. The design therefore continues to provide a stringent test of the conjunction cost.

Procedure

The procedure for Experiment 5 was the same as that in all the previous experiments. At the beginning of testing, the participants were encouraged to conceptualize the tone locations as values along a continuum from extreme left to extreme right and to similarly conceptualize the tone frequencies as values along a continuum from extreme high to extreme low. The targets in all the conditions were defined as “any extreme high tone in the extreme left speaker and any extreme low tone in the extreme right speaker.” As in all the previous experiments, trial order was randomized within each block, and condition order and assignment of preferred hand to target response were counterbalanced across participants.

Results

Of the scores, 8.56% were eliminated owing to incorrect responses, and a further 2.37% of the scores were removed as outliers. RT analyses were conducted upon correct responses in the experimental blocks only. Table 6 provides summary statistics of the data for the conditions of interest.

Both RT and error rate data were analyzed via two-way repeated measures ANOVAs, with the first factor of condition having three levels (feature, control, and conjunction) and the second factor of response having two levels (target and nontarget). From the analysis of RTs, the critical finding was that responses to conjunction targets were significantly longer than RTs to targets in both the feature and the control conditions (Tukey's HSD test, $p < .05$). Responses to the targets in the latter two conditions were not statistically different from each other ($p > .05$). In addition, analysis of nontarget responses revealed that RTs to the nontargets in the feature condition were significantly shorter than RTs to nontargets in the control and con-

A FIRST FEATURE CONDITION				
Ex. 5	ELeft	Left	Right	ERight
EHigh	T1	NT1		
High				
Low				NT2
ELow				T2

B SECOND FEATURE CONDITION				
Ex. 5	ELeft	Left	Right	ERight
EHigh	T1			
High	NT1			
Low				
ELow			NT2	T2

C CONTROL CONDITION				
Ex. 5	ELeft	Left	Right	ERight
EHigh	T1	NT1		
High	NT2			
Low				NT3
ELow			NT4	T2

D CONJUNCTION CONDITIONS				
Ex. 5	ELeft	Left	Right	ERight
EHigh	T1			NT1
High				
Low				
ELow	NT2			T2

Figure 4. Schematic representation of the stimulus sets used in Experiment 5. T1 and T2 refer to the target stimuli, and NT1, NT2, NT3, and NT4 refer to the nontargets, where appropriate. Each stimulus set was defined on two dimensions, and each dimension comprised four feature values. The columns represent the values of the location dimension, and the rows represent the values of the frequency dimension. Any dimensional value prefixed with "E" denotes *extreme*.

junction conditions ($p < .05$). RTs across the latter two types of nontargets were not statistically different ($p > .05$). The relatively long RTs to nontargets in the control condition may reflect the fact that, in this condition, there were more nontargets (i.e., four) than in the other two conditions (i.e., two).

Overall, the error rate data revealed the presence of two independent effects. First, responses to targets were more inaccurate than responses to nontargets. The increase in error rates between targets and nontargets (approximately 2%) was consistent across all conditions and was also independent of RT differences between targets and nontargets. Second, there was an overall increase in error rate (approximately 3%) in the control condition, relative to the feature and the conjunction conditions. Although there is no obvious reason for the first effect, the second effect may be a result of the increased stimulus uncertainty in the control condition. Whereas in both the feature and the conjunction tasks there were two targets and two nontargets presented per block, in the control condition there were two targets and four nontargets (see Experiment 1 for a similar account). Taken together, though, across the

RT and error analyses, there was no evidence of any compromising speed/error tradeoff in Experiment 5.

Discussion

As in Experiments 2–4, the basic conjunction cost was also observed in Experiment 5. That is, the participants were slower in making target responses in the conjunction condition than in the feature condition. Of critical interest, though, was performance in the new control condition. Now, target responses in the control condition were as fast as target responses in the feature condition but were faster than target responses in the conjunction condition. Therefore, the central result is of a conjunction cost for target responses gauged relative to both the feature and the new control conditions.

The control condition was specifically designed to test a simple feature-checking account of performance in the tasks. The conjunction cost found in the present experiments (i.e., Experiments 2–4) may simply have reflected a reduction in the number of feature checks demanded by the feature condition relative to the conjunction condition. This stands in contrast to the alternative FIT view, which

Table 6
Summary Statistics for the Six Stimulus Types in Experiment 5

Stimulus Type	RT	SE	%Error	%Outlier
Feature				
Targets	662	27	9.03	2.19
Nontargets	678	27	7.09	2.31
Control				
Targets	660	26	11.59	2.31
Nontargets	724	27	9.88	2.41
Conjunction				
Targets	697	24	8.52	2.12
Nontargets	714	24	7.00	2.77

Note—Reaction times (RTs) and standard errors (SEs) are given in milliseconds.

states that at least part of the cost reflects additional feature-binding operations demanded by the conjunction task. To try to help adjudicate between these two accounts, the control condition in Experiment 5 was designed so that the same number of feature checks was required for target detection in both the control and the conjunction tasks. Despite this, a conjunction cost was still uncovered, which seems to suggest that additional binding processes were needed in order to make correct target responses in the conjunction condition.

Of additional interest is the fact that target responses were no slower in the control condition than in the feature condition. This result seriously undermines any simple serial feature-checking account of performance and is also difficult to square with a coactivation account of performance in the control condition. According to the latter view, target RTs should have been shorter in the control condition than in the feature condition. Since such an effect did not occur, the simple coactivation account of performance in the control condition is discredited. These data do suggest, though, that some form of parallel feature-checking account of performance, in which evidence for more than one target feature can be checked for at the same time, is plausible, given that the participants were no slower in checking for two target features than they were in checking for one. Such an account sits comfortably with the application of an FIT framework to the present data, in which a featural stage of analysis precedes a later stage of feature binding.

GENERAL DISCUSSION

Collectively, in the five experiments reported here, performance in speeded auditory classification tasks has been systematically investigated. An aim was to provide some insights into feature processing and feature binding in the auditory modality.

Experimental Rationale and Summary of the Main Findings

Experiment 1 employed the case in which participants were asked to classify auditory stimuli, using location and vocal identity. As a first exploratory step, the stimulus-

response contingencies between feature and conjunction conditions were taken directly from a previous visual search experiment (Quinlan & Humphreys, 1987, Experiment 1). In contrast to the standard conjunction cost found in the visual search literature, the participants were significantly faster in responding to targets defined by a conjunction of location and identity, relative to targets defined by location or identity alone. This so-called conjunction benefit apparently replicated a similar finding reported previously by Woods et al. (1998).

The observation of a conjunction benefit is interesting, since it is difficult to reconcile with accounts of auditory processing based on FIT (Treisman & Gelade, 1980). According to such an account, the recovery of conjunction information occurs *after* an earlier stage of feature detection; hence, conjunction classification responses should be no faster than responses to the most slowly processed constituent feature. However, given the observation of a conjunction benefit, Woods et al. (1998) proposed an alternative, parallel interactive account of auditory stimulus processing. Under this alternative framework, several auditory features are processed in parallel at a very early stage of processing. These component feature processes are assumed to interact in their operations, and as a consequence, auditory conjunctions can be registered during the same stage as their constituent features (after Hillyard & Münte, 1984). Although it is clear that any conjunction benefit might be taken as support for this parallel interactive account, in the present case it was considered to be important to examine a number of methodological issues arising from Experiment 1 before attempting to draw such a conclusion.

Specifically, it became apparent that in the conjunction conditions of Woods et al. (1998) and Experiment 1, the participants need not have based their responses on a representation that contained bound feature information. Rather, the conjunction benefit may have arisen simply because the participants were responding on the basis of the number of target features present. When conceived of in this way, performance can be accounted for in terms of processes of feature coactivation, as has been discussed by Miller (1982) and Mordkoff and co-workers (Mordkoff & Yantis, 1991; Mordkoff et al., 1990). On such a view, the outputs of independent feature analyzers are summed prior to response execution, and it is the amount of target feature activation that is critical in determining the speed of response. More activation will be produced when a conjunction target is presented than when a feature target is presented; hence, a conjunction benefit will arise. However, it is clear that the coactivation account fails to capture the notion of feature binding set out by FIT (see Mordkoff et al., 1990, p. 164, for a similar argument). That is, it is not simply the amount of target feature activity that is important, but the fact that one feature has been combined with another to give rise to a representation that codes a particular conjunction. This may appear to be a subtle distinction, but when the results from the first and the later experiments are compared, the difference is striking.

In Experiment 2, the stimuli were again defined relative to location and vocal identity, and in Experiment 3, location and frequency were the defining dimensions. In these experiments, a condensation task was used as the conjunction condition, so that the participants were forced to decide about the presence of particular combinations of features. Under these conditions, the conjunction benefit was replaced with a conjunction cost, in that the participants were now slower to make responses in the conjunction condition relative to the feature condition. Experiment 4 went on to demonstrate that the conjunction cost was also independent of any differences in NT–NT similarity between the conditions.

Although Experiments 2–4 controlled for several confounding variables identified in Experiment 1, alternative accounts of the data remained. For example, there was still a potential imbalance in the number of feature checks between the feature and the conjunction conditions that could account for the different speeds of response. Specifically, from a simple serial feature-checking account, half the number of feature checks were necessary to make target-present responses in the feature condition than in the conjunction condition. In order to test this hypothesis, a new control condition was developed in Experiment 5. This control condition was designed so that, logically, the number of feature checks was the same as that required in the conjunction case. The critical difference now was that target responses in the conjunction condition could be based only on the detection of a particular combination of features, whereas in the control condition individual feature checks would suffice.

The data from Experiment 5 revealed a conjunction cost measured with respect both to a standard feature task and to the new control condition. Interestingly, there was no difference in RTs to target stimuli between the feature and the control conditions. Such a finding fits well with the idea of a parallel feature-checking stage in which tests for more than one feature can be carried out simultaneously. Such an idea, in turn, is central to the FIT account on human visual processing. More important, though, the data strongly suggest that performance in the conjunction condition cannot solely be the result of any simple feature-checking process; participants must have recovered an integrated representation of auditory features in order to make a response. The general conclusion seems to be that when participants are forced to decide about the presence of particular combinations of features, such decisions take time—time that is reflected in the conjunction cost described here.

Other Related Work

Although the observation of a conjunction cost in audition is a relatively novel finding, a similar cost has also recently been described by Woods, Alain, Diaz, Rhodes, and Ogawa (2001). Here, a variant of the simple go/no-go task was reported in which only two locations (i.e., left and right) and two frequencies (i.e., 1500 and 250 Hz) were used. As a result, RTs were substantially reduced, in com-

parison with Woods et al.'s (1998) study. However, the absence of binaural (i.e., central) stimuli resulted in an RT reduction in responding to location targets that was much larger than the reduction in RT for frequency targets, owing to the absence of middle-frequency (i.e., 612 Hz) stimuli. Now, RTs to feature targets defined relative to a distinctive frequency and feature targets defined relative to a distinctive location were shorter than RTs to conjunction targets defined as a combination of frequency and location.

In order to explain their original conjunction benefit, Woods et al. (1998) developed the parallel interactive account of processing, in which assumptions about the relative difference in processing speeds of the constituent features of an auditory conjunction were critical. Specifically, they argued that if the analysis of the most salient feature revealed the presence of a target feature, top-down excitation would be fed back to facilitate the processing of the less salient feature. Given that such excitation occurs for conjunction targets, and not for feature targets, the conjunction benefit is explained. In almost all respects, this kind of model is identical to the coactivation models discussed by Mordkoff and Yantis (1991). The central assumption of both models is that constituent features are processed via specific feature analyzers but that these analyzers interact in their operations. Whereas both the parallel interactive model and the coactivation model comfortably account for the conjunction benefit, neither model handles the presence of a conjunction cost without making additional assumptions.

In order to accommodate the observation of both conjunction costs and benefits, Woods et al. (2001) argued that a conjunction benefit is most likely to arise when at least one of the constituent features (typically, the location value) takes a relatively long amount of time to be recovered. In contrast, when both features are recovered quickly, a conjunction cost will arise. Woods et al. (2001) argued that although facilitatory interactive feature analysis does take place, this facilitation occurs only after a substantial delay. Consequently, any increase in task difficulty or the concomitant slowing of feature analysis should increase the likelihood of facilitation in the conjunction condition.

Although the coactivation account described above could also be embellished with further assumptions about the speed of constituent featural analysis, neither account provides a useful framework for thinking about the overall pattern of results described here. Conjunction costs consistently occurred in Experiments 2–5, and the presence of these effects does not appear to be related to either the speed of constituent feature analysis or overall task difficulty in the manner predicted by Woods et al. (2001). For instance, it was shown throughout Experiments 1–4 that RTs to conjunction targets cannot be predicted from the relative speed of responding on the constituent feature target trials. Even though the experimental series did not reveal a reliable difference in RTs between distinctive location and distinctive identity/frequency feature targets, Experiment 3 provided an isolated example whereby responses to conjunction targets were significantly different

from those to only one type of feature target (i.e., distinctive location). Although according to Woods et al. (2001), this should increase the likelihood of facilitatory feature processing, Experiment 3 demonstrated an overall conjunction cost, as opposed to a conjunction benefit. Moreover, conjunction *benefits* for combinations of features that had equivalent speeds of processing were also reported in the present experimental series (Experiment 1). Therefore, the differential effects on conjunction trials cannot be attributed to changes in the relative speed of the corresponding constituent featural analysis. Similarly with respect to task difficulty, there appears to be no systematic relationship between the absolute speed of the task and the likelihood of observing a certain type of conjunction effect (i.e., cost or benefit).⁹ Consequently, attempts of how best to characterize performance in simple classification experiments fall somewhat short of providing a more general account of the pattern of performance described in the present experiments.

Overall, therefore, quite a complicated picture of auditory processing has emerged. Although evidence can be provided to support the idea that auditory conjunctions may be available concurrently with auditory features at a preattentive stage of analysis, the degree to which these auditory conjunction tasks necessitate feature binding may be questioned. Even though a stimulus may be defined as a conjunction of two constituent features, the simple classification of that stimulus need not reflect the operation of mechanisms that are intrinsically concerned with feature binding as defined in FIT.¹⁰ It has been established that when the task demands are such that participants must identify particular combinations of component features, these processes incur a time penalty that reveals itself as a conjunction cost. Therefore, the *identification* of particular combinations of auditory features appears to demand attentional processes that operate after an initial stage of parallel (although not necessarily interactive) feature detection. Data such as these (see also Dyson & Quinlan, 2002) indicate that when the modalities are considered at this higher level of analysis, both the processing of auditory stimuli and that of visual stimuli appear to exhibit similar constraints. The identification of further modality-independent constraints remains an exciting goal for future research.

REFERENCES

- CHUN, M. M., & WOLFE, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, **30**, 39-78.
- COHEN, J., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.
- DAY, R. S. (1972). Interactions between linguistic and nonlinguistic processing [Abstract]. *Journal of the Acoustical Society of America*, **51**, 79.
- DUNCAN, J., & HUMPHREYS, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.
- DUNCAN, J., & HUMPHREYS, G. W. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 578-588.
- DYSON, B. J., & QUINLAN, P. T. (2002). Within- and between-dimensions processing in the auditory modality. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 1483-1498.
- FITTS, P. M., & BIEDERMAN, I. (1965). S-R compatibility and information reduction. *Journal of Experimental Psychology*, **69**, 408-412.
- GARNER, W. R. (1970). The stimulus in information processing. *American Psychologist*, **25**, 350-358.
- GARNER, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, **8**, 98-123.
- GOMES, H., BERNSTEIN, R., RITTER, W., VAUGHAN, H. G., JR., & MILLER, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, **34**, 712-716.
- HALL, M. D., PASTORE, R. E., ACKER, B. E., & HUANG, W. (2000). Evidence for auditory feature integration with spatially distributed items. *Perception & Psychophysics*, **62**, 1243-1257.
- HANSEN, J. C., & HILLYARD, S. A. (1983). Selective attention to multi-dimensional auditory stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 1-19.
- HILLYARD, S. A., & MÜNTE, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, **36**, 185-198.
- JOLICŒUR, P. (1999). Concurrent response-selection demands modulate the attentional blink. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1097-1113.
- LOCKHEAD, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review*, **79**, 410-419.
- MELARA, R. D., & MARKS, L. E. (1990a). Dimensional interactions in language processing: Investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 539-554.
- MELARA, R. D., & MARKS, L. E. (1990b). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition*, **18**, 477-495.
- MILLER, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, **14**, 247-279.
- MONDOR, T. A., ZATORRE, R. J., & TERRIO, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 66-79.
- MORDKOFF, J. T., & YANTIS, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 520-538.
- MORDKOFF, J. T., YANTIS, S., & EGETH, H. E. (1990). Detecting conjunctions of color and form in parallel. *Perception & Psychophysics*, **48**, 157-168.
- NÄÄTÄNEN, R., PORKKA, R., MERISALO, A., & AHTOLA, S. (1980). Location vs. frequency of pure tones as a basis of fast discrimination. *Acta Psychologica*, **44**, 31-40.
- QUINLAN, P. T., & HUMPHREYS, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints of feature and conjunction searches. *Perception & Psychophysics*, **41**, 455-472.
- RHODES, G. (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics*, **42**, 1-14.
- SCHRÖGER, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology*, **32**, 55-65.
- SMITH, M. C. (1968). Repetition effect and short-term memory. *Journal of Experimental Psychology*, **77**, 435-439.
- SPRINGER, S. P. (1973). Memory for linguistic and nonlinguistic dimensions of the same acoustic stimulus. *Journal of Experimental Psychology*, **101**, 159-163.
- SUSSMAN, E., GOMES, H., NOUSAK, J. M. K., RITTER, W., & VAUGHAN, H. G., JR. (1998). Feature conjunctions and auditory sensory memory. *Brain Research*, **793**, 95-102.
- TREISMAN, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 652-676.
- TREISMAN, A. (1992). Spreading suppression or feature integration? A reply to Duncan and Humphreys (1992). *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 589-593.
- TREISMAN, A., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

- VAN SELST, M., & JOLICŒUR, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, **47A**, 631-650.
- WINER, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.
- WOLFE, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-73). Hove, U.K.: Psychology Press.
- WOODS, D. L., ALAIN, C., DIAZ, R., RHODES, D., & OGAWA, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 65-74.
- WOODS, D. L., ALAIN, C., & OGAWA, K. H. (1998). Conjoining auditory and visual features during high-rate serial presentation: Processing and conjoining two features can be faster than processing one. *Perception & Psychophysics*, **60**, 239-249.

NOTES

1. As Douglas Creelman correctly pointed out with respect to an earlier draft of this paper, the terms *feature target* and *conjunction target* are, under a certain reading, misleading, since both types of targets are usually defined relative to the combination of two or more features. Rather, it is the task that determines whether the target undergoes feature or conjunction analysis. Although we acknowledge this fact, these terms will be used as shorthand for describing the predefined stimuli in target/nontarget discrimination tasks that require feature or conjunction information, respectively.

2. These, in turn, were essentially taken from the earlier experiments of Treisman and Gelade (1980).

3. Winer (1962) recommended that when the data are expressed as proportions, it is advisable to use the arcsin transform. It should be noted, however, that the analyses of the transformed data did not differ substantially from those carried out on the original error scores. Following standard practice, only the analysis of the transformed data is reported.

4. This procedure is described in detail by van Selst and Joliceur (1994), but in summary, it involves an iterative process whereby, on each iteration, outlying scores are identified, then eliminated. At each iteration, scores in each cell are rank ordered, and a new mean is calculated in the absence of the most extreme score (defined relative to the old mean). Cutoff points are then established according to the revised mean, the revised standard deviation, and a value determined by the number of observations within that cell. If any scores fall outside these limits, they are excluded, and the next iteration is initiated. If no extreme values are identified, the cell mean and standard deviation are calculated from the remaining set of observations, and the process stops.

5. A similar imbalance in the number of feature and conjunction target tokens was also present in Experiment 2 reported by Woods et al. (1998). Although both feature and conjunction targets represented one ninth of all stimuli, there were three tokens for each feature target type (i.e., left type: left, 250 Hz; left, 612 Hz; and left, 1500 Hz) and only one

token for the single conjunction target type (e.g., left, 1500 Hz). Therefore, it may be that the conjunction benefit observed in Woods et al. (1998) and in Experiment 1 had the same source—namely, that a greater number of target tokens were used in the feature condition than in the conjunction condition.

6. It will be noted that Woods et al.'s (1998) study also failed to preserve the contingencies between stimuli as in a standard visual search conjunction condition. Whereas traditionally, conjunction nontargets must share one feature with the target, in any one of the conjunction blocks in Woods et al. (1998), only half of the nontargets shared a critical target feature. Therefore, performance in the conjunction condition may have been facilitated by reduced false alarm rates for conjunction nontargets when no target feature was shared. Indeed, Woods et al. (1998) reported the occurrence of significantly more illusory conjunctions when the nontarget shared one target feature than when the nontarget did not.

7. A subtle point regarding interstimulus similarity is that different metrics of similarity should be employed according to the interaction between stimulus dimensions. That is, if the dimensions are separable, the similarity relations are best represented in terms of a city-block metric. Conversely, integral dimensions preserve a more euclidean metric of similarity (Garner, 1970, 1976). However, since the dimensional interaction between location and frequency is ambiguous (compare Näätänen, Porkka, Merisalo, & Ahtola [1980] with Mondor et al. [1998]), it is difficult to arrive at the best characterization of NT-NT similarity. However, Figure 2 demonstrates that irrespective of whether a city-block or a euclidean metric of interstimulus similarity is employed, the observation of increased NT-NT similarity in the feature condition for Experiments 2 and 3 still holds.

8. We thank David Woods for suggesting this alternative account.

9. A detailed comparison of the RTs reported in Woods et al.'s (2001) study and the present investigations reveals much faster responding in the former case. This apparent discrepancy in speed of responding between the two studies can be attributed to the use of a go/no-go task in Woods et al.'s (2001) investigation and the use of a T/NT discrimination task in the present experimental series. Indeed, other studies that have reported choice RTs to auditory stimuli constructed from combinations of frequency and location show ranges of responding similar to those found in the present experiments (cf. Mondor et al., 1998). Although it may be hypothesized that the relatively slow RTs observed in the present experiments account for the conjunction costs in Experiments 2-5, this seems unlikely given the idea that when constituent feature processing is slow, the likelihood of a conjunction *benefit* is increased (Woods et al., 2001).

10. Indeed, a similar argument may be leveled at studies into mismatch negativity that also support the preattentive availability of auditory conjunctions (Gomes, Bernstein, Ritter, Vaughan, & Miller, 1997; Sussman, Gomes, Nousak, Ritter, & Vaughan, 1998). This is not to deny that participants are unable to automatically *detect* a deviant stimulus that differs from standard stimuli on the basis of a unique combination of features alone, but simply to say that this representation fails to preserve the notion of a conjunction as defined by visual stimulus processing.

APPENDIX
Results of Analyses of Variance for Experiments 1-5

Source	Reaction Times	Errors
	Experiment 1	
Condition (C)	$F(1,11) = 13.78, MS_e = 8,633.88, p < .01$	$F(1,11) = 10.93, MS_e = 0.01, p < .01$
Response (R)	$F(1,11) = 12.38, MS_e = 1,782.29, p < .01$	$F(1,11) = 7.68, MS_e = 0.01, p < .05$
C × R	$F(1,11) = 0.61, MS_e = 847.66$	$F(1,11) = 14.10, MS_e = 0.01, p < .01$
	Experiment 2	
C	$F(1,15) = 25.19, MS_e = 14,334.54, p < .001$	$F(1,15) = 24.84, MS_e = 0.01, p < .001$
R	$F(1,15) = 30.56, MS_e = 2,068.40, p < .001$	$F(1,15) < 0.01, MS_e = 0.01$
C × R	$F(1,15) = 6.84, MS_e = 2,428.62, p < .05$	$F(1,15) = 1.06, MS_e = 0.01$
	Experiment 3	
C	$F(1,15) = 11.05, MS_e = 6,842.84, p < .01$	$F(1,15) = 1.76, MS_e = 0.01$
R	$F(1,15) = 17.30, MS_e = 2,153.20, p < .001$	$F(1,15) = 1.07, MS_e = 0.01$
C × R	$F(1,15) < 0.01, MS_e = 1,880.56$	$F(1,15) = 0.53, MS_e = 0.01$
	Experiment 4	
C	$F(1,15) = 29.35, MS_e = 6,843.16, p < .001$	$F(1,15) = 18.39, MS_e = 0.01, p < .001$
R	$F(1,15) = 14.67, MS_e = 1,934.36, p < .01$	$F(1,15) = 0.22, MS_e = 0.01$
C × R	$F(1,15) = 25.02, MS_e = 851.72, p < .001$	$F(1,15) < 0.01, MS_e = 0.02$
	Experiment 5	
C	$F(2,62) = 1.73, MS_e = 11,957.64$	$F(2,62) = 10.26, MS_e = 0.02, p < .001$
R	$F(1,31) = 18.06, MS_e = 2,817.16, p < .001$	$F(1,31) = 12.47, MS_e = 0.02, p < .001$
C × R	$F(2,62) = 10.71, MS_e = 1,097.58, p < .001$	$F(2,62) = 0.39, MS_e = 0.01$

(Manuscript received October 16, 2000;
revision accepted for publication September 16, 2002.)