

Evidence for auditory feature integration with spatially distributed items

MICHAEL D. HALL

University of Nevada, Las Vegas, Nevada

and

RICHARD E. PASTORE, BARBARA E. ACKER, and WENYI HUANG

State University of New York, Binghamton, New York

Recent auditory research using sequentially presented, spatially fixed tones has found evidence that, as in vision for simultaneous, spatially distributed objects, attention appears to be important for the integration of perceptual features that enable the identification of auditory events. The present investigation extended these findings to arrays of simultaneously presented, spatially distributed musical tones. In the primary tasks, listeners were required to search for specific cued conjunctions of values for the features of pitch and instrument timbre. In secondary tasks, listeners were required to search for a single cued value of either the pitch or the timbre feature. In the primary tasks, listeners made frequent errors in reporting the presence or absence of target conjunctions. Probability modeling, derived from the visual search literature, revealed that the error rates in the primary tasks reflected the relatively infrequent failure to correctly identify pitch or timbre features, plus the far more frequent illusory conjunction of separately presented pitch and timbre features. Estimates of illusory conjunction rate ranged from 23% to 40%. Thus, a process must exist in audition that integrates separately registered features. The implications of the results for the processing of isolated auditory features, as well as auditory events defined by conjunctions of features, are discussed.

Feature integration theory (FIT) is one prominent modern theory that describes the role of attention in visual object identification (e.g., Treisman, 1992; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Treisman, Kahneman, & Burkell, 1983; Treisman & Schmidt, 1982; Treisman & Souther, 1986). The sequence of operations leading to object identification, as proposed by FIT, is summarized in Figure 1. The theory argues that individual features (e.g., size, color/contrast, and shape) are first preattentively derived in parallel from the various objects in a simultaneous visual array, with the position of each feature roughly coded in a master map of spatial locations. The notion of the early, preattentive, parallel processing of features is quite typical in theories of visual object recognition (e.g., Marr, 1982). Thus, the demonstration of FIT requires a spatial distribution of objects or events that differ in values of two separable features. The identification of objects at each location (e.g., in Figure 1, a large, filled square or a small, outlined triangle) then is typically enabled by a capacity-limited attentional process

moving in a serial fashion from one location to another. FIT also assumes that the processing of particular conjunctions of features, which define objects, may become automatic given sufficient experience with those objects. In such instances, the overlearned conjunction of features would be processed in parallel and, thus, in the manner observed with individual features.

Critical support for feature integration in visual processing comes from speeded search tasks in which observers try to rapidly report the presence or absence of a single cued feature, or the conjunction of two features, within a visual array. One major type of evidence for feature integration is the report of illusory conjunctions, which are the perception of incorrect combinations of accurately registered features. For example, Treisman and Schmidt (1982) demonstrated that illusory conjunctions are quite common for color and shape (e.g., perceiving a red O and a green X when presented with a red X and a green O), as well as for size and solidity (e.g., filled vs. outlined shapes). As would be expected from a limited capacity system, illusory conjunction rate was found to increase with increasing attentional load, such as that associated with increased array size or complexity of features. Thus, presumably, given sufficient stimulus complexity, it becomes difficult to rapidly focus the attention needed to accurately integrate features; with inadequately focused attention, illusory conjunctions often appear either as the random coupling of features or as the conjunction of separate features to match expectations.

This article is based on work supported by Grants F49620931003 and F49609310327 from the Air Force Office of Scientific Research to the second author. Opinions, findings, conclusions, and recommendations are the authors' and do not necessarily reflect views of the granting agencies. The authors gratefully acknowledge William F. Thompson and an anonymous reviewer for their many helpful comments and suggestions on the manuscript. Correspondence concerning this article should be sent to M. D. Hall, Psychology Department, University of Nevada, Box 455030, Las Vegas, NV 89154-5030 (e-mail: hallm@nevada.edu).

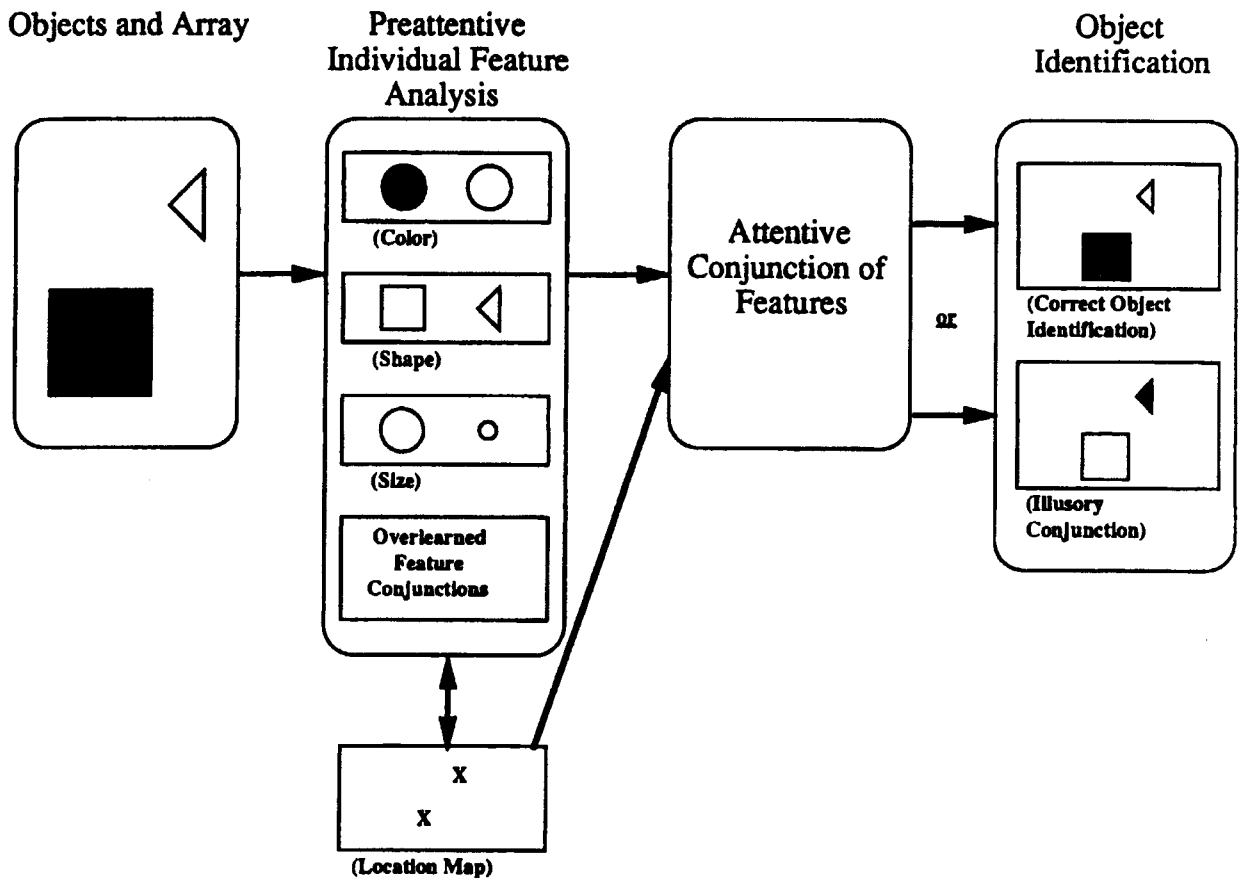


Figure 1. Sequence of events leading to object identification in feature integration theory, including parallel processing of individual features and integration of features with focused attention to perceive (appropriate or illusory) objects.

Some of the original assertions of FIT, such as specific assumptions about parallel and serial processing of items (e.g., Duncan & Humphreys, 1989; Wolfe, Cave, & Franzel, 1989), as well as the posited role of attention in the integration process (Ashby, Prinzmetal, Ivry, & Maddox, 1996; Prinzmetal, Henderson, & Ivry, 1995), may not be completely accurate. However, it is generally agreed that some form of feature integration does occur in vision.¹

FEATURE INTEGRATION AND AUDITION

Recently, the concept of feature integration also has been generalized to the processing of auditory stimuli. However, direct behavioral evidence for auditory feature integration is extremely limited.² Some attempts to demonstrate auditory feature integration, based on analyses of processing speed and perceptual errors for sequences of tones, have yielded mixed results. For example, Woods, Alain, and Ogawa (1998) measured reaction times (RTs) and errors for auditory search tasks with the rapid serial presentation (RSP) of tones, where listeners monitored a sequence of single tones for a specific pitch, a specific location, or a conjunction of pitch and location. RT values

were shortest for the pitch search, longest for the location search, and intermediate for the conjunction task, with significant differences across conditions. In contrast to the original assumptions of FIT, this finding was taken to indicate parallel, rather than serial, processing of the conjunction of auditory features (pitch and location). Furthermore, false alarms were minimal in the conjunction task, indicating that illusory conjunctions probably did not occur in the conditions evaluated. However, use of these findings to reject the notion of auditory feature integration may not be appropriate, given complicating factors inherent in the sequential or serial nature of the RSP tasks, as well as the strategy of treating location as one of the two available features of events. These factors contrast with the possibility for parallel processing when spatially distributed stimuli are presented simultaneously, as in typical visual search tasks. (These differences will be discussed in more detail below.)

A different pattern of results was obtained by Dehaene (1993), who observed distinct patterns of peaks and troughs in the distribution of RTs under different task conditions. Specifically, a longer period between peaks in the distributions was obtained for identifying the order of a pair of 50-msec pitches (presumably, a conjunction task) than

for identifying an isolated pitch (a feature task). This pattern of results could reflect greater task difficulty in the nominal conjunction task and, thus, the need for greater processing resources. However, since the tones in this nominal conjunction task were again sequential and were distinguished on the basis of a single feature (pitch), the nominal conjunction task did not require any perceptual conjunction of features and could be argued to represent a more difficult type of single feature (i.e., pitch) task.

Thompson (1994) also used temporally distributed auditory events, but in what seems to be a more appropriate analogue to visual search tasks. Listeners indicated whether they perceived alterations in repeating two-tone sequences, or *textures*, under four conditions: *no change* within the texture, *simple change* (change in either the pitch or the duration of both tones), *disjunct* (change in the pitch of one tone and the duration of the other), and *switch* (original durations were swapped across pitches in the texture). Error rates were greatest in the switch condition, suggesting that illusory conjunctions of pitch and duration frequently occurred. A similar pattern of results also was obtained for corresponding changes within longer, five-tone sequences (scalar melodies). Consistent with a limited capacity process, sensitivity to the combination of pitch and duration was significantly reduced when attention was distracted or available attentional capacity was reduced by the simultaneous performance of a verb/noun labeling task. Thus, it appears that attentional or processing resources are critical not only for the integration of visual features, as was originally argued in FIT, but also for the integration of the auditory features of pitch and duration.

It could be argued that temporally distributed events represent the most logical stimuli with which to study and demonstrate auditory feature integration. After all, it has generally been argued that time is the most critical dimension for the segregation of auditory events (e.g., Kubovy, 1988). However, the temporal distribution of events and of the features that define those events represents a fundamental change in the conceptualization of the processes described by FIT. For visual stimuli, FIT describes the parallel, and thus simultaneous, identification of all the features of every object (e.g., all the colors, shapes, and sizes of all the objects in the visual array), each coupled loosely with a spatial location map. All the features thus are simultaneously present and available to the subsequent serial (i.e., temporally distributed) conjunction process. With sequentially presented events, the features that define a single event might be processed simultaneously, although even within a single event, some features may not be simultaneously available. (For example, whereas the pitch of a tone can be determined after its first 20 to 40 msec, its duration definitely cannot be determined until the tone ends.) However, the initial processing of the features from different sequential events must be serial, and these features are simultaneously available only to the degree that the earlier events are accurately available in memory. Therefore, although some

form of feature integration appears to be needed for sequential events, the nature of the perceptual processes may be quite different from those typically described for vision and implied in FIT. A fundamental question, then, is whether processes similar to those found in simultaneous visual objects also exist for auditory events.

Like visual objects, auditory events may be both simultaneous and spatially distributed, with the spatial and/or temporal distributions of attributes being important for the perceptual grouping and segregation of events (see, e.g., Bregman, 1990; Handel, 1988a, 1988b; Woods et al., 1998). If the goal of research is to demonstrate an initial parallel stage of processing, then the stimulus set must be distributed in a manner that allows for the possibility of initial parallel processing, which then can be followed by a serial process moving across that distribution. The spatial distribution of auditory or visual stimuli clearly allows for this possibility, whereas the sequential distribution of stimuli is inherently serial. Furthermore, the idea that auditory information can be selected by location in auditory space is not new, with early work on attention treating each ear as a separate input channel (Broadbent, 1958; Moray, 1970; Treisman, 1969). Focal attention to a specific spatial location in an auditory array also has been used in investigations of processing capacity (e.g., Darwin, Turvey, & Crowder, 1972). It therefore should be possible to evaluate feature integration processes by using a search task with simultaneously presented, spatially distributed, auditory events. Providing a direct analogue to visual FIT within such an evaluation constitutes the goal of the present investigation.

Our motivation for demonstrating feature integration with simultaneously presented stimuli came from a perceptual study in our laboratory involving dichotic musical stimuli (Hall & Pastore, 1992a). In this earlier study, the effect on perceptual organization of stimulus complexity, as defined by the number of distinct tones contributing to a chord, was investigated. In one condition, the tone that distinguished a specific major chord from its corresponding minor chord (E and E_b, respectively) was presented to one ear, with the remainder of the chord (e.g., C-G, C-G-A, C-G-B-D) being presented to the other ear. This dichotic condition resulted in the frequent perceptual fusion of the separate stimuli to create a major or a minor chord, as had been reported in earlier studies of musical duplex perception (e.g., Collins, 1985; Hall & Pastore, 1992b; Pastore, Schmuckler, Rosenblum, & Szczesiul, 1983). In another condition, the E and E_b stimuli were presented simultaneously, but to separate ears, with one tone presented in isolation and the other mixed with the remainder of the chord. Independent of where the remainder of the chord was presented, responses often indicated that listeners had perceived the E and E_b tones in the wrong ears, thus providing evidence for migration of the pitch feature across ears. Similar patterns of perceptual change involving migrated or incorrectly conjoined features have been reported for speech stimuli (e.g., Cutting, 1976; Kolinsky, 1992; Repp, 1978a, 1978b).

Since both fusion and tone migration represent the mislocalization and incorrect combination of presented elements, both types of perceptual organization could be considered examples of illusory conjunctions. Furthermore, in our earlier study, the probability of both fusion and migration increased with increasing stimulus complexity (number of tones presented in the chord ear). Therefore, perceptual errors were more probable in the context of many, rather than few, stimulus elements, as is typically expected and observed for illusory conjunctions in vision.

The present experiment was conducted to provide direct evidence of illusory conjunctions with simultaneous, spatially distributed auditory events. We developed auditory search tasks that were analogous to the visual search tasks used by Treisman and colleagues (e.g., Treisman & Gelade, 1980; Treisman & Schmidt, 1982). In the primary tasks, listeners searched arrays of simultaneous but differentially lateralized tones for the conjunction of cued values for the features of musical pitch and instrument timbre. Error rates in these conjunction search tasks were used to estimate the probabilities of feature misperceptions and illusory conjunctions, with the latter providing evidence of feature integration. Separate search tasks for pitch and timbre features also were included, with listeners searching for a cued value of a single target feature. Error rates from these secondary tasks provided alternative estimates of the probability of misperceiving each type of feature and thus were used to confirm the probability estimates based on the conjunction search tasks.

METHOD

Participants

For the participants in the present study, it is important that the two features, pitch and instrument timbre, be reasonably separable dimensions. Furthermore, the participants must possess the analytic listening abilities required for the separate search tasks for these features, which, in turn, requires a functional knowledge about both pitch and instrument timbre. Since these are typical characteristics of experienced musicians (Pitt, 1994), we used participants with at least 5 years of ongoing training on a musical instrument. The 10 participants were students attending the State University of New York at Binghamton (5 undergraduates, and 5 graduate students, including authors M.H. and B.A.) who had studied the piano or one of various orchestral instruments. All the participants reported not having any known hearing problem, and their absolute thresholds, measured using a two-alternative forced-choice task at 1000 and 2000 Hz, were all within normal ranges.

Stimuli

The values of two features of the stimuli were systematically varied by factorially combining five values of instrument timbre with five values of pitch. The stimuli were constructed from 2-sec, single-tone recordings of five experienced musicians playing their primary musical instrument (violin, flute, clarinet, trombone, and piano). Recordings for four of the instruments were originally on a 1/2-track Tandberg TD 20A tape deck; the piano timbre was digitally sampled from a Yamaha AWM Sound Expander EMT-10 and recorded on a high-bias, chrome cassette. Thus, timbre was specified both by the distribution of partials and by attack and decay characteristics. All the recorded tones were digitized (12-bit, 20-kHz sample rate, with 4-kHz antialiasing).³

Five pitches were recorded from each instrument: 262 Hz (C_4), 370 Hz ($F\#_4$), 509 Hz (halfway between B_4 and C_5), 762 Hz (halfway between $F\#_5$ and G_5), and 1078 Hz (halfway between C_6 and $C\#_6$).⁴ All the pitches were inharmonically related and were separated by more than a critical band (see, e.g., Fletcher, 1972). These precautions were taken to reduce possible fusion tendencies, thereby helping to maintain the separation of auditory events. Finally, it was necessary to attenuate the intensity of a few isolated tokens in order to equate loudness (subjectively evaluated by authors M.H. and B.A.) across all tokens.

Auditory cues and arrays (defined below) were presented to the listeners at a comfortable listening level (peak amplitude of 80 dB[A]) over TDH49-10Z headphones. Unique, localized positions for each array tone were achieved by manipulating the interaural time disparity of individual tokens. Simultaneous presentation of the same tone to each ear (without interaural time disparity) resulted in a centered percept. All the auditory cues were presented in this centered location. The selected levels of interaural time disparity for the array tones were derived from the values used by Kubovy (1981), with perception of the intended lateralized positions of the tokens verified through extensive listening by the authors. Leading a tone to one ear by 0.5 msec resulted in a shift in perceived location toward that ear; leading by 1.5 msec produced greater lateralization. Thus, there were four possible distinct tone locations for a given array tone: two strongly lateralized to the right or left (*far* from centered) and two moderately lateralized to the right or left (*near* center).

Arrays with two tones and four tones were constructed by digitally mixing two or four tokens, each with unique interaural time disparities across the stimuli to be presented to the left and right earphones. The two-tone arrays were divided into *near* and *far* conditions, differing solely in the degree to which the two tones were lateralized toward different ears (i.e., 0.5- or 1.5-msec interaural time disparity, respectively). This manipulation enabled an evaluation of the possible influence of distance between events on both the misperception of features and the illusory conjunction rate, as has been done for vision (e.g., Ashby et al., 1996; Treisman & Schmidt, 1982). Finally, in order to maximize object separability, no specific pitch or timbre could occur at more than one location in any given array. There was a total of 13 different two-tone near arrays, 13 different two-tone far arrays, and 7 different four-tone arrays.

Procedure

The listeners participated alone or in pairs in commercial acoustic chambers. All listeners completed the two feature (pitch and timbre) search tasks in one 2-h session and the two conjunction search tasks (described below) in a separate 2-h session. The order of the sessions was counterbalanced across listeners, as were the two tasks within each session. At the beginning of each session, the listeners were familiarized with all of the possible pitches and timbres. This was accomplished through the binaural presentation of each pitch token for each instrument. The familiarization sequence was repeated until the given listener reported being comfortable in identifying all the stimuli; no listener requested hearing the stimulus set more than three times prior to beginning a task.

Experimental trials with data collection followed the familiarization sequence. An experimental trial consisted of an initial 2-sec stimulus cue, a 500-msec interstimulus interval, and a subsequent 2-sec array of stimuli. The stimulus cue indicated the target property, or properties, to be evaluated in the given search task. On each trial, the listeners responded with buttonpresses to indicate the presence or absence in the array of the single cued property in feature search tasks or of the cued conjunction of properties in conjunction search tasks. Cues were either *valid* (i.e., a perfect match to the cue was in the array) or *invalid* (a perfect match was not in the array) with equal probability, and with every cue equally probable across trials. The listeners were instructed to respond as quickly and accu-

rately as possible but could respond any time within 4 sec after the onset of the stimulus array. Responses were recorded by a PC. The listeners received rest breaks after half of the 330 randomized trials in each task, as well as after completing the initial feature or conjunction search task within each session.

Conjunction search. Table 1 provides examples of the different types of cuing conditions in conjunction search for a two-tone array. In all of the examples, the search is in an array consisting of (1) a 262-Hz violin tone and (2) a 509-Hz trombone tone. On valid trials, the initial stimulus cue matched a tone that was in the subsequent array; in Table 1, the cue matches Array Tone 1. In contrast to the feature search task (described below), there were two types of invalid trials in the conjunction search conditions, and these occurred with equal probability. On invalid(-) trials, the array contained neither the cued pitch nor the cued timbre, as is shown in Table 1 for cuing with the 370-Hz clarinet tone. On invalid(+) trials, arrays contained both the cued pitch and the cued timbre, but in separate locations and not as an actual conjunction of these features. In other words, the cued pitch and cued timbre were presented on different array tones, as is illustrated in Table 1 by a cue consisting of the 262-Hz pitch and the trombone timbre. In the actual study, the listeners were cued in the conjunction search tasks with each of the (25) individual stimuli, representing the factorial conjunctions of the five possible values of the pitch and timbre features.

There were two different versions of the conjunction search task. In the Y/N task, listeners indicated target presence or absence with a binary *yes/no* response. In true instances of illusory conjunctions, the listeners should demonstrate confidence that the cued, or target, stimulus was accurately perceived in the array (see, e.g., Treisman & Schmidt, 1982). Therefore, the conjunction task was repeated, using a 7-point confidence rating scale. A rating of 1 indicated *strong confidence of target presence*, a 4 indicated *no confidence of either target presence or absence*, and a 7 indicated *strong confidence of target absence*. It was made clear to the listeners that the extreme responses should be made only when they were sure that the cued pitch and cued timbre were definitely present or absent as a single tone at one location in the array.

Feature search. Search tasks for the individual pitch and timbre features were included as control conditions to provide what was expected to be a more direct evaluation of feature perception. In each feature search task, the listeners evaluated arrays for the presence or absence of a single feature, either a cued pitch (pitch search task) or a cued timbre (timbre search task). Sample cuing conditions for each type of feature search task are provided in Table 2.

Cues for the pitch search (Table 2, top) were isolated sine tones that specified the target frequency. In the absence of any upper partials and with a rectangular temporal envelope, the timbre associated with these pitch cues was simplified and was very distinct, relative to the timbre of all the stimuli in the arrays. The listeners reported no difficulty in focusing solely on the pitch feature when presented with such cues. On valid trials, the cued feature was pres-

ent in the array (e.g., the 262-Hz pitch in Table 2, top, matches the pitch of Array Tone 1 from Table 1). An invalid(+) condition, as was described for the conjunction search tasks, with both cued features present but not conjoined, is incompatible with a feature search. Therefore, for both the pitch and the timbre search tasks, there was only one type of invalid trial, representing a type of invalid(-) trial. On these trials, the cued feature was absent from the array (e.g., in Table 2, 370 Hz).

An example of the cuing conditions for the timbre search task is summarized in Table 2 (bottom). Because it is impossible to present an auditory stimulus that possesses a timbre but not a pitch, visual cues were used to indicate timbre. In this way, the pitch feature was completely eliminated from timbre cues. Timbre cues were four-character displays in the response boxes (i.e., PINO for piano, VIOL for violin, TRMB for trombone, FLUT for flute, and CLAR for clarinet). Prior to the timbre search condition, each listener was familiarized with the meaning of these cues and felt confident in their use. In timbre search, the relationship between cued features and arrays on both valid and invalid(-) trials was the same as that for pitch search. Thus, on valid trials, the cued timbre was in the array (e.g., the violin in Table 2, bottom), and on invalid(-) trials, the cued timbre was not in the array (e.g., the flute in Table 2). All other procedures for feature search, including the nature of responses, were identical to those described for the Y/N task in conjunction search.

RESULTS

Conjunction Search

The model used to estimate the probability of illusory conjunctions (described below) is based on error rates. Thus, individual error rates from the Y/N and rating versions of the conjunction search task were calculated for each of the tone conditions (two-tone near, two-tone far, four-tone) as a function of cue validity [valid, invalid(-), invalid(+)]. Results from the confidence rating task were transformed into error rates by treating ratings of 1-3 as *yes* responses and ratings of 5-7 as *no* responses. Ratings of 4, which indicated that the listener was not sure of target presence or absence, were always treated as correct responses [i.e., *yes* on valid trials, and *no* on both invalid(-) and invalid(+) trials]. With this intentional bias against errors, the model provides conservative estimates of illusory conjunction rate.

Mean error rates under each condition in each version of the conjunction search task are shown in Table 3 (top).⁵ For the two-tone conditions, there were no substantial differences in mean error rates between near and far con-

Table 1
Example Cuing Conditions for a Single Two-Tone Array,
Assuming Search for a Conjunction of Pitch (in Hertz) and Timbre

Condition	Tone Cue		Example Array			Responses	
	Pitch	Timbre	Features in Array?		Conjoined?	Yes	No
			Pitch	Timbre			
Valid	262	violin	yes	yes	yes	C	E
Invalid(-)	370	clarinet	no	no	n/a	E	C
Invalid(+)	262	trombone	yes	yes	no	E	C

Note—C, correct; E, error.

Table 2
Example Cuing Conditions for the
Two-Tone Array of Table 1, Assuming Search
for an Isolated Pitch (Top) or Timbre (Bottom)

Condition	Cue	Feature in Array?	Responses	
			Yes	No
Search for Pitch				
Valid	262 Hz	yes	C	E
Invalid(-)	370 Hz	no	E	C
Search for Timbre				
Valid	VIOLin	yes	C	E
Invalid(-)	FLUTE	no	E	C

Note—In the search for pitch, the cue was auditory; in the search for timbre, the cue was visual. C, correct; E, error.

ditions ($p > .05$), indicating that lateralized distance between items was not a critical variable. In addition, error rate tended to be higher for the larger (four-tone) array size than for either of the two-tone conditions. Separate analyses of variance (ANOVAs) with tone condition (two-tone near, two-tone far, four-tone) and cue validity [valid, invalid(-), and invalid(+)] as factors, for each version of the task confirmed this tendency, with significant main effects of tone condition [for Y/N and rating tasks, respectively, $F(2,18) = 36.25$ and $30.56, p < .0001$].⁶ (All subsequent statistics were evaluated with an ANOVA, then supplemented with post hoc, pairwise comparisons [Tukey tests].) This effect of tone condition was significant for both valid trials [$F(2,18) = 33.53$ and $15.31, ps < .001$] and invalid(-) trials [$F(2,18) = 47.98$ and $84.41, ps < .001$], with significantly more errors in the four-tone condition than in the two-tone conditions ($ps < .01$).

In the two-tone conditions, the error rate was very low on invalid(-) trials and also was low on valid trials. On invalid(+) trials, where the two cued features were separately present in the array but were not physically conjoined, error rate was considerably higher. This pattern of error rates across the three types of trials is reflected in significant main effects of cue validity for both the Y/N and the rating task [$F(2,18) = 9.35, p < .01$; $F(2,18) = 33.81, p < .0001$], with significantly more errors obtained

on invalid(+) trials, relative to either valid ($p < .05$) or invalid(-) ($p < .01$) trials. In fact, this effect of cue validity was significant ($p < .01$) for all the tone conditions in the rating task and, in the Y/N task, for both two-tone conditions, but not for the four-tone condition ($p > .05$).

In addition to the error (Y/N) transformation analysis, the actual confidence ratings also were evaluated. The mean confidence ratings, which are summarized in Table 3 (middle), indicate a clear rank ordering of results across conditions. This ordering is reflected in a significant main effect of cue validity [$F(2,18) = 203.72, p < .0001$], so that any change in cue validity produced a significant change in rating ($p < .01$). The combination of low numerical ratings (indicating high positive confidence) for valid trials and high numerical ratings (indicating high negative confidence) for invalid(-) trials confirms that the listeners were confident in identifying conditions in which the cued pitch and timbre either were both present and conjoined or were both absent. A similar tendency was not observed for invalid(+) trials, where mean ratings were intermediate to those for the valid and the invalid(-) conditions.

Feature Search

Table 4 provides a summary of mean error rates from both feature search tasks as a function of task (pitch, timbre), tone condition (two-tone near, two-tone far, four-tone), and cue validity [invalid(-), valid]. The mean error rates from the feature search tasks are relatively high, as compared with the error rates from the corresponding conditions in conjunction search, particularly considering the use of well-trained musicians. The high error rate was most pronounced for the four-tone condition, especially on invalid(-) trials. This increase in error rate is reflected in a main effect of tone condition [$F(2,18) = 31.90, p < .0001$], an interaction between tone condition and cue validity [$F(2,18) = 6.20, p < .01$], and a three-way interaction across all factors [task \times tone condition \times cue validity: $F(2,18) = 6.11, p < .01$].

The error rates in Table 4 provide mixed evidence about the equivalence of task difficulty for the pitch and the timbre search tasks and the difficulty of these tasks rel-

Table 3
Mean Error Rates as a Function of Tone Condition and Cue Validity for Each Task in Conjunction Search

Task	Two-Tone Near						Two-Tone Far						Four-Tone					
	Valid		Invalid(-)		Invalid(+)		Valid		Invalid(-)		Invalid(+)		Valid		Invalid(-)		Invalid(+)	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Error Rates From Conjunction Search																		
Y/N	.09	.04	.06	.05	.29	.04	.11	.04	.05	.05	.33	.05	.30	.04	.21	.04	.30	.05
Rating	.06	.02	.02	.01	.30	.05	.06	.01	.03	.01	.37	.05	.16	.02	.21	.02	.34	.05
Actual (Untransformed) Confidence Ratings																		
	1.74	0.13	6.69	0.08	5.06	0.26	1.78	0.12	6.65	0.13	4.78	0.26	2.67	0.17	5.54	0.16	4.73	0.22
Rates of Confidence of Veridical Perception for Error Responses																		
	.76		.80		.64		.71		.86		.65		.56		.52		.49	

Note—For the rating task, mean ratings also are displayed, along with the percentages of errors that received ratings indicating confidence of veridical perception. Where applicable, standard errors are provided.

Table 4
Mean Error Rates (and Corresponding Standard Errors)
From Feature Search for Pitch or Timbre

Task	Two-Tone Near				Two-Tone Far				Four-Tone			
	Valid		Invalid(-)		Valid		Invalid(-)		Valid		Invalid(-)	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Pitch	.10	.02	.21	.04	.10	.02	.19	.04	.25	.06	.35	.05
Timbre	.26	.03	.18	.03	.24	.03	.18	.02	.29	.03	.50	.04

ative to the feature perception aspects of the conjunction search task. On invalid(-) trials, error rates did not significantly differ for pitch and timbre search [$F(1,18) < 1$]. However, on valid trials, error rate was greater for timbre search than for pitch search [$F(1,18) = 7.82, p < .05$]; this difference also contributed to a main effect of task [$F(1,18) = 7.44, p < .05$] and to a marginal interaction of task and cue validity [$F(1,18) = 3.61, p < .07$]. A potential explanation for this difference in error rates, as well as for the high overall error rates in feature search tasks, will be provided in the discussion.

Finally, the effect of lateralized distance between items on error rates was evaluated by comparing data from the two-tone near and the two-tone far conditions. With an average difference in the probability of an error as a function of distance being only .01 (see Table 4), the effect of distance in statistical analyses did not approach significance (all $F_s < 1$). Thus, as with the conjunction search results, it appears that distance between tones/features was not a critical factor in feature search.

DISCUSSION

The relatively low error rates on valid and invalid(-) trials with two-tone arrays, as well as ratings that reflect the confident and accurate report of the relationship between cue and array features, imply that the misperception of individual features seldom occurred in the conjunction search tasks. Assuming that the trained musician participants, after correctly perceiving the physically conjoined cues features, will not then misconjoin these features,⁷ an incorrect (i.e., *no*) response on a valid trial should reflect the misperception of at least one cued feature as a noncued value. Using the example of the specific cue and array in Table 1, it is conceivable that the listener could misperceive the matching 262-Hz pitch in the array as 509 Hz and/or misperceive the matching violin timbre as a flute. Likewise, an incorrect (*yes*) response on an invalid(-) trial, where neither cued feature was present in the array, could arise only from misperceiving both pitch and timbre features in the array as cued values (for the array in Table 1, misperceiving the 262-Hz pitch as 370 Hz and the violin timbre as a clarinet). One should expect the misperception of pitch or timbre to be infrequent when, as in the present study, the listeners are well-trained musicians. Thus, it should not be surprising

that error rates for two-tone arrays were low for either type of trial. (An account of the increase in error rates for four-tone arrays will be provided below.)

In contrast, error rates were relatively high on invalid(+) trials. These high error rates may reflect the combination of several perceptual events. An incorrect (*yes*) response on this type of trial could reflect the misperception of a noncued feature value as a cued value (e.g., in Table 1, perceiving the violin timbre for both array tones). Alternatively, an incorrect response could reflect the illusory conjunction of the physically separate, but accurately perceived, cued features (e.g., in Table 1, perceiving a 262-Hz pitch with the trombone timbre and a 509-Hz pitch with the violin timbre).

Some support for the occurrence of illusory conjunctions is provided by an examination of confidence ratings. For example, the probabilities of making an incorrect response while reporting confidence of veridical perception (invalid trials: ratings of 1 or 2; valid trials: ratings of 5 or 6) are summarized in Table 3 (bottom). Treisman and Schmidt (1982) utilized an examination of corresponding probabilities from visual search tasks as a means of ensuring that conjunction errors reflected true illusory conjunctions. If high confidence errors are infrequent on valid and invalid(-) trials, the confidence ratings can be viewed as indicating that the responses reflect the clarity of perception. Since illusory conjunctions should be actual (but incorrect) combinations of features, with those combinations indistinguishable from perception on valid trials, there should be a higher probability of high confidence errors on invalid(+) trials. In our auditory conjunction search task, the probabilities of high confidence errors were relatively low on valid and invalid(-) trials; this was particularly true for invalid(-) trials, where these probabilities were below .02 and .03, respectively, for the two-tone near and far conditions. It is only on invalid(+) trials (where the cued features were present, but were not conjoined) that there were substantial probabilities of confident but erroneous responses; these probabilities exceeded .24 in the two-tone far condition. (The slightly reduced magnitude of this probability in the four-tone condition reflects an overall reduction in reported confidence in the four-tone condition; see below.) The intermediate mean ratings obtained on invalid(+) trials thus reflect a mixture of these low numerical ratings (the confident report of the incorrect conjunc-

tion of cued features) and high numerical ratings (the confident and accurate report that the cued features were not conjoined).

Although the increased rate of errors (accompanied by high ratings of confidence) on invalid(+) trials could be argued to reflect the frequent illusory conjunction of the cued pitch and timbre, no reasonable assessment of the actual rate of illusory conjunctions can be made solely from the error rates in each condition. Such an assessment requires the use of a model of search performance under each condition to estimate the probabilities of the underlying perceptual events, including illusory conjunction rate.

A Model of Auditory Illusory Conjunction Rate

Our model of auditory search performance is equivalent to the original, multinomial model for visual search proposed by Treisman and Schmidt (1982). This model has two serial stages, with the feature identification processes being completed before the processes conjoining the perceived features. Our auditory model, including all probability formulae and a derivation of those formulae, is provided in the Appendix. Like the model of Treisman and Schmidt, this auditory model assumes that incorrect responses reflect only two classes of perceptual events: illusory conjunctions and various types of feature misperceptions.⁸ (The probabilities of other types of perceptual errors are assumed to be very small and, thus, are ignored.) The feature misperceptions include the perception of a noncued feature as having a cued value [resulting in incorrect responses on invalid(+) and invalid(-) trials in the conjunction search task] and the perception of a cued feature as having a noncued value (resulting in incorrect responses on valid trials). Since expectations might increase the tendency to misperceive a feature as the cued value, our model differentiates between these two types of feature misperceptions. Thus, there are three probabilities (feature misperceptions as cued and noncued values, plus illusory conjunction rate) estimated from error rates under the three conditions of cue validity [valid, invalid(-), and invalid(+)].

The development of an appropriate perceptual model of search performance requires consideration of the nature of search tasks for spatially distributed auditory events. For example, in existing models of visual search performance, the similarity of features is one important factor that has been argued to influence the probabilities of both feature misperceptions and illusory conjunctions (e.g., Ivry & Prinzmetal, 1991). As only an initial effort to evaluate the importance of feature integration processes for spatially distributed auditory events, the present study neither directly evaluated nor manipulated the similarity of pitch and timbre feature values. However, since listeners in our study reported that some pairs of timbres were more similar than other pairs, we need to consider the implications of differences in feature similarity across conditions.

Feature similarity could account for the effects of array size on error rate in auditory search performance. Al-

though the present discussion will focus on timbre, our analysis could apply equally to differences in pitch feature similarity. In the two-tone conditions, with the cued value in the stimulus array, two of the five possible timbre values were sampled on each valid and invalid(+) trial. On each invalid(-) trial, with the cued value not in the array, three values were sampled (the timbre value for the cue and each of the two array tones). In the four-tone condition, the number of sampled feature values increased by two. Thus, four out of five timbres were sampled on valid and invalid(+) trials, and all five timbres were sampled on invalid(-) trials. Because the four-tone condition had a much higher probability of sampling perceptually similar feature values on any given trial, the four-tone condition was probably inherently more difficult than the two-tone conditions at the initial feature stage of processing. Therefore, the change from a two-tone to a four-tone array probably represents both an increase in attentional demand, as conjectured by FIT, and an increase in difficulty due to the higher probability of sampling similar features. An assessment of the contribution of feature similarity to the increase in error rate for the larger array size is beyond the scope of the present investigation. The sampling problem can be minimized by confining probability modeling to the two-tone conditions. In modeling performance in these conditions, any differences in feature similarity that contributed to perceptual errors will tend to be reflected in an increased probability of misperceiving a feature [i.e., on invalid(-) trials]. The resulting inflated feature error estimate, in turn, will result in a lower estimated probability of illusory conjunctions in our probability model.

In visual search performance, the probability of an illusory conjunction is an inverse function of the distance between items containing the cued features (see, e.g., Ashby et al., 1996). Thus, interitem distance is an important factor in current models of visual search performance. However, interitem distance did not affect auditory search performance in the two-tone conditions. Therefore, lateralized distance between items does not seem to be important to the binding of spatially distributed auditory features and, thus, was not explicitly coded as a factor in our model. Despite this null pattern of results, the potential influence of distance on illusory conjunctions still was evaluated by separately modeling illusory conjunction rates for the two-tone near and far conditions.

Modeling illusory conjunction rate using only data from conjunction search. Two versions of the model were developed; each version is summarized separately in the Appendix. The first version is based solely on data from the conjunction search tasks and thus does not differentiate between the misperception of pitch features and of timbre features. For each listener in each version of the conjunction search task (Y/N and rating), the probability of each type of feature error (to and from cued values) was first estimated, with these estimates then used to assess the probability of a conjunction error. The av-

Table 5
Mean Estimated Probabilities (and Corresponding Standard Errors) for Different Types of Perceptual Errors Based on Error Rates From Conjunction Search

Task	Two-Tone Near		Two-Tone Far	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Misperception of a Cued Feature as a Noncued Feature				
Y/N	.05	.02	.06	.02
Rating	.03	.01	.03	.01
Misperception of a Noncued Feature as a Cued Feature				
Y/N	.06	.05	.07	.05
Rating	.05	.02	.08	.02
Illusory Conjunctions				
Y/N	.23	.05	.26	.06
Rating	.24	.05	.25	.05

average values (and standard errors) of these probabilities are summarized in Table 5, with the top and middle sections of the table reflecting the misperception of features as noncued values and as cued values, respectively. Probabilities of feature misperception generally were low, with mean estimates not exceeding 8%. Although our modeling differentiates between the probabilities of misperceiving features as cued and noncued values, means for the two types of probabilities did not significantly differ [task (Y/N, rating) \times distance (near, far) \times type of feature misperception (as noncued value, as cued value), $F(1,9) = 1.41, p > .10$].

In contrast to the low probabilities for the misperception of features, mean illusory conjunction rates (Table 5, bottom) were greater than 23% and thus were substantial.⁹ Thus, as with spatially coherent, temporally distributed auditory events (Thompson, 1994), illusory conjunctions with temporally coherent, spatially distributed auditory events appear to be very common. Illusory conjunction rate did not significantly differ with task [task (Y/N, rating) \times distance (near, far), $F(1,9) < 1$]. Furthermore, in contrast to recent evidence from visual search tasks for the role of location in the binding of features (e.g., Ashby et al., 1996), distance between items (near vs. far conditions) did not influence illusory conjunction rate.

Modeling illusory conjunction rate using data from feature search. The modeling results based solely on data from the conjunction search tasks provide strong evidence for relatively high rates for the illusory conjunction of auditory features in spatially distributed auditory events. However, since the data on which the model is based always reflected the overall responses to the cued conjunction of pitch and timbre features, the model requires the simplifying assumption that the probabilities of feature misperception are equivalent for pitch and timbre. This assumption actually maximizes the estimated probability of misperceiving a feature and thus minimizes the estimated probability of an illusory conjunction.¹⁰

There was some evidence that the probabilities of misperceiving pitch and timbre features may not be equal. Specifically, differences in error rate were obtained for

valid trials in pitch and timbre feature search tasks. We therefore believed it worthwhile to include a second version of the model that uses feature search results to obtain a more direct evaluation of the probability of feature misperception.

In generating the alternative set of probabilities, each listener's error rates from the pitch and timbre feature search tasks first were used to separately estimate the probability of each type of feature misperception (for a derivation of the formulae, see the Appendix). The resulting probabilities, averaged across listeners, are summarized in Table 6. As was found in the conjunction search results, lateralized distance was not an important variable. Specifically, within each task, the estimated probabilities for misperceiving features in the far condition were either equal to or only slightly ($\leq 2\%$) smaller than those in the near condition ($p > .05$). Listeners also misperceived features as noncued values (Table 6, top) significantly more often than as cued values [middle; task (timbre, pitch) \times distance (near, far) \times type of feature misperception (as cued value, as noncued value), $F(1,9) = 41.26, p < .0001$].

A comparison of the estimated probabilities of feature misperception obtained from the feature search tasks (Table 6) with those obtained from the conjunction search tasks (Table 5) reveals consistently higher average estimated probabilities from the feature search tasks. This difference is relatively small for misperceptions from noncued to cued values (.09 vs. .07), but quite large for misperceptions from cued to noncued values (.17 vs. .04). An explanation of the higher estimated probabilities of feature misperception from feature search can be found through an examination of the error rates that were used to derive these estimates (Table 4). On invalid trials in the pitch search task, consistently higher error rates were obtained than in corresponding trials in the conjunction search task, yet errors in the conjunction tasks reflect the sum of feature and conjunction errors. The high error rates from feature search probably are due to the nature of the cues used to indicate the target feature. Whereas cues for the conjunction search tasks were the actual tones

Table 6
Mean Estimated Probabilities (and Corresponding Standard Errors) as a Function of Tone Condition for Different Types of Perceptual Errors Based on Error Rates From Feature Search

Task	Two-Tone Near		Two-Tone Far	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Misperception of a Cued Feature as a Noncued Feature				
Pitch	.10	.02	.10	.02
Timbre	.26	.03	.24	.03
Misperception of a Noncued Feature as a Cued Feature				
Pitch	.10	.01	.09	.02
Timbre	.09	.01	.09	.01
Illusory Conjunctions				
Y/N	.26	.07	.34	.10
Rating	.30	.10	.40	.09

that either were or were not in the subsequent array, cues to feature search were never physically identical to a feature in an array stimulus. Thus, the feature search tasks could not be performed as a simple identity match between an array feature and a memory trace of the cued target feature. In the pitch search task, the sine tone cues lacked the reinforcement of pitch provided by the upper partials of the cues to conjunction search. Furthermore, the sine tone cues occasionally differed slightly from the intended pitch of isolated tokens from the natural instruments (by a maximum of 2–3 Hz). In timbre search, the reliance on visual cues, specifically orthographically presented names of instruments, meant that listeners had to perform the task on the basis of representations of the cued instrument in long-term memory. For both types of feature search, the reduction in cue specificity relative to that in the conjunction search conditions should have increased task difficulty, resulting in elevated error rates.

There also appeared to be a general tendency to report the cued value as being present for pitch and absent for timbre. This tendency was reflected in a simple effect of type of feature misperception for timbre search [$F(1,9) = 26.11, p = .0001$], which directly contributed to a main effect of task [$F(1,9) = 10.98, p < .01$] and an interaction of task with the type of feature misperception [$F(1,9) = 38.57, p < .0001$]. This tendency could be argued to reflect either a feature-specific response bias or an increased probability on valid trials of the aforementioned problem of sampling perceptually similar timbres. In timbre search, an incorrect response on a valid trial could reflect the misperception of the cued timbre as any of four noncued values (either the noncued value that was presented or one of the three values not in the stimulus array). In contrast, an incorrect response on an invalid(–) trial could only reflect the misperception of either of the two presented timbres as the single cued value. Thus, there are twice as many ways that an incorrect response could be obtained from the misperception of a cued feature as a noncued value (on valid trials) than from the misperception of a noncued feature as a cued value [on invalid(–) trials]. This increase in the number of features that could be misperceived as a presented value to produce an incorrect response on a valid trial, relative to an invalid(–) trial, should be accompanied by an increased probability that one of the noncued timbres is perceptually similar to the cued value. As a result, the error rate on valid trials should be elevated, as was obtained. Likewise, the estimated rate of misperceiving a cued feature as a noncued value, which is derived from the error rate on valid trials, also would be elevated.

The separate probabilities for misperceiving pitch and timbre features (e.g., see the top and middle panels of Table 6) were used with the invalid(+) results from the conjunction search tasks to estimate illusory conjunction rates for each listener, with the average probability of a conjunction error summarized in Table 6 (bottom). These alternative estimates of illusory conjunction rates (averaging .33, with a range from .26 to .40) are even greater

than the estimates based solely on conjunction search performance (Table 3; averaging .24). Although the inflated probability of feature misperceptions due to less precise cuing in the feature search tasks would have reduced the estimated probability of an illusory conjunction, the unequal probabilities for misperceiving pitch and timbre features should have increased the estimated probability of an illusory conjunction. However, with significant individual differences, the mean illusory conjunction rates of Table 6 are not significantly higher than the corresponding rates of Table 3 [type of model (initial, alternative) \times task (Y/N, rating) \times distance (near, far), $F(1,9) < 1$]. Thus, across conditions, the average probability of illusory conjunctions appears to be somewhere between .23 and .40.

Interpretation of modeling estimates. A major implication for the finding of the frequent illusory conjunction of pitch and timbre is that these two classes of perceptual properties must initially be processed independently before they can subsequently be conjoined. This implication is consistent with recent demonstrations that experienced musicians can treat pitch and timbre as perceptually separable attributes (Pitt, 1994). Furthermore, when coupled with the finding of increased error rates in conjunction search tasks with increases in array size (from two to four tones, but qualified by stimulus sampling considerations), our evidence of illusory conjunctions seems at least to suggest that, as is conjectured in FIT for vision, auditory feature integration is capacity limited. This interpretation of the present findings lies in direct contrast to that of Woods et al. (1998) for auditory search in their RSP task. Woods et al. (1998) found little evidence of errors that represented potential illusory conjunctions and, thus, little evidence of a feature integration process when listeners searched for specific conjunctions of pitch and location. Also, the increased speed of processing in the RSP conjunction task, relative to an RSP “feature” search for location, was taken to indicate the parallel, rather than the serial, processing of a conjunction of (pitch and location) features, as well as the general difficulty associated with processing the location feature as an isolated property.

Two major factors probably contributed to the contrast between the Woods et al. (1998) results with sequential tones and those of the present investigation with simultaneous tones: the nature of the RSP task in relation to the concepts of FIT and the operational definition of location in the RSP task. The first factor reflects the basic difference between spatially and sequentially distributed stimuli. In Figure 1, which summarizes FIT for visual stimuli, n spatially distributed objects are presented simultaneously in a single array. Each object has a distinct value on each of f features. In Figure 1, there are two objects ($n = 2$), each with a distinct value on three features ($f = 3$). The ($2 \times 3 =$) six simultaneously available feature values are assumed to be processed in parallel and weakly associated with a location map that has two locations. Parallel processing is assumed to occur both within and across features, with rough location information im-

PLICITLY assumed to be processed in parallel with the features. The attentive conjunction of features then operates serially across the location map. FIT does not treat location as a feature but, rather, as information critical to the operation of the serial conjunction process. Assuming that the six features are correctly perceived, the conjunction process can result in either the accurate perception of the two objects or the illusory conjunction of features (i.e., a mismatch between physical and perceptual objects).

In the present study, n equals two or four events (i.e., two-tone or four-tone conditions) and f equals two (pitch and timbre) features, thus representing four or eight feature values distributed across either two or four locations. Therefore, either two or four conjunctions are required for the perception of all of the auditory events. From this same perspective, each "trial" in the RSP task consisted of a single auditory event ($n = 1$) defined by its pitch ($f = 1$) that is roughly represented in the location map that has only a single entry. With only one feature value to be processed on each trial, the concept of parallel versus serial processing of features is irrelevant. One could argue that a second (conjunction) processing stage is necessary to firmly associate the pitch feature to a position in the location map, but with only a single pairing of pitch and location, the concept of an illusory conjunction of features also is irrelevant for this processing stage. Alternatively, one could argue that feature integration occurs across sequentially adjacent tones in the RSP task and, therefore, that illusory conjunctions can occur across such tones (as either a memory or a perceptual process). In fact, Thompson (1994) had demonstrated the illusory conjunction of sequential events. However, implicit in this discussion is the fact that the initial processing of feature values across sequentially presented tones must be serial.

The second factor concerns the operational definition of location for the auditory experiment in Woods et al. (1998). In the auditory perception literature, spatial location (i.e., localization or lateralization) of auditory events is defined by the interaural disparity in time and/or intensity between the binaural stimuli. In the Woods et al. (1998) study, location is represented by (1) a monaural tone to the left ear, (2) a monaural tone to the right ear, and (3) a diotic tone (i.e., an identical pair of monaural tones in each ear). These three experimenter-defined "locations" perceptually represent (1) diffuse left, (2) diffuse right, and (3) precise central lateralization, with each location derived after initial filtering, as reflected in the concept of critical bands. If one treats each ear and critical band as a separate input channel, then the conjunction search task requires detecting the presence of energy in only one specific critical band. However, the location feature search task requires that the detection processes be performed at that location for the critical bands corresponding to each possible pitch. The location feature search task therefore has greater stimulus uncertainty than the conjunction search task, and, independent of whether the processes across critical bands occur serially or in parallel, one should expect greater processing and longer

decision times for the more difficult location feature search. Consistent with this interpretation is the possibility raised by Woods et al. (1998) that listeners may have constrained the number of attended locations in conjunction search by searching for a specific target pitch. Similar arguments have been made for visual search, where feature information can be used to guide conjunction search (e.g., Duncan & Humphreys, 1989; Ivry & Prinzmetal, 1991; Wolfe et al., 1989).

CONCLUSIONS

The present study clearly demonstrates the frequent illusory conjunction of features in spatially distributed auditory events. This finding strongly implies that, like the visual system, the auditory system first must separately analyze the features of simultaneous events and then must combine these features. The features are probably combined with reference to their perceived location, in a manner consistent with that described by Treisman and colleagues in FIT (e.g., Treisman & Gelade, 1980; Treisman & Schmidt, 1982). Our findings of illusory conjunctions thus represent direct evidence that Treisman's notion of feature integration applies to audition, as was conjectured by Deutsch (1982), who stated:

When we hear a tone, we attribute a fundamental pitch, a loudness, a timbre; and we hear the tone at a given location. Each tonal percept may therefore be described as a bundle of attribute values. If our perception is veridical, this bundle reflects the location and characteristics of the sound emitted. We shall see, however, that in situations where more than one tone is presented at a time, these bundles of attribute values may fragment and recombine in other ways, so that illusory percepts result. Perceptual grouping in music is therefore not simply a matter of linking different sets of stimuli together; rather it involves a process whereby the stimuli are fragmented into their separate attributes, followed by a process of perceptual synthesis in which the different attribute values are recombined. (p. 101)

The type of auditory illusory conjunctions described by Deutsch (1982) and empirically demonstrated in the present report seems contrary to the assumption that the perceptual system generally provides an accurate reflection of the physical signal. However, illusory conjunctions with spatially distributed events become logically possible with the notion that features of an event are independently perceived and that event recognition requires an additional process of limited capacity for their integration. Such illusory conjunctions are probably quite common. Anyone who has attended a performance by a symphony orchestra has probably experienced a mislocalization of timbre or pitch that cannot be simply and logically explained by the reverberant properties of the concert hall or room. For example, one might perceive a pitch in a melody as having been performed by a brass instrument when the target pitch was carried by strings simultaneously sounding a different pitch. Likewise, several laboratory-based phenomena reflect the mislocal-

ization of properties and thus could be interpreted as examples of the illusory integration of simultaneously presented auditory features. For speech and music, these phenomena include fusion (Cutting, 1976), duplex perception (e.g., Collins, 1985; Hall & Pastore, 1992b; Mattingly & Liberman, 1988; Pastore et al., 1983; Whalen & Liberman, 1987), and the related *streaming* phenomenon (Ciocca & Bregman, 1989). Examples specific to music include Deutsch's octave (e.g., 1974) and scale (1975) illusions.

When the numerous anecdotal and empirical examples given above are interpreted in light of our evidence for the frequent illusory conjunction of pitch and timbre, it is apparent that they probably reflect a general perceptual process of feature integration. Furthermore, when taken together with evidence for illusory conjunctions with sequentially presented auditory features (Thompson, 1994), our evidence for illusory conjunctions with simultaneous features strongly suggests that feature integration is a fundamental process for the identification of all auditory events. Our ongoing investigations thus are sensitive to the fact that feature integration can be demonstrated for audition using both methods of stimulus event presentation (i.e., sequential and simultaneous) and various types of features (duration, pitch, timbre). In this way, similarities, as well as important differences, in the perception of sequential and simultaneous auditory events will be revealed, and a more thorough understanding of the nature of auditory feature integration will be achieved.

REFERENCES

- ASHBY, F. G., PRINZMETAL, W., IVRY, R., & MADDOX, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review*, **103**, 165-192.
- BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- BROADBENT, D. E. (1958). *Perception and communication*. London: Pergamon.
- CIOCCA, V., & BREGMAN, A. S. (1989). The effects of auditory streaming on duplex perception. *Perception & Psychophysics*, **46**, 39-48.
- COLLINS, S. C. (1985). Duplex perception with musical stimuli: A further investigation. *Perception & Psychophysics*, **38**, 172-177.
- CUTTING, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, **83**, 114-140.
- DARWIN, C. J., TURVEY, M. T., & CROWDER, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, **3**, 255-267.
- DEHAENE, S. (1993). Temporal oscillations in human perception. *Psychological Science*, **4**, 264-270.
- DEUTSCH, D. (1974). An auditory illusion. *Nature*, **251**, 307-309.
- DEUTSCH, D. (1975). Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, **57**, 1156-1160.
- DEUTSCH, D. (1982). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (pp. 99-134). New York: Academic Press.
- DUNCAN, J., & HUMPHREYS, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**, 433-458.
- FLETCHER, H. (1972). *Speech and hearing in communication*. New York: Krieger.
- HALL, M. D., & PASTORE, R. E. (1992a). Effects of base complexity in musical duplex perception [Abstract]. *Journal of the Acoustical Society of America*, **91**, 2339.
- HALL, M. D., & PASTORE, R. E. (1992b). Musical duplex perception: Perception of figurally good chords with subliminal distinguishing tones. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 752-762.
- HANDEL, S. (1988a). No one analogy is sufficient: Rejoinder to Kubovy. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 321.
- HANDEL, S. (1988b). Space is to time as vision is to audition: Seductive but misleading. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 315-317.
- IVRY, R. B., & PRINZMETAL, W. (1991). Effect of feature similarity on illusory conjunctions. *Perception & Psychophysics*, **49**, 105-116.
- KOLINSKY, R. (1992). Conjunction errors as a tool for the study of perceptual processes. In J. Algeria, D. Holender, J. Junca de Moraes, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 133-149). Amsterdam: North-Holland.
- KUBOVY, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55-98). Hillsdale, NJ: Erlbaum.
- KUBOVY, M. (1988). Should we resist the seductiveness of the space:time:vision:audition analogy? *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 318-320.
- MARR, D. (1982). *Vision*. New York: Freeman.
- MATTINGLY, I. G., & LIBERMAN, A. M. (1988). *Speech and other auditory modules* (Haskins Laboratories Status Report on Speech Research No. SR-93/94). New Haven, CT: Haskins Laboratories.
- MORAY, N. (1970). *Attention*. New York: Academic Press.
- NAVON, D. (1990a). Does attention serve to integrate features? *Psychological Review*, **97**, 453-459.
- NAVON, D. (1990b). Treisman's search model does not require feature integration: Rejoinder to Treisman (1990). *Psychological Review*, **97**, 464-465.
- PASTORE, R. E., SCHMUCKLER, M. A., ROSENBLUM, L., & SZCZESIUL, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics*, **33**, 469-474.
- PITT, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 976-986.
- PRINZMETAL, W., HENDERSON, D., & IVRY, R. B. (1995). Loosening the constraints on illusory conjunctions: Assessing the roles of exposure duration and attention. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1362-1375.
- REPP, B. H. (1978a). *Stimulus dominance and ear dominance in fused dichotic speech and nonspeech stimuli* (Haskins Laboratories Status Report on Speech Research No. SR-55/56). New Haven, CT: Haskins Laboratories.
- REPP, B. H. (1978b). *Stimulus dominance in fused dichotic syllables* (Haskins Laboratories Status Report on Speech Research No. SR-55/56). New Haven, CT: Haskins Laboratories.
- THOMPSON, W. F. (1994). Sensitivity to combinations of musical parameters: Pitch with duration, and pitch pattern with durational pattern. *Perception & Psychophysics*, **56**, 363-374.
- TREISMAN, A. (1969). Strategies and models of selective attention. *Psychological Review*, **76**, 282-299.
- TREISMAN, A. (1990). Variations on the theme of feature integration: Reply to Navon (1990). *Psychological Review*, **97**, 460-463.
- TREISMAN, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, **47**, 862-875.
- TREISMAN, A., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
- TREISMAN, A., & GORMICAN, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, **95**, 15-48.
- TREISMAN, A., KAHNEMAN, D., & BURKELL, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, **33**, 527-532.
- TREISMAN, A., & SCHMIDT, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, **14**, 107-141.
- TREISMAN, A., & SOUTHER, J. (1986). Illusory words: The roles of attention and of top-down constraints in conjoining letters to form words. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 3-17.
- WHALEN, D., & LIBERMAN, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, **237**, 169-171.
- WOLFE, J. M., CAVE, K. R., & FRANZEL, S. L. (1989). Guided search:

An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 419-433.

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.

WOODS, D. L., ALHO, K., & ALGAZI, A. (1994). Stages of auditory feature conjunction: An event-related brain potential study. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 81-94.

NOTES

1. Navon (1990a, 1990b) has argued that the registration of all features of an object as occurring in one location eliminates the need for feature integration. However, as Treisman (1990) has suggested, this logic does not anticipate illusory conjunctions. The finding of illusory conjunctions would seem to indicate that features are separately perceived and, thus, that features must subsequently be integrated to enable object identification.

2. Woods, Alho, and Algazi (1994) used patterns of event-related brain potentials to argue that processing conjunctions of frequency and location begins prior to completion of an analysis of individual features.

3. In theory, sample rate must be at least twice the maximum frequency. However, lowering the upper cutoff (i.e., maximum) frequency provided a more accurate specification of the stimuli. The 4-kHz antialiasing filter may have slightly reduced the intensity of upper partials for some stimuli at the highest pitch but did not significantly alter attack and decay attributes of the timbres.

4. Mistunings for the flute and clarinet, which are generally considered to be instruments that produce a fixed set of pitches, were made possible by slight modifications in the functional length of the tube either at the mouthpiece or at the joints. Corresponding mistunings for the piano timbre were made possible by the use of a digital sampler with an electronic keyboard, which allowed for dial control of tuning. Also, since the trombonist was unable to produce the 1078-Hz tone, this pitch was obtained by doubling the sample rate (to 20 kHz) of a recorded 539-Hz tone with a 4-sec duration. Finally, because the 767-Hz trombone token varied slightly in pitch, the variable portion was excised and replaced with a copy of a portion of the steady-state with more stable pitch, thus maintaining the attack, decay, and spectral attributes of timbre while more precisely defining pitch. All cutting and pasting were done at zero crossings in the waveform. The modified trombone stimuli sounded natural and constant in pitch. The actual frequencies of the different stimuli were confirmed both by listening and by examining the average period of the digitized samples.

5. Although we focus on mean values, similar results were obtained with median error rates.

6. In the present report, an alpha level of .05 was used for all the statistical tests. Also, as a matter of convention, in instances where the results of analyses are summarized together for both conjunction search tasks, *F*-ratios will be listed first for the *Y/N* task and then for the rating task.

7. If any such errors do exist, our treating them as being due to feature processing, rather than to the conjunction of features, would reduce our estimate of the probability of conjunction errors. Further justification of this assumption is provided in the Appendix.

8. One fundamental difference between our modeling approach and current models of visual attention is that our model does not explicitly include the effects of guessing about feature values in analyzing error rates. This difference in model characteristics is primarily a consequence of the fact that the participants in the present study were not required to respond by labeling individual pitch and timbre features in the arrays. However, consideration of the role of various guessing strategies on responses for invalid(+) trials (the critical trials for observing illusory conjunctions) reveals that guessing should not adversely affect our estimates of illusory conjunction rate. Our model thus treats guessing due to the failure to register a feature as a feature error. For example, one viable strategy would be to respond *no* (correctly indicating target absence) on any trial where a participant was unsure about the presented

features. Since our model is based on error rates, this type of guessing would never be counted as a possible illusory conjunction. Alternatively, on trials where one or more features were not registered, listeners could be biased with some probability toward responding *yes*, erroneously indicating target presence. In such instances, guessing errors should be coded as misperceived features. For example, on invalid(+) trials, it is unlikely that the musically trained participants would misperceive both pitch and timbre features to match the cue when the array already contained both cued features. Even if listeners correctly guessed the identity of one or both of the cued features in the array, an illusory conjunction still would be required to erroneously match the cued tone.

9. A few of the individual estimates of illusory conjunction rate that were obtained were slightly less than 0, and one estimate was slightly greater than 1. These estimates were respectively treated as 0 (meaning that no illusory conjunctions occurred) or 1 (reflecting an illusory conjunction on every trial). The inclusion or exclusion of the original estimates did not alter the magnitude of the estimated mean conjunction rates.

10. The equations used to estimate the rate of illusory conjunctions (see Equations A2 and A3 in the Appendix) reflect both the sum and the product of the probabilities of misperceiving pitch and timbre features, with the product expressed as the square of the average probability of misperceiving a feature. If the actual probabilities of the two types of feature misperceptions are not equal (e.g., .10 and .35), the sum of the two probabilities (.45) will be unchanged, but the product of the two probabilities will be less than the square of their average (.035 vs. .063). Thus, when solving for unknowns, the assumption of equivalent error rates for pitch and timbre will overestimate the proportion of errors attributed to the misperception of features and will thereby underestimate errors from illusory conjunctions.

APPENDIX

Derivation of Formulae

Modeling Illusory Conjunction Rate Using Only Data From Conjunction Search

Each response (i.e., *yes* or *no*) in any given condition of the conjunction search task can reflect several possible combinations of perceptual events. Tables A1, A2, and A3, respectively, summarize the various combinations of perceptual events and the corresponding responses that can occur on valid, invalid(-), and invalid(+) trials. Particularly improbable combinations of perceptual events (described separately for each condition below) are excluded from the model and, thus, also from these tables. As a matter of convention, several abbreviated terms are used to represent the probabilities of particular perceptual events. The uppercase symbol *IC* represents illusory conjunction rate. The upper case symbol *FE* represents the probability of misperceiving a presented noncued feature as having a cued value. The lower case symbol *fe* represents the probability of misperceiving a presented cued feature as having a noncued value. Since our model of illusory conjunction rate is based on error rates, the description of the model will focus on those combinations of

Table A1
Combined Probabilities of Perceptual Events Contributing to Each Type of Response on Valid Trials in Conjunction Search

	Tone 1		Tone 2		Relevant Illusory Conjunction	Response
	Cued	Timbre	Cued	Timbre		
<i>fe</i>	<i>fe</i>	1	1	1	1	E
	<i>1 - fe</i>	1	1	1	1	E
<i>1 - fe</i>	<i>fe</i>	1	1	1	1	E
	<i>1 - fe</i>	1	1	1	1	C

Note—C, correct; E, error.

perceptual events (in Tables A1–A3) that result in an incorrect response. This modeling of error rates is summarized individually for each condition below.

As is depicted in Table A1, an incorrect response on a valid trial could only reflect the misperception of the cued timbre feature as a noncued value [i.e., $fe \cdot (1 - fe)$], the similar misperception of the cued pitch feature [$fe \cdot (1 - fe)$], or the misperception of both cued features (fe^2). Such responses do not depend on the nature of perception for the remaining, noncued tone; in Table A1, this is shown for Tone 2 as a combined probability of 1. Summing the probabilities of perceptual events resulting in an incorrect response thus yields the following equation, which can be solved for fe :

$$p(\text{error}|\text{valid}) = 2 \cdot fe(1 - fe) + fe^2. \quad (A1)$$

The modeling for valid trials is simplified by two assumptions about perception by musically trained listeners. First, on the array tone that matches the cue, there should be a zero probability of an illusory conjunction that separates the cued pitch and timbre. Second, correct responses should not reflect the misperception of a cued feature value as having a noncued value (on the tone that matches the cue) combined with a pairing of subsequent errors (on the remaining tone) to restore the match to the cue. The pairing of subsequent errors includes (1) the misperception of both noncued feature values as the cued pitch and timbre and (2) the combination of the misperception of the noncued feature as the cued value (to compensate for the misperception of the presented cued feature) coupled with an illusory conjunction (e.g., misperceiving the cued pitch on Tone 1, misperceiving the noncued pitch on Tone 2 as having the cued value, and then conjoining the cued values from the two tones). Based on the probability of feature and conjunction errors in Table 3, the probability of this latter sequence of events is .001 or less. If both assumptions are correct, then the only types of perceptual errors that matter will be those involving the misperception of a cued feature as a noncued value. If either (or both) of the underlying assumptions is incorrect, we will have overestimated the probability of feature misperceptions and, thus, underestimated the illusory conjunction rate.

An incorrect response on an invalid(–) trial requires the misperception of both pitch and timbre as cued values, as is de-

scribed in Table A2. If both feature misperceptions occur on a single array tone (FE^2), an incorrect response should be obtained regardless of the nature of perception for the other array tone. Perception for this other array tone is shown in Table A2 as a combined probability of 1 (when Tone 2 represents the other array tone) or as the separate probabilities for each possible combination of perceptual events, $2 \cdot FE \cdot (1 - FE)$ and $(1 - FE)^2$ (when Tone 1 represents the other array tone). The joint feature misperceptions can occur on either array tone and are treated as mutually exclusive events (i.e., $2 \cdot FE^2$). Illusory conjunctions were assumed not to occur after such joint feature misperceptions to match the cue. The model also ignores the possibility that both array tones are misperceived as matching the cue (i.e., FE^4). Given the use of well-trained musicians, the probability of such an occurrence should approach zero (based on a maximum estimated FE value of .09 from the conjunction search data, the maximum value of FE^4 is less than .0001).

Alternatively, the same misperception of pitch and timbre features as the cued values could occur on separate tones [$FE \cdot (1 - FE) \cdot FE \cdot (1 - FE)$], requiring an illusory conjunction (IC) to match the cued tone. As Table A2 indicates, the misperception of timbre (or pitch) could occur on either array tone—that is, $2 \cdot IC \cdot [FE \cdot (1 - FE)]^2$. Summing the probabilities of the perceptual events that lead to an incorrect response on invalid(–) trials results in the equation

$$p(\text{error}|\text{invalid}(-)) = 2 \cdot FE^2 + 2 \cdot IC \cdot [FE \cdot (1 - FE)]^2. \quad (A2)$$

As Table A3 depicts, incorrect responses on invalid(+) trials can occur when the cued feature on one array tone is perceived correctly and the noncued feature of the same tone is misperceived as the other cued feature. This type of feature misperception can occur on either tone in the array [$2 \cdot (1 - fe) \cdot FE$], with perception of the remaining tone not influencing the response. Perception of the remaining tone is shown in Table A3 for Tone 2 as a combined probability of 1 and for Tone 1 as the separate probabilities for each possible combination of perceptual events [$FE \cdot fe$, $(1 - FE) \cdot fe$, and $(1 - FE)(1 - fe)$]. As with invalid(–) trials, for invalid(+) trials the model ignores the very remote possibility that illusory conjunctions could occur after the pitch and the timbre of a tone were initially perceived to match the cue.

An incorrect response also can occur on an invalid(+) trial when the two cued features are perceived correctly and are erroneously conjoined; this second type of error additionally requires that the other (noncued) features not be perceived as cued values; thus, $IC \cdot [(1 - FE) \cdot (1 - fe)]^2$ (see Table A3). The model ignores the very remote possibility that an incorrect response would result from misperceiving cued features as noncued values, then misperceiving noncued features as cued values and integrating them to match the cue [i.e., $IC \cdot (FE \cdot fe)^2$ or $IC \cdot FE \cdot fe \cdot (1 - fe)(1 - FE)$]. Summing the two types of perceptual errors that produce an incorrect response on invalid(+) trials yields the equation

$$p(\text{error}|\text{invalid}(+)) = 2 \cdot (1 - fe) \cdot FE + IC \cdot [(1 - FE) \cdot (1 - fe)]^2. \quad (A3)$$

Modeling Illusory Conjunction Rate Using Data From Feature Search

In the feature search task, an incorrect response occurs on a valid trial whenever a cued feature is misperceived as a noncued value. Thus, fe should be equivalent to the error rate on valid trials. (We ignore the joint probability of misperceiving a feature as

Table A2
Combined Probabilities of Perceptual Events Contributing to Each Type of Response on Invalid(–) Trials in Conjunction Search

	Tone 1		Tone 2		Relevant Illusory Conjunction	Response
	Timbre	Pitch	Timbre	Pitch		
FE	FE	1	1	1	1	E
		FE	FE	FE	1	E
	$1 - FE$	FE	$1 - FE$	$1 - FE$	1	C
		$1 - FE$	FE	$1 - FE$	IC	E
$1 - FE$	FE	1	1	1	1	C
		FE	$1 - FE$	$1 - FE$	IC	E
	$1 - FE$	FE	$1 - FE$	$1 - FE$	1	C
		$1 - FE$	FE	$1 - FE$	1	C

Note—C, correct; E, error.

Table A3
Combined Probabilities of Perceptual Events
Contributing to Each Type of Response on
Invalid(+) Trials in Conjunction Search

Tone 1		Tone 2		Relevant Illusory Conjunction	Response	
Timbre	Cued Pitch	Cued Timbre	Pitch			
<i>FE</i>	<i>fe</i>	<i>fe</i>	<i>FE</i>	1	C	
			$1 - FE$	1	C	
	$1 - fe$	1	<i>FE</i>	1	E	
			$1 - FE$	1	C	
$1 - FE$	<i>fe</i>	<i>fe</i>	<i>FE</i>	1	C	
			$1 - FE$	1	C	
		$1 - fe$	1	<i>FE</i>	1	E
				$1 - FE$	1	C
	$1 - fe$	<i>fe</i>	<i>FE</i>	1	C	
			$1 - FE$	1	C	
		$1 - fe$	1	<i>FE</i>	1	E
				$1 - FE$	1	C
			<i>IC</i>	1	E	
			$1 - IC$	1	C	

Note—C, correct; E, error.

the cued value that corrects for the misperception of the cued value as a noncued value; using the estimated rate of feature misperception from Table 3, we estimate that this joint probability is .006 or less.)

Incorrect responses on invalid(-) trials reflect the misperception of a feature as the cued value (*FE*). This could occur on either array tone, resulting in

$$p(\text{error}|\text{invalid}(-)) = 2 \cdot FE. \tag{A4}$$

Illusory conjunction rate then is reevaluated by substituting the individual estimates of the probabilities of misperceiving a pitch and timbre feature in an expanded version of Equation A3. This expanded equation reflects separate probabilities for misperceiving pitch and timbre features. This is shown as Equation A5,

$$IC = \{p(\text{error}|\text{invalid}(+)) - [FE_p \cdot (1 - fe_t) + FE_t \cdot (1 - fe_p)]\} / [(1 - fe_p) \cdot (1 - FE_p) \cdot (1 - fe_t) \cdot (1 - FE_t)], \tag{A5}$$

where the subscripts p and t indicate that the rates were for pitch and timbre, respectively.

(Manuscript received June 4, 1996;
 revision accepted for publication September 3, 1999.)