Dissociation of perceptual judgments of "what" and "where" in an ambiguous auditory scene

Andrew H. Schwartz and Barbara G. Shinn-Cunningham[a)]

Hearing Research Center, Boston University, Boston, Massachusetts 02215

Whenever an acoustic scene contains a mixture of sources, listeners must segregate the mixture in order to compute source content and/or location. Some past studies have explored whether perceived location depends on which sound elements are perceived within a source. However, no direct comparisons have been made of "what" and "where" judgments for the same sound mixtures using the same listeners. The current study tested if the sound elements making up an auditory object predict that object's perceived location. Listeners were presented with an auditory scene containing competing "target" and "captor" sources, each of which could logically contain a "promiscuous" tone complex. In separate blocks, the same listeners matched the perceived spectro-temporal content ("what") and location ("where") of the target. Generally, as the captor intensity decreased, the promiscuous complex contributed more to both what and where judgments of the target. However judgments did not agree either quantitatively or qualitatively. For some subjects, the promiscuous complex consistently contributed more to the spectro-temporal content of the target than to its location while for some it consistently contributed more to target location. These results show a dissociation between the perceived spectro-temporal content of an auditory object and where that object is perceived. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3495942]

I. INTRODUCTION

Listeners in everyday settings, from a bird pinpointing a familiar call to a human hearing his or her name across the room at a cocktail party, are constantly distinguishing between multiple sound sources that overlap in time and frequency. Many acoustic cues promote "grouping" different components of a sound together (Bregman, 1990), encouraging the listener to perceive these elements as belonging to the same source. These cues can include monaural cues (such as co-modulation, frequency continuity, and harmonicity) as well as binaural cues (interaural time and level differences, known as ITDs and ILDs, respectively).

A complex listening situation, such as a noisy cocktail party, presents a problem for determining the location of a target sound source. For most sound sources in simple settings (i.e., when only that source is present), this task is relatively straightforward and can be robustly accomplished using a combination of ITD, ILD, and other spectral cues. For instance, in determining the lateral position of a single broadband sound source, the binaural system integrates such spatial cues across frequency (Buell and Hafter, 1991; Dye, 1990; Hill and Darwin, 1996; Trahiotis and Stern, 1989). However, when multiple sources from different locations overlap in time and frequency, the auditory system must somehow tease apart the spatial cues due to the target from those due to competing sources in order to accurately estimate a target's location. The question we pose here is how this separation of spatial cues between competing sources

and integration of cues across frequency is performed in complex auditory scenes with multiple competing sources.

One plausible hypothesis, which we refer to as the "consistent-object" hypothesis, posits that the auditory system first analyzes available grouping cues (both monaural and binaural) in an auditory scene to determine the current spectral components of the target source. The system then integrates the spatial cues across the estimated target frequencies to produce an estimate of the target's location (e.g., Best *et al.*, 2007; Darwin and Hukin, 1999; Hill and Darwin, 1996).

Of course, sound elements are not allocated all-or-none to a particular object in a complex scene (e.g., Darwin, 1995; Shinn-Cunningham *et al.*, 2007; Warren *et al.*, 1972). Instead, the *degree* to which a sound element contributes to the spectro-temporal content of a perceived auditory object is affected by the combination of available cues in a sound mixture. Thus, the consistent-object hypothesis posits that the degree to which a sound object contributes to an auditory object's perceived spectral content should predict the degree to which it contributes to its perceived location. Recent experiments that manipulated the monaural grouping cues in an auditory scene between a high-frequency target and low-frequency interferer qualitatively support the consistent-object hypothesis (Best *et al.*, 2007). In conditions that encouraged grouping between the interferer and target, subjects tended to combine the spatial cues of both target and interferer to estimate the target's location. In conditions that promoted segregation of the two objects, this interference was greatly reduced.

Other recent work, however, calls the consistent-object hypothesis into question. Specifically, sound mixtures with ambiguous grouping cues, in which sound elements could

---

[a)]Author to whom correspondence should be addressed. Electronic mail: shinn@cns.bu.edu

logically be part of either of two competing objects, can show an apparent disconnect between the degree to which an interfering tone contributes to an auditory object's perceived spectro-temporal content (Shinn-Cunningham *et al.*, 2007) versus the object's perceived spatial location (Lee *et al.*, 2009). These experiments employed a synthetic vowel and a stream of "captor" tones designed to promote segregation of the third harmonic from the vowel (the "ambiguous tone"). The vowel, captors, and ambiguous tone were played with various simulated source locations, and subjects were asked to judge either the vowel's spectrum or location. In the absence of the captors, the ambiguous tone contributed strongly to the perceived spectrum of the vowel, even when the ambiguous tone and the other vowel components had different spatial cues. Estimates of the vowel location in this situation generally fell in between the reported locations of the ambiguous tone alone and the vowel alone (without the ambiguous third harmonic). This result is consistent with the idea that localization is determined by integrating cues across the sound components that are perceived as making up the target object. In trials with the captor stream, listeners heard the ambiguous tone as contributing only weakly, if at all, to the spectrum of the vowel. However, despite this large change in the spectral judgment of the vowel due to the presence of the captors, the captors had a very small influence on the perceived location, which was always strongly affected by the spatial cues in the ambiguous tone.

Although these results appear to contradict the consistent-object hypothesis, it is still possible that the hypothesis holds in these conditions. For instance, a sound element that only contributes a small amount to the perceived spectrum of an object could still greatly influence the perceived location of the object. Thus, the consistent-object hypothesis could explain these results if the perceived location of the vowel is influenced strongly by the presence of a component sound element that has a relatively low intensity.

Here we explicitly test the consistent-object hypothesis in two-object mixtures with ambiguous grouping cues by quantifying the level of contribution of a "promiscuous" tone complex (which could logically be grouped into either of two competing auditory objects) to both the perceived spectro-temporal content and the perceived location of a target tone complex (one of the two competing objects). We used a repeating captor tone complex, similar to the captor stream in previous experiments (Lee *et al.*, 2009; Shinn-Cunningham *et al.*, 2007) to promote segregation of the promiscuous complex from the target complex. We then manipulated the level of the captor stream and promiscuous complex to change the degree to which the promiscuous complex contributed to the perception of the target's spectral content and perceived location.

First, to understand how the level of the promiscuous complex influenced the perceived spectral content or location of the target complex, we asked listeners, in separate sessions, to match either the spectral content or location of the target complex for different intensities of the promiscuous complex in trials where the captors were absent ("no-captor" trials). We then had subjects perform the same perceptual judgments in trials with the captors present ("ambiguous-

mixture" trials). We used data from the no-captor trials to estimate the "effective level" of the promiscuous complex contributing to subjects' matches in the ambiguous mixtures. By combining the results from the no-captor trials with the results from the ambiguous mixtures, we directly tested the consistent-object hypothesis. We found that the contribution of the promiscuous complex to the target's spectral content did not quantitatively predict the level of contribution to its perceived location, contradicting the consistent-object hypothesis.

II. METHODS

In separate tasks, subjects were trained to match either the spectral content/timbre ("what" task) or the location ("where" task) of a repeating target tone complex. Each trial repeatedly alternated back and forth between the stimulus to be matched (the target) and a "match" stimulus. Both the target and match stimuli lasted three seconds each time they were played; they alternated repeatedly, allowing subjects to compare them back to back. During each presentation of the match stimulus, listeners could adjust either the match stimulus' spectral content (in the "what" task) or its laterality (in the "where" task) to perceptually match the corresponding attribute of the target, as described below. There was no limit on the number of target/match presentations per trial; when satisfied with the match, the subject ended the trial by pressing a button.

All tone complexes were 75 ms in duration with 6 ms linear on/off ramps. The target had a fundamental frequency of 110 Hz and had harmonics as described in Fig. 1(A) (solid
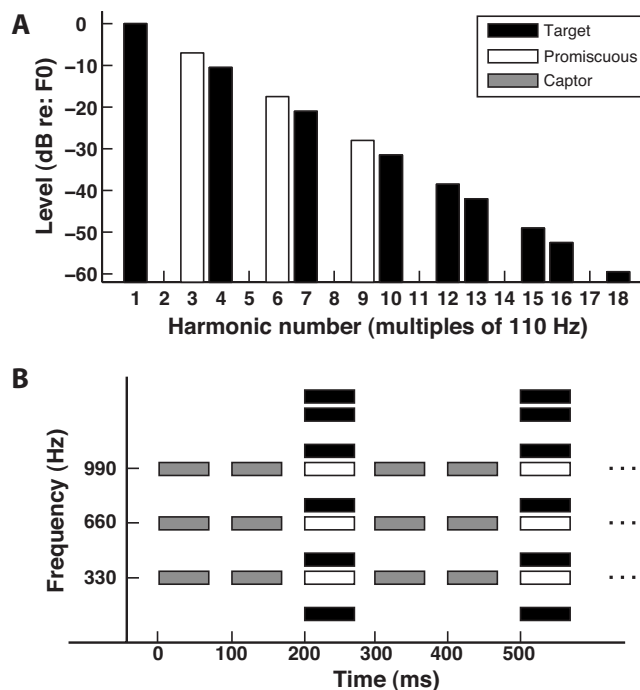

FIG. 1. Stimulus setup. (A) Frequency distribution of target and promiscuous/captor tone complexes. (B) Temporal sequence of stimuli during the presentation period. The promiscuous complex could be grouped with the captors to form an isochronous tone-complex sequence and/or with the harmonically related target complex, thereby adding to the target-complex' spectral content.

Schwartz and Shinn-Cunningham: Dissociation of "what" and "where"

bars). The target was repeated every 300 ms during the target presentation portion of a trial (corresponding to 10 repetitions). A second "promiscuous" tone complex [Fig. 1(A), open bars] had a fundamental frequency of 330 Hz (the third harmonic of the target) and was presented simultaneously with the target. In ambiguous-mixture trials, two "captor" complexes, identical to the promiscuous complex, were presented sequentially prior to each presentation of the promiscuous complex and target. These captor complexes were separated by 100 ms, creating an isochronous sequence of the 330-Hz-fundamental complex [Fig. 1(B)]. Thus, the promiscuous complex could be perceptually grouped with either the target or the captors, or possibly with both or with neither (Shinn-Cunningham *et al.*, 2007). In no-captor conditions, only the target and simultaneous promiscuous complex were present, forming a single harmonic complex with a fundamental frequency of 110 Hz.

Stimuli were played on commercially available hardware (Tucker Davis Technologies, Alachua, FL) and delivered to subjects via Etymonic ER-1 insert earphones. The level of the combined target+promiscuous tone complex was roughly 73 dB SPL prior to any attenuation of the promiscuous tone complex. Thirteen subjects participated in both the main "what" task and the "where" task. Six of these thirteen subjects (selected based on availability, rather than any other criterion) participating in a follow-up control task to ensure that "what" matches were based on perceived spectral content rather than loudness. All subjects, ranging in age from 18–30, had clinically normal hearing. Subjects gave written consent, as overseen by the Boston University Charles River Campus IRB, and were compensated $10/h for their participation.

A. "What" task

In the "what" task, the match stimulus was a diotic tone complex consisting of the target and promiscuous tone complexes. To reset any buildup of streaming between the test and match tone complexes, both the match and the presentation period were followed by a white noise burst at roughly 60 dB SPL whose length varied randomly from 600–1000 ms. This noise burst was preceded by a pause whose length varied randomly from 400 to 700 ms. The noise burst presented after the match stimulus had a 50 ms silent gap starting at 200 ms to help subjects avoid confusing the test and match stimuli. Subjects could control the spectral content of the match stimulus by pressing two buttons, one that raised and one that lowered the level of the promiscuous complex within the match stimulus (similar to the procedure used in Lee *et al.*, 2008). The changes were made in real time, with a sufficiently small increment size that the spectrum of the match stimulus changed continuously and smoothly while a button was depressed. The level of the promiscuous tone within the match stimulus was restricted to levels between +2 and −20 dB relative to that depicted in Fig. 1(A).

Prior to gathering the actual data, all subjects underwent training. In training, all stimuli were presented diotically and only no-captor trials were used. In each trial in a training run, the level of the promiscuous complex in the test stimulus

was set randomly to be 0, −3, −6, or −12 dB relative to the level depicted in Fig. 1(A). This meant that listeners could match the spectral content of the 110 Hz tone complex formed by the target and promiscuous tone complexes exactly with the proper adjustment of the promiscuous tone level in the match stimulus. Training was conducted in 40-trial-long runs consisting of five repetitions of each of the four diotic target stimuli. Typically, each training run lasted 20–30 min. At the start of each experimental session, subjects trained until they completed one training run in which the RMS difference between the attenuation of the promiscuous complex in the test and match stimuli was less than 3.5 dB. On the initial day of the "what" task, most subjects met the training criterion within three training runs. Subjects who completed training quickly enough occasionally performed the main "what" task the same day, while most subjects came back on a subsequent day to perform the main experiment. Each day that a listener came back, they were re-tested and, if needed, re-trained to meet the criterion afresh. Most subjects met the training criterion in their first run of the second session. This improvement from the initial day to the subsequent day shows that listeners were relatively stable in how they performed the spectral matching task that we asked them to undertake, and that the training was effective across days.

During the main "what" task trials, subjects were given both no-captor trials and ambiguous trials. Subjects were informed that trials in the main task should sound essentially identical to those heard in training, except that half of the trials would have an additional, higher-pitched, faster-repeating series of tones (the captors) in addition to the target. Subjects were instructed to ignore these added tones, and to perform the same task as in the training sessions, matching the timbre or brightness of the slower-repeating target complex. The level of the promiscuous complex in the test stimulus was set randomly for each trial to be 0, −3, −6, or −12 dB relative to the level depicted in Fig. 1(A). In ambiguous mixtures, the level of the captors always equaled the level of the promiscuous complex. The target was given a 600 μs, right-leading ITD, while the captors and promiscuous complex had 0 μs ITD.

Data from the no-captor trials determined the relationship between the relative level of the promiscuous tones and their perceptual contribution to the target's spectral content in an "unambiguous mixture." We used these data to interpret responses in the ambiguous-mixture trials, determining the "effective level" of the promiscuous complex in the perceived spectral content of the target in such a mixture. If the captors promoted the segregation of the promiscuous complex from the target, we expected subjects' matches in the ambiguous-mixture trials to yield a greater attenuation of the promiscuous tones than in the no-captor trials.

Subjects completed all of the trials for the main "what" task in a single test session (not including training runs) that lasted 1.5 h, on average. Each test session included eight repetitions of each kind of trial, including all combinations of trial type (ambiguous-mixture or no-captor) and promis-

cuous tone attenuations (four values), for a total of 64 trials per session. The 64 trials in each session were randomly ordered, separately for each subject.

B. "Where" task

For the "where" task, the match stimulus was an acoustic ILD pointer (Bernstein and Trahiotis, 1985; Trahiotis and Stern, 1989; Buell and Hafter, 1991; Best *et al.*, 2007) made up of a 200-Hz-wide band of noise centered at 2 kHz, played at a level of roughly 55 dB SPL. Subjects could change the ILD of the pointer during presentation of the match stimulus by pressing two buttons (the same as those used in the "what" task) to increase or decrease the ILD between ± 30 dB. As in the "what" task, ILD changes were performed in real time with a sufficiently small step size that the ILD of the pointer changed continuously and smoothly as long as a button was depressed.

Subjects were trained to use the ILD pointer before the main data collection began. During training, the test stimulus was a 200-Hz-wide band of noise centered at 2 kHz with a 2 Hz, 50%-depth sinusoidal modulation envelope that distinguished it from the match stimulus. We used this noise stimulus rather than the tone complexes used in the main experiment (which contained ITDs) so that we could objectively verify that subjects could reliably match the stimulus ILD with the pointer ILD. The test stimulus was given a fixed ILD selected from a uniform distribution between ± 20 dB. Subjects were asked to match the lateral location of the test noise stimulus by adjusting the ILD of the match stimulus. Training was conducted in 40-trial-long runs, each of which typically lasted 20–30 min. At the start of each experimental session, subjects trained until they completed one training run in which the RMS difference between the ILDs of the test and match stimuli was less than 4 dB. Subjects typically met this criterion after 2–3 training runs on the initial day of "where" testing. Subjects who completed training quickly enough occasionally performed one "where" session on the initial day, while most subjects came back on a subsequent day to begin the main experiment. Each day that a listener came back, they were re-tested and, if needed, re-trained to meet the criterion afresh.

Each subject completed the main "where" task in three separate experimental sessions, typically performed on separate days. The first of the main "where" sessions consisted only of no-captor trials. The promiscuous complex was either absent ($-\infty$ dB attention, or "target-only" trials) or was played at 0, -3, -6, or -12 dB relative to the level depicted in Fig. 1(A). The target had a 600 μs, right-leading ITD, and the promiscuous complex had an ITD of 0 μs. Subjects were instructed to match the perceived laterality of the target by adjusting the ILD of the response pointer, as they had done in training. Data from this no-captor session quantified how a promiscuous complex of different intensities altered the perceived location of the target-plus-promiscuous-complex object in an "unambiguous mixture."

In the second session, ambiguous mixtures (with captors present) were played. The same promiscuous complex attenuation conditions were used; just as in the "what" task,

the level of the captors always equaled the level of the promiscuous complex. The spatial configurations also were identical to those used in the "what" task (the captors and promiscuous complex had ITDs of 0 μs, while the target had a 600 μs, right-leading ITD). As in the "what" task, subjects were informed that some trials would contain higher-pitched, faster-repeating captor tones; they were instructed to ignore these and simple match the location of the slower-repeating target. If the target and promiscuous complexes were grouped together, the consistent-object hypothesis predicts that the perceived location of the target-plus-promiscuous-complex object should depend on the integration of the ITD of the target (0 μs) and promiscuous (600 μs) complexes, resulting in perceived locations closer to midline than when the target was presented alone (e.g., Best *et al.*, 2007; Buell and Hafter, 1991; Dye, 1990; Stern *et al.*, 1988). Conversely, if the target and promiscuous complexes were not perceived as part of the same object, the consistent-object hypothesis predicts that the matched location should be close to that of the target alone (600 μs), or perhaps displaced even further laterally due to "repulsion" by the diotic captor stream (e.g., see Lee *et al.*, 2009).

Results from the first experimental "where" session were later used to interpret the results of the second experimental "where" session. Specifically, the relationship between the promiscuous complex level in the no-captor trials and the corresponding perceived target location (from session one) determined the "effective level" of the promiscuous complex in the ambiguous-mixture trials of session two. The consistent-object hypothesis posits that this computed "effective level" should equal the effective level of the promiscuous complex in the perception of the target's spectral content, determined in the "what" task.

In the third session, only the captors (0 μs ITD) and target (600 μs ITD) were presented, with no promiscuous complex. The level of the captors was set to the same levels used in the previous two sessions, with the level randomly chosen from trial to trial. Data from this session measured spatial interactions between the captors and the target, without any influence of the promiscuous complex, allowing us to quantify any spatial repulsion (Lee *et al.*, 2009).

As in the "what' task, subjects performed eight repetitions of each kind of trial for each captor/promiscuous complex attenuation condition (including the infinite-attenuation/ target-only condition, which was not included in the "what" task). Sixteen additional control trials were presented in each session in which the target (normally at 600 μs ITD) was diotic: eight of these trials used a 0 dB captor/promiscuous complex attenuation, while the remaining eight had no captor or promiscuous complex. These control trials served to define the subject's "midline" response, which might vary from subject to subject, or even from session to session for a particular subject (Bernstein and Trahiotis, 1985; Lee *et al.*, 2009).

C. Control task for "what" judgments

Although we instructed subjects to match the perceived timbre/spectral content of the target in the "what" task, it is

Schwartz and Shinn-Cunningham: Dissociation of "what" and "where"
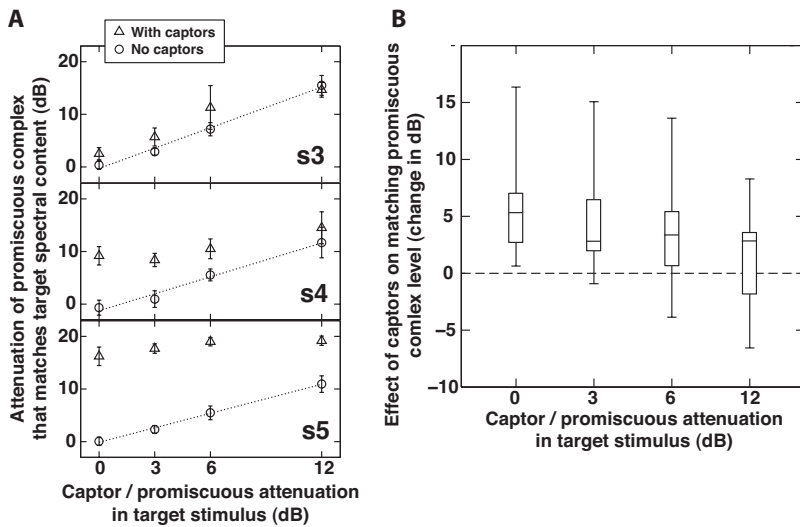
FIG. 2. Results of main "what" task. (A) Individual results showing the mean attenuation of the promiscuous complex that matched the perceived spectral content of the target stimuli. In each panel, the *x*-axis shows the attenuation of the promiscuous complex present in the target stimulus. Results are shown for no-captor trials (circles) and ambiguous-mixture trials (triangles). Error bars represent the 95% confidence intervals of the matches assuming responses are normally distributed for a given stimulus condition. Dotted lines show the linear least-squares regressions fit to no-captor data (circles), used to estimate the effective attenuation of the promiscuous complex in "what" judgments for the ambiguous-mixture trials (triangles). Results are shown for three example listeners selected to show the range of responses observed. (B) Distribution of the mean change in the matching promiscuous complex level for conditions with captors relative to the corresponding no-captor conditions from the 11 subjects who successfully completed the study. Plots show the median, inter-quartile range (boxes) and full range (whiskers) as a function of the attenuation of the promiscuous complex present in the target stimulus.

possible that they instead matched some other target attribute. Specifically, listeners may have adjusted the level of the promiscuous complex in the match stimulus to equate the perceived loudness of the target and match stimuli, rather than spectral shape.

To test for this possibility, six of the original subjects participated in a control task. This experiment was similar to the main "what" task, except that (1) for brevity, we presented target stimuli with only two attenuations (0 and 3 dB) of the captors and promiscuous complex, and (2) in half of the trials, the subject controlled the attenuation of the target complex (within the match stimulus), rather than the attenuation of the promiscuous complex (leaving the promiscuous complex level in the match stimulus unchanged). To encourage listeners to use the same strategy, regardless of whether they adjusted the promiscuous complex or the target complex, we intermingled trials randomly. A total of eight different trial types were used: all combinations of two target-stimulus attenuations (either 0 or 3 dB), two stimulus elements adjusted by subjects during the matching task (either target complex or promiscuous complex), and two types of target stimuli (captor either present or absent). Subjects performed six matches for each trial type, for a total of 48 trials per subject. Each subject completed these trials in a single, brief (less than 1 h long) session, following completion of both the main "what" task and the "where" task. They were instructed to perform this task just as they had performed the main "what" task.

If listeners were matching target loudness and if the presence of the captors decreased the target loudness (for instance), then the presence of the captors would cause listeners to attenuate whichever components they controlled in the match stimulus, whether they could adjust the promiscuous-complex level or the target-complex level. However, if they were matching perceived target timbre, *opposite* adjustments would be needed in the two matching paradigms. Specifically, if captors decreased the perceptual contribution of the promiscuous complex to the target and listeners matched the target spectral content as instructed,

they would attenuate the match stimulus' promiscuous complex if they controlled its level, but they would increase the match stimulus' target complex if they controlled its level. Moreover, such a pattern of adjustment would change the overall loudness of the match stimulus in opposite directions in the adjust-promiscuous-complex and adjust-target-complex trial. Therefore, if this pattern is observed, it rules out the possibility that listeners matched loudness.

III. RESULTS

A. "What" task

Subjects generally set the intensity of the promiscuous complex in the match stimulus in the no-captor trials close to the physical intensity of the promiscuous complex in the test stimulus [Fig. 2(A), circles], even though the promiscuous complex and target had different ITDs. The mean difference between subjects' responses and the true, physical attenuation of the promiscuous complex was 1.4 dB, with a standard deviation of 2.6 dB. These results confirm that subjects were able to reliably match the spectral content of the target-plus-promiscuous-complex object using our procedures. Moreover, these results are consistent with results from past studies suggesting that listeners do not segregate sound elements on the basis of ITD alone (Culling and Summerfield, 1995; Darwin and Hukin, 1999).

These data were fit by a least-squares linear regression [Fig. 2(A), dotted lines] representing the relation between the true physical level of the promiscuous complex (at 0 μs ITD) and the perceived level of the promiscuous complex grouped with the target complex (at 600 μs ITD). To ensure response reliability, subjects whose least-squares fit yielded an RMS error of over 7 dB (twice the training criterion) were excluded from analysis (two out of 13 subjects were excluded for not meeting this criterion). Note that this RMS error refers only to the prediction error between the linear fit and the underlying data, and does not make the explicit assumption that subjects will "correctly" match the true promiscuous complex intensity, which differed in ITD from
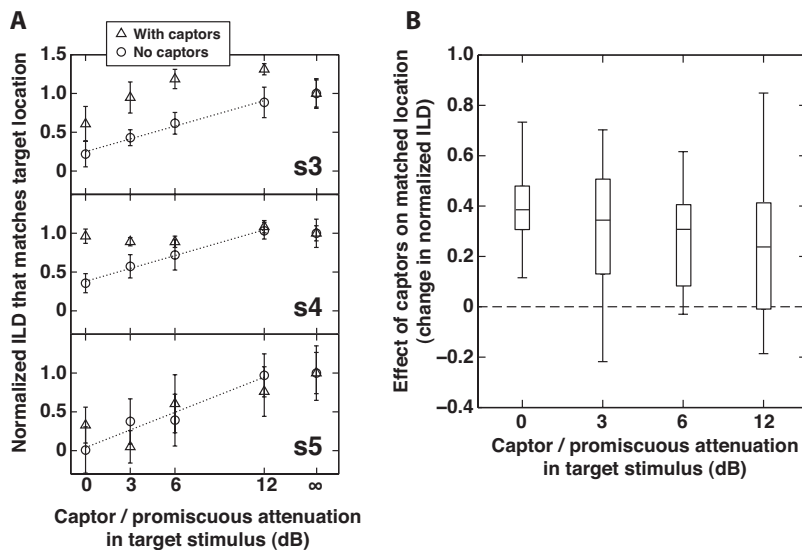
FIG. 3. Results of the "where" task for no-captor and ambiguous-mixture sessions. (A) Individual results showing the mean, normalized ILD that matched the perceived location of the target stimuli. In each panel, the x axis shows the attenuation of the promiscuous complex present in the target stimulus. Results are normalized for ease of comparison across subjects: zero represents a given subject's mean ILD match for target stimuli that are diotic (presumably heard to the center of the head); one corresponds to the ILD used to match the location of a target presented in isolation with a 600 μs (the ∞ dB attenuation condition, shown on the right of each panel). Results are shown for no-captor trials (circles) and ambiguous-mixture trials (triangles). Error bars represent the 95% confidence intervals of the matches assuming responses are normally distributed for a given stimulus condition. Dotted lines show the linear least-squares regressions fit to no-captor data (circles), used to estimate the effective attenuation of the promiscuous complex in "where" judgments for the ambiguous-mixture trials (triangles). Results are shown for the same three example listeners presented in Fig. 2(A). (B) Distribution of the mean change in the normalized ILD that matches the target for conditions with captors relative to the corresponding no-captor conditions from the 11 subjects who successfully completed the study. Plots show the median, inter-quartile range (boxes) and full range (whiskers) as a function of the attenuation of the promiscuous complex present in the target stimulus.

tones in the target complex. However, it was generally true that the perceived contribution of the (straight ahead) promiscuous complex to the composite target-plus-promiscuous-complex object was very close to the full intensity of the promiscuous complex.

Subjects generally set the level of the promiscuous complex in the match stimulus to be lower in the with-captor trials than in the no-captor trials [in Fig. 2(A), triangles lie above circles]. This result is consistent with the captors reducing the contribution of the promiscuous complex to the target. However, the degree of this reduction was inconsistent across subjects, demonstrated in the three panels of Fig. 2(A). For s3, the captors had little effect on the perceived spectral content of the target (circles and triangles fall almost on top of each other in the top panel). For s5, the presence of the captors reduced the promiscuous complex contribution by roughly 15 dB (triangles fall well above the circles in the bottom panel). For s4, the captors had a modest effect (triangles fall above the circles in the middle panel).

To summarize results across subjects, we quantified the effect of the captors on the perceived spectral content of the target by computing the difference between the response in the presence of the captors [triangles in Fig. 2(A)] and in their absence [circles in Fig. 2(A)]. Full and inter-quartile ranges of this difference are plotted in Fig. 2(B) for the subjects who responded consistently (i.e., who passed the RMS criterion described above). Overall, the captors tended to reduce the contribution of the promiscuous complex to the target [values in Fig. 2(B) generally tend to be positive or near zero]. The effect tended to decrease as the level of the captors and promiscuous complex decreased. It is worth noting that this decrease may in part be due to the limited response range available (subjects could set the promiscuous complex level in the match stimuli between +2 and −20 dB).

For the 6 and 12 dB conditions, the upper range of the response attenuations reaches the maximum possible value. In all conditions there was large inter-subject variability in the effect of the captors. Yet despite this variability the individual subjects were very reliable in their responses, with some subjects showing large, consistent shifts in the perceived intensity of the promiscuous complex contributing to the target's spectral content due to the captors [e.g., s5 in Fig. 2(A)].

B. "Where" task

To account for possible shifts between sessions in subjects' maps from ITD and ILD to perceptual space and to enable more direct comparisons across subjects (who may not use the ILD scale identically; e.g., Bernstein and Trahiotis, 1985; Best et al., 2007), responses in the "where" task were shifted and normalized within each session so that zero represents the mean ILD response to the diotic control trials and one represents the mean ILD match to the 600 μs ITD target alone (the largest ITD used).

Figure 3(A) plots the normalized laterality of the target as a function of the promiscuous-complex attenuation for the same three subjects whose "what" task results are shown in Fig. 2(A). The results for the no-captor trials [circles in Fig. 3(A)] show that increasing the level of the diotic promiscuous complex increases its influence on the perceived location of the target (i.e., responses are closer to midline/zero for higher promiscuous-complex intensities). This is consistent with the idea that spatial cues are integrated across sound elements that make up an auditory object, weighted by relative intensity.

The data describing the relation between the physical attenuation of the promiscuous complex and the perceived
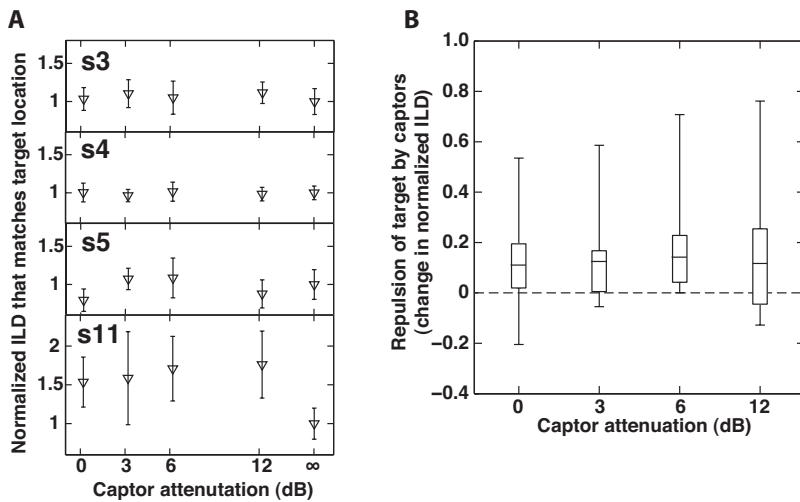
FIG. 4. Results of the "where" task for the no-promiscuous-tones session used to estimate across-object localization repulsion. (A) Individual results plotted as in Fig. 3(A) for the three example subjects shown in Figs. 2(A) and 3(A), as well as one additional subject who had even greater repulsion (bottom panel). Values exceeding 1.0 correspond to cases in which the captors caused the perceived target location to be repelled and heard farther off center than a target-alone stimulus with a 600 μs ITD. Error bars represent the 95% confidence intervals of the matches assuming responses are normally distributed for a given stimulus condition. (B) Distribution of the mean normalized ILD, relative to the reference value of 1.0 (the match for an isolated target complex with an ITD of 600 μs) estimating the "repulsion" of the target complex by the captors from the 11 subjects who successfully completed the study. Plots show the median, interquartile range (boxes) and full range (whiskers) as a function of the attenuation of the promiscuous complex present in the target stimulus.

target location data are well summarized by a straight line [e.g., see the correspondence of dotted lines and circles in Fig. 3(A)], justifying the use of a least-squares linear regression to fit these responses. The resulting linear relationship maps the physical attenuation of the promiscuous complex to a perceived location of the target auditory object.

The presence of the captors reduced the influence of the promiscuous complex on the perceived location of the target (see also Best *et al.*, 2007; Lee *et al.*, 2009); the perceived locations tend to be more lateral with the captors than without [triangles in Fig. 3(A) tend to fall above the circles]. However, the size of this effect differs across subjects. For s5 (bottom panel), the captors have a negligible influence on the perceived target location, while this effect is moderate for s3 and s4 (top two panels). Moreover, as discussed at greater length below, these individual differences are not predicted by the individual differences in the "what" task.

As in the "what" task, we quantified the effect of the captors by computing the difference between the response in the presence of the captors [Fig. 3(A), triangles] and in their absence [Fig. 3(A), circles]. The across-subject full and inter-quartile ranges of this difference are plotted in Fig. 3(B). Values above zero indicate that the addition of the captors caused the target to be perceived more laterally than in the no-captor condition (i.e., to be less influenced by the 0 μs ITD of the promiscuous complex). Despite the fact that inter-subject differences were large, Fig. 3(B) shows that most subjects perceived the target as more lateral when the captors were present compared to when the captors were absent (there was typically a positive shift in perceived location due to the captors).

In addition to affecting the influence of the promiscuous tones on the target, the captors also had some influence on the perceived location of the target itself. To address this influence, we included a session in which subjects matched the target location when the target and captors were present, but the promiscuous complex was absent. In these trials, subjects tended to perceive the target further to the side when the captors were present than when the target was played alone (see also Best *et al.*, 2005; Braasch and Hartung, 2002; Lee *et al.*, 2009; Lorenzi *et al.*, 1999). This resulted in nor-

malized localization responses greater than one, demonstrating "repulsion" between the captors and the target (see, for example, Lee *et al.*, 2009).

Figure 4(A) shows data for these no-promiscuous-complex trials for the same three subjects as in Figs. 2(A) and 3(A), as well as for one additional subject who showed an even stronger repulsion effect than any of the other three example subjects. As with the results for "what" and "where," these measures reveal large subject differences. Repulsion was negligible for s3, s4, and s5 [triangles fall near one in the top three panels of Fig. 4(A)]; however, it was significant for s11 when the captors were present, regardless of their level [triangles are above one in the bottom panel of Fig. 4(A), except when the captor had infinite attenuation, in the far right of the panel]. We quantified the repulsion for each stimulus condition by computing the difference between responses in the presence of the captors and responses to the lateral target alone (whose mean value was always one due to normalization). Full and inter-quartile ranges of this difference are shown in Fig. 4(B). These values tend to be positive but small; however, as with the other results, some subjects showed strong, consistent effects (e.g., s11). Interestingly, the magnitude of repulsion was not significantly affected by the captor stream attenuation for the conditions tested.

C. Comparing "what" and "where" judgments

The linear regressions from the no-captor trials provide us with individualized maps that summarize the effect of a promiscuous complex of a given physical level on the target's perceived timbre or location. The inverses of these functions, where defined (i.e., over the range of responses observed in the no-captor trials), map the reported perceptual attribute of the target object (perceived timbre or location) to an "effective level" of the promiscuous complex. This effective level equals the intensity that the promiscuous complex had to be in the no-captor condition to produce equivalent spectral or spatial judgments of the target. Specifically, we used the linear fits to the no-captor results to compute the effective level of the promiscuous complex in spectral or
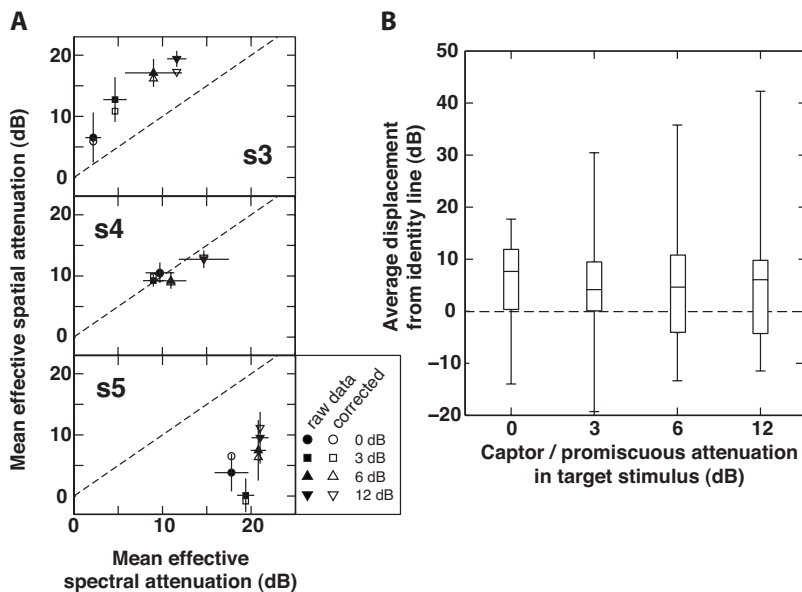
FIG. 5. Comparison of "what" and "where" results, testing the consistent-object hypothesis. (A) Example results for the three individual subjects whose results are also shown in Figs. 2(A), 3(A), and 4(A). Scatter plot of the mean "effective attenuation" of the promiscuous complex in the "where" task plotted against its effective attenuation in the "what" task. Points that fall on the dashed lines (the identity lines) fit the consistent-object hypothesis. Different symbols represent different target stimuli, with different captor/promiscuous tone attenuations. Filled symbols represent the basic results; open symbols include a subject-specific correction factor for the "repulsion" effect of the captors on the target alone, seen in Fig. 4 (see text). Error bars represent the 95% confidence intervals in either dimension. (B) Distribution of the mean discrepancy between the effective spectral attenuation and the effective spatial attenuation (points in panel A) and the identity line (dashed line in panel A) from the 11 subjects who successfully completed the study. Positive values correspond to greater effective attenuation of the promiscuous tone when matching location than when matching spectral content of the target. Plots show the median, inter-quartile range (boxes) and full range (whiskers) as a function of the attenuation of the promiscuous complex present in the target stimulus.

spatial judgments of the target in ambiguous mixtures. If, in computing the target's location, binaural cues in different frequency components are weighted according to their contribution to the target's spectral content (the consistent-object hypothesis), then the effective levels computed from the "what" and "where" tasks for a given subject should be equal.

The effect of the captors was often large enough that, for stimuli in which the promiscuous complex was attenuated by 6 and 12 dB, responses were outside the range of responses observed in corresponding no-captor trials. For these responses, computing an effective level of the promiscuous complex would require extrapolation of the linear fits to the no-captor data [e.g., see Fig. 3(A), top panel; triangles for the 6 and 12 dB attenuations lie well above the range of y-axis values described by the dotted line]. Moreover, the relationship between physical attenuation of the promiscuous complex and the normalized perceived location cannot be linear over an infinite range (e.g., once the promiscuous complex is attenuated enough that it has no measurable impact on the target, further attenuation will not change the perception of the target). Therefore, we restricted all statistical analyses that used effective levels of the ambiguous matches to the 0 and 3 dB attenuations, where such extrapolation was not generally needed.

We can quantify how well our data fit the consistent-object hypothesis by plotting, for each subject and stimulus condition, the effective attenuation of the promiscuous complex in the "where" task (hereafter, "effective spatial attenuation") against the effective attenuation of the same stimulus in the "what" task ("effective spectral attenuation"). The consistent-object hypothesis predicts that these quantities should be equal, so values plotted this way should lie along the identity line.

Figure 5(A) shows the means and 95% confidence intervals of the effective attenuations in the two tasks for the same three subjects from previous figures (filled symbols).

For subject s3, the effective spatial attenuation is larger than the effective spectral attenuation (all points fall above the diagonal in the top panel). The opposite is true for subject s5, where the effective spectral attenuation is larger than the effective spatial attenuation (all points fall below the diagonal in the bottom panel). Other subjects' results lie in between these extremes; for instance, data for s4 fall on the diagonal, in accordance with the consistent-object hypothesis [middle panel of Fig. 5(A)]. Although the inter-subject differences are large, intra-subject differences are small, demonstrating response reliability. Thus, though the consistent-object hypothesis may describe results for some subjects in our population [e.g., s4, in the middle panel of Fig. 5(A)], it does not generally hold for all subjects (see below for a more thorough statistical analysis).

Figure 5(B) summarizes the group data by plotting, for each stimulus condition, the full and inter-quartile ranges of the difference between the mean effective spatial attenuation and mean effective spectral attenuation for each subject. Although the mean displacement is near zero when averaged across all subjects (suggestive of the consistent-object hypothesis), this occurs because some subjects reliably demonstrate larger effective spatial attenuation than effective spectral attenuation [s3 in Fig. 5(A)], some reliably demonstrate larger effective spectral attenuation than effective spatial attenuation [s5 in Fig. 5(A)], while others demonstrate roughly equal effective attenuations in both tasks [s4 in Fig. 5(A)].

As discussed above, the captor tones tend to repel the perceived location of the target. To see if this repulsion helped explain the failure of the consistent object hypothesis, we analyzed results after correcting for repulsion. Specifically, we subtracted the amount of repulsion determined from the third "where" task (captors present, but no promiscuous tones) from all results with captors and promiscuous complexes present. This correction assumes that the presence of the promiscuous complex does not significantly alter the strength of this repulsion (beyond its direct influence on the
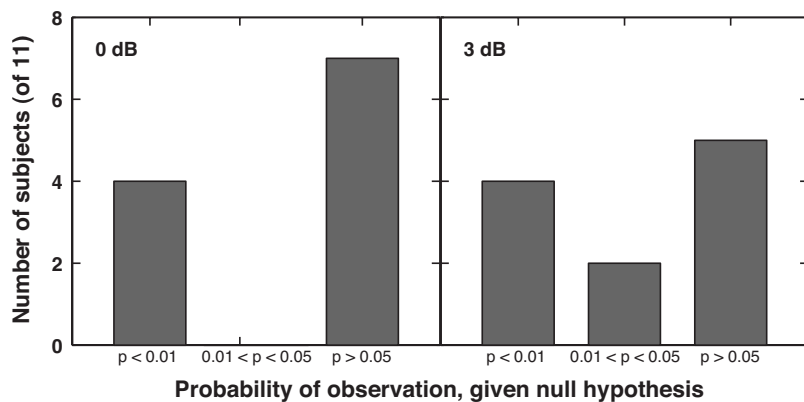
FIG. 6. Histograms showing the distribution of likelihoods of observing the paired "what" and "where" match results of the 11 subjects who successfully completed the study, given that the consistent-object hypothesis is true (i.e., the consistent-object hypothesis is the null hypothesis). Each bar shows the number of subjects whose results yielded a p-value in the corresponding range, based on a two-tailed t-test of the "effective attenuation" data of "what" and "where" matches (Fig. 5). Only the 0 and 3 dB cases were tested (left and right panels, respectively) to limit the need to extrapolate results when estimating effective attenuation.

target), which may be incorrect; however, this analysis gives a first-order correction for repulsion effects. We used the resulting corrected lateralization values to calculate a corrected effective promiscuous complex level [open symbols in Fig. 5(A)].

As seen in Fig. 5(A), correcting for repulsion does not improve the fit of the effective level points to the identity line. Moreover, given that such a correction will generally tend to reduce the effective spatial attenuation of the promiscuous complex, and given that some subjects already demonstrate less effective spatial attenuation than effective spectral attenuation [e.g., s5 in Fig. 5(A)], we conclude that repulsion cannot account for the observed departure from the consistent-object hypothesis.

Although inter-subject variability is large, individual subjects are relatively consistent in how they respond. We therefore analyzed individual results to see if we could reject the consistent-object hypothesis for individual subjects. Assuming response variations for a given subject and given stimulus condition can be accurately modeled as Gaussian-distributed noise, we can test for significant differences in the effective spectral and spatial attenuations (which, being affine transformations of subjects' responses, are also normally distributed) using a paired t-test against the null hypothesis (the consistent-object hypothesis). Specifically, the null hypothesis posits that the effective spatial attenuation and effective spectral attenuation have the same distribution. As discussed above, effective attenuations computed from extrapolated values of the linear fits are unreliable; therefore, we only analyzed results for the 0 and 3 dB conditions.

Figure 6 plots histograms of the resulting p-value distributions for the individual subjects' data. The small p values for some subjects in our population suggest that the consistent-object hypothesis can be rejected for several, but not all, of our subjects (four out of eleven subjects in the 0 dB condition and six out of eleven subjects in the 3 dB condition). However, in such a population analysis, a small fraction of trials (here, individual subject's t-test results) will turn out to be significantly different simply by chance. To assess the significance of our t-test p-value distribution, we can treat the outcome of each t-test as a Bernoulli trial, where the event "$p < 0.05$" defines a success. The probability of obtaining four or more such successes (the result of the 0 dB condition) in eleven independent trials is approximately 1.6×10^{-3}. For six or more successes (the result of the 3 dB

condition), this probability is approximately 5.8×10^{-6}. Note that for the 0 dB condition, all four significant p-values were below 0.01; defining "$p < 0.01$" as a success gives a stricter probability of approximately 3.1×10^{-6}. Hence, we conclude that these failures of the null hypothesis across the population of tested subjects were not observed by chance. Thus, although results for some subjects are well fit by the consistent-object hypothesis, we can reject the consistent-object hypothesis as describing a general property that holds for all subjects in the population at large.

D. Control task for "what" judgments

In the main "what" task, the captors' presence caused listeners to attenuate the level of the promiscuous tones to match the target stimuli. However, these results could arise if the captors reduced the perceived loudness of the target and listeners matched loudness rather than spectral shape. Our control experiment tested for this possibility by asking listeners to perform the same matching task as in the main "what" task, both when controlling the level of the match stimulus' promiscuous complex and when controlling the level of the match stimulus' target complex.

In the control experiment, there were four different target stimuli to be matched, corresponding to all combinations of attenuations (0 and 3 dB) and captor status (present or absent). For each of these stimuli, listeners performed six matches each when adjusting the promiscuous-complex attenuation and six matches when adjusting the target-complex attenuation.

One easy way to assess whether listeners were matching loudness, rather than timbre, is to plot the mean attenuation of the promiscuous complex against the mean attenuation of the target complex when listeners were matching the same physical target stimulus. If the listeners were matching loudness of the target stimuli, then these attenuations should be positively correlated, since listeners would decrease the intensity of whatever components they controlled to decrease the match-stimulus loudness, or increase the intensity of whatever components they controlled to increase the match-stimulus loudness. However, if they were matching the perceived timbre, they would increase the level of the target complex to match the spectral shape of a target stimulus that led them to decrease the level of the promiscuous complex level. Thus, if listeners followed our instructions and

matched the spectral shape of the target stimulus, the mean response attenuations of the promiscuous complex should be negatively correlated with the mean response attenuations of the target complex.

For each subject, we computed the Pearson's correlation coefficient relating the attenuations of the promiscuous complex to attenuations of the target complex for the same target stimulus. These Correlation coefficients ranged from -0.67 to -0.98, with a mean of -0.87. From this, we conclude that subjects were matching the timbre of the tone complexes rather than overall stimulus loudness.

IV. SUMMARY AND DISCUSSION

Recent experiments showed that a reduction in the contribution of a tone to a target vowel's spectrum did not appear to result in an equivalent reduction in its contribution to the vowel's location (Lee *et al.*, 2009; Shinn-Cunningham *et al.*, 2007). Yet it could be that the "effective levels" were reduced by equal amounts if a large change in the tone's spectral contribution to the target vowel were to cause only a small change in the vowel's perceived location. Moreover, this past study did not directly compare "what" and "where" judgments in the same group of subjects, but looked only at across-subject average results.

Here we explicitly tested whether an individual listener's "what" and "where" judgments of the same object in the same sound mixture are quantitatively consistent with one another. We manipulated the degree to which a promiscuous complex contributed to a target by adding captors. Presumably because the captors grouped with the promiscuous complex, they reduced the contribution of the promiscuous complex to the perceived spectro-temporal content of the target complex, consistent with similar past studies (Best *et al.*, 2007; Lee *et al.*, 2008; Shinn-Cunningham *et al.*, 2007). Moreover, the captors also reduced the contribution of the promiscuous complex to the perceived location of the target. However, we found that the effective level of the promiscuous complex contributing to the target in the "what" task often differed from the effective level of the promiscuous complex contributing to the target in the "where" task.

These data do not support the consistent-object hypothesis. Instead, we found that many of our subjects depart reliably from the predictions made by the consistent-object hypothesis, even though the direction of this departure varies across subjects (Fig. 5). Although across-subject differences are large, the deviations from the consistent-object hypothesis for particular subjects are robust and repeatable. Thus, we show that the perceptual contribution of a sound element to the spectral content of an auditory object *cannot* reliably predict its contribution to the object's perceived location: we reject the consistent-object hypothesis. This result suggests that the process determining what elements comprise an auditory object is somewhat independent of the process determining where an object is perceived in space.

The reduced contribution of the promiscuous complex to the target spectrum or location in the presence of the captors may be in part caused by peripheral adaptation due to the captors, which have the same frequency content as the promiscuous complex. Adaptation, however, would not result in a deviation from the consistent-object hypothesis, as any reduction in the representation of the promiscuous complex due to peripheral adaptation is necessarily the same no matter the perceptual task used to measure the effective level of the promiscuous complex. Adaptation can be viewed simply as one mechanism that contributes to the decreased contribution of the promiscuous complex to the target due to the captors, affecting both the contributions to the target's spectral content as well as contributions to the target's perceived location.

Our results show that the presence of the captors changes how the promiscuous complex influences perceived attributes of the target stimulus, affecting perceived target spectral content as well as perceived target location. In line with many related studies (Darwin, 1995; Shinn-Cunningham *et al.*, 2007; Lee *et al.*, 2009), we have discussed our results in terms of grouping, suggesting that when the captors are present, some portion of the energy in the promiscuous complex is perceptually grouped with the captors, which then reduces the perceptual contribution of the promiscuous complex to the target. However, it is possible that the presence of the captors enhances the perceptual salience of the target complex through some process other than perceptual grouping. In other words, just as adaptation may be a contributing factor in our experiments, some other, generic form of "perceptual enhancement" of the target complex by the captors may be at play. Although the current results cannot rule out such an explanation, this does not change our main finding. Specifically, (1) the presence of the captors reduced the perceptual contribution of the promiscuous complex to the target in both "what" and "where" tasks, (2) for many of our subjects, the effect of the captors on "what" and "where" judgments was quantitatively inconsistent, allowing us to conclude that (3) the consistent-object hypothesis is violated.

When presented only with the captors and the target, subjects tended to perceive the target as more lateral than in the target-only condition. Whereas spatial cues of perceptually grouped sound elements seem to be integrated when localizing the grouped object, this "repulsion" is thought to occur between distinct, segregated auditory objects (Best *et al.*, 2005; Braasch and Hartung, 2002; Lee *et al.*, 2009). As repulsion between captors and target will change the perceived location of a target object in addition to any changes in the location caused by the promiscuous tones, we computed an adjustment to our effective level data [plotted in Fig. 5(A)] by subtracting the mean repulsion for each subject and stimulus condition from that subject's "where" task data. Notably, correcting for repulsion will tend to reduce the effective spatial attenuation [lowering data points along the y-axis direction in Fig. 5(A)]. Since some subjects' data points already lay significantly below the diagonal, this type of correction did not account for the observed discrepancies between effective spatial and spectral attenuations.

Schwartz and Shinn-Cunningham: Dissociation of "what" and "where"

V. CONCLUSIONS

1. Two harmonically related tone complexes (target and promiscuous complexes) were grouped together as a single object despite having different ITDs. The perceived location of the composite, grouped object varied with the level of the promiscuous complex, consistent with object location being determined by a weighted integration of binaural cues across perceptually grouped frequency components.

2. Adding a captor stream reduced the effective level of the promiscuous complex for both judgments of the target's spectrum and of the target's location.

3. The reduction in the contribution of the promiscuous complex to the perceived location of the target differed from the reduction in the contribution to the perceived spectral content of the target. Although some individual subjects' spatial and spectral judgments were influenced in the same way by the captors, effects varied markedly between subjects. These results contradict the consistent-object hypothesis, and show a dissociation between how the brain computes what an object is and where it is located.

4. Some subjects showed "repulsion" between the captors and the target. However, this cannot account for the discrepancies observed between the effective spectral and spatial contributions of the promiscuous complex to the target.

ACKNOWLEDGMENTS

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities," J. Acoust. Soc. Am. **77**, 1868–1880.

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**). "Binaural interference and auditory grouping," J. Acoust. Soc. Am. **121**, 1070–1076.

Best, V., van Schaik, A., Jin, C., and Carlile, S. (**2005**). "Auditory spatial perception with sources overlapping in frequency and time," Acta. Acust. Acust. **91**, 421–428.

Braasch, J., and Hartung, K. (**2002**). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data," Acta. Acust. Acust. **88**, 942–955.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA), pp. 1–790.

Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Darwin C. J. (**1995**). "Perceiving vowels in the presence of another sound: A quantitative test of the 'old-plus-new' heuristic," in *Levels in Speech Communication: Relations and Interactions: A Tribute to Max Wajskop*, Elsevier, Amsterdam, pp. 1–12.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Dye, J. (**1990**). "The combination of interaural information across frequencies: Lateralization on the basis of interaural delay," J. Acoust. Soc. Am. **88**, 2159–2170.

Hill, N. I., and Darwin, C. J. (**1996**). "Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning," J. Acoust. Soc. Am. **100**, 2352–2364.

Lee, A. K., Deane-Pratt, A., and Shinn-Cunningham, B. G. (**2009**). "Localization interference between components in an auditory scene," J. Acoust. Soc. Am. **126**, 2543–2555.

Lee, A. K. C., Babcock, S., and Shinn-Cunningham, B. G. (**2008**). "Measuring the perceived content of auditory objects using a matching paradigm," J. Assoc. Res. Otolaryngol. **9**, 388–397.

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.

Shinn-Cunningham, B. G., Lee, A. K. C., and Oxenham, A. J. (**2007**). "A sound element gets lost in perceptual competition," Proc. Natl. Acad. Sci. U.S.A. **104**, 12223–12227.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.

Trahiotis, C., and Stern, R. M. (**1989**). "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," J. Acoust. Soc. Am. **86**, 1285–1293.

Warren, R. M., Obusek, C. J., and Ackroff, J. M. (**1972**). "Auditory Induction: Perceptual synthesis of absent sounds," Science **176**, 1149–1151.