Localization interference between components in an auditory scene

Adrian K. C. Lee
Hearing Research Center, Boston University, Boston, Massachusetts 02215 and Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Ade Deane-Pratt
Hearing Research Center, Boston University, Boston, Massachusetts 02215

Barbara G. Shinn-Cunningham[a)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215 and Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Some past studies suggest that when sound elements are heard as one object, the spatial cues in the component elements are integrated to determine perceived location, and that this integration is reduced when the elements are perceived in separate objects. The current study explored how object localization depends on the spatial, spectral, and temporal configurations of sound elements in an auditory scene. Localization results are interpreted in light of results from a series of previous experiments studying perceptual grouping of the same stimuli, e.g., Shinn-Cunningham *et al.* [Proc. Natl. Acad. Sci. U.S.A. **104**, 12223–12227 (2007)]. The current results suggest that the integration (pulling) of spatial information across spectrally interleaved elements is obligatory when these elements are simultaneous, even though past results show that these simultaneous sound elements are not grouped strongly into a single perceptual object. In contrast, perceptually distinct objects repel (push) each other spatially with a strength that decreases as the temporal separation between competing objects increases. These results show that the perceived location of an attended object is not easily predicted by knowledge of how sound elements contribute to the perceived spectro-temporal content of that object.
© 2009 Acoustical Society of America. [DOI: 10.1121/1.3238240]

I. INTRODUCTION

In everyday life, the sound arriving at our ears is the sum of energy from multiple acoustical events in the environment, typically originating from many different sources at different locations in space. The cognitive process of interpreting the sound energy coming from different sound sources and forming the sound in the mixture into distinct perceived objects is known as auditory scene analysis or ASA (Bregman, 1990). Such objects can then be attended, allowing a listener to process an object of interest and judge its content. While it is important to be able to understand the spectro-temporal content of a signal of interest (i.e., "what" you are listening to), the spatial location of that source is also behaviorally important (i.e., "where" an auditory event comes from). For example, in a cocktail party, you not only need to able to hear your name when it is spoken (Cherry, 1953), but you also want to know the location of the person calling you.

A number of studies have investigated how ASA influences the ability to understand an attended signal (Darwin and Hukin, 1999; Freyman *et al.*, 1999; Arbogast *et al.*, 2002; Shinn-Cunningham *et al.*, 2005a). However, there are relatively few studies investigating how competing sources in a sound mixture affect localization of the perceived objects in an auditory scene. Moreover, results of these past studies show that presenting multiple sound components in a mixture can cause many different effects on sound localization.

Although some simultaneous sounds can be localized quite accurately, without strong perceptual interference between the resulting objects (Good and Gilkey, 1996; Lorenzi *et al.*, 1999; Best *et al.*, 2005), other studies suggest that simultaneous sound elements coming from different locations interfere with localization of a target element, even if the interfering elements and the target element are spectrally remote from one another (McFadden and Pasanen, 1976; Best *et al.*, 2007). Moreover, the literature on how spatial perception is affected by interactions between competing sound elements contains evidence for spatial "pulling" and "pushing" effects, as defined below (e.g., see Gardner, 1969).

Pulling (also known as "integration" or "attraction") occurs when spatial information from different sound elements is perceptually combined. Pulling causes the perceived spatial location of a target sound to be displaced toward the

---

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

location at which the competing elements would be perceived if they were presented in isolation. Pulling has been observed, for instance, when subjects localize a source in the presence of an interfering stimulus delivered monaurally (Butler and Naunton, 1964). Another robust example of pulling is the precedence effect, in which the perceived location of a target sound closely following a preceding sound is dominated by the spatial cues in the preceding sound (see Litovsky et al., 1999 for a review). A recent review of studies in which pulling occurs for sources that are spectrally remote (Best et al., 2007) rekindled the idea that the degree of integration of spatial cues in different sound elements is directly affected by auditory grouping (see also Woods and Colburn, 1992).

Specifically, pulling seems to occur when sound elements are perceived as coming from the same auditory object, but this integration is reduced when grouping cues promote perceiving the spectrally remote elements in distinct auditory objects (i.e., spatial cues are perceptually integrated across only those sound elements making up a target object).

Pushing (also known as "repulsion") occurs when the perceived location of a target is displaced away from the location at which competing elements would be perceived if they were presented in isolation (Lorenzi et al., 1999; Braasch and Hartung, 2002). In contrast to pulling, pushing is thought to arise when competing sounds are perceived as coming from distinct auditory objects, each of which is heard at a unique position (Best et al., 2005).

To test the hypothesis that pulling occurs within objects and pushing occurs between objects, we measured the perceived laterality of auditory objects using stimuli identical to those used previously to explore the influence of spatial cues on perceived object content in a sound mixture (Shinn-Cunningham et al., 2007; Lee and Shinn-Cunningham, 2008b). Briefly, when presented with a sound mixture containing a slowly repeating harmonic complex, a single target harmonic is perceived as part of the complex (Darwin and Hukin, 1997; Shinn-Cunningham et al., 2007). However, if there are intervening tones that, together with the target, form an isochronous sequence of tones identical to the target, the target is typically no longer heard as part of the simultaneous harmonic complex (Darwin and Hukin, 1997; Shinn-Cunningham et al., 2007). Most importantly, when the spatial cues of the target and intervening tones are manipulated, the manipulation strongly influences the perceived rhythm of the rapidly repeating tone sequence (the contribution of the target to the tone stream), but not the perceived content of the harmonic complex (not the contribution of the target to the simultaneous complex; Shinn-Cunningham et al., 2007; Lee and Shinn-Cunningham, 2008b). Specifically, the tone stream is most often perceived with a galloping rhythm (the target harmonic is not part of the tone stream) when spatial cues promote (1) grouping the target with the simultaneous complex and (2) segregating the target tone and intervening harmonic tones. However, the tone stream is most often perceived with an even rhythm (the target is heard in the tone stream) when spatial cues promote (1) segregating the target tone from the complex and (2) integrating the target tone into the tone stream. Thus, spatial cues strongly affect how much

the target contributes to the intervening tone stream. However, the spatial cues have only a weak effect on the perceived contribution of the target to the harmonic complex: in the presence of the tone stream, the target never strongly contributes to the complex, regardless of the spatial cues (Shinn-Cunningham et al., 2007; Lee and Shinn-Cunningham, 2008b). It is worth noting that the target tone is never heard as a distinct object in these mixtures. Instead, all of these mixtures are perceived as containing only two perceptual objects: the tone stream and the harmonic complex. Manipulating the spatial cues of the sound elements simply changes the degree to which the target tone contributes to the perceived spectro-temporal content of the two objects in the scene.

The perceptual organization of mixtures of this sort (containing a rapidly repeating tone stream and a more slowly repeating harmonic complex, each of which competes for "ownership" of an ambiguous target element) is robust. If the salience of the spatial cues is reduced by adding ordinary reverberant energy, grouping results are similar, but the degree to which spatial cues modulate the perceptual contribution of the target to the tone stream is reduced (Lee and Shinn-Cunningham, 2008b). If the tone stream is changed from a simple pure tone to a complex tone containing multiple harmonics, the ambiguous target (now also a complex tone, with rich harmonic structure) contributes more to the harmonic complex, but the perceptual contribution of the target to the objects in the scene is still modulated by the spatial cues of the constituent sound elements (Lee et al., 2008). If the frequency of the repeating tones vying for ownership of the target is offset from the frequency of the ambiguous target tone, the degree to which the target contributes to the tone stream decreases as the frequency disparity increases, but the same general trends are seen (i.e., spatial cues have a strong effect on the perceived content of the tone stream, but have a weaker effect on the perceived content of the harmonic complex; Lee and Shinn-Cunningham, 2008a). Thus, although how listeners group complex sound mixtures can be hard to measure precisely, multiple studies investigating mixtures like those used in the current study support the notion that there is a consistent, natural way to group these mixtures that depends on the balance of all of the various factors that affect perceptual grouping, including the spatial cues of the components in the mixture. Most importantly, for mixtures identical to those investigated here, a simple target tone is never heard strongly as part of the simultaneous harmonic complex, regardless of the spatial cues. However, the contribution of the target to the tone stream depends strongly on the spatial cues of the elements making up the mixture.

In the current study, we measured *where* subjects perceived the intervening tones and harmonic complex for stimuli identical to those used in Shinn-Cunningham et al. (2007). Specifically, using an interaural level difference (ILD) pointer, subjects matched the perceived laterality of either the intervening tones or the harmonic complex. Taken together with the results of our previous experiments, we find evidence for obligatory integration of spatial cues in elements presented simultaneously, even across sound elements that are not strongly perceived to be in the same au-
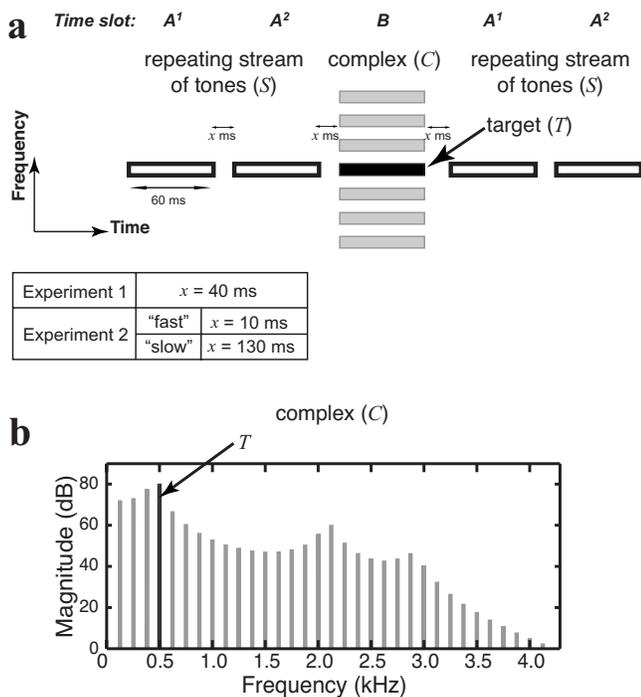
FIG. 1. (A) The two-object stimulus consists of a three-part sequence: a pair of pure tones followed by a harmonic complex (in the form of A^1A^2B). In the basic configuration, the pure tones in time slots A^1 and A^2 (S) are at 500 Hz. Time slot B is made up of two components: a target tone at 500 Hz (T) and a harmonic complex (C) with an $F0$ of 125 Hz (with the fourth harmonic 500 Hz omitted). In Experiment 1, each time slot is 100 ms in duration (60-ms-long acoustic events with 40-ms-long silent gaps between). In Experiment 2, the time slot duration is 70 ms in the "fast" block and 190 ms in the "slow" block (with silent gaps of duration 10 and 130 ms, respectively). See text for details. (B) The complex spectral envelope was shaped to sound like the vowel /ɛ/ when the target was heard as part of the complex.

ditory object. This result demonstrates a dissociation between how sound elements contribute to the perceived spectro-temporal content of objects in a scene and how the spatial cues in constituent sound elements contribute to the perceived locations of those objects. We also observe a spatial repulsion between the perceived location of competing objects. Finally, we show that the strength of the across-object repulsion decreases as the temporal separation between competing objects increases, but the across-element integration of simultaneous elements is unaffected by the temporal separation between competing objects. This final result is further evidence that how auditory elements are grouped into perceptual objects (which is strongly influenced by the temporal separation of the elements) does not always predict how spatial information is combined across elements to determine perceived object location.

II. EXPERIMENT 1: SINGLE REPETITION RATE

Stimuli consisted of a sequence of two repeating tones (S) and a harmonic complex (C) that repeated at one-third the rate of the tones. A 500-Hz tone known as the target (T) could logically belong to both the stream of tones and the harmonic complex [see Fig. 1(a)]. We manipulated the spatial content of the repeating-tone stream, the complex, and the target to explore how spatial cues influence the localization of the perceived objects. Comparison with results of

grouping experiments using identical stimuli (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b), let us explore whether there was a direct relationship between grouping and localization, as has been previously posited (Woods and Colburn, 1992; Best *et al.*, 2007). We hypothesized that if the target was strongly grouped with the simultaneous complex, the perceived location of the complex would be strongly pulled by the spatial cues of the target. However, we predicted that when the target was not heard as part of the complex, its perceived location would have little influence on the perceived location of the complex, consistent with recent results for binaural interference stimuli (Best *et al.*, 2007). We expected the repeating tone stream to be pulled by the spatial cues in the target when it was heard as part of the tone stream, but not when it was not heard as part of the tone stream. Finally, we predicted that there would be pushing between the two objects (complex and tone stream) if they were perceived in different locations (Best *et al.*, 2005).

A. Methods

1. Stimuli

The frequency of the pair of repeating tones making up the tone stream was 500 Hz [Fig. 1(a)]. The harmonic complex was filtered so that its spectral structure was vowel shaped, to enable direct comparison with our companion studies of perceptual organization using identical stimuli [Fig. 1(b); Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b]. The target was a 500-Hz tone that had the same onset/offset as the complex and that, when taken together with the repeating tones, formed an isochronous stream of identical 500-Hz tones.

The amplitudes of the target and the tones were equal, and matched the level that the fourth harmonic of the complex would have, given the spectral shaping applied to the complex. This basic pattern, a pair of repeating tones followed by the harmonic complex and target, was repeated to produce a stimulus that was perceived as two streams: an ongoing stream of tones and a repeating complex occurring at a rate one-third as rapid.

The tones, the harmonic complex, and the target were all gated with a Blackman window of 60-ms duration. There was a 40-ms-long silent gap between each tone and the simultaneous harmonic complex, creating a regular rhythmic pattern with an event occurring every 100 ms. In order to control build-up of streaming, which is known to affect perceptual grouping (Bregman, 1978; Anstis and Saida, 1985; Carlyon *et al.*, 2001; Cusack *et al.*, 2004), we kept the presentation time of these stimuli fixed at three seconds (i.e., ten repetitions of the pair-of-tones-and-complex triplet).

A number of past studies used an ILD acoustic pointer to get repeatable measures of perceived object laterality (Bernstein and Trahiotis, 1985; Trahiotis and Stern, 1989; Buell *et al.*, 1991; Heller and Trahiotis, 1996; Bernstein and Trahiotis, 2003; Best *et al.*, 2007). Therefore, we used a 200-Hz-wide band of noise, centered at 2 kHz, as an acoustic pointer, which listeners used to indicate the perceived laterality of the attended object in each trial. Subjects adjusted the

ILD of the pointer using one button to increase and another button to decrease the pointer's ILD. We used this procedure to quantify the perceived location of object whose constituent sound elements had spatial cues from pseudo-anechoic head-related transfer functions (HRTFs) measured on a KEMAR manikin at a distance of 1 m in the horizontal plane (see Shinn-Cunningham *et al.*, 2005b for details). In general, subjects did not find our HRTF-processed stimuli particularly well externalized, but they nonetheless found it intuitively easy to match the intra-cranial location of each object with the ILD pointer. Moreover, as in past studies, the ILD matches we obtained were very consistent and repeatable.

2. Task

The same physical stimuli were presented in two experimental blocks. In one block, subjects matched the perceived location of the repeated tones with the acoustic pointer. In the other block, subjects matched the perceived location of the harmonic complex. The order of stimuli was a different random sequence for each subject and each block to mitigate any learning effects.

3. Equipment

All stimuli were generated offline using MATLAB software (Mathworks Inc.). Sources were processed to have spatial cues consistent with a source from a position straight ahead (0° azimuth), 45° to the left, or 45° to the right of the listener.

Digital stimuli were generated at a sampling rate of 25 kHz and sent to Tucker-Davis Technologies (TDT) hardware for D/A conversion and attenuation before presentation over headphones (Etymotic ER-1 insert earphones). Presentation of the stimuli was controlled by a personal computer, which selected the stimulus to play on a given trial. A different random attenuation level (0–14 dB) was applied to both the stimulus and the acoustic pointer in each trial in order to minimize any influence of presentation level on localization. Subjects were seated in a sound-treated booth and responded via a button-box (TDT Bbox), which was directly connected to the hardware. All signals were presented at a listener controlled, comfortable level (maximum value 80 dB sound pressure level).

4. Subjects

Nine subjects (four male, five female, aged 18–31) took part in the experiment. All participants had pure-tone thresholds in both ears within 20 dB of normal-hearing thresholds at octave frequencies between 250 and 8000 Hz, and within 15 dB of normal-hearing thresholds at 500 Hz. All subjects gave written informed consent to participate in the study, as overseen by the Boston University Charles River Campus Institutional Review Board and the Committee On the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology.

B. Procedures

1. Training

At the beginning of each experimental block, all listeners received 15 min of practice to familiarize themselves with the experimental procedures and task, which were identical to those of the main experiment (described below). During these practice sessions, subjects were encouraged to explore the full range of acoustic pointer positions they could achieve, and diagrams were presented on screen to help emphasize the difference between the repeating tones and the harmonic complex. No feedback was provided either during training or during the main experiment.

2. Matching procedure

Each trial began with a presentation of the 3-s-long stimulus. This was followed by a 3-s-long presentation of the acoustic pointer, during which subjects could adjust its ILD. A right button press caused the ILD to increase by one step while a left button press caused it to decrease (achieved by symmetrically adjusting the level to the right ear upward and the level to the left ear downward by the same amount). Updates occurred at a rate of 25 kHz. The constant step-size was set to be very small ($|\Delta|=1.5\times10^{-3}$ dB), so that listeners perceived, in real time, an essentially continuous sound image moving along the intracranial axis as they adjusted the pointer ILD.

Presentations of the stimulus and pointer alternated every 3 s until the subject was satisfied that the pointer laterality matched the perceived laterality of the attended object. To indicate their satisfaction, they pressed a third button, which caused the current pointer ILD to be stored and the next trial to be initiated. The initial pointer ILD was set to a random value (between −20 and +20 dB) at the start of each trial. Typically, subjects cycled through three to four iterations of the listening-matching sequence for each trial before signaling satisfaction with their response.

3. Blocking

Each block of the experiment included seven single-object conditions that served as controls (see left panels of Fig. 2). In three of these conditions, the target and the object to be localized (either the complex or the repeating tones) were both from the same location, either at 0°, 45° (to the right), or −45° (to the left). In the four other single-object conditions, the target and the object to be localized were from different locations (but there were no other competing objects). Seven two-object conditions (which were identical in the "match-tones" and "match-complex" blocks) were intermingled with the appropriate single-object conditions in each block (right panel in Fig. 2). In these conditions, the complex always originated from 0° azimuth. In one two-object control condition, the target and the repeating tones were co-located with the complex. The other six conditions consisted of two conditions in which only the target came from the side (either left or right) and the complex and repeating tones were straight ahead; two with only the repeating tones coming from the side (either left or right) and the complex and target from straight ahead; and two with the

Pictogram Legend

time · azimuth re: head

S repeating tones
C complex
T target
· empty time / location

Bold denotes object being matched

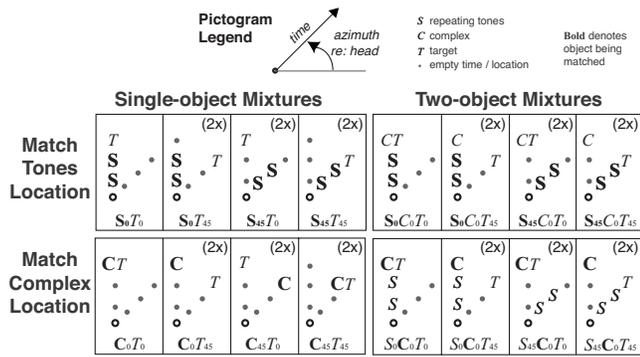|	Single-object Mixtures				Two-object Mixtures			
Match Tones Location	S_0T_0	S_0T_{45} (2x)	$S_{45}T_0$ (2x)	$S_{45}T_{45}$ (2x)	$S_0C_0T_0$	$S_0C_0T_{45}$ (2x)	$S_{45}C_0T_0$ (2x)	$S_{45}C_0T_{45}$ (2x)
Match Complex Location	C_0T_0	C_0T_{45} (2x)	$C_{45}T_0$ (2x)	$C_{45}T_{45}$ (2x)	$S_0C_0T_0$	$S_0C_0T_{45}$ (2x)	$S_{45}C_0T_0$ (2x)	$S_{45}C_0T_{45}$ (2x)

FIG. 2. Summary of the spatial configurations tested (mirror-symmetric versions of the three right-most conditions of each group were presented, for a total of seven stimuli of each type). Single-object conditions are shown on the left for the match-tones (top) and match-complex conditions (bottom). Two-object conditions, presented in both the match-tones and match-complex blocks, are shown on the right. The radial dimension in each diagram denotes time, while the azimuthal angle of each component relative to the listener is denoted by the angle relative to the top-down view of the head.

repeating tones and the target both coming from the same side (either left or right) and the complex from ahead.

Because conditions were mirror symmetric and there were no significant differences in either the main effect of side of presentation (tones: $F_{1,7}=0.449$; complex: $F_{1,7}=0.757$) or the interaction between side of presentation and condition (tones: $F_{7,52}=0.494$; complex: $F_{3,20}=0.534$), results from left/right symmetric conditions were combined in all subsequent analysis. The resulting configurations are denoted by the shorthand $S_xC_yT_z$, where x, y, and z denote the locations of the stream of tones (S), complex (C), and target (T) and can either be 0 for center (components from 0° azimuth) or 45 for side (components from either ±45° azimuth). Bold font highlights the component of the mixture that listeners were asked to match in a given condition. This leads to four unique configurations for the single-object match-tones conditions (S_0T_0, S_0T_{45}, $S_{45}T_0$, $S_{45}T_{45}$; see Fig. 2, top left panel), four configurations for the single-object,

match-complex conditions (C_0T_0, C_0T_{45}, $C_{45}T_0$, $C_{45}T_{45}$; see Fig. 2, bottom left panel), and four two-object configurations that were presented in both match complex ($S_0C_0T_0$, $S_0C_0T_{45}$, $S_{45}C_0T_0$, and $S_{45}C_0T_{45}$) and match-tone conditions ($S_0C_0T_0$, $S_0C_0T_{45}$, $S_{45}C_0T_0$, and $S_{45}C_0T_{45}$; see Fig. 2, right panel).

Each subject completed two experimental blocks, each consisting of 8 repetitions of each of the 14 stimuli in random order, for a total of 224 trials per block. The order of the blocks was counter-balanced across sessions and listeners. Each session lasted no longer than 1.5 h.

C. Results

In all data analysis, results from mirror-symmetric configurations were combined by reversing the sign of the ILD match for configurations with elements to the left and then averaging these values with the corresponding results for configurations with elements to the right. Table I summarizes results of all the statistical tests performed on results from Experiment 1, discussed in detail below.

1. Single-object mixtures

a. Tones. When tone stream and target were both from 0° [S_0T_0: left-most data point in Fig. 3(a)], the mean ILD was near zero (i.e., when both target and tones were from straight ahead, the tones were perceived at midline). Results were similar when the target was shifted to the side and the tones remained in the center [S_0T_{45}; compare the first and second data points from the left in Fig. 3(a)]. When tones and target were both to the side [$S_{45}T_{45}$; right-most data point in Fig. 3(a)], the mean ILD was large and pointing to the expected side. When the tones were from the side and the target was from the center [$S_{45}T_0$: third data point from left in Fig. 3(a)], the perceived location of the tones was shifted toward midline compared to when both tones and target were to the side [compare the two right-most data points in Fig.

TABLE I. Summary of paired-sample *t*-test and Wilcoxon signed-rank tests (comparisons shown in italics) performed on the group mean of Experiment 1 after collapsing across left-right symmetric configurations (see also Fig. 3). *Post-hoc* adjusted significance levels, using the Dunn–Sidak factors, are reported here and are denoted by the subscript DS. Ellipses denote comparisons for which both conditions may have been affected by the limited response range and whose significance was therefore not tested.[1]

|		Effect of target location		Effect of competing object		Effect of tones location	
		Conditions	Significance	Conditions	Significance	Conditions	Significance
Single-object	Tones	$S_0T_{45}-S_0T_0$	$t_8=-1.327$ $p_{DS,2}=0.113$				
		$S_{45}T_{45}-S_{45}T_0$	$Z=-3.4187$ $p_{DS,2}=0.001$				
	Complex	$C_0T_{45}-C_0T_0$	$t_8=-3.267$ $p_{DS,2}=0.023$				
		$C_{45}T_{45}-C_{45}T_0$	⋯ ⋯				
Two-object	Tones			$S_0C_0T_{45}-S_0T_{45}$	$t_{17}=-2.211$ $p_{DS,3}=0.118$		
		$S_0C_0T_{45}-S_0C_0T_0$	$t_8=2.689$ $p_{DS,2}=0.054$	$S_{45}C_0T_0-S_{45}T_0$	$Z=2.025$ $p_{DS,3}=0.123$		
		$S_{45}C_0T_{45}-S_{45}C_0T_0$	⋯ ⋯	$S_{45}C_0T_{45}-S_{45}T_{45}$	⋯ ⋯		
	Complex			$S_0C_0T_{45}-C_0T_{45}$	$t_{17}=2.144$ $p_{DS,3}=0.134$		
		$S_0C_0T_{45}-S_0C_0T_0$	$t_8=-3.638$ $p_{DS,2}=0.013$	$S_{45}C_0T_0-C_0T_0$	$t_{17}=-5.347$ $p_{DS,3}=0.002$	$S_{45}C_0T_0-S_0C_0T_0$	$t_8=4.321$ $p_{DS,2}=0.005$
		$S_{45}C_0T_{45}-S_{45}C_0T_0$	$t_8=-3.358$ $p_{DS,2}=0.007$	$S_{45}C_0T_{45}-C_0T_{45}$	$t_{17}=-5.184$ $p_{DS,3}=0.001$	$S_{45}C_0T_{45}-S_0C_0T_{45}$	$t_{17}=5.349$ $p_{DS,2}<0.001$
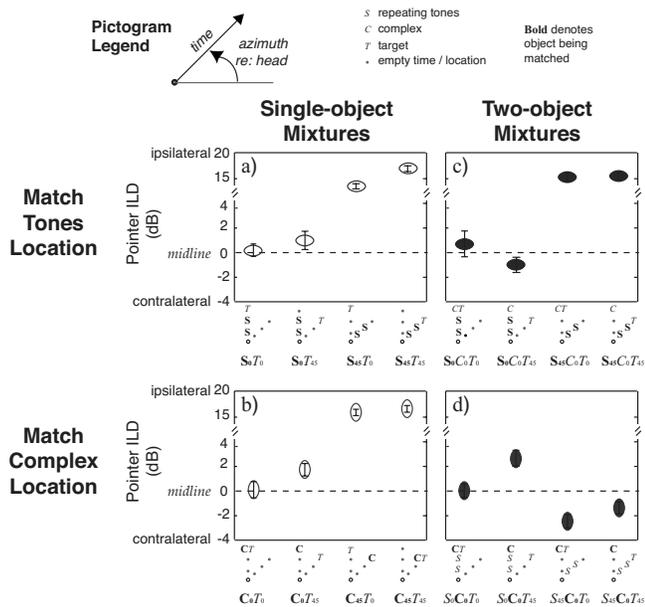
FIG. 3. Across-subject average of the matched ILD for all conditions in Experiment 1, collapsed across mirror-symmetric conditions (all but the left-most condition). Matches to the tones are denoted by horizontal ellipses (top panels) and matches to the complexes are denoted by vertical ellipses (bottom panels). Open symbols represent single-object conditions [(A) and (B)] and filled symbols represent two-object conditions [(C) and (D)]. Error bars show the standard error of the mean across subjects.

$3(a)$; this effect was significant according to a Wilcoxon signed-rank test, $p < 0.05$[1]; see comparison of $S_{45}T_0$ and $S_{45}T_{45}$ in Table I.

b. Complex. When the complex and target were in the same location, results are as expected: near zero for C_0T_0 [left-most data point in Fig. $3(b)$] and large and toward the expected side for $C_{45}T_{45}$ [right-most data point in Fig. $3(b)$]. When the target was from the side and the complex was from the center (C_0T_{45}), the perceived location of the complex was pulled toward the side of the target [compare the two left-most data points in Fig. $3(b)$, $p < 0.05$ for the comparison of C_0T_0 and C_0T_{45} in Table I]. When the complex was simulated from the side, its perceived location was far off to the expected side, both when the target was in front and when the target was to the side ($C_{45}T_0$ and $C_{45}T_{45}$).[1]

2. Two-object mixtures

a. Tones. When the tone stream, the target, and the complex were all from the center [$S_0C_0T_0$: leftmost data point Fig. $3(c)$], the judged tone location was close to zero, as expected. There was a trend for the target location to affect the perceived location of the tones, but this trend did not reach statistical significance. Specifically, when the target was to the side and all of the other components were straight ahead ($S_0C_0T_{45}$), there was a trend for the perceived tone stream location to be displaced slightly away from midline, away from the side of the target, compared to when all components were from in front [compare two left-most data points in Fig. $3(c)$; $p = 0.054$, as shown in Table I]. When the tones and target were to the side and the complex was straight ahead ($S_{45}C_0T_{45}$: right-most data point), the tones were heard far to expected side. Similarly, when the tones

were to the side but the target and complex were from straight ahead, the perceived tone stream location was far to the side in the expected direction [$S_{45}C_0T_0$: third data point from the left in Fig. $3(c)$].[1]

The effect of adding the complex to the sound mixture can be discerned by comparing single-object and two-object judgments [corresponding open and filled horizontal ellipses in Figs. $3(a)$ and $3(c)$, respectively]. These comparisons give no strong evidence for an effect of the complex on the perceived location of the tones. When all components (target, tones, and complex) were from the front, there was no effect of adding the complex: the tones continue to be heard from midline [S_0T_0 versus $S_0C_0T_0$; compare left-most data points in Figs. $3(a)$ and $3(c)$]. Adding the complex from in front had no statistically significant effect on the perceived location of the tones either when the tones were in front and the target was to the side [S_0T_{45} versus $S_0C_0T_{45}$; the second data points from the left in Figs. $3(a)$ and $3(c)$, respectively] or when the tones were from the side and the target was from in front [$S_{45}T_0$ versus $S_{45}C_0T_0$; the third data points from the left in Figs. $3(a)$ and $3(c)$, respectively]. (When the tone stream and the target were to the side, responses may have been affected by the response range, so no statistical tests were performed.[1])

b. Complex. As expected, when the tones, the target, and the complex were all from the center, the perceived location of complex was near zero [$S_0C_0T_0$: see left-most data point in Fig. $3(d)$]. The perceived location of the complex was influenced by the location of the tones in the two-object mixtures. When the complex and tones were in front and the target was to the side, the perceived location of the complex was displaced toward the side of the target compared to when all three sound elements were from in front [compare the two left-most data points in Fig. $3(d)$; $S_0C_0T_0$ and $S_0C_0T_{45}$ differ significantly, with $p < 0.05$, in Table I]. When the complex was in front and the tones were to the side, the perceived location of the complex also was displaced towards the side of the target when the target was moved from midline [compare the two right-most data points in Fig. $3(d)$; $S_{45}C_0T_0$ and $S_{45}C_0T_{45}$ differ significantly, with $p < 0.05$, in Table I]. Thus, the effect of moving the target location was to shift the perceived location of the complex in the direction of the target in the two-mixture conditions ($S_0C_0T_{45}$ versus $S_0C_0T_0$ and $S_{45}C_0T_{45}$ versus $S_{45}C_0T_0$).

In addition to being affected by the location of the target, the perceived location of the complex could be influenced by the simple presence of the tones. When the complex and target were from the center and the tones were from the side, the perceived location of the complex was displaced from midline, away from the tones [compare the third data point from the left in Fig. $3(d)$ with the left-most data point in Fig. $3(b)$; $S_{45}C_0T_0$ and C_0T_0 differ significantly, with $p < 0.05$, in Table I]. Similarly, when the complex was from the center and the target and tones were from the side, the perceived location of the complex was displaced away from the tones compared to the perceived location of the complex without the tones present [compare the right-most data point in Fig.

Lee *et al.*: Localization interference between auditory scene components

3(d) with the second data point from the left in Fig. 3(b); $S_{45}\mathbf{C}_0T_{45}$ and $\mathbf{C}_0T_{45}$ differ significantly, with $p<0.05$, in Table I].

Finally, the location of the competing tones also influenced the perceived location of the complex when comparing two-object conditions. Changing the location of the tones from the center to the side caused the perceived location of the complex to be displaced away from midline into the hemifield opposite the location of the tones, both when the complex and target were in the center [$S_0\mathbf{C}_0T_0$ versus $S_{45}\mathbf{C}_0T_0$; compare first and third data points in Fig. 3(d)] and when the complex was in the center and the target was to the side [$S_0\mathbf{C}_0T_{45}$ versus $S_{45}\mathbf{C}_0T_{45}$; compare the second and fourth data points from the left in Fig. 3(d)]. In both of these cases, these results, which are consistent with the competing tones repelling the perceived location of the complex, were statistically significant ($p<0.05$ for both comparisons in Table I).

D. Discussion

1. Single-object mixtures

a. Tones. In a single-object mixture, the location of a target at midline pulled the perceived location of a tone stream ($\mathbf{S}_{45}T_0$ was closer to midline than $\mathbf{S}_{45}T_{45}$). Thus, in the absence of any competing object, there can be across-time integration of spatial cues that affects the perceived location of the repeated tones. This effect is consistent with the well-known phenomenon of "binaural sluggishness" (Grantham and Wightman, 1978; Culling and Summerfield, 1998; Culling and Colburn, 2000), which is thought of as an obligatory across-time integration of spatial cues. Such sluggishness should depend strongly on the repetition rate of the stimuli, with more integration at faster rates (larger pulling from the target) and less at slower rates. In Experiment 2, we directly tested this hypothesis by comparing localization of the tones with the target at two different repetition rates.

b. Complex. The perceived location of the complex presented without the tone stream tended to be pulled toward the location of the simultaneous target ($\mathbf{C}_0T_{45}$ was pulled from midline toward the side of the target compared to $\mathbf{C}_0T_0$). Because the pulling of the complex by the target depends only on integration of simultaneously presented elements, this pulling should not be affected by repetition rate. We tested this hypothesis in Experiment 2.

2. Two-object mixtures

Just as the perceived location of the complex was pulled by the target in single-object conditions, the complex was pulled toward the location of the target when the tone stream was present in the mixture. In contrast, the target sometimes pulled the perceived location of the tones in single-object mixtures, but there was no evidence for across-time integration of the target and the tones when the complex was present. Indeed, in the two-object mixtures with the complex and tones both from in front, the perceived location of the tones had a tendency to be displaced away from the side of the target, rather than pulled toward the target. Thus, results suggest the target always pulls the perceived location of the

simultaneous complex, but that the target spatial information is not integrated with the perceived location of the tones when there is a simultaneous complex present in the mixture. Given these results, we expected the manipulations of the repetition rate undertaken in Experiment 2 to have little effect on how strongly the target pulled either the perceived location of the complex (which presented simultaneously with the target) or the perceived location of the tones (which was never pulled by the target in the two-object mixtures).

When there were two competing objects in the mixture, there was a tendency for the competing objects to repel each other. Specifically, adding tones from the side caused a significant displacement of the complex away from the side of the tones ($S_{45}\mathbf{C}_0T_0$ versus $\mathbf{C}_0T_0$ and $S_{45}\mathbf{C}_0T_{45}$ versus $\mathbf{C}_0T_{45}$). Similarly, moving the tones to the side caused the perceived location of the complex to be displaced away from side of the tones, as if the tones repelled the complex ($S_0\mathbf{C}_0T_0$ versus $S_{45}\mathbf{C}_0T_0$ and $S_0\mathbf{C}_0T_{45}$ versus $S_{45}\mathbf{C}_0T_{45}$). Across-object repulsion could also explain the trend for the perceived location of the tones to be displaced away from the side of the target in condition $\mathbf{S}_0C_0T_{45}$. In the limit, if the tones and complex are separated by a large inter-stimulus interval, any across-object repulsion must disappear. Thus, we hypothesized that repulsion would be stronger when the repetition rate was faster and weaker when the rate was slower, an idea tested in Experiment 2.

III. EXPERIMENT 2: VARYING REPETITION RATES

In this experiment, three hypotheses were tested:

(1) In the tones-only conditions, the strength of the pulling of the tone stream by the target will increase with increasing repetition rate.
(2) In the complex-only conditions, the pulling of the target on the complex will be independent of the repetition rate.
(3) In the two-object conditions, the strength of across-object repulsion will increase with increasing repetition rate.

A. Methods and procedures

Stimuli were identical to those in Experiment 1 except that the length of the silent gap between each tone and complex varied [Fig. 1(a)]. In the "fast" block of the experiment, the silent gap was 10 ms and an acoustic event occurred every 70 ms. In the "slow" block of the experiment, the silent gap was set at 130 ms, with events every 190 ms. In both experimental blocks, presentation time was fixed at 3 s. As a result, stimuli in the fast block consisted of 14 repetitions of the repeating-tone-complex triplet, while in the slow block it consisted of five repetitions.

Nine subjects (four male, five female, aged 18–31) took part in this experiment. Eight out of these nine subjects also participated in Experiment 1. The training and matching procedures were identical to those used in Experiment 1. Each subject completed four experimental blocks (localization of the repeating tones and the complex at two different rates) on two separate days. On any given day, each subject completed two blocks of tone stream localization and two blocks of
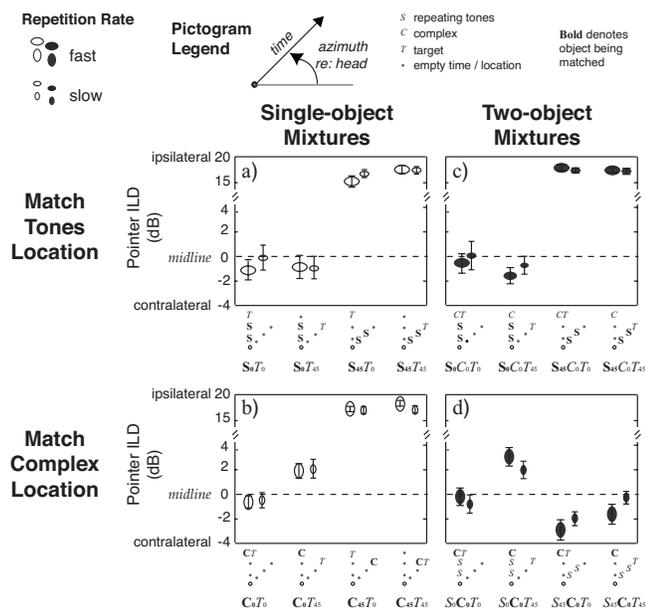
FIG. 4. Across-subject average of the matched ILD for all conditions in Experiment 2 collapsed across mirror-symmetric conditions. Matches to the tones are denoted by horizontal ellipses (top panels) and matches to the complexes are denoted by vertical ellipses (bottom panels). Open symbols represent single-object conditions [(A) and (B)] and filled symbols represent two-object conditions [(C) and (D)]. Big symbols represent the "fast" repetition rate and small symbols represent the "slow" repetition rate. Error bars show the standard error of the mean across subjects.

complex localization, one each for slow and fast repetition rates. The order of the stimulus rate and the order of the task were counter-balanced across subjects.

B. Results

1. Single-object mixtures

a. Tones. In general, repetition rate had little effect on the localization of the tones [compared the bigger and the smaller markers of each condition in Fig. 4(a)]. As expected, when tones and target were both from 0° [S_0T_0; left-most pair of data points in Fig. 4(a)], the mean ILD was near zero, independent of the repetition rate. When the target was from the side and the repeated tones from center, the tones were still perceived near center [S_0T_{45}; second pair of data points from left in Fig. 4(b)]. A two-way, repeated-measures analysis of variance (ANOVA) was conducted on the mean data exploring the effect of repetition rate fast versus slow and

target location (S_0T_0 versus S_0T_{45}) on localization judgments when the repeating tones were from center. Neither of the main effects nor their interaction was statistically significant [compare the two left-most pairs of data points in Fig. 4(a); see first row of Table II].

When tones and target were both to the side [$S_{45}T_{45}$: right-most pair of data points in Fig. 4(a)], the mean ILD was large and to the expected side, independent of repetition rate. When the repeating tones were to the side and the target was in front, the perceived laterality of the tones was also far to the expected side [$S_{45}T_0$: third pair of data points from left in Fig. 4(a); because the response range may have affected these results, no statistical tests were performed to compare single-object tone conditions[1]].

b. Complex. The perceived location of the complex was as expected in the control conditions. When the complex and target were both from midline [C_0T_0; left-most pair of data points in Fig. 4(b)], the ILD was near zero. When the complex and target were both to the side [$C_{45}T_{45}$: right-most pair of data points in Fig. 4(b)], the ILD was large, and in the expected direction.

Consistent with results of Experiment 1, the target tended to pull the perceived location of the midline complex. When the complex was from in front and the target was to the side, judgments were displaced from midline toward the target side [C_0T_{45}; pair of data points second from the left in Fig. 4(b)]. When the complex was from the side, judgments were far to the side for all matches [see the two right-most pairs of data points in Fig. 4(b)].

Consistent with our hypothesis, repetition rate had little effect on the localization of the complex, independent of the exact spatial configuration of the target and complex. This observation was supported by a two-way, repeated-measures ANOVAs with factors of target location and repetition rate (fast and slow) for the complex coming from the center: the main effect of target location was significant, but neither the effect of repetition rate nor the two-way interaction was statistically significant [the left-most pair of data points is lower than the pair of data points second from the left in Fig. 4(b), but within each pair, the points are similarly valued; see third line of Table II]. (When the complex was from the side, results may have been affected by the response range, so statistical tests were not performed.[1])

TABLE II. Summary of the two-way repeated ANOVA tests performed on the group mean of the single-object conditions in Experiment 2 after collapsing across left-right symmetric configurations (see also Fig. 4). Ellipses denote comparisons that were excluded, as results in both conditions may have been artificially limited by the response range.[1]

| | | Single-object (two-way, repeated ANOVA) Main Effect | | | | | |
	Conditions	Target location		Rate		Interaction	
Tones	$S_0T_{45}-S_0T_0$	$F_{1.8}=0.092$	$p=0.348$	$F_{1.8}=3.487$	$p=0.099$	$F_{1.8}=0.929$	$p=0.363$
	$S_{45}T_{45}-S_{45}T_0$	...	...	...	...	...	...
Complex	$C_0T_{45}-C_0T_0$	$\mathbf{F_{1.8}=16.987}$	$\mathbf{p=0.003}$	$F_{1.8}=0.16$	$p=0.902$	$F_{1.8}=0.027$	$p=0.874$
	$C_{45}T_{45}-C_{45}T_0$	...	...	...	...	...	...

Lee *et al.*: Localization interference between auditory scene components

TABLE III. Summary of the one-tailed Wilcoxon signed-rank tests of the effect of repetition rate on the group mean of the two-object conditions in Experiment 2 (see also Fig. 4). Ellipses denote comparisons that were excluded, as results in both conditions may have been artificially limited by the response range.[1]

Two-object (Wilcoxon signed-rank tests on rate of repetition)				
Conditions	Tones	$S_0C_0T_{45}$	$Z=1.677$	$p=0.047$
		$S_{45}C_0T_0$	⋯	⋯
		$S_{45}C_0T_{45}$	⋯	⋯
	Complex	$S_0C_0T_{45}$	$Z=2.286$	$p=0.011$
		$S_{45}C_0T_0$	$Z=1.633$	$p=0.051$
		$S_{45}C_0T_{45}$	$Z=2.112$	$p=0.035$

2. Two-object mixtures

a. Tones. When the tone stream, target, and complex were all from the center ($S_0C_0T_0$), the mean ILD for the tone stream match was near zero, independent of repetition rate, as expected [see the left-most pair of data points in Fig. 4(c)]. From the results in Experiment 1, we expected any across-object repulsion to be evident only when the competing tones and complex were perceived in different locations and the responses were not far to the side (and thus not affected by a response ceiling effect[1]). Therefore, we only expected to see an effect of rate on the repulsion of the tones in configurations $S_0C_0T_{45}$. A Wilcoxon signed-rank test was performed to test for the effect of rate in condition $S_0C_0T_{45}$.[2] There was a significant effect of the repetition rate on the perceived location of the tones, consistent with there being repulsion of the tones that was significantly smaller for the slower repetition rate in condition $S_0C_0T_{45}$ [see the pair of data points second from the left in Fig. 4(c) and the first line of Table III].

b. Complex. As expected, when the tones, the target, and the complex were all from the center ($S_0C_0T_0$), the mean ILD for complex localization was near zero, independent of repetition rate. However, rate affected localization of the complex. In particular, spatial repulsion was consistently stronger for the faster repetition rate than for the slower repetition rate: within each of the three right-most pairs of data points in Fig. 4(d), the right data point in each pair is closer to the perceived location of the complex without the tones present [shown by the two left-most pairs of matches in Fig. 4(b)] than the left data point in the pair. Moreover, all of these judgments of the perceived location of the complex are displaced away from the perceived location of the tones in that mixture [shown by the corresponding results in Fig. 4(c)]. Using three separate Wilcoxon signed-rank tests, the repulsion of the harmonic complex was found to be significantly smaller for the slower repetition rate in conditions $S_0C_0T_{45}$ [the pair of data points second from the left in Fig. 4(d); see fourth line of Table III] and $S_{45}C_0T_{45}$ [the pair of data points third from the left in Fig. 4(d); see final line of Table III]. Although the effect of repetition rate on the localization of the complex failed to reach statistical significance in condition $S_{45}C_0T_0$, there was a trend for repulsion by the tones to be weaker at the slower repetition rate even in this condition [right-most pair of data points in Fig. 4(d); in fifth line of Table III, $p=0.051$].

C. Discussion

1. Single-object mixtures

a. Tones. In the single-object conditions, there was little evidence for an effect of the target location on the perceived location of the tones. These results suggest that there is relatively little integration of the target spatial cues when judging the location of the repeated tones, even at the fastest repetition rate. This result is interesting, especially in light of the fact that listeners strongly perceive the target as part of the repeated tone object in related experiments investigating what objects listeners perceived in these kind of sound mixtures (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b).

b. Complex. When listeners judged the laterality of the complex in single-object configurations, the target significantly pulled the perceived location of the complex when the complex originated from the center. This pulling was not significantly influenced by the rate of repetition, consistent with our hypothesis. These spatial judgments suggest that there is an obligatory integration of the spatial information in the target with the spatial cues in the simultaneously present complex, and are consistent with the fact that the spatially displaced target is heard as part of the complex when there is no other object competing for ownership of the target (Darwin and Hukin, 1997; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b).

2. Two-object mixtures

Across-object spatial repulsion was generally stronger at the faster repetition rate and weaker when the competing objects were more separated in time, consistent with our hypothesis. This effect of repetition rate was found to be statistically significant both when localizing the tones (in condition $S_0C_0T_{45}$) and when localizing the harmonic complex ($S_0C_0T_{45}$ and $S_{45}C_0T_{45}$; there was a trend for repulsion to be weaker at the slower repetition rate in condition $S_{45}C_0T_0$).

In two-object mixtures, the perceived location of the complex depends both upon pulling by the simultaneous target and repulsion by the competing tone stream. For instance, in condition C_0T_{45}, the side target pulls the perceived location of the complex away from midline. Adding tones from midline to the mixture causes the complex and tones to repel one another, so that the complex is heard even farther to the side of the target than when the tones are not present ($S_0C_0T_{45}$ versus C_0T_{45}). As noted above, the repulsion between the complex and the tones decreases as the temporal separation between complex and tones increases, as expected. These observations highlight the fact that both pulling and pushing can occur in the same conditions.

3. Build-up of streaming

Many studies have shown that when listeners hear a repeating sequence of elements, the way in which the listeners perceptually organize the sound mixture changes, or "builds up," over time (Bregman, 1978; Anstis and Saida, 1985; Carlyon *et al.*, 2001; Cusack *et al.*, 2004). To the extent that any build-up of streaming depends on the number of presentations, build-up should be greater in the fast block than in the

slow block. Conversely, if build-up of streaming depends on absolute time rather than the number of stimulus presentation, build up should be similar in the two blocks. Further work is necessary to investigate how streaming build-up may influence the perceived location of objects in an auditory scene. However, the current results show that across-object interactions influence where listeners perceive objects in a complex scene.

IV. GENERAL DISCUSSION

A. Pulling (integration)

Spatial information in the target influenced localization of both across-time objects (the tone stream) and objects grouped across frequency (the harmonic complex) when there were no competing objects in the mixture. The target weakly pulled the tone stream location, having an observable influence only when the tones were from the side and the target was from the center. Moreover, this pulling did not depend significantly on repetition rate. This suggests that binaural sluggishness does not cause a strong, obligatory temporal integration of spatial cues across time for the tones and target used in these experiments. However, such effects might arise in similar experiments if the temporal gap between events was smaller (or, equivalently, if the repetition rate was greater).

In both experiments, in single-object conditions containing the target and complex, the target significantly pulled the complex when the complex came from the center. This result shows that the target spatial information was integrated with the spatial information in the simultaneously presented complex in single-object conditions.

When there are two objects present in the scene, the target spatial cues did not ever pull the perceived location of the tone stream significantly, but always pulled the perceived location of the complex. There was no strong effect of repetition rate on the pulling of the complex by the target. As noted above, this makes intuitive sense, given that the complex and the target are simultaneous, suggesting that the integration of their spatial information should be independent of any temporal parameters.

The observed obligatory integration of the target spatial cues with the complex in the presence of the tones is surprising in light of past studies of perceptual organization of auditory mixtures like those used here. Specifically, when listeners are asked whether the target contributes to what elements the harmonic complex contains, the target usually is not heard as part of the complex for these mixtures (Darwin and Hukin, 1997; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). Conversely, the target contributes to the tone stream in some cases, but not others, depending on the spatial cues in these stimuli (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). In contrast, we find that the target, which is *never* heard strongly as part of the complex, *always* contributes to the perceived location of the complex. Moreover, the target never contributes to the perceived location of the tones in the two-object mixtures, regardless of the spatial configuration of the elements; however, the spatial configuration has a dra-

matic effect on whether or not the target is heard as part of the tone stream.

These results are interesting when taken in conjunction with results of past studies that suggest that the perceived location of an object depends on integrating spatial cues contained only in the sound elements that are heard as part of the attended object (Woods and Colburn, 1992; Best *et al.*, 2007). For example, in one recent study, listeners did not show obligatory integration of spatial cues from simultaneously presented elements. Only when the simultaneous low-and high-frequency elements were perceived as part of the same object was integration observed. However, unlike in the current study, the elements that were perceived in different objects were not comprised of interleaved frequency components; instead, the low-and high-frequency elements were far removed from each other, spectrally. Nonetheless, manipulations that altered how strongly listeners grouped the low-and high-frequency elements together altered the amount of spatial-cue integration. If it were generally true that listeners only integrate spatial information across elements that are perceived as making up an object, then the perceived location of the complex should not be pulled by the target in the current experiment, since the target is not heard strongly as part of the complex. Thus, taken with past studies of how listeners group the current sound mixtures, we show here that integration of spatial cues contained in simultaneous elements (the target and the complex) is not predicted by whether the elements are heard as part of the same object. Instead, we find that across-frequency integration of spatial cues for the current simultaneous, spectrally interleaved elements is obligatory, regardless of whether or not the simultaneous elements are perceived as one object.

It is possible that subjects used in the current experiments actually heard the target as part of the complex, given that this was not explicitly measured here. However, this is unlikely based on the robustness of results from our past studies of how listeners group these kinds of sound mixtures (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b). Another alternative is that there is no link between grouping and localization. However, this is inconsistent with previous observations showing that grouping influences what spatial cues listeners integrate when determining the perceived location of an object (Woods and Colburn, 1992; Hill and Darwin, 1996; Best *et al.*, 2007). Another possibility is that even a small contribution of an element to what an object sounds like can yield a large shift in where that object is perceived. This idea can be tested in future experiments by measuring whether the perceived location of simultaneous complex presented with an attenuated target is strongly pulled toward the location of a target even when it has little energy. However, preliminary tests in our laboratory suggest this is not the case (Schwartz and Shinn-Cunningham, 2008).

We believe that the most likely possibility is that listeners generally integrate spatial cues from only those sound elements that are part of an attended object. However, (1) listeners cannot perfectly filter out spatial cues in simultaneously presented components and (2) the degree to which they can filter out spatial cues in competing elements de-

pends on the frequency separation between within-object elements and interfering sound elements. In other words, just as the binaural system may be sluggish in processing spatial cues in the temporal dimension (Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999), binaural analysis may also be coarse in frequency compared to monaural analysis (Holube *et al.*, 1998). One possible experiment that can test this conjecture would be to use similar stimuli but increase the fundamental frequency of the harmonic complex, so that the frequency separation of the harmonics is large enough that the spatial cues of the target are spectrally distinct from those of the complex. In this case, one could predict that the target will no longer influence the perceived location of the complex when the target is heard as part of the complex.

B. Pushing (repulsion)

There is no evidence of pushing in any of our single-object results. Instead, the target pulled both tone stream and complex in the absence of a competing object. However, when two objects were present in the scene and were perceived at different locations, they generally repelled one another. For instance, the perceived location of the complex was always displaced away from the tones location when tones were added to the mixture ($S_{45}C_0T_0$ versus C_0T_0 and $S_{45}C_0T_{45}$ versus C_0T_{45}). Similarly, the effect of moving the location of the tones in the two-object mixtures was to displace the perceived location of the complex in the opposite direction from the displacement of the tones ($S_0C_0T_0$ versus $S_{45}C_0T_0$ and $S_0C_0T_{45}$ versus $S_{45}C_0T_{45}$). Experiment 2 showed that across-object repulsion was often stronger when the stimulus rate was faster and weaker when the rate was slower, both when localizing the tones (in condition $S_0C_0T_{45}$) and when localizing the harmonic complex (in conditions $S_0C_0T_{45}$ and $S_{45}C_0T_{45}$). This rate effect supports the idea that objects repel one another spatially, but that this effect decreases as the objects are more separated in time.

In the current study, only two objects were heard (the tone stream and the harmonic complex), and they were separated in time. One might, therefore, postulate that only temporally segregated objects repel one another. However, previous evidence for repulsion has been reported when objects overlap in time (Lorenzi *et al.*, 1999; Best *et al.*, 2005). Therefore, we suggest that repulsion occurs between objects (as opposed to within an object), and that this across-object repulsion weakens with temporal separation between the competing objects.

C. Assessing auditory segregation through pulling and pushing

If spatial information is integrated within an object, but objects repel one another, then spatial perception can be used to assess the perceptual segregation of elements comprising an auditory scene.

Kubovy and van Valkenburg (2001) argued that "perceptual boundaries" are important for the formation of auditory objects. In vision, perceptual boundaries, or edges, are determined by spatio-temporal discontinuities (Adelson and Ber-

gen, 1991). In audition, spectro-temporal structure determines how objects form (Bregman, 1990; Darwin and Carlyon, 1995; Darwin, 1997; Van Valkenburg and Kubovy, 2003; Griffiths and Warren, 2004; Shinn-Cunningham, 2008). By parametrically varying the spectro-temporal features of sound in a mixture (e.g., in dimensions such as onset synchrony, harmonicity, common amplitude modulation, etc.), perceptual segregation can be manipulated, which should impact localization judgments. Specifically, if listeners judge sound elements to be coming from different locations, then the elements must belong to different perceptual objects. Conversely, if sound elements are heard as part of the same object, then listeners are likely to integrate the spatial information in the elements, leading to a pulling effect.

However, auditory objects are not always distinct (Rand, 1974; Liberman *et al.*, 1981; Moore *et al.*, 1986; Darwin, 1995; McAdams *et al.*, 1998; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b; Shinn-Cunningham and Wang, 2008). These results are consistent with the fact that sound making up an auditory scene is transparent, with sounds from different sources adding together rather than obscuring each other (Bregman, 1990). Some studies suggest that it is necessary, but not sufficient, for sound elements to be perceived in distinct objects for them to be perceived as coming from distinct locations (Litovsky and Shinn-Cunningham, 2001; Best *et al.*, 2007). The current results (taken together with our past results investigating perceptual organization of these mixtures; Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b) demonstrate that integration of spatial cues can occur across elements that are not perceived within the same object. In other words, while spatial repulsion across elements may prove that the elements are heard in different objects, the current results show that integration of spatial cues across element does *not* prove that the elements are heard in the *same* object. Still, using a continuous measure such as perceived location can provide a bound on when different objects are perceived as distinct, even though it cannot rule out cases when two objects are heard, but are perceived at the same location. Future work can assess the degree to which spatial measures of across-element spatial integration and across-object spatial repulsion give insights into auditory scene analysis.

V. CONCLUSIONS

These data show that there is repulsion between the perceived locations of auditory objects, and that this repulsion tends to decrease with increasing temporal separation of the objects. Moreover, spatial cues in one sound element are often integrated with other spatial cues, pulling the perceived location of an attended object toward the location of the individual element, both for single- and two-object sound mixtures. We observed some weak integration of spatial cues across time, consistent with binaural sluggishness. However, this across-time pulling was only present in some single-object mixtures and was not observed for any of the two-object mixtures used here. We found evidence for an obligatory integration of the spatial cues in a simultaneous target element with those of a spectrally interleaved harmonic com-

plex, an effect that was independent of repetition rate. Taken together with our companion grouping experiments using the same sets of stimuli (Shinn-Cunningham *et al.*, 2007; Lee and Shinn-Cunningham, 2008b), and in contrast with previous results (Woods and Colburn, 1992; Hill and Darwin, 1996; Best *et al.*, 2007), the current results show that spatial cues in elements that do not contribute strongly to the perceived content of an auditory object can nonetheless strongly pull the perceived location of that object. We suggest that spatial repulsion is an effect observed between objects, while spatial cue integration is an effect observed either within an object or across sound elements whose spatial cues cannot be resolved in spatial computations.

ACKNOWLEDGMENTS

[1]Based on pilot data, we selected the allowable range of ILD matches to have a maximum magnitude of 20 dB. While this range was sufficiently large to guarantee that our pilot subjects never reached the maximum allowable value, this was not the case for all of the subjects tested in the formal experiment. In general, we used simple, paired *t*-tests to check for the statistical significance of effects of interest. However, *t*-tests assume Gaussian-distributed matches. Thus, for some comparisons involving matches to objects perceived to the side, the response distributions of the ILD matches might be skewed due to the response limitations, violating the assumptions of a parametric *t*-test. Therefore, we performed nonparametric Wilcoxon signed-rank tests on comparisons involving the perceived locations of objects originating from the side. These comparisons are denoted with italics in Table I. Taking a fairly conservative approach, we further excluded comparisons between pairs of conditions when more than a quarter of the matches in each of the conditions to be compared had magnitudes equal to or greater than 18 dB (within 2 dB of the maximum allowable response) to prevent over-interpreting the results. Comparisons that were excluded due to responses near the allowable maximum are indicated by ellipses in the tables summarizing the statistical comparisons.
[2]We hypothesized that the matched location would be further away from midline for the faster repetition rate (with no underlying distribution assumed for the amount of repulsion). Therefore, when testing the significance of rate on repulsion in these two-object conditions, we used one-tailed Wilcoxon signed-rank tests.

Adelson, E. H., and Bergen, J. R. (**1991**). *The Plenoptic Function and the Elements of Early Vision* (MIT, Cambridge, MA).

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A binaural analog of gap detection," J. Acoust. Soc. Am. **105**, 2807–2820.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of sinusoidally amplitude-modulated tones: effects of spectral locus and temporal variation," J. Acoust. Soc. Am. **78**, 514–523.

Bernstein, L. R., and Trahiotis, C. (**2003**). "Enhancing interaural-delay-based extents of laterality at high frequencies by using "transposed stimuli,"" J. Acoust. Soc. Am. **113**, 3335–3347.

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.

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.

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. (**1978**). "Auditory streaming is cumulative," J. Exp. Psychol. Hum. Percept. Perform. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, Mass.).

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1991**). "Lateralization of low-frequency tones—Relative potency of gating and ongoing interaural delays," J. Acoust. Soc. Am. **90**, 3077–3085.

Butler, R. A., and Naunton, R. F. (**1964**). "Role of stimulus frequency and duration in the phenomenon of localization shifts," J. Acoust. Soc. Am. **36**, 917–922.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). "Effects of attention and unilateral neglect on auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **27**, 115–127.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F., and Summerfield, Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.

Cusack, R., Deeks, J., Aikman, G., and Carlyon, R. P. (**2004**). "Effects of location, frequency region, and time course of selective attention on auditory scene analysis," J. Exp. Psychol. Hum. Percept. Perform. **30**, 643–656.

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*, edited by C. Sorin, J. Mariani, and H. Meloni (Elsevier, Amsterdam), pp. 1–12.

Darwin, C. J. (**1997**). "Auditory grouping," Trends Cogn. Sci. **1**, 327–333.

Darwin, C. J., and Carlyon, R. P. (**1995**). in *Hearing*, edited by B. C. J. Moore (Academic, San Diego, CA), p. 387.

Darwin, C. J., and Hukin, R. W. (**1997**). "Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity," J. Acoust. Soc. Am. **102**, 2316–2324.

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.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gardner, M. B. (**1969**). "Image fusion, broadening, and displacement in sound location," J. Acoust. Soc. Am. **46**, 339–349.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Griffiths, T. D., and Warren, J. D. (**2004**). "What is an auditory object?," Nat. Rev. Neurosci. **5**, 887–892.

Heller, L. M., and Trahiotis, C. (**1996**). "Extents of laterality and binaural interference effects," J. Acoust. Soc. Am. **99**, 3632–3637.

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.

Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.

Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward-masking-evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

Kubovy, M., and Van Valkenburg, D. (**2001**). "Auditory and visual objects," Cognition **80**, 97–126.

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.

Lee, A. K. C., and Shinn-Cunningham, B. G. (**2008a**). "Effects of frequency disparities on trading of an ambiguous tone between two competing auditory objects," J. Acoust. Soc. Am. **123**, 4340–4351.

Lee, A. K. C., and Shinn-Cunningham, B. G. (**2008b**). "Effects of reverberant spatial cues on attention-dependent object formation," J. Assoc. Res.

Otolaryngol. **9**, 150–160.

Liberman, A. M., Isenberg, D., and Rakerd, B. (**1981**). "Duplex perception of cues for stop consonants—Evidence for a phonetic mode," Percept. Psychophys. **30**, 133–143.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Litovsky, R. Y., and Shinn-Cunningham, B. G. (**2001**). "Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression," J. Acoust. Soc. Am. **109**, 346–358.

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

McAdams, S., Botte, M. C., and Drake, C. (**1998**). "Auditory continuity and loudness computation," J. Acoust. Soc. Am. **103**, 1580–1591.

McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

Rand, T. C. (**1974**). "Dichotic release from masking for speech," J. Acoust. Soc. Am. **55**, 678–680.

Schwartz, A., and Shinn-Cunningham, B. G. (**2008**). "The influence of ambiguous grouping cues on an auditory object's perceived spectral content and location," in Mid-Winter Meeting of the Association for Research in Otolaryngology, Phoenix, AZ.

Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cogn. Sci. **12**, 182–186.

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acta. Acust. Acust. **91**, 967–979.

Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (**2005b**). "Localizing nearby sound sources in a classroom: Binaural room impulse response," J. Acoust. Soc. Am. **117**, 3100–3115.

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.

Shinn-Cunningham, B. G., and Wang, D. (**2008**). "Influences of auditory object formation on phonemic restoration," J. Acoust. Soc. Am. **121**, 295–301.

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.

Van Valkenburg, D., and Kubovy, M. (**2003**). "In defense of the theory of indispensable attributes," Cognition **87**, 225–233.

Woods, W. S., and Colburn, H. S. (**1992**). "Test of a model of auditory object formation using intensity and interaural time difference discrimination," J. Acoust. Soc. Am. **91**, 2894–2902.