JARO 9: 388–397 (2008)
DOI: 10.1007/s10162-008-0124-0

JARO
Journal of the Association for Research in Otolaryngology

Measuring the Perceived Content of Auditory Objects Using a Matching Paradigm

Adrian K. C. Lee,[1,2] Steve Babcock,[1] and Barbara G. Shinn-Cunningham[1,2,3]

[1]*Hearing Research Center, Boston University, Boston, MA, USA*

[2]*Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA*

[3]*Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Room 311, Boston, MA 02215, USA*

Received: 9 January 2008; Accepted: 21 April 2008; Online publication: 13 May 2008

ABSTRACT

Two previous studies manipulated spatial cues to alter the perceptual organization of a sound mixture containing an ambiguous sound element (a pure tone; the "target") that could belong to two competing auditory objects (a sequential tone stream and a simultaneous harmonic complex). In both studies, the sum of the contributions of the target to the two objects was less than the physical target level in the mixture. However, many listeners had difficulties making consistent judgments about the perceptual contribution of the target to the harmonic complex. The current study used stimuli similar to those used in the previous study, but with a target made up of five tones rather than a single pure tone. In addition, listeners performed a direct matching task to indicate the perceptual contribution of the target to the competing objects rather than relying on an indirect mapping procedure. The matching task proved to be efficient and reliable. However, the complex-tone target was perceptually stronger in the harmonic complex and weaker in the sequential tone stream than in past studies. As a result, the sum of the target contributions to the two objects roughly equaled the physical target level for all tested spatial configurations, unlike in the previous studies.

Correspondence to: Barbara G. Shinn-Cunningham · Department of Cognitive and Neural Systems · Boston University · 677 Beacon Street, Room 311, Boston, MA 02215, USA. Telephone: +1-617-3535764; fax: +1-617-3537755; email: shinn@cns.bu.edu

Keywords: auditory scene analysis, segregation, grouping, streaming, method of adjustment

INTRODUCTION

The sound arriving at our ears is a sum of acoustical energy from all the acoustic sources in the environment. Determining which energy came from which physical source is intrinsically an ill-posed problem. Nonetheless, in most circumstances, we are able to estimate the content of the physical sources in the environment with relative ease. The challenge of estimating the content of acoustic sources is known as the "cocktail party problem" (Cherry 1953), which we solve through "auditory scene analysis" (Bregman 1990).

In a sound mixture in which ambiguous sound elements could belong to more than one object, it seems logical to suppose that when the ambiguous elements contribute more to one object they contribute less to the competing object, obeying a form of perceptual trading (Darwin 1995; Lee and Shinn-Cunningham 2008a, b; McAdams et al. 1998; Shinn-Cunningham et al. 2007). A special form of energy trading occurs if the sum of the perceived energy that the ambiguous elements contribute to the various objects present in a complex mixture equals the physical energy of the ambiguous elements in the mixture (what we have called "energy conservation"; see Lee and Shinn-Cunningham 2008a). Results from some past studies are consistent with perceptual trading, if not with the more strict energy conserva-

tion (Darwin 1995; Lee and Shinn-Cunningham 2008b; McAdams et al. 1998). However, we performed two recent studies of how listeners perceive ambiguous sound mixtures in which we manipulated spatial cues to alter how an ambiguous scene was perceptually organized into objects and found that even perceptual trading failed (Lee and Shinn-Cunningham 2008a; Shinn-Cunningham et al. 2007).

In these studies, there was an ambiguous target tone that could logically be either (1) the third tone in an isochronous, repeating stream of identical tones or (2) the fourth harmonic in a repeating harmonic complex that occurred simultaneously with the ambiguous target. In different blocks, listeners identified either the perceived rhythm of the isochronous tone stream, which was "even" when the target tone was heard in the stream and "galloping" when it was not, or the perceived vowel identity of the harmonic complex, which was more like "/ɪ/ as in 'bit'" when the tone was heard in the complex and "/ɛ/ as in 'bet'" when it was not. Spatial cues were manipulated to alter how the acoustic mixture was formed into objects, whereas results for intermingled single-object control trials were measured to ensure that listeners were able to label prototype stimuli consistently. We found that spatial cues altered how the mixture was perceptually organized. In most spatial configurations, the target contributed strongly to the perception of the tone stream and contributed weakly or not at all to the harmonic complex. However, when spatial cues encouraged grouping the target with the harmonic complex but discouraged grouping the target with the tone stream, the target contributed very little to either object, showing a complete breakdown in perceptual trading of the ambiguous target between the objects in the scene.

The current study was designed to extend these findings, addressing two weaknesses with the previous experimental design.

First, although listeners in the past studies were generally consistent in judging the presence or absence of the target in the tone stream, many were unable to maintain accurate vowel labels for the harmonic complex because the perceptual changes were relatively subtle (more than a third of the subjects had to be dismissed because of poor performance in identifying single-object vowel prototypes). The current stimuli were similar to the stimuli used in the previous experiments, but the ambiguous target was a complex harmonic tone rather than a simple (single-frequency) tone to make the contribution of the target to the harmonic complex more salient.

Second, in the past experiments, an indirect mapping procedure was adopted to quantify the perceived contribution of the target to the objects in the scene. Listeners performed an auxiliary experiment in which they categorized single-object stimuli containing a target element that was attenuated to different degrees from trial to trial. The resulting psychometric functions were used to map the percent responses in the main experiment to an effective perceptual contribution of the target to the corresponding two-object stimuli. The current experiment tests a direct matching procedure to measure the effective level that the target contributes to the competing objects, bypassing the need for the indirect mapping used in the past.

METHODS

Participants

Ten subjects (eight males, two females, aged 18–33 years) took part in the experiments. All participants had pure-tone thresholds of 20 dB HL or better at octave frequencies in the range from 250–8,000 Hz in both ears with thresholds at 500 Hz of 15 dB HL or better. All subjects gave 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.

Stimuli

Two-object stimuli consisted of a 3-s-long sequence, composed of ten identical repetitions of three items: two harmonic complexes of fundamental frequency 300 Hz (the *rapidly repeating pair* or RRP) followed by a *slowly repeating complex* (SRC) with fundamental frequency 200 Hz (see left panel of Fig. 1A). The ambiguous target was a harmonic complex whose spectrotemporal content was identical to the harmonic complexes making up the RRP but that occurred simultaneously with the SRC. Two-object control stimuli contained only the RRP and SRC with no target. The tone complexes making up the RRP, SRC, and target (if present) were all gated with a 60-ms-long Blackman window. Each temporal event was followed by a silent gap of 40 ms duration. Thus, the three-item sequence lasted 300 ms with one event each 100 ms (see Fig. 1A).

Single-object stimuli (used as controls during the main experiment as well as during training, described below) consisted of either the RRP or the SRC (see middle and right panels of Fig. 1A, respectively), along with the target. Single-object stimuli differed in the intensity of the target, which could take on one of six levels: $-\infty$ (no target present), −12, −8, −4, 0, or +4 dB (relative to the target level in the ambiguous, two-object stimuli).

Spectrally, the RRP and target consisted of harmonics of frequencies 300, 600, 900, 1,200, and

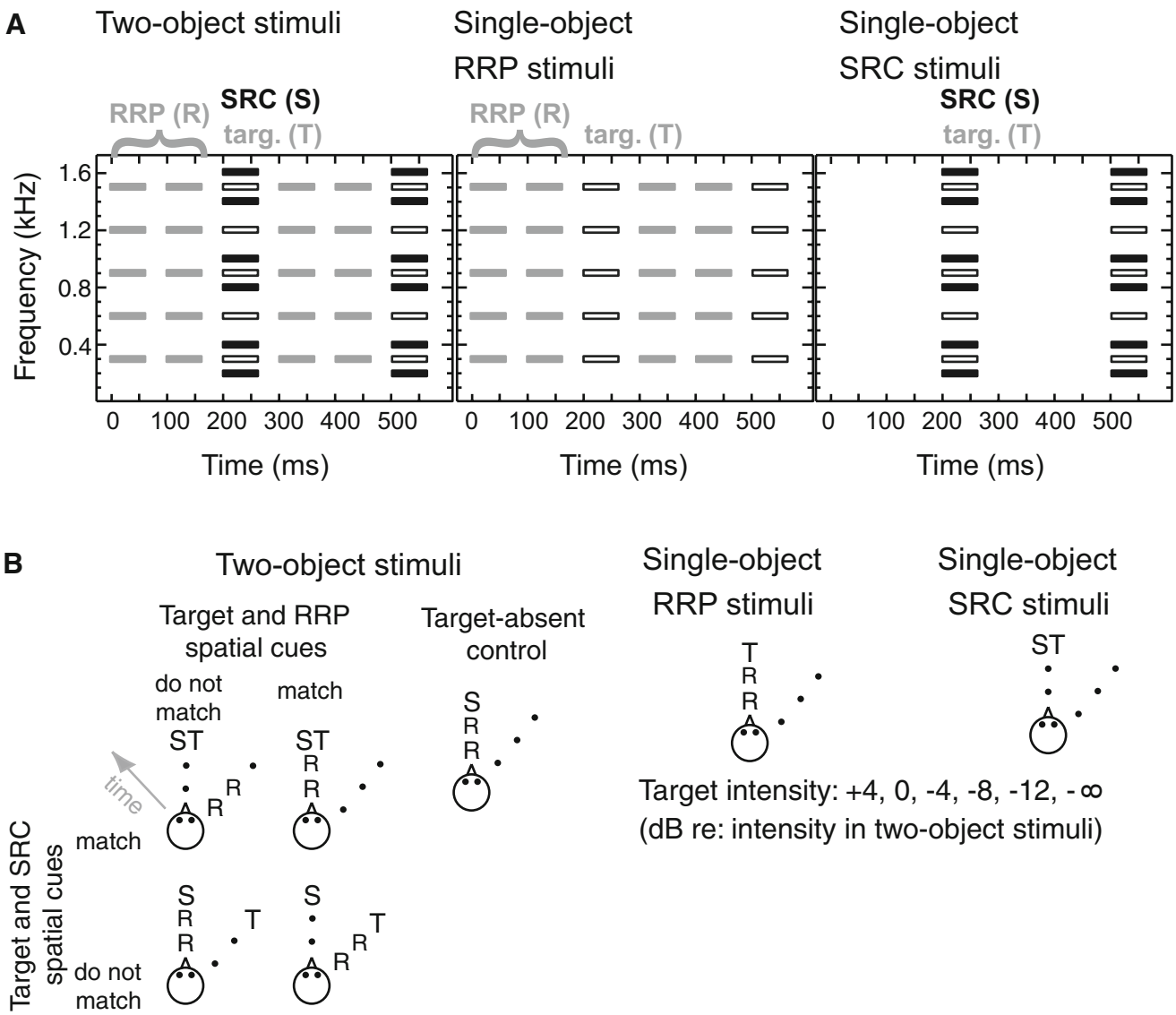

FIG. 1. Stimulus structure and spatial configurations. **A** Two-object stimuli present a repetition of a three-item sequence that consisted of two harmonic complexes of fundamental frequency 300 Hz (the RRP) followed by a harmonic complex with fundamental frequency 200 Hz (the SRC). The target was presented at the same time as the SRC, but had the same spectral content as the harmonic complexes in the RRP (see *left panel*). Single-object stimuli consisted either of the RRP (*middle panel*) or the SRC (*right panel*), along with the target. **B** In the two-object stimuli, the SRC was always simulated from straight ahead; the RRP was either from straight ahead or 45° azimuth; and the target was either from straight ahead or 45° azimuth (*leftmost panel*). A two-object, target-absent control stimulus was also presented with the SRC and the RRP always simulated from straight ahead (*second panel from the left*). Elements making up the single-object stimuli were always presented from straight ahead (*two rightmost panels*).

1,500 Hz (gray rectangles and open rectangles, respectively, in the left and middle panels of Fig. 1A). The SRC consisted of harmonics of frequencies 200, 400, 800, 1,000, 1,400, and 1,600 Hz (black tones in the left and right panels of Fig. 1A). All complexes were filtered by a second-order Butterworth low-pass filter with a cutoff frequency of 500 Hz, which was determined during piloting to maximize the sensitivity of the matching tasks while simultaneously reducing the perceptual salience of any edge pitch (Kohlrausch et al. 1992).

The repeating, three-item sequence resulted in a percept of two distinct auditory objects: a stream consisting of a rapidly repeating harmonic complex with a pitch of 300 Hz and a slowly repeating harmonic complex occurring at one third that rate (the spectrotemporal structure of these objects can be seen in the middle and right panels of Fig. 1A, respectively, which show the corresponding single-object stimuli). Because of the stimulus structure, the rhythm of the stream containing the RRP depended on the degree to which the target contributed

perceptually to that stream. When the target was heard as part of the stream containing the RRP, the stream's perceived rhythm was even; conversely, when the target did not contribute significantly to the stream, its rhythm was galloping. Similarly, both the spectral density and the perceived pitch of the SRC depended on the degree to which the target was heard as part of the SRC. When the target was heard in the SRC, listeners perceived a complex with a dense spectral composition and a pitch of 100 Hz (corresponding to the missing fundamental of the harmonics of 200, 300, 400, 600, 800, 900, 1,000, 1,200, 1,400, 1,500, and 1,600 Hz). When the target was not heard as part of the SRC, its perceived pitch was an octave higher (200 Hz) and its perceived spectral density was sparser.

The levels of both the test and matching stimuli were adjusted by different random amounts (over a 20-dB range) before presentation to discourage using loudness as a cue in the matching task. All signals were presented at a listener-controlled, comfortable level that had a maximum value of 80 dB SPL.

Spatial cues

Stimuli were generated offline using MATLAB software (Mathworks). Signals were processed with pseudoanechoic head-related transfer functions (HRTFs; for details, see Shinn-Cunningham et al. 2005) measured for a manikin head located in the center of the room with the sources 1 m away, either originating from 0° or 45° to the right of the manikin (the same spatial configurations used in our companion studies; Lee and Shinn-Cunningham 2008a; Shinn-Cunningham et al. 2007).

For two-object stimuli containing the target, four different spatial configurations were tested, differing in whether the spatial cues of the RRP and/or the SRC matched those of the target (see left portion of Fig. 1B). Specifically, the simulated RRP and target location varied from trial to trial with each either simulated from 0° or 45° azimuth, whereas the SRC was always simulated from 0° azimuth (see the left panel of Fig. 1B). In the two-object control stimuli in which there was no target, the RRP and SRC both came from 0° azimuth (see second panel from left in Fig. 1B).

As in our past studies using similar stimuli (Lee and Shinn-Cunningham 2008a, b; Shinn-Cunningham et al. 2007), subjectively, the target was never heard as a distinct object, separate from the RRP and the SRC. Instead, when present, the target was heard as part of one or both of the main objects in the scene.

In single-object stimuli (both in the main experiment and used during training; see below), all elements were simulated from 0° azimuth (see the two rightmost panels in Fig. 1B).

Equipment

Digital stimuli were generated at a sampling rate of 12 kHz and sent to Tucker-Davis Technologies hardware for D/A conversion and attenuation. A PC selected the stimulus to play on a given trial. Stimuli were presented over insertion headphones (Etymotic ER-1) to subjects seated in a sound-treated listening booth. Subjects responded via a button box (TDT Bbox).

Task

Subjects used the method of adjustment to match the perceived spectrotemporal content of the attended object (either the RRP or the SRC), either presented alone (in single-object control trials in the main experiment as well as during training) or as part of a two-object mixture. Each trial began by presenting a 3-s-long test stimulus, which could either be a single- or two-object stimulus, depending on the kind of trial. This was followed by a 3-s-long, single-object matching stimulus consisting of an adjustable-level target and a fixed-level complex (either the RRP or the SRC, depending on the block; see the two rightmost panels in Fig. 1B).

During the presentation of the matching stimulus, subjects could adjust (in real time) the target by pressing one button to increase its intensity and a different button to decrease its intensity. At the start of each trial, the level of the target in the matching stimulus was random, in the range of −14 to +6 dB relative to the physical level of the target in the ambiguous, two-object stimuli. Subjects could adjust the target level of the matching stimulus to fall between −60 and +6 dB relative to its level in the two-object stimuli containing a target.

In each trial, the overall level was set to a random value (over a 20-dB range) to discourage using loudness as a cue in the matching task.

Three-second-long test and matching stimuli alternated until the subject was satisfied that the perceived spectrotemporal content of the attended object in the test stimulus matched its content in the single-object matching stimulus. When the subject was satisfied with the match, she/he pressed a third button, which stored the results of that trial and initiated the next trial in the block. Typically, subjects cycled through three to four iterations of the test-matching sequence before proceeding to the next trial.

Procedures

Each subject performed two experimental sessions on two different days, each of which lasted roughly 1 h. In one session, they matched the perceived spectro-

temporal content of the RRP; in the other, they matched the perceived spectrotemporal content of the SRC. The order of sessions was counterbalanced across subjects.

Each session began by training the subjects to ensure that they were able to match the appropriate unambiguous, single-object stimuli (either the RRP or the SRC). After training with the appropriate type of single-object stimuli, listeners performed the main matching experiment for the attended object.

Training consisted of two phases in which subjects matched the perceived spectrotemporal content of single-object stimuli: familiarization and testing. In both phases of training, test stimuli consisted of either the RRP or the SRC and a target whose level varied from trial to trial (taking on one of six levels, from $-\infty$ to +4 dB; see middle and right panels of Fig. 1A and B). These test stimuli alternated with matching stimuli that had similar spectrotemporal content and an adjustable-level target. Thus, it was possible for listeners to objectively match the content of an unambiguous test stimulus by appropriately manipulating the target level in the matching stimulus.

During the initial, familiarization phase of training, real-time visual feedback was provided to the subject as they adjusted the target level in the matching stimulus. A graphical user interface displayed a face whose mouth shape and color gave the listeners information about the difference between the spectral content of the matching stimulus and the test stimulus. When the relative intensity of the target in the matching stimulus was adjusted to be within ±3 dB of its intensity in the test stimulus, the face turned yellow and the mouth turned upward into a smile (to indicate success). Whenever the relative target intensity was outside of this ±3 dB range, the face turned gray and the mouth turned down into a frown; moreover, the curvature of the downward arc was proportional to the absolute difference between the relative target intensity in the test and matching stimuli (to indicate the degree of mismatch). Subjects performed as many trials as they wished in the familiarization phase until they were satisfied that they understood how to match the spectrotemporal content of the attended object.

Once subjects were ready to continue, the testing phase of training began. Trials during the testing phase were similar to those in the familiarization phase, but without real-time graphical feedback. Subjects adjusted the intensity of the target in the matching stimulus until they were content that its spectrotemporal content matched that of the test stimulus. Once they indicated that they were satisfied with their response, they received visual feedback. The testing phase was organized into runs of 12 trials (two repetitions of each of the six possible stimuli, differing in the target intensity in the test stimuli). To proceed to the main experiment, subjects had to come within an average of 3 dB of matching the spectrotemporal content of each possible single-object stimulus in each run. Typically, subjects required less than two runs of testing to reach criterion.

After listeners achieved the criterion level of performance on the familiarization task, they performed the main experiment. During the main experiment, listeners matched the perceived spectrotemporal content of ambiguous two-object stimuli (four different spatial configurations), control target-absent two-object stimuli, and single-object stimuli (six different target levels), all of which were intermingled in the same main-experiment block. Subjects performed eight repetitions of each of the 11 possible stimuli in random order (different for each subject), for a total of 88 trials per session.

RESULTS

Subject screening

We excluded the data from any subject who failed to match single-object stimuli consistently. For each type of stimulus, we calculated the mean and standard error of the mean (SEM) of the eight matches. If the mean of more than half of these matches had an SEM larger than 1.5 dB, we excluded that subject from all subsequent analysis. One of the ten subjects was unable to reliably perform the matching task, generating SEMS of 1.5 dB or more for all of the single-object SRC control trials and three of the six RRP control trials. All subsequent results are from the remaining nine subjects.

To ensure that subjects actively matched each trial, we did not analyze any trials in which the subjects made no adjustments to the target level in the matching stimulus before signaling to go on to the next trial. This occurred only rarely, on less than 5% of all trials.

Single-object matches

Figure 2 summarizes the single-object matching results for both the RRP (Fig. 2A) and the SRC (Fig. 2B). The panels show the final intensity of the target in the matching stimulus plotted as a function of the target level in the single-object test stimulus, averaged across subjects (here, and throughout the rest of the manuscript, the matching target intensity results are plotted in dB relative to the physical level of the target presented in two-object test stimuli containing the target). Error bars show the standard error of the across-subject mean. In these stimuli, perfect physical matches of the test stimulus spectrotemporal content would fall along the dashed line.

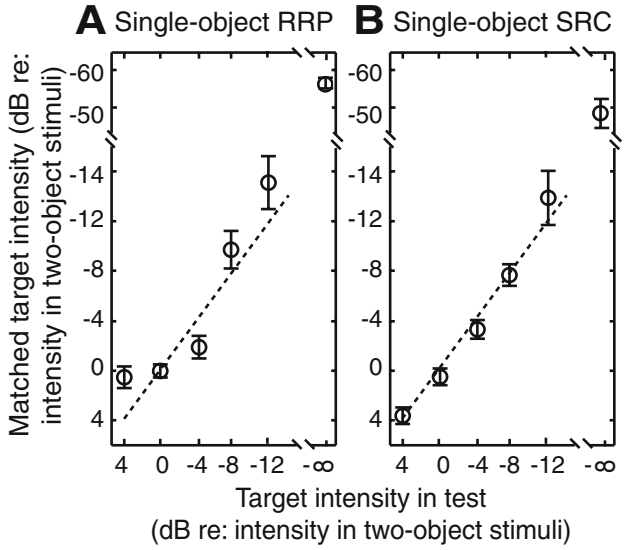

FIG. 2. Results for single-object control stimuli. **A** Matches for single-object RRP trials. The across-subject average of the target intensity in the matching stimulus is plotted as a function of the target intensity in the RRP single-object test stimuli. The *error bars* show the SEM. **B** Matches for single-object SRC trials. The across-subject average of the target intensity in the matching stimulus is plotted as a function of the target intensity in the SRC single-object test stimuli. The *error bars* show the SEM.

Subjects were generally consistent in matching the content of the RRP (Fig. 2A). However, when the test stimulus target intensity was −4 dB, subjects tended to set the target intensity of the matching stimulus too low (one-sample t test, $p<0.005$ with Bonferroni post hoc correction applied). For the target-absent RRP prototype, subjects set the target intensity of the matching stimulus at about −55 dB (recall that the maximum possible attenuation was 60 dB). In general, the target intensity that subjects set in the matching stimulus increased monotonically as the target intensity in the test stimulus increased.

For all six single-object SRC stimuli (Fig. 2B), the matching stimulus target intensity was not statistically different from the test stimulus target intensity (none of the post hoc adjusted t tests reached significance at $p=0.05$). For the target-absent SRC, subjects set the target intensity of the matching stimulus to about −48 dB. In general, the group mean of the matching stimulus target was within one standard error of the test stimulus target intensity.

Two-object matches

Figure 3A summarizes the results for the two-object stimulus matches. Each two-object stimulus was pre-

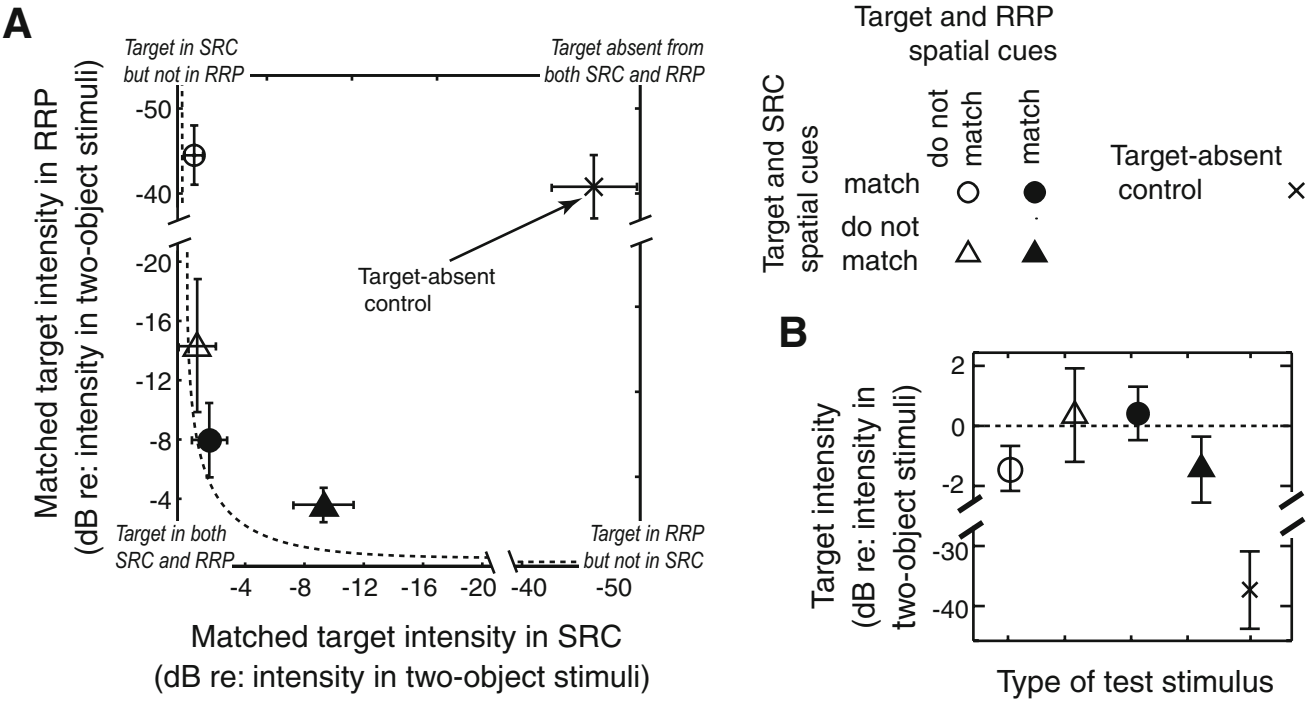

FIG. 3. Results for two-object stimuli. **A** Scatter plot of the matched target intensity for two-object stimuli when subjects matched the RRP content versus when they matched the SRC content for identical stimuli. The *dashed line* shows where results would fall if energy conservation occurred (see text). To aid in interpretation, the *italic* phrases in the four corners of the scatter plot describe the percepts that would lead to results in the corresponding region of the plot. **B** The perceived target intensity relative to the intensity of the target present in the two-object stimuli, calculated as the sum of the target intensity perceived in the RRP and the SRC. For the two-object stimuli, values near zero are consistent with energy conservation (see text).

sented in both a "match the RRP" and in a "match the SRC" session. Figure 3A plots the across-subject average of the target intensity that subjects matched when attending to the RRP (vertical axis) against the target intensity that they matched when attending to the SRC for the same physical stimulus (horizontal axis). The dashed curve plots the trading relationship that would be observed if energy conservation holds (Darwin 1995; Lee and Shinn-Cunningham 2008a, b; McAdams et al. 1998; Shinn-Cunningham et al. 2007).

As expected, the results for the target-absent control stimulus fall near the upper-right corner of the plot with the perceived target contribution to both the RRP and the SRC attenuated by more than 40 dB (X in Fig. 3A). In the conditions for which the target was present, spatial cues altered the perceived spectrotemporal content of both the RRP and the SRC with the target contributing more to an object when their spatial cues matched and less when their spatial cues differed. Moreover, the contribution of the target to the two objects generally obeys a trading relationship with the target contributing more to one object when it contributes less to the competing object (data fall along a monotonically decreasing curve in Fig. 3A; however, data tend to fall to the right and above where expected if energy conservation held). When the target and the SRC are collocated and the RRP is from a different location, the target is heard as part of the SRC but contributes almost nothing to the perceived content of the RRP (Fig. 3A, open circle). When the location of the target differs from the location of both the RRP and the SRC, it contributes less to the RRP but still contributes very little to the SRC (Fig. 3A, open triangle). When the locations of the target, RRP, and SRC all match, the target contributes to both the RRP and to the SRC, but has a stronger contribution to the SRC than to the RRP (Fig. 3A, filled circle). Finally, when the spatial cues of the target and RRP match and the spatial cues of the target and the SRC do not match, the target contributes to both the RRP and to the SRC, but has a stronger contribution to the RRP than to the SRC (Fig. 3A, filled triangle).

The total effective intensity of the target for a given two-object stimulus is the sum of the target intensities subjects set when matching the RRP content and when matching the SRC content for the same stimulus. To quantify the "trading relationship," we compared the total effective intensity to the actual physical target intensity in the two-object conditions. These values are shown in Figure 3B, plotted in dB relative to the actual target intensity in the two-object test stimuli that had a target present.

For the four two-object stimuli containing the target, the total contribution of the target to the RRP and SRC roughly equals the physical target intensity in the stimuli with results falling between ±2 dB. The only two-object stimulus whose effective intensity is statistically significantly different from the physical target level reference, based on two-tailed, one-sample t tests, is the target-absent control ($p<0.05$), which actually had no target energy in the to-be-matched stimulus. These results are similar to those of past studies that show that the total perceptual contribution of ambiguous sound elements are generally between 0 and 3 dB less than the actual physical intensity of the ambiguous element in the mixture (Darwin 1995; Lee and Shinn-Cunningham 2008b; McAdams et al. 1998). However, these results contrast with those of our own previous studies that used very similar stimuli (Lee and Shinn-Cunningham 2008a; Shinn-Cunningham et al. 2007). Specifically, when the spatial cues of the target matched those of the SRC and did not match those of the RRP, the total effective target intensity is nearly equal to the physical target intensity (see open circle in Fig. 3). This result is in direct contrast with results in our previous studies using a pure-tone target and a two-alternative-forced-choice method where the total effective target intensity for the corresponding spatial configuration was 6–10 dB less than the physical target intensity in the mixture (Lee and Shinn-Cunningham 2008a; Shinn-Cunningham et al. 2007).

DISCUSSION

Effectiveness of the stimulus design

One goal of the current study was to make the perceptual contribution of the target to the SRC salient because nearly half of all naive subjects in our previous tests were unable to reliably categorize single-object stimuli without feedback (Shinn-Cunningham et al. 2007). To make the contribution of the target to the SRC more salient, we increased the number of pure-tone components making up the target. We also selected frequencies of the tonal elements making up the target and the SRC so that when the target was heard as part of the SRC, the perceived pitch of the composite object shifted (from 200 Hz when the SRC was presented alone to 100 Hz when the SRC and the target were heard as one object).

The redesigned stimuli made it easier for the subjects to tell whether or not the target was heard in the SRC, just as we had hoped. Informal reports confirmed that listeners had no difficulty telling the difference between the SRC played without the target and the SRC played with the target at full intensity.

Objectively, for the single-object control stimuli, listeners adjusted the target level in the matching stimuli accurately, so that the spectrotemporal con-

tent of the matching stimuli and test stimuli were very close (see Fig. 2), even though the overall level of the stimuli varied randomly over a 20-dB range from trial to trial. In fact, subjects were, if anything, more accurate in matching the spectrotemporal content of the SRC (corresponding to the vowel that proved difficult to reliably label in our previous studies) than in matching the content of the RRP (data in Fig. 2B are closer to the diagonal than data in Fig. 2A). Thus, the redesigned stimuli achieved our goal of increasing the salience of the target's contribution to the simultaneous harmonic complex.

Effectiveness of the matching paradigm

A second goal of the current study was to develop a paradigm that would allow us to directly assess the effective level that the target contributed to each of the competing objects in the two-object stimuli, instead of relying on a categorization task and using a mapping procedure to map response percentages into effective target levels (Lee and Shinn-Cunningham 2008a, b; Shinn-Cunningham et al. 2007).

Listeners had no trouble using the matching procedure when asked to match the perceived spectrotemporal content of either auditory object in the two-object mixture. Not only did listeners find the task simple to understand and easy to perform, results were generally consistent across trials. For the nine listeners who passed our screening, the SEMs of their matches were on the order of 3 dB in the RRP task and on the order of 1.5 dB in the SRC task. This result, combined with the fact that listeners had comparable response variability when matching single-object control stimuli, shows that the matching paradigm yields repeatable, reliable measures.

Subjects generally did not set the target intensity to match the stimulus with the maximum target attenuation level (i.e., −60 dB). Subjects set the target intensity of the matching stimulus at about −55 dB for the target-absent RRP prototype and at about −48 dB for the target-absent SRC prototype. This undershoot may reflect a reluctance on the part of the subjects to use the most extreme values of attenuation. Moreover, when the target is attenuated by more than 30 dB, it may be nearly masked (especially when presented with the simultaneous SRC), producing little discernible effect on the perceived spectrotemporal content of the total stimulus, reflecting a limit on the maximal attenuation that is perceptually meaningful.

We conclude that other than an effective floor that limits the maximal attenuation that listeners used, the matching procedure is a reliable and effective method for measuring the perceived content of objects in a sound mixture.

Direct comparison with previous results

Two of our previous studies manipulated spatial cues to alter perceptual organization in the same way that spatial cues were manipulated in the current study, using very similar stimuli. The primary difference in the stimuli between the current study and these past studies is that the earlier studies used pure tones for the target and for the repeating tones (corresponding to the complex-tone target and RRP in the current study).

In the first spatial-manipulation study, spatial cues were generated from the pseudoanechoic HRTFs employed here (Shinn-Cunningham et al. 2007). In the second spatial-manipulation study, the same stimuli were presented, but room reverberation from a classroom was included in the spatial simulation (Lee and Shinn-Cunningham 2008a). In both studies, manipulating the spatial cues in the repeating tones and target changed how the objects in the mixture were perceived. The reverberation reduced the influence of spatial cues on performance, but the results of the two studies were otherwise similar.

Because the current study used pseudoanechoic simulations, the most relevant comparisons are between the current study and the initial study (without reverberation). Figure 4 plots the effective level of the target, using the same format as Figure 3, for all of the

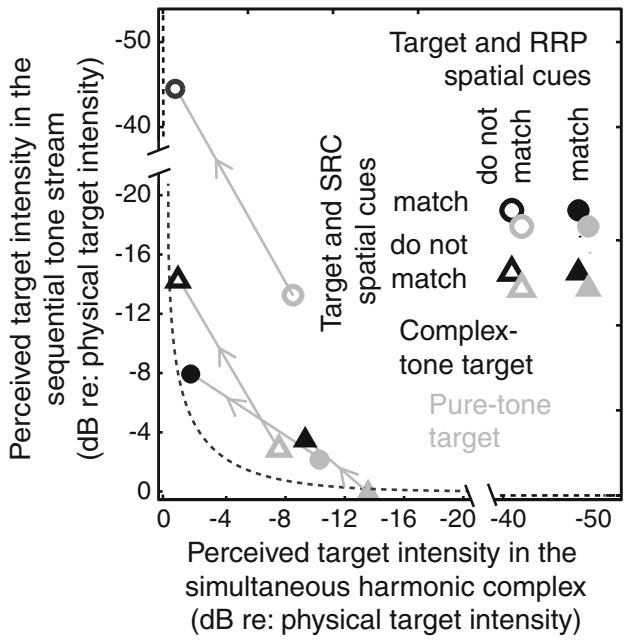

FIG. 4. Comparison of results from the current study and a similar study in which the target was a pure, rather than complex tone. The scatter plot shows the perceived target intensity in the sequential tone stream as a function of the perceived target intensity in the simultaneous harmonic complex using the same format as Figure 3. *Black symbols* are from the current study (repeated from Fig. 3). *Gray symbols* are from the previous study using a pure-tone target.

two-object stimuli that contained a target for both the current study (shown in black) and the previous study using a pure-tone target (shown in gray). Lines connect the data for the corresponding spatial conditions in the two studies for easier comparison.

The results from the two experiments are similar, but for all points, data from the current experiment fall to the left of and above the corresponding data from the current experiment. The fact that the current data are above the corresponding points in the previous study shows that the contribution of the target to the sequential tone stream (here, the RRP) is smaller in the current study compared to the previous study. The fact that the current data fall to the left of the corresponding points in the previous study shows that the contribution of the target to the simultaneous harmonic complex (here, the SRC) is greater in the current study compared to the previous study.

It is possible that the changes in the methods for estimating the perceived contribution of the target to the two objects and differences in the cues that listeners used in deciding how to respond contribute to these differences. In the previous study, subjects were asked to categorize percepts (into one of two rhythmic categories for the RRP and into one of two vowel categories for the SRC). In the current study, the direct matching task is primarily based on rhythm perception for the sequential stream, but subjects could also use loudness to judge the target level and perform the task. Moreover, by changing the target from a simple tone to a complex comprised of multiple frequency components, listeners may have focused on different attributes when judging the SRC content. In the earlier categorization studies, the presence of the target could alter the timbre and overall level of the SRC, but little else. In this study, the target also altered the perceived pitch of the SRC. Additional study is required to clarify whether differences in the task or in the strategy of the subjects contribute to differences in the degree to which trading is observed.

However, we believe that a more parsimonious explanation is that the target's contribution to the simultaneously presented harmonic complex is more than that in the previous studies because of the increased number of components comprising the target. This, in turn, should increase the relative influence of simultaneous grouping cues on perception, weakening the contribution of the target to the RRP and strengthening the contribution of the target to the SRC. The net result of such changes can explain the upward and leftward shifts in the results plotted in Figure 4.

This interpretation is supported by the similarities across studies. For instance, in both studies, the effect of manipulating the spatial cues causes larger changes in the contribution of the target to the sequential stream (an object for which spatial cues should play a strong role in perceptual organization; Darwin 1997; Darwin and Hukin 1999; 2000a, b; Shinn-Cunningham et al. 2007) than to the simultaneous harmonic complex (an object for which spatial cues should have a relatively weak influence on perceptual organization; Darwin 1997; Darwin and Hukin 2000a, b; Shackleton and Meddis 1992; Shinn-Cunningham et al. 2007). In the current study, the effective attenuation of the target in the tone stream ranges from about 4 dB (black filled triangle) to essentially infinite (48 dB; black open circle). In the study using a pure-tone target, the effective attenuation of the target in the tone stream ranges from about 0 dB (gray filled triangle) to about 15 dB, which was near the maximum obtainable attenuation using the previous methodology (gray open circle). In contrast, the effective target attenuation in the simultaneous complex in the current study ranges over only about 8 dB, from a minimum of 2 dB (black open circle) to a maximum of 10 dB (black filled triangle). In the pure-tone target study, the effective target attenuation in the simultaneous complex ranges over only about 6 dB, from a minimum of 8 dB (gray open triangle) to a maximum of 14 dB (gray filled triangle).

In the original study (shown in gray), we found a breakdown of trading when manipulating spatial cues. When the target spatial cues matched the spatial cues of the simultaneous complex and did not match the spatial cues of the repeating tones, the total effective target intensity was very low with the target contributing very little to either the tone stream or to the simultaneous complex (gray open circle in Fig. 4). However, in the current results, the total effective target intensity is close to the actual physical target intensity for all stimuli, including the corresponding condition in the current study (black open circle). As shown in Figure 3B, the mean difference between the sum of the effective target contributions to the RRP and SRC and the actual target energy is less than 2 dB when the spatial cues in the target match those of the SRC and do not match those of the RRP.

We conclude that increasing the number of target components increases the strength of simultaneous grouping cues, which increases the contribution of the target to a simultaneous complex and decreases the contribution of the target to the sequential stream. As a result of this manipulation, the physical target intensity trades between the RRP and the SRC when spatial cues are manipulated to alter perceptual grouping, unlike in our previous studies using pure-tone targets.

CONCLUSIONS

A direct spectrotemporal matching paradigm can be used reliably by subjects to indicate the perceived content of different objects in a sound mixture.

Spatial cues influence how strongly an ambiguous sound element (the target) contributes both to a sequential stream of similar elements and to a simultaneous complex of harmonically related elements. However, spatial cues generally have a stronger effect on the perceived content of an object that groups sequentially across time (here, the RRP) compared to an object that occurs simultaneously with the target (here, the SRC).

Unlike in similar past studies, in the current study, the target contributes more to one object when it contributes less to the competing object. Differences in the tasks used (e.g., categorization versus matching, etc.) may contribute to the observed differences in the degree to which perceptual trading is observed. However, we believe that these differences are most simply explained by the fact that, in the current study, the perceptual contribution of the target to the simultaneous object was generally larger than in past studies. Specifically, in this study, we used a target composed of multiple tones, which should increase the relative importance of simultaneous grouping cues compared to similar past studies that used a target comprised of a single pure tone (Darwin 1995; Darwin et al. 1995; Lee and Shinn-Cunningham 2008a, b; Shinn-Cunningham et al. 2007). Thus, compared to in previous studies, the target may simply contribute more strongly to the simultaneous harmonic complex and less strongly to the sequential stream. The end result is that the total contribution of the target to the two objects in the sound mixture is larger than in our past studies, and is essentially equal to the physical target intensity in the mixture for all spatial configurations tested. Further tests are necessary to test these alternative explanations.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Office of Naval Research (N00014-04-1-0131) to BGSC. Tim Streeter assisted with hardware and software development and designed and implemented the "friendly feedback" graphical user interface.

REFERENCES

Bregman AS. Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, MA, MIT, 1990.

Cherry EC. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am 25:975–979, 1953.

Darwin CJ. Perceiving vowels in the presence of another sound: A quantitative test of the "Old-plus-new" Heuristic. In: Sorin C, Mariani J, Meloni H, Schoentgen J (eds) Levels in Speech Communication: Relations and Interactions: A Tribute to Max Wajskop. Amsterdam, Elsevier, pp. 1–12, 1995.

Darwin CJ. Auditory grouping. Trends Cogn. Sci 1:327–333, 1997.

Darwin CJ, Hukin RW. Auditory objects of attention: The role of interaural time differences. J. Exp. Psychol. Hum. Percept. Perform 25:617–629, 1999.

Darwin CJ, Hukin RW. Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. J. Acoust. Soc. Am 107:970–977, 2000a.

Darwin CJ, Hukin RW. Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. J. Acoust. Soc. Am 108:335–342, 2000b.

Darwin CJ, Hukin RW, al-Khatib BY. Grouping in pitch perception: Evidence for sequential constraints. J. Acoust. Soc. Am 98:880–885, 1995.

Kohlrausch A, Houtsma AJM, Evans EF. Pitch related to spectral edges of broad-band signals. Philos. Trans. R. Soc. Lond. B Biol. Sci 336:375–382, 1992.

Lee AKC, Shinn-Cunningham BG. Effects of reverberant spatial cues on attention-dependent object formation. J. Assoc. Res. Otolaryngol 9:150–160, 2008a.

Lee AKC, Shinn-Cunningham BG. Effects of frequency disparities on trading of an ambiguous tone between two competing auditory objects. J. Acoust. Soc. Am. in press, 2008b.

McAdams S, Botte MC, Drake C. Auditory continuity and loudness computation. J. Acoust. Soc. Am 103:1580–1591, 1998.

Shackleton TM, Meddis R. The role of interaural time difference and fundamental-frequency difference in the identification of concurrent vowel pairs. J. Acoust. Soc. Am 91:3579–3581, 1992.

Shinn-Cunningham BG, Kopco N, Martin TJ. Localizing nearby sound sources in a classroom: Binaural room impulse response. J. Acoust. Soc. Am 117:3100–3115, 2005.

Shinn-Cunningham BG, Lee AKC, Oxenham AJ. A sound element gets lost in perceptual competition. Proc. Natl. Acad. Sci. U. S. A 104:12223–12227, 2007.