Journal of Experimental Psychology: General
2014, Vol. 143, No. 1, 71–76

0096-3445/14/$12.00 DOI: 10.1037/a0032296

BRIEF REPORT

Attentional Rhythm: A Temporal Analogue of Object-Based Attention

Julian De Freitas, Brandon M. Liverence, and Brian J. Scholl
Yale University

The underlying units of attention are often discrete visual objects. Perhaps the clearest form of evidence for this is the *same-object advantage*: Following a spatial cue, responses are faster to probes occurring on the same object than they are to probes occurring on other objects, while equating brute distance. Is this a fundamentally spatial effect, or can same-object advantages also occur in time? We explored this question using independently normed rhythmic temporal sequences, structured into phrases and presented either visually or auditorily. Detection was speeded when cues and probes both lay within the same rhythmic phrase, compared to when they spanned a phrase boundary, while equating brute duration. This same-phrase advantage suggests that object-based attention is a more general phenomenon than has been previously suspected: Perceptual structure constrains attention, in both space and time, and in both vision and audition.

Keywords: object-based attention, rhythm, music perception, auditory perception

Supplemental materials: http://dx.doi.org/10.1037/a0032296.supp

The most fundamental feature of our visual experience may be space itself, since it appears to be the medium in which all other visual representations exist. Perhaps for this reason, much of the classic work on visual attention was grounded in spatial metaphors (Fernandez-Duque & Johnson, 1999), likening attention to a spotlight or zoom lens (for a review, see Cave & Bichot, 1999). However, more recent research in cognitive science has revealed that a wide range of mental processes, including visual attention, operate over units that are fundamentally discrete.

Object-Based Attention

A perfect example of such processing is *object-based attention*, a class of effects in which discrete visual objects act as units of selection, constraining otherwise equated shifts of spatial attention (for reviews, see Chen, 2012; Scholl, 2001). Perhaps the most direct evidence for this comes from demonstrations of *same-object advantages* in spatial shifts of attention. In a classic demonstration of this effect (Egly, Driver, & Rafal, 1994), observers viewed two vertically oriented rectangles, with their attention cued by a luminance change to an end of one rectangle (see Figure 1A). After a brief delay, a probe appeared at an end of one of the rectangles, and observers pressed a key in response. The cue validly predicted the location of the probe on most trials, but on invalid trials the probe occurred at either the opposite end of the cued rectangle (within object) or at an equidistant point on the neighboring rectangle (between object). There was a same-object advantage: On invalidly cued trials, observers responded faster to within-object than between-object probes. The mechanisms underlying such effects have been explored and debated in dozens of subsequent studies (Chen, 2012; Scholl, 2001).

Structure in Space

Despite the name *object-based attention*, several results suggest that objects per se are not required. This type of effect has also been demonstrated with groups (e.g., Dodd & Pratt, 2005), parts (e.g., Barenholtz & Feldman, 2003; Vecera, Behrmann, & McGoldrick, 2000), surfaces (e.g., He & Nakayama, 1995), and texture flows (e.g., Ben-Shahar, Scholl, & Zucker, 2007)—surely reflecting the same underlying general influence of structure on attention. Similarly, studies of individual cues to objecthood have revealed same-“object” advantages even to types of structure that lack key intuitive features of objects, such as closure (Avrahami, 1999; Marino & Scholl, 2005) and connectedness (Ben-Shahar et al., 2007; Feldman, 2007). We may conclude that “object-based attention” is really a more general phenomenon, in which spatial attention is influenced by visual structure (of many kinds).

Structure in Time

But how general? Object-based attention has traditionally been conceptualized in terms of a specific modality (*visual* structure) and dimension (influencing *spatial* attention). Here we explore whether object-based attention may reflect an even more abstract

This article was published Online First April 15, 2013.
Julian De Freitas, Brandon M. Liverence, and Brian J. Scholl, Department of Psychology, Yale University.
For helpful comments and conversation we thank Mike Dodd, Chaz Firestone, Alexandria Marino, Greg McCarthy, and Ève Poudrier. Readers may visit the authors' website at http://www.yale.edu/perception/
Correspondence concerning this article should be addressed to Brian J. Scholl, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205. E-mail: brian.scholl@yale.edu

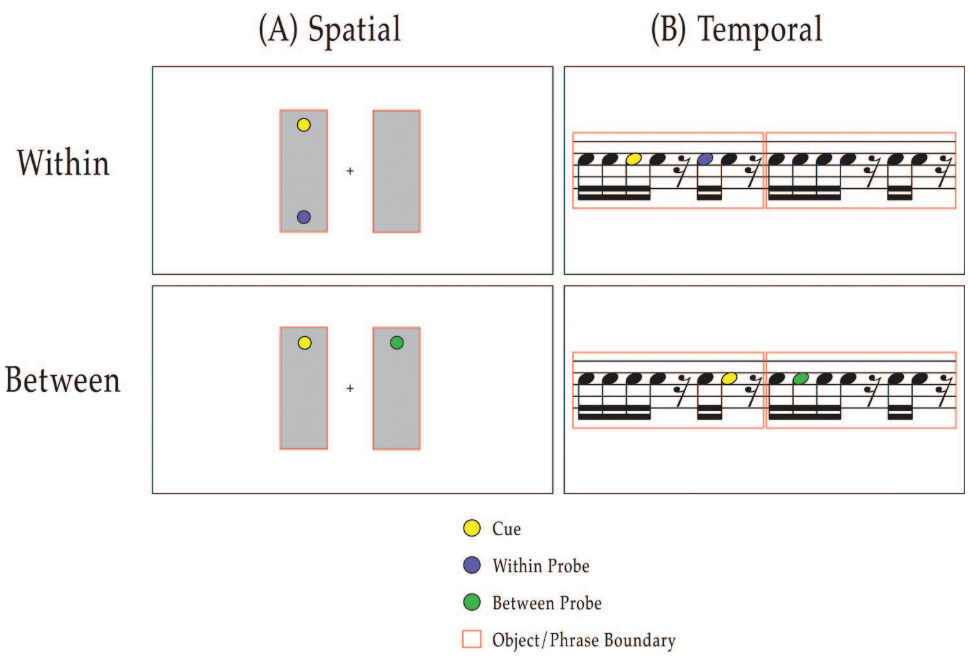

Figure 1. Experimental designs used to explore the influence of discrete representations on selection, in the context of (A) visuospatial object-based attention (Egly et al., 1994) and (B) auditory and visual temporal attention constrained by rhythmic phrases (in the present study).

influence of structure on attention. In particular, we explore whether rhythmic structure in temporal sequences of auditory or visual stimuli influences the allocation of attention through time in the form of a same-"phrase" advantage.

A rich body of research in event perception has emphasized that the mind automatically segments dynamic experience into discrete event representations. For example, when observers are asked to explicitly segment a movie or story at "major boundaries," they show widespread agreement regarding the placement of the boundaries (e.g., Zacks & Tversky, 2001). Event segmentation also influences memory and perception: It is more difficult to remember details from a past event than from an ongoing event (Swallow, Zacks, & Abrams, 2009), and perceived temporal durations are shortened by the presence of event segmentation cues (Liverence & Scholl, 2012).

Previous work also has revealed that segmentation in dynamic scenes can interact in rich ways with attention. For example, target detection in visual stimuli is improved at event boundaries (Newtson & Engquist, 1976), and such boundaries also attract eye movements (Smith, 2012). Auditory sequences can be segmented via cues such as pitch or timbre (for a seminal review see Bregman, 1990), and the resulting "auditory objects" (Griffiths & Warren, 2004) can influence attention in several ways (for reviews see Kubovy & Van Valkenburg, 2001; Shinn-Cunningham, 2008). For example: (a) Paying attention to one feature of an auditory object necessitates attending to its other features as well (Mondor, Zatorre, & Terrio, 1998); (b) it is easier to remember information from one auditory object than from multiple objects (Dyson & Ishfaq, 2008); and (c) observers respond more quickly to target events aligned with the meter of a rhythm (Jones, Moynihan, MacKenzie, & Puente, 2002; Miller, Carlson, & McAuley, 2013).

The Current Study: Attentional Rhythm?

Despite extensive research on the relationship between attention and segmentation in dynamic scenes and on the influence of meter on temporal attention, no previous work to our knowledge has ever investigated whether object-based attention also encompasses temporal sequences in the form of same-object advantages.[1] Here we explore this possibility, using rhythmic structure in auditory tone sequences (Experiment 1) and in visual animations of a moving object (Experiment 2). In particular—and by analogy to the initial demonstrations of same-object advantages in visual cuing (Egly et al., 1994)—we explore whether responses to cued probes are speeded when the cue and probe both occur within the same rhythmic phrase, compared to when they span a phrase boundary (equating duration; see Figure 1B).

[1] One study of attention and segmentation did allude to same-object advantages with auditory sequences (Dyson & Ishfaq, 2008) but used this term to refer to a distinct phenomenon having to do with objects' surface features: Encoding one feature of an object leads to automatic encoding of other (simultaneous) features of that object in memory. By comparison, the present study contrasts temporally extended rhythmic "objects" with absolute durations, allowing us to investigate the influence of such structures on shifts of attention in time. The current stimuli can also be differentiated from other conceptualizations of "auditory objects," for example, in the form of brief tones, with different harmonic relationships to one another, played simultaneously (e.g., Alain, Theunissen, Chevalier, Batty, & Taylor, 2003; Leung, Jolicoeur, Vachon, & Alain, 2011), or alternating tone sequences that perceptually segregate into "auditory streams" (e.g., Bregman, 1990).

Experiments 1A and 1B: Auditory Rhythm-Based Attention

Initial studies of same-object advantages in spatial cuing defined "visual objects" intuitively. This seemed fair, since what could be more intuitively object-like than a simple outlined rectangle? Nevertheless, subsequent research into just what factors define "objecthood" for this purpose has added considerable nuance to this picture (e.g., Feldman, 2007; Scholl, Pylyshyn, & Feldman, 2001). For example, having a closed contour is not necessary (Marino & Scholl, 2005), and same-object advantages can occur even with ungrouped parallel lines (Avrahami, 1999).

Our goal was not to work out which cues define rhythmic phrase boundaries for attention (which would require many subsequent studies) but simply to replicate the spirit of the initial visual study—demonstrating the effect using a case study of rhythmic structure that seemed especially intuitive. Our stimuli consisted of sequences of auditory tones of a single frequency, with tone durations and pause durations arranged to yield rhythmic structures. We created four base rhythmic phrases (see Figure 2), with each individual phrase played repeatedly. Each phrase ended with a brief pause (a rest), which was naturally interpreted as signaling a boundary between phrase repetitions (i.e., between the moments when the current phrase ended and the next phrase began). This was always matched by an intraphrase rest (of identical duration) that was naturally interpreted as part of the ongoing phrase itself.

Despite their intuitiveness, we first normed each of the rhythmic phrases, to ensure that subjects agreed about the temporal locations of the phrase boundaries. These methods and results are described in the online supplemental materials and are summarized in Figure 2—where the blue shading (with horizontal extent indicating 95% confidence intervals) indicates where naïve observers judged the phrase boundaries to be. As can be appreciated from the figure, the points where subjects indicated phrase boundaries were impressively consistent and mirrored our initial intuitions.

Next, we asked whether rhythmic phrases act as units of temporal attention in the same way that objects act as units of spatial attention. To do so, we utilized a paradigm modeled on the classic "two-rectangles" study (Egly et al., 1994), but substituting time for space (see Figure 1B). Whereas prior work used a cue to indicate the upcoming spatial location of a probe, we used a cue (a note from the phrase, played at a higher than usual frequency) to indicate the upcoming onset of a probe (a second note played at a different higher frequency). For between-phrase probes, the cue and probe spanned a phrase boundary (instead of spatial object contours), whereas for within-phrase probes, the cue and probe lay within the same phrase (instead of the same rectangle). As those prior studies equated brute cue–probe spatial distance, we equated brute cue–probe duration across within-phrase and between-phrase conditions (while also equating the number and duration of all rests). But whereas those classic studies employed mostly valid cues (so as to include more unpredictability about where the probe would appear), we used multiple cue–probe durations (so as to include more unpredictability about *when* the probe would appear, with this degree of unpredictability always equated across within-phrase and between-phrase conditions).

Method

Subjects. In Experiment 1A, 12 subjects (students and other members of the Yale and New Haven, CT community; five male and seven female, mean age 21.5 years) participated in exchange for course credit or $10. In the absence of any previous studies that had used the paradigm created for the present experiments, we began with the heuristic assumption that the resulting effect size would be comparable to visuospatial same-object advantages, for which several previous studies had used similar sample sizes (e.g., Dodd & Pratt, 2005; Egly et al., 1994).

Stimuli and procedure. Auditory stimuli were created in MATLAB using the Psychtoolbox libraries (Brainard, 1997; Pelli, 1997) and played on a Macintosh computer through headphones. On each trial subjects heard 48 repetitions of a

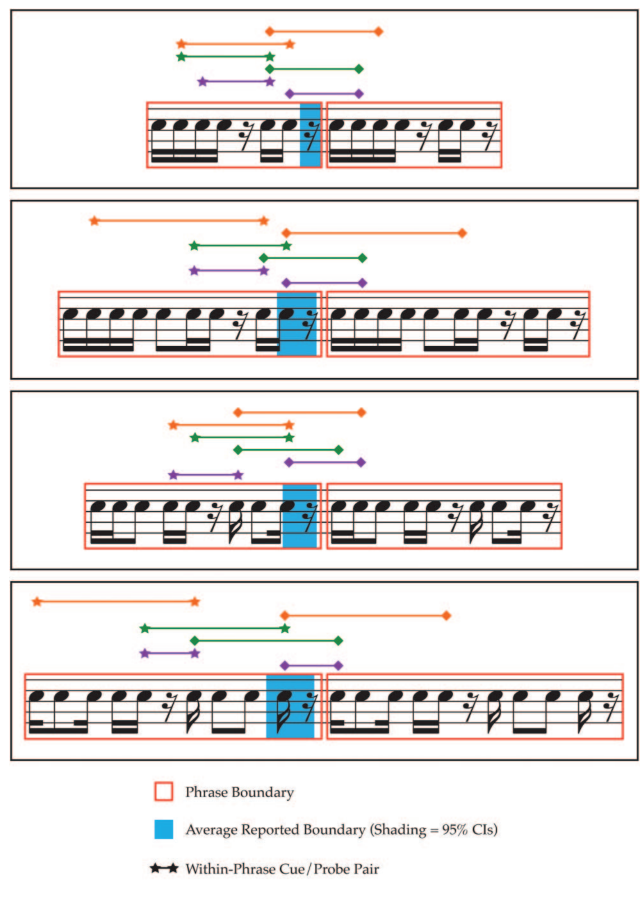

Figure 2. The four rhythmic phrases used in the present experiments, each repeated twice, expressed in standard musical notation. The red boxes indicate the phrase boundaries as defined intuitively in the design of the experiment, whereas the blue shading indicates naïve subjects' reported judgments of where the boundaries occurred while listening to (or viewing) the rhythms (with the horizontal extent of the shading indicating 95% confidence intervals [CIs] surrounding the mean). For each rhythm, the stars indicate the three possible within-phrase cue/probe pairs, and the diamonds indicate the three possible between-phrase cue/probe pairs, with matching colors indicating matched pairs in terms of brute duration.

randomly chosen one of the four rhythmic phrases (for trial durations of approximately 133.9, 155.6, 139.8, and 170.7 s, respectively). To ensure that responses were not influenced by which portion of the phrase subjects heard first, the phrase gradually faded in from one of four potential starting points (counterbalanced across subjects—and also identical to the starting points used during the initial rhythm norming).[2] On average, the fade-in period lasted 7.4 s (approximately 2.4 phrase repetitions), during which the volume linearly increased from silence to full volume.

Cue/probe pairs occurred 12 times per trial, during randomly chosen repetitions of the phrase. Half of the cue/probe pairs spanned a phrase boundary (thus using up two phrase repetitions), while the other half occurred within a single phrase repetition. The exact timings of cues and probes were different for each phrase and are depicted in Figure 2 by stars (for within-phrase pairs) and diamonds (for between-phrase pairs). Of the 12 cue/probe pairs per trial, each cue–probe duration was used four times (twice for within-phrase pairs, twice for between-phrase pairs).

Cue and probe notes were identical to the other notes, except that they were played at pitches higher than the baseline (279 Hz): On 50% of the cue/probe presentations, the cue was played at 314 Hz and the probe at 332 Hz, and on the other 50% these values were reversed. Despite the complexity of this design (necessary to ensure careful balancing), the task was exceedingly simple: Subjects just listened for pairs of tones with higher than usual pitches, and pressed a key as soon as they heard the second such tone (i.e., the probe). Each session consisted of eight experimental trials, with two separate trials for each of the four rhythmic phrases (in a different random order for each subject). These were preceded by a single, unrecorded practice trial (using a randomly chosen phrase).

As illustrated in Figure 2, between-phrase probes nearly always occurred closer to the beginning of the phrase than within-phrase probes—a regularity necessitated in practice by the fact that we controlled for absolute cue–probe duration. We know of no reason to think that absolute probe placement of this sort would influence detection speed, and our stimuli did not yield any experience of tone capture (e.g., Bregman & Rudnicky, 1975). Nonetheless, to rule out this possibility, 12 additional subjects participated in Experiment 1B (eight male and four female, mean age 22.4 years), which was identical except that no cues were presented.

Results and Discussion

Responses were excluded from analysis if they occurred before a target onset (9%) or from a phrase repetition without a probe (1%). There was a significant same-phrase advantage in Experiment 1A: Subjects responded faster to within-phrase probes than to between-phrase probes (427 ms vs. 449 ms), $t(11) = 3.29$, $p < .01$, $d = 0.95$. No such difference was observed in Experiment 1B (656 ms vs. 653 ms), $t(11) = 0.307$, $p = .76$, $d = 0.09$, with the magnitude of the same-phrase advantage (or lack thereof) differing significantly by experiment (22.13 ms vs. −2.96 ms), $t(22) = 2.136$, $p = .04$, $d = 0.87$, indicating that the same-phrase advantage in Experiment 1A was not driven by absolute probe position (which was equated in the two experiments).

Because we held cue–probe durations constant between conditions, within-phrase cues necessarily occurred earlier on average than between-phrase cues. It is possible that systematic differences in cue location could have led to anticipation effects that might have contributed to the observed same-phrase advantage. For example, since cues occurring very early in a phrase are fully predictive of within-phrase probes, subjects might have been better able to prepare for these probes than for those occurring in the middle of the phrase (where cues could signal either a within-phrase or a between-phrase probe). While this could contribute to a same-phrase advantage, it could not explain it away, since the reverse is also true for very late cues (vs. phrase-middle cues) in the between-phrase condition. In other words, while it's true that some cues are more predictive than others about when the probes will arrive (only relative to the phrase boundaries, of course), those cues are equally split between the within-phrase and between-phrase trials—and so they cannot explain the same-phrase advantage results, in principle. Nevertheless, we also analyzed our data for such effects in practice: To determine whether such anticipation effects were present in our study, we conducted linear regressions (separately for each condition) with normalized cue position as an independent variable and response time as the dependent variable. If phrase-relative cue position influences response times, then there should be a positive correlation between normalized cue position and response time in the within-phrase condition (e.g., early cues causing the fastest responses) and a negative correlation in the between-phrase condition (e.g., late cues causing the fastest responses). Contrary to these predictions, cue position did not significantly predict response time in either the within-phrase condition ($R^2 < .01$; $\beta = .01$), $t(11) = 0.05$, $p = .965$, or the between-phrase condition ($R^2 = .18$; $\beta = .42$), $t(11) = 1.46$, $p = .175$.

Experiments 2A and 2B: Visual Rhythm-Based Attention

Having established that rhythmic segmentation can influence temporal attention in audition, we next asked whether this effect generalizes to vision. These new experiments utilized the same design as Experiments 1A and 1B but replaced sequences of tones with rhythmic sequences of an animated bar (see Grahn, 2012).

Method

Subjects. Twelve new subjects (five male and seven female, mean age 19.1 years) participated for course credit or $10. This sample size was chosen to match that of Experiment 1A and can be justified post hoc based on the results of that experiment: If we heuristically assume that any resulting same-phrase advantage will have a similar effect size, then a power analysis reveals that we would need 11 subjects to find a significant effect (with $\alpha = .05$ and a desired power of .80).

Stimuli and procedure. These experiments were identical to Experiments 1A and 1B, except as noted here. Stimuli consisted of an animated gray bar (red [R] = 50, green [G] = 50, blue [B] = 50) with a small crossbar for fixation, presented against a black

[2] The fade-in began with the following notes for each pair of rhythmic phrases (see Figure 2). Rhythm 1: 1, 8, 12, and 14 (the second note of the third phrase repetition, not depicted). Rhythm 2: 1, 8, 12, 14. Rhythm 3: 1, 7, 11, 13. Rhythm 4: 1, 7, 10, 12.

background. The bar "seesawed" up and down, switching orientations at every new tone in the rhythm (such that the line oscillated between +18° and −18° from horizontal). The rhythms were played slower than before (for durations of approximately 3.47, 3.95, 3.47, and 4.05 s, respectively), to account for the relatively poor temporal resolution of vision (e.g., Guttman, Gilroy, & Blake, 2005; Repp & Penel, 2002). The cues and probes were brief, bright flashes of the bar; for 50% of instances, the cue flashed 2.4 times brighter (R = 120, G = 120, B = 120) and the probe 2.6 times brighter (R = 130, G = 130, B = 130) than usual, and for the other 50%, these values were reversed. The bar gradually faded in at the start of the trial, as in Experiments 1A and 1B. To control for absolute probe placement, an additional 12 subjects (two male and 10 female, mean age 20.7 years) participated in Experiment 2B, in which no cue flashes were presented.

Results and Discussion

Responses were excluded from analysis if they occurred before a target onset (9%) or from a phrase repetition without a probe (none in this experiment). There was a same-phrase advantage in Experiment 2A (402 ms vs. 417 ms), $t(11) = 3.10$, $p = .01$, $d = 0.89$, but no reliable difference in Experiment 2B, and even a slight trend in the opposite direction (554 ms vs. 541 ms), $t(11) = 1.69$, $p = .12$, $d = 0.49$, with these two differences themselves differing significantly (14.71 ms vs. −13.20 ms), $t(22) = 3.05$, $p = .006$, $d = 1.25$. Phrase-relative cue position did significantly predict response time in the between-phrase condition, with later cues associated with longer response times ($R^2 = .40$; $\beta = .63$), $t(11) = 2.56$, $p = .029$. This could not account for same-phrase advantages without appeal to the within- versus between-phrase manipulation, however, since phrase-relative cue position did not predict response time in the within-phrase condition ($R^2 = .01$; $\beta = -.07$), $t(11) = 0.23$, $p = .822$. And in any case, as with the auditory experiments, any differential predictability associated with phrase-relative cue timing was equally distributed across within-phrase and between-phrase trials.

General Discussion

One of the clearest demonstrations that objects can act as units of visual selection is the same-object advantage in spatial cuing: Responses to probes on the same object as a cue are prioritized over responses to spatially equated probes on other identical nearby objects. Here we report a novel, and entirely analogous, phenomenon in the temporal domain, using both auditory and visual stimuli: Responses to probes in the same rhythmic phrase as a cue are prioritized over responses to temporally equated probes in other identical adjacent rhythmic phrases. Although these results are to our knowledge the first demonstration of the same-object advantage in audition, more important for present purposes is that they are (also) the first such demonstration over durations of *time* in any modality.

And just as further studies of visuospatial object-based attention have exploited that phenomenon to begin working out the precise cues that define objecthood, so too may future studies employ the present phenomenon to help work out the precise cues that define "rhythmhood," perhaps in synergy with music cognition research (e.g., Lerdahl & Jackendoff, 1983). For now we conclude that object-based visuospatial attention may not require objects, may not be restricted to visual processing, and may not even be fundamentally spatial. Instead, it may reflect a broader phenomenon in which perceptual structure (in vision or audition, in space or time) constrains attention.

References

Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., & Taylor, M. J. (2003). Developmental changes in distinguishing concurrent auditory objects. *Cognitive Brain Research, 16,* 210–218. doi:10.1016/S0926-6410(02)00275-6

Avrahami, J. (1999). Objects of attention, objects of perception. *Perception & Psychophysics, 61,* 1604–1612. doi:10.3758/BF03213121

Barenholtz, E., & Feldman, J. (2003). Visual comparisons within and between object parts: Evidence for a single-part superiority effect. *Vision Research, 43,* 1655–1666. doi:10.1016/S0042-6989(03)00166-4

Ben-Shahar, O., Scholl, B. J., & Zucker, S. W. (2007). Attention, segregation, and textons: Bridging the gap between object-based attention and texton-based segregation. *Vision Research, 47,* 845–860. doi:10.1016/j.visres.2006.10.019

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Bregman, A. S., & Rudnicky, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1,* 263–267. doi:10.1037/0096-1523.1.3.263

Cave, K. R., & Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. *Psychonomic Bulletin & Review, 6,* 204–223. doi:10.3758/BF03212327

Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics, 74,* 784–802. doi:10.3758/s13414-012-0322-z

Dodd, M. D., & Pratt, J. (2005). Allocating visual attention to grouped objects. *European Journal of Cognitive Psychology, 17,* 481–497. doi:10.1080/09541440440000122

Dyson, B. J., & Ishfaq, F. (2008). Auditory memory can be object based. *Psychonomic Bulletin & Review, 15,* 409–412. doi:10.3758/PBR.15.2.409

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123,* 161–177. doi:10.1037/0096-3445.123.2.161

Feldman, J. (2007). Formation of visual "objects" in the early computation of spatial relations. *Perception & Psychophysics, 69,* 816–827. doi:10.3758/BF03193781

Fernandez-Duque, D., & Johnson, M. L. (1999). Attention metaphors: How metaphors guide the cognitive psychology of attention. *Cognitive Science, 23,* 83–116. doi:10.1016/S0364-0213(99)80053-6

Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Experimental Brain Research, 220,* 51–61. doi:10.1007/s00221-012-3114-8

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5,* 887–892. doi:10.1038/nrn1538

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science, 16,* 228–235. doi:10.1111/j.0956-7976.2005.00808.x

He, Z. J., & Nakayama, K. (1995). Visual attention to surfaces in 3-D space. *Proceedings of the National Academy of Sciences USA, 92,* 11155–11159. doi:10.1073/pnas.92.24.11155

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13,* 313–319. doi:10.1111/1467-9280.00458

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80,* 97–126. doi:10.1016/S0010-0277(00)00155-4

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Leung, A. W. S., Jolicoeur, P., Vachon, F., & Alain, C. (2011). The perception of concurrent sound objects in harmonic complexes impairs gap detection. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 727–736. doi:10.1037/a0021932

Liverence, B. M., & Scholl, B. J. (2012). Discrete events as units of perceived time. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 549–554. doi:10.1037/a0027228

Marino, A. C., & Scholl, B. J. (2005). The role of closure in defining the "objects" of object-based attention. *Perception & Psychophysics, 67,* 1140–1149. doi:10.3758/BF03193547

Miller, J. E., Carlson, L. A., & McAuley, J. D. (2013). When what you hear influences when you see: Listening to an auditory rhythm influences the temporal allocation of visual attention. *Psychological Science, 24,* 11–18. doi:10.1177/0956797612446707

Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 66–79. doi:10.1037/0096-1523.24.1.66

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12,* 436–450. doi:10.1016/0022-1031(76)90076-7

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1085–1099. doi:10.1037/0096-1523.28.5.1085

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46. doi:10.1016/S0010-0277(00)00152-9

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80,* 159–177. doi:10.1016/S0010-0277(00)00157-8

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12,* 182–186. doi:10.1016/j.tics.2008.02.003

Smith, T. (2012). The relationship between overt attention and event perception during dynamic social scenes [Abstract]. *Journal of Vision, 12*(9), article 407. doi:10.1167/12.9.407

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138,* 236–257. doi:10.1037/a0015631

Vecera, S. P., Behrmann, M., & McGoldrick, J. (2000). Selective attention to the parts of an object. *Psychonomic Bulletin & Review, 7,* 301–308. doi:10.3758/BF03212985

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3–21. doi:10.1037/0033-2909.127.1.3

Received November 9, 2012
Revision received February 18, 2013
Accepted February 18, 2013 ■

Supplemental Information: Norming Rhythm Boundaries

[De Freitas et al. (2013), *JEP:General*]

JULIAN DE FREITAS, BRANDON M. LIVERENCE, & BRIAN J. SCHOLL
Yale University

In order to confirm that naïve observers would perceive boundaries in our rhythmic stimuli in the same places that we did when designing them, we normed these stimuli in an independent group of subjects. While listening to each repeating phrase, subjects simply pressed a key whenever they perceived a boundary between repetitions of the phrase.

METHOD

Subjects

144 subjects (students and other members of the Yale and New Haven community) participated in exchange for $1. We chose this relatively large number of subjects since (a) each subject made only a few responses in a between-subjects design; (b) we were using a type of measure and stimuli here that to our knowledge had never before been used (so that we had no a priori guide as to what sample size would be required, unlike the experiments reported in the main text); and (c) this experiment was only meant for norming stimuli for the primary experiments, rather than as a test for our hypothesis about same-'object' advantages in time rather than space.

Stimuli

Auditory stimuli were created in MATLAB using the PsychToolbox libraries (Brainard, 1997; Pelli, 1997), and played on a Macintosh computer through headphones. Subjects listened to repeating sequences of 279 Hz tones (at moderate volume), arranged into four rhythmic phrases (approximately 2.78, 3.24, 2.91, and 3.56 s, respectively), with each subject hearing only one of the four phrases. The four phrases are each depicted in musical notation in Figure 2 of the main text. To ensure that responses were not influenced by which portion of the phrase subjects heard first, the phrase gradually faded in from one of four potential starting points (counterbalanced across subjects). On average, the fade-in period lasted 7.4 s (approximately 2.4 phrase repetitions), during which the volume linearly increased from silence to full volume. The fade-in began with the same notes for each pair of rhythmic phrases as was noted in Footnote 2 of the main text (see also Figure 2).

Procedure and Design

Subjects were asked to press a key at the end of each repetition of the rhythmic phrase. (Because it was also possible to hear 'sub-phrases' — just as visual objects can have parts — the instructions clarified that keypresses were only to be made at the end of each "largest pattern that you hear repeating".) Subjects first listened to eight practice repetitions of the rhythmic phrase, after which a brief high-pitched (450 Hz) tone signaled when they should begin responding. So that the tone did not by itself yield a segmentation cue, it occurred at one of three possible points (counterbalanced across subjects) during the last two practice repetitions. The phrase then repeated six more times, while keypresses were recorded.

RESULTS AND DISCUSSION

36 subjects participated for each of the 4 rhythmic phrases. Since the task featured only six repetitions of a single phrase per subject, we excluded 16 additional subjects for making either fewer than four or greater than eight keypresses total. The average moment of the keypresses is depicted in Figure 2 of the main text via blue shading (horizontal extent indicates 95% confidence intervals). As can be appreciated from the figure, the points where subjects indicated phrase boundaries were impressively consistent, and mirrored our initial intuitions. Statistically, the average offset between keypress and Between-Phrase rest was only 512ms, whereas the average offset between keypress and Within-Phrase rest was almost twice as long (959ms) — a highly significant difference ($t(143)=5.68$, $p<.001$, $d=.47$).