A sound element gets lost in perceptual competition

Barbara G. Shinn-Cunningham*†‡, Adrian K. C. Lee*†, and Andrew J. Oxenham§

*Boston University Hearing Research Center, 677 Beacon Street, Boston, MA 02215; †Speech and Hearing Bioscience and Technology Program, Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, 77 Massachusetts Avenue, Cambridge, MA 02139; and §Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, MN 55455

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved June 5, 2007 (received for review May 18, 2007)

Our ability to understand auditory signals depends on properly separating the mixture of sound arriving from multiple sources. Sound elements tend to belong to only one object at a time, consistent with the principle of disjoint allocation, although there are instances of duplex perception or coallocation, in which two sound objects share one sound element. Here we report an effect of ''nonallocation,'' in which a sound element ''disappears'' when two ongoing objects compete for its ownership. When a target tone is presented either as one of a sequence of tones or simultaneously with a harmonic vowel complex, it is heard as part of the corresponding object. However, depending on the spatial configuration of the scene, if the target, the tones, and the vowel are all presented together, the target may not be perceived in either the tones or the vowel, even though it is not perceived as a separate entity. This finding suggests an asymmetry in the strength of the perceptual evidence required to reject vs. to include an element within the auditory foreground, a result with important implications for how we process complex auditory scenes containing ambiguous information.

auditory scene analysis | attention | auditory objects | spatial hearing | streaming

Many species, including birds (1), frogs (2), and mammals (3), must hear out important communication calls from a background of competing sounds to procreate and survive. Whether in a raucous penguin colony in Antarctica (4) or a crowded cocktail party in Europe (5), listeners are adept at analyzing the acoustic mixture to determine what sound sources are present.

Successful sound-source identification requires that the individual sound elements within a mixture be assigned to the correct ''auditory objects.'' Many spectrotemporal features in the sound mixture promote grouping of sound elements into auditory objects, including common onsets, common amplitude modulation, harmonicity, continuity over time, frequency proximity, and common spatial cues such as interaural time differences (6, 7). Listener experience and expectations can also influence how the scene is analyzed, suggesting that ''top-down'' processes interact with low-level ''bottom-up'' stimulus features in auditory object formation (8–10).

The perceptual grouping principle of exclusive or disjoint allocation states that a single sound element, such as a pure tone, cannot be assigned simultaneously to more than one auditory object (6). Although this principle has fairly general applicability, there are some exceptions. For instance, a frequency glide, presented to the opposite ear from the rest of a speech sound, can influence the perceived phonetic content of the speech sound while at the same time being heard as a separate object (11). Similarly, a mistuned harmonic within a harmonic complex tone can be heard as a separate tone while at the same time influencing the perceived pitch of the overall complex tone (12). These situations, in which a sound element contributes to more than one auditory object, are examples of duplex perception.

The term ''duplex perception'' suggests that a single sound element can be assigned independently to more than one object. However, a more parsimonious explanation may be that, in fact, the energy of the element can simply be shared between sound objects. Physically, if a frequency component is present in two independent sound sources then, on average, the total energy of that frequency in the mixture should equal the sum of the energies in the constituent sound sources. Thus, a veridical perceptual representation would divide the total sound energy of each frequency component across the two objects. Although many past studies have considered the question of trading, few have explicitly measured the perceptual contribution of the target to both competing objects (13–16).

Here we adapt an earlier paradigm (15) to assess directly the relative contribution of a pure-tone element to each object. This technique allows us to quantify the degree to which perceptual trading of energy holds when two objects compete for a sound element. We generated rhythmically repeating stimuli consisting of two auditory objects: a sequence of rapidly repeating tones and a synthetic vowel, repeating at a slower rate (see Fig. 1*A*). In this mixture, an ambiguous tone, known as the target, could logically be a member of each (or both) of the two perceived objects, either as another tone in the repeating sequence of tones or as the fourth harmonic in the vowel. Importantly, the formation of the tones stream depends primarily on perceptual organization across time, in which spatial cues are known to have a large influence (17), whereas the organization of the vowel depends primarily on a local spectrotemporal structure, where spatial cues should have a weaker effect (6).

We manipulated the spatial cues of the sound elements in the mixture and measured how perceptual organization was affected. Identical stimuli were presented in two separate blocks, one in which listeners attended to the tones and one in which they attended to the vowel. In each block, we measured whether the target was perceived as part of the attended object. When attending to the tones, listeners identified the rhythm of the stream. If the target was perceived as part of the tones stream, then the perceived rhythm was even; otherwise, it was galloping. Similarly, listeners identified the perceived vowel category, which depended on whether or not the target was perceived as part of the harmonic complex (9, 18). If the target was heard as part of the vowel, it was labeled /ɛ/ (as in ''bet''); when the target was not part of the vowel, it was labeled /I/ (as in ''bit''; see Fig. 1*B* and *Methods*). The spatial cues in the target could either match or differ from the spatial cues in the tones and the vowel (see Fig. 1*C Left* and *Methods*) to either promote or discourage grouping of the target with the attended object. Intermingled control conditions presented single-object prototype stimuli in which only the attended object was presented (with and without the target; see Fig. 1*C Right*) to confirm that listeners were able to consistently label unambiguous stimuli properly with the

Author contributions: B.G.S.-C., A.K.C.L., and A.J.O. designed research; A.K.C.L. performed research; B.G.S.-C. and A.K.C.L. analyzed data; and B.G.S.-C., A.K.C.L., and A.J.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

‡To whom correspondence should be addressed. E-mail: shinn@cns.bu.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0704641104/DC1.

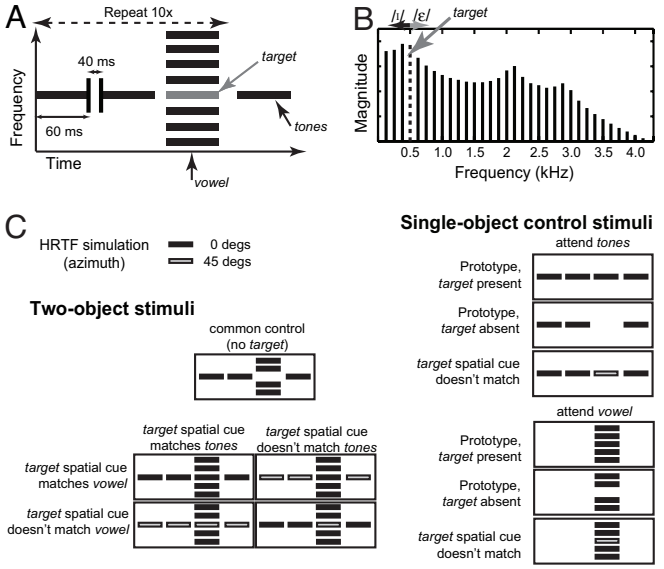

Fig. 1. Stimuli and conditions. (*A*) Time-frequency schematic of stimuli. Stimuli consist of a 3-s-long repeating sequence of 500-Hz tones, a synthetic vowel with fundamental frequency of 125 Hz, and a 500-Hz target that logically could group with either object. (*B*) Spectra shape of vowel. The spectrum of the synthetic vowel was shaped to sound like /I/. If the target tone was perceived as part of the vowel, the perceived vowel category shifted toward /ɛ/. (*C*) Spatial configurations of stimuli. The simulated spatial locations of the target and repeating tones varied across conditions. Control conditions included a case without a target and conditions in which there was only one attended object.

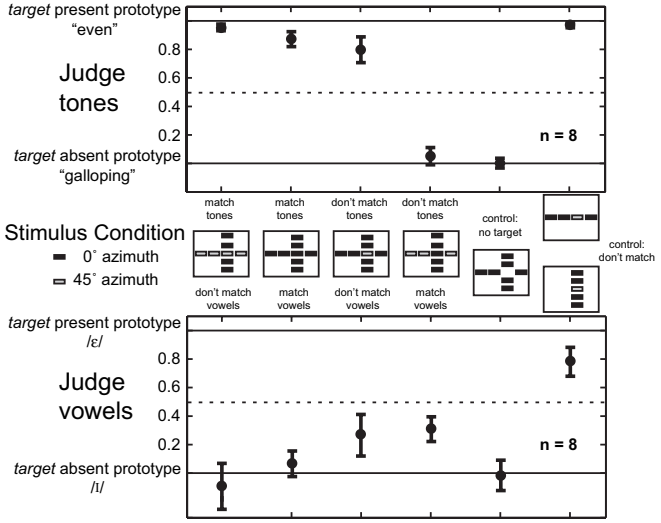

Fig. 2. Spatial cues have a large effect on perception of the tones but have very little effect on perception of the vowel. Across-subject averages of the relative perceptual distance of each condition to single-object prototypes are shown (error bars show the standard error of the mean).

even/galloping or ɛ/I labels. Finally, another control condition presented the two competing objects without any target to ensure that attended objects in this two-object condition were not perceived as if they contained target energy.

To obtain a more direct interpretation of the category responses in the main experiment, an auxiliary experiment was conducted with single-object stimuli. The same categorical judgments (even vs. galloping or /ɛ/ vs. /I/) were measured when only one object was present (either the tones or the vowel) and the level of the target was varied systematically, with attenuations ranging from 0 dB (no change) to 14 dB. This procedure allowed us to quantify how the response probabilities for the two categories (target-present and target-absent) mapped to the physical energy present in the target. In turn, the individual-subject results from this auxiliary experiment enabled us to map the response percentages in the main experiment to an "effective target energy" (the level of the target, in dB, that would lead to the observed response percentages in the auxiliary, single-object experiments).

Results

For both the tones and vowel single-object prototype-control stimuli in the main experiment, subjects responded as if the target was part of the object, whether or not the target location matched that of the attended object. Thus, mismatching target and object spatial cues are insufficient to perceptually remove the target from either the tones or the vowel object if there is no perceptual competition for the target.

The percentage of "even" vs. "galloping" and /ɛ/ vs. /I/ responses in the different two-stream conditions were compared with the corresponding percentages for the prototype-control stimuli in the main experiment (either the tones or vowels in isolation) with and without the target present. We extended traditional signal-detection theory approaches to determine whether the responses to an ambiguous, two-source stimulus were closer to responses for a single-source stimulus with or without the target present (see *Methods*). Fig. 2 uses the relative perceptual distance between the ambiguous stimuli and the target-absent prototype to summarize responses. A value of 0 indicates that response percentages for a particular condition equalled the response percentages for the target-absent prototype. A value of 1 indicates that the stimulus was perceived like a spatially unambiguous target-present prototype (i.e., a single-object stimulus in which the target location matched that of the attended vowel or tones; see Fig. 2 and *Methods*).

When asked to judge the rhythm of the tones in the presence of the vowels, the spatial cues had a large effect, in line with earlier studies showing a large influence of spatial cues when grouping sounds across time (17, 19). The target contributed strongly to the tones sequence whenever the simulated target location matched that of the tones, regardless of the vowel location. When the target location matched neither that of the tones nor that of the vowel, subjects still perceived the target as part of the tones sequence. However, when the target location matched that of the vowel but not the tones, the target was no longer perceived as part of the tones sequence, and listeners heard a galloping rhythm. Thus, spatial cues are strong enough to overcome other grouping cues and perceptually remove the target from the attended tones stream. However, this only occurs in the most extreme case when the spatial cues in the target match those in the background vowel and do not match those in the attended foreground tones.

When asked to identify the perceived vowel in mixtures containing competing tones, spatial cues had a much less pronounced effect. Moreover, listeners never heard the target as strongly present in the vowel (they consistently responded /I/ more often than /ɛ/, and the response percentages were much more similar to the response percentages for the /I/ prototype than for the /ɛ/ prototype). Even when the spatial cues in the target matched those in the vowel and not the tones, the attended vowel was perceptually more like the /I/ (without the target) than /ɛ/ (with the target; see fourth condition from the left in Fig. 2 *Bottom*).

The results so far suggest that the contribution of the target to the tones does not predict its contribution to the vowel, in apparent contradiction to the trading hypothesis outlined in the introduction. However, if the perceptual judgments of galloping–

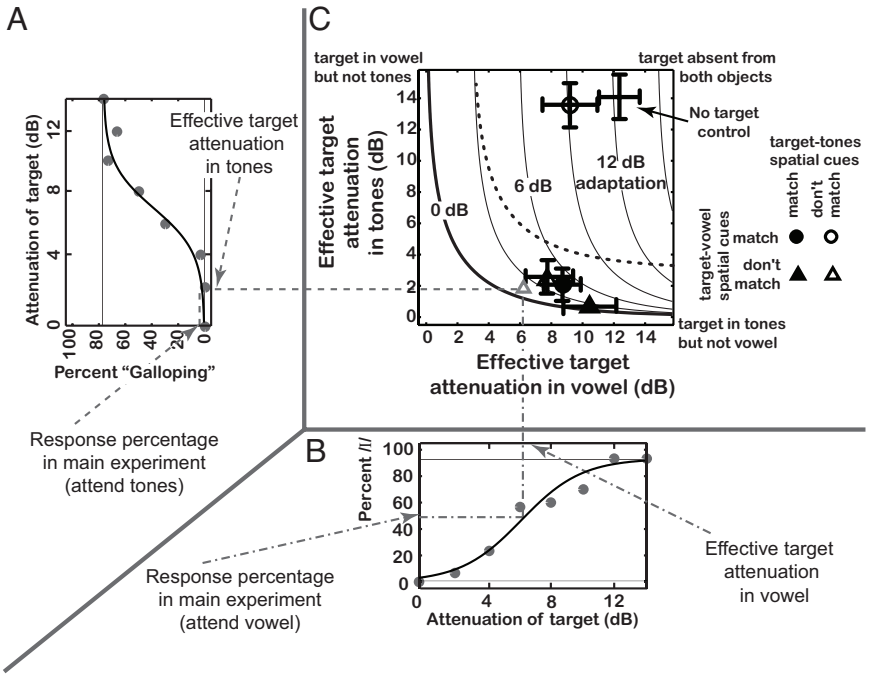

Fig. 3. The perceived target energy in the two objects does not always account for all of the physical target energy present in the mixture. (*A* and *B*) Example psychometric functions map the response percentages from the main experiment to effective target attenuations for the tones (*A*) and the vowel (*B*). These mappings are used to derive data plotted in *C*. (*C*) Effective target attenuation when listeners attend to the tones vs. when they attend to the vowel. Across-subject means are plotted, with error bars showing the standard error of the mean. The solid thick line shows the expected trading relationship when the perceived energy accounts for the physical target energy. Thin lines show the trading relationships that would occur if peripheral adaptation reduces the contribution of the target to the vowel (shown for 3-dB increments). The dashed line shows the predictions if target amplitude, rather than energy, trades.

even and /I/–/ε/ have a different dependence on the target level (e.g., if the category boundary is steep as a function of target level for one judgment and shallow for the other), then the response category percentages in the two tasks will not trade quantitatively, even if the energy-trading hypothesis holds. The auxiliary experiment allowed us to quantify the degree to which the results in the first experiment obeyed the trading hypothesis. Using results from the auxiliary experiment, we calculated the effective intensity of the target corresponding to the raw response percentages (i.e., even–galloping response percentage for the tones or /ε/–/I/ response percentage for the vowel) in the main experiment. The percentage of trials in the auxiliary experiment in which listeners responded even vs. galloping (in the tones condition) or /ε/ vs. /I/ (in the vowel condition) provided a subject-specific mapping between target attenuation and a corresponding response percentage. These psychometric functions relating response percentages to the physical attenuation of the target were generally well behaved; increasing target attenuation systematically increased the probability that the listeners responded as if the target were absent from the attended object (see Fig. 3 *A* and *B* for example psychometric functions from the tones and vowels control experiments, respectively).

For each subject and stimulus condition, we mapped the percent responses obtained in the main experiment to an "effective target attenuation" and compared the resulting effective attenuations for physically identical stimuli in the attend-tones and attend-vowel blocks (see Fig. 3, in which *A* and *B* demonstrate how the effective attenuations of the target are obtained for one subject, listening to either the tones or the vowel, for one example condition; these values are projected to the ordinate and abscissa in *C*, respectively, producing the open gray triangle).

A pure energy trading relationship for the target's contribution to the vowel and the tones would produce data points that lie along the solid thick curve in Fig. 3*C*. Data would fall along the dashed line if an amplitude-trading relationship holds (15, 16). If the tones caused an effective attenuation of the target that reduced its contribution to the vowels (20, 21) through, for instance, neural adaptation, the data would fall along one of the family of curves shown by the thin lines in Fig. 3*C* (see *Discussion*). None of these predictions can fully account for our results.

When the spatial location of the target matched that of the tones but not the vowel (Fig. 3*C*, filled triangle), the target was perceived almost exclusively as part of the tones sequence, in line with expectations that are based on energy trading. When the spatial location of the target matched that of both the tones and the vowel (filled circle) or matched neither (open triangle), there was a tendency to assign more of the target to the vowel and less to the tones (i.e., the effective target attenuation decreases in the vowel task and increases in the tones task). Results for these two conditions can be fit well by assuming that the tones cause adaptation that effectively reduces the target level by ≈4 dB. However, when the spatial location of the target matched that of the vowel but not the tones (Fig. 3*C*, open circle), the effective level of the target was attenuated by 9 dB or more both when the listeners attended to the tones and when they attended to the vowel. In fact, in both tasks, responses were similar to the responses to the control condition in which the target was physically absent (compare open circle and cross in Fig. 3*C*).

Discussion

We find that in situations of perceptual competition, the perceptual coherence between a sound element (the target) and one object (the vowel) can be sufficient to prevent the target from binding with another object (the tones) but still insufficient to bind the element with the first object. In one of our conditions, this results in the target element falling into a form of perceptual limbo, where it belongs to neither competing object. The finding

provides an interesting counterpart to duplex perception, or coallocation, whereby a single sound element contributes to two perceptual objects at once (12, 22). In contrast, we observe nonallocation, whereby the element does not strongly contribute to either object and is also not heard as an independent object. It is important to note, however, that the target is not undetectable: subjects can easily distinguish between sequences that contain the target and those that do not, even when the target fails to bind with either tones or vowel. We next consider some possible explanations for this effect.

Certain forms of neural adaptation may contribute to our results. If the preceding sequence of tones reduces the target's "internal" level, the target contribution to the vowel will be reduced. However, because all of the tones would be perceived at the same level as the adapted target, the contribution of the target to the tones would be unaffected. Thus, if adaptation were the only effect present, a skewed form of energy trade would occur and results would fall along one of the thin lines in Fig. 3*C*. Data for three conditions are consistent with peripheral adaptation reducing the effective target level by ≈4 dB. However, this level of adaptation cannot account for the condition in which the target spatial cues match those of the vowel but not the tones. Moreover, many studies have shown that the loudness of targets is not reduced by preceding tones of the same intensity (like those in our experiment), making it unlikely that our results are due solely to an internal attenuation of the target (23–25). In addition, earlier similar studies have also concluded that adaptation cannot account for the effects of a preceding sequential stream on perception (26). However, to address the issue more directly, we undertook a supplemental control experiment [see supporting information (SI) *Text*].

In the supplemental experiment, the vowel of the main experiment was replaced by a harmonic complex with a fundamental frequency (F0) of 200 Hz (henceforth, the simultaneous complex) and the target was itself a harmonic complex with a fundamental frequency of 300 Hz (see *Methods for Supplemental Experiment* in *SI Text*). When the simultaneous complex and target are presented together in quiet, a single harmonic complex with an F0 of 100 Hz and a dense spectral profile is heard. As in the main experiment, when the target is preceded by an isochronous pair of matching 300-Hz complexes (the complex stream, replacing the tones of the main experiment), the contribution of the target to the simultaneous complex decreases by an amount that depends on spatial cues. However, unlike in the main experiment, the target contributes significantly to the perceived spectral content of the simultaneous complex in all conditions (see SI Fig. 4*B*), presumably because across-frequency grouping cues are stronger for these stimuli than for a single-frequency target. The fact that the effective attenuation of the target in the simultaneous complex is near zero in many conditions in the supplemental experiment suggests that there is not obligatory adaptation of the target response for stimuli repeating at the rates and levels used in our experiments.

Another possible explanation relates to auditory spatial processing. Rapid changes in location can result in a diffuse spatial percept, attributed to "binaural sluggishness" in spatial processing (27, 28), raising the possibility that when target and tones have different spatial cues, the target is spatially diffuse. Target diffuseness, in turn, could cause the target to contribute relatively little to the perceived content of the vowel and help explain why trading hypotheses fail. Again, however, no such effect is observed in the supplemental experiment, even though the spatial cues change dynamically at the same rate as in the main experiment. Thus, there is no evidence that the perceptual contribution of the target is reduced because it is spatially diffuse.

To explain our finding of nonallocation, we suggest that the auditory system favors efficient processing over veridical representation of the entire auditory scene. In particular, the perceptual organization of the auditory background (here, the unattended object) may not be as fully elaborated as that of the foreground (29). This suggestion implies that sound elements that are rejected from the auditory foreground are not necessarily assigned to auditory objects within the unattended background. Interpreted in this way, perceptual nonallocation may reflect a figure-ground asymmetry, with stronger perceptual cues necessary to pull an element into the auditory foreground than are needed to push the same element into the (unattended) background.

Our results cannot answer the question of whether the target was part of the unattended object in the background or whether it was isolated in some "perceptual limbo." In informal listening, when listeners attempted to attend to both objects at once, they perceived no salient change in the perceived organization compared with when they actively attended to the tones or vowel. However, it was difficult to attend to both objects simultaneously; listeners felt that they rapidly shifted attention from object to object, rather than simultaneously attending to both objects (30). From these reports, we cannot rule out the possibility that perceptual organization in our experiment is bistable, changing so that the target moves to the background whenever attention shifts between objects.

In many everyday acoustic settings, competition for attention between auditory objects may be the most important problem facing a listener (31, 32). In vision, this problem has long been recognized, and theories of how stimulus attributes interact with top-down processes to mediate competition for attention are well developed (33, 34). Similar mechanisms may work to resolve competition for attention in complex auditory environments (35, 36). In both vision and audition, attention appears to operate on perceived objects, rather than simple features of the visual or acoustic scene (33, 35, 37). This suggests that the ability to direct attention in a complex, multisource auditory scene is directly affected by the way in which objects are formed. Past work demonstrates that both bottom-up factors and top-down attention influence the perceptual organization of sound (6). The current results hint that the ultimate interpretation of the acoustic scene may depend on what object a listener attends, just as attention can alter perception of objects in a visual scene (38). The organization of the scene in turn impacts how well the listener can reduce interference from unwanted objects and understand an attended object. The current results show that spatial cues can affect the perceptual organization of ambiguous sound mixtures, which can then cause the interesting phenomenon in which not all of the physical energy in a sound mixture is allocated to the identifiable objects.

Methods

Stimuli. Stimuli consisted of a 3-s-long sequence, composed of 10 identical presentations of three 100-ms-long elements: two 500-Hz tone bursts (tones) followed by a synthetic vowel with fundamental frequency of 125 Hz (see Fig. 1*A*). The target was a 500-Hz tone presented simultaneously with the vowel. All tones, target, and the harmonics of the vowel were gated with a Blackman window (60-ms duration), followed by a silent gap of 40 ms. The sequence of repeating tones and vowel caused a percept of two distinct auditory objects (rapidly repeating tones and a slower sequence of repeating vowels).

The vowel consisted of individual random-phase harmonics of the fundamental frequency 125 Hz, spectrally shaped like the vowel /I/ (formant peaks at frequencies 490, 2,125, and 2,825 Hz; see Fig. 1*B*). The vowel did not contain any energy in the fourth harmonic, the frequency of the target. When the target was present and perceived as part of the vowel, the perceived vowel quality shifted from /I/ (target-absent, or not part of the vowel) toward /ɛ/ (target heard as part of the vowel; refs. 8–10), presumably by shifting the perceived frequency of the first formant peak.

Spatial cues in the tones and target were controlled by

processing the sounds with head-related transfer functions (HRTFs) measured on a mannequin (39). This processing simulates the interaural time and level differences and spectral cues that would arise for sources from a particular location relative to the listener. Sources were processed to have spatial cues consistent with a source either from straight ahead (azimuth = 0°) or 45° to the right of the listener. In all trials, the simulated vowel azimuth was zero. Four different spatial configurations were tested, differing in which component's spatial cues matched those of the target (see Fig. 1*C*). Various control trials ensured that we only included listeners who could reliably identify the tones rhythm or the vowel identity for unambiguous single-object stimuli (see Fig. 1*C*). In two-object control trials, there was no target and both tones and vowel were simulated from straight ahead. In single-object control trials, the attended object was simulated from straight ahead; the target was simulated from either 0° or 45° azimuth, or was not present.

Procedures. In the main experiment, two-object stimuli (with single-object controls intermingled) were presented in two blocks of trials differing only in the instructions to the subjects. In tone blocks, subjects identified the perceived tones rhythm as even (an evenly spaced sequence of 500-Hz tones, one every 100 ms) or galloping (a pair of tones 100 ms apart, followed by a 100-ms silent gap). In vowel blocks, subjects identified the perceived vowel identity as /ɛ/ or /I/. Thirty trials of each condition were presented in a different random order for each block of trials.

In the single-object control experiment, trials consisted of the attended object and the target, both simulated from straight ahead (0° azimuth). On each trial, the target was attenuated by a random amount ranging from 0 to 14 dB, in 2-dB steps. As in the main experiment, in separate blocks, subjects judged either the rhythm of the tones or the identity of the vowel.

Analysis. The data from the main experiment were analyzed using a decision theory model. The internal decision variable was assumed to be a unidimensional, Gaussian-distributed random variable whose mean depended on the stimulus and whose variance was independent of the stimulus. A single criterion value was assumed to divide the decision space into two regions, corresponding to target-present or target-absent responses. The probability of responding "target present" was calculated for each condition, then used to estimate the distances between the underlying means of the corresponding conditional probability density functions and the mean of the distribution for the target-absent prototype in units of standard deviation (d'). These d' measures were normalized by the d' separation between the target-present and target-absent prototypes to estimate the relative perceptual distance between the condition and the single-object prototypes. By definition, the resulting statistic was zero for the target-absent prototype and one for the spatially unambiguous, target-present prototype. The across-subject means and standard errors of these relative perceptual distances were computed for each stimulus and are presented in Fig. 2.

In the single-object control study, the percent responses consistent with the "target present" generally decreased monotonically with increasing attenuation of the target. These functions were fit with a sigmoidal function with free parameters of slope, threshold, and upper and lower asymptotes. The fitted curves were used to map the raw percentage of responses for each stimulus to an effective attenuation of the target in the main experiment (see Fig. 3 *A* and *B*). If the response percentage for a given condition was less than the lower asymptote or greater than the upper asymptote of the psychometric function fit to the auxiliary results, the effective attenuation was set to 0 dB or 16 dB, respectively.

Subjects. Eight subjects were selected on the basis of their ability to reliably distinguish between the single-object prototypes in a similar pilot experiment. In the prior experiment, subjects had to achieve both (*i*) a d' of 0.7 or greater between the target-present and target-absent prototypes in the main experiment and (*ii*) a slope of 10% correct/dB attenuation to the fit of their responses in the single-object control experiment. All naïve subjects met the criteria for the tones stimuli in the prior experiment. The 8 current subjects were recruited from the 10 of 20 naïve subjects who reached the performance criterion for the vowel control stimuli. Seven of the eight subjects had greater d' values here than in the pilot experiment, presumably from experience with the task. In the current experiment, all eight subjects achieved d' scores of 1.5 or better on both tones and vowel tasks.

We thank Steven Babcock for assistance with data collection in the supplemental experiment and Christophe Micheyl for helpful comments on the manuscript. This work was supported by Office of Naval Research Grant N00014-04-1-0131 and National Institutes of Health Grants R01 DC 05778-02 (to B.G.S.-C.) and R01 DC 05216 (to A.J.O.).

1. Hulse SH, MacDougall-Shackleton SA, Wisniewski AB (1997) *J Comp Psych* 111:3–13.
2. Endepols H, Feng AS, Gerhardt HC, Schul J, Walkowiak W (2003) *Behav Brain Res* 145:63–77.
3. Cherry EC (1953) *J Acoust Soc Am* 25:975–979.
4. Jouventin P, Aubin T, Lengagne T (1999) *Animal Behav* 57:1175–1183.
5. Darwin CJ, Brungart DS, Simpson BD (2003) *J Acoust Soc Am* 114:2913–2922.
6. Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
7. Darwin CJ, Carlyon RP (1995) in *Hearing*, ed Moore BCJ (Academic, San Diego, CA), p 387.
8. Darwin CJ (1984) *J Acoust Soc Am* 76:1636–1647.
9. Darwin CJ, Hukin RW (1997) *J Acoust Soc Am* 102:2316–2324.
10. Darwin CJ, Hukin RW (1998) *J Acoust Soc Am* 103:1080–1084.
11. Liberman AM, Isenberg D, Rakerd B (1981) *Percept Psychophys* 30:133–143.
12. Moore BCJ, Glasberg BR, Peters RW (1986) *J Acoust Soc Am* 80:479–483.
13. Turgeon M, Bregman AS, Ahad PA (2002) *J Acoust Soc Am* 111:1819–1831.
14. Turgeon M, Bregman AS, Roberts B (2005) *J Exp Psychol Hum Percept Perform* 31:939–953.
15. Darwin CJ (1995) in *Levels in Speech Communication: Relations and Interactions: A Tribute to Max Wajskop*, eds Sorin C, Mariani J, Meloni H, Schoentgen J (Elsevier, Amsterdam), pp 1–12.
16. McAdams S, Botte MC, Drake C (1998) *J Acoust Soc Am* 103:1580–1591.
17. Darwin CJ, Hukin RW (1999) *J Exp Psychol Hum Percept Perform* 25:617–629.
18. Darwin CJ, Pattison H, Gardner RB (1989) *Percept Psychophys* 45:333–342.
19. Darwin CJ, Hukin RW (2000) *J Acoust Soc Am* 107:970–977.
20. Javel E (1996) *J Acoust Soc Am* 99:1040–1052.
21. Smith RL (1977) *J Neurophysiol* 40:1098–1111.
22. Rand TC (1974) *J Acoust Soc Am* 55:678–680.
23. Elmasian R, Galambos R, Bernheim A (1980) *J Acoust Soc Am* 67:601–607.
24. Mapes-Riordan D, Yost WA (1999) *J Acoust Soc Am* 106:3506–3511.
25. Zwislocki JJ, Sokolich WG (1974) *Percept Psychophys* 16:87–90.
26. Darwin CJ, Hukin RW, al-Khatib BY (1995) *J Acoust Soc Am* 98:880–885.
27. Dye RH, Jr, Brown CA, Gallegos JA, Yost WA, Stellmack MA (2006) *J Acoust Soc Am* 120:3946–3956.
28. Joris PX, van de Sande B, Recio-Spinoso A, van der Heijden M (2006) *J Neurosci* 26:279–289.
29. Cusack R, Deeks J, Aikman G, Carlyon RP (2004) *J Exp Psychol Hum Percept Perform* 30:643–656.
30. Best V, Gallun FJ, Ihlefeld A, Shinn-Cunningham BG (2006) *J Acoust Soc Am* 120:1506–1516.
31. Freyman RL, Helfer KS, McCall DD, Clifton RK (1999) *J Acoust Soc Am* 106:3578–3588.
32. Shinn-Cunningham BG, Ihlefeld A, Satyavarta, Larson E (2005) *Acta Acustica* 91:967–979.
33. Desimone R, Duncan J (1995) *Ann Rev Neurosci* 18:193–222.
34. Peers PV, Ludwig CJH, Rorden C, Cusack R, Bonfiglioli C, Bundesen C, Driver J, Antoun N, Duncan J (2005) *Cereb Cortex* 15:1469–1484.
35. Scharf B (1998) in *Attention*, ed Pashler H (Psychology, Hove, UK), p 75.
36. Tata MS, Ward LM (2005) *Neuropsychologia* 43:509–516.
37. Cusack R, Carlyon RP, Robertson IH (2000) *J Cog Neurosci* 12:1056–1065.
38. Carrasco M, Ling S, Read S (2004) *Nat Neurosci* 7:308–313.
39. Shinn-Cunningham BG, Kopco N, Martin TJ (2005) *J Acoust Soc Am* 117:3100–3115.