Cogn Affect Behav Neurosci (2014) 14:336–353
DOI 10.3758/s13415-013-0207-1

# Resources required for processing ambiguous complex features in vision and audition are modality specific

**Morgan D. Barense · Jonathan Erez · Henry Ma · Rhodri Cusack**

Published online: 11 September 2013


**Abstract** Processing multiple complex features to create cohesive representations of objects is an essential aspect of both the visual and auditory systems. It is currently unclear whether these processes are entirely modality specific or whether there are amodal processes that contribute to complex object processing in both vision and audition. We investigated this using a dual-stream target detection task in which two concurrent streams of novel visual or auditory stimuli were presented. We manipulated the degree to which each stream taxed processing conjunctions of complex features. In two experiments, we found that concurrent visual tasks that both taxed conjunctive processing strongly interfered with each other but that concurrent auditory and visual tasks that both taxed conjunctive processing did not. These results suggest that resources for processing conjunctions of complex features within vision and audition are modality specific.

**Keywords** Visual feature conjunctions · Auditory feature conjunctions · Visual object perception · Auditory object perception · Crossmodal processing · Perirhinal cortex

**Electronic supplementary material** The online version of this article (doi:10.3758/s13415-013-0207-1) contains supplementary material, which is available to authorized users.

M. D. Barense (✉) · J. Erez · H. Ma
Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada
e-mail: barense@psych.utoronto.ca

M. D. Barense
Rotman Research Institute, Toronto, Canada

R. Cusack
Brain and Mind Institute, University of Western Ontario, Toronto, Canada

Detecting visual objects is an important part of many everyday tasks. When a visual object is seen, regions in the ventral visual stream within the occipital and temporal lobes extract many features that describe it. As we search for a target object in some contexts, a single feature may reliably signal the presence of the target (e.g., if it has a unique color). In other contexts, any given feature may be ambiguous and occur in both relevant and irrelevant objects, and only a particular conjunction of features can be taken to signal the target. A similar problem exists in audition. The early auditory system is also thought to extract a number of features, such as pitch, bandwidth, and modulation. In some contexts, a single feature will identify a target object, but in others, only a conjunction of features will suffice (Cusack & Carlyon, 2003).

Processing visual information is often described according to a simple-to-complex hierarchy in the *ventral visual stream*, a pathway that extends ventrally through the inferotemporal cortex toward anterior temporal regions (e.g., Desimone & Ungerleider, 1989; Martinovic, Gruber, & Muller, 2008; Riesenhuber & Poggio, 1999; Tanaka, 1996). Low-level inputs are transformed into more complex representations through successive stages of processing. For example, whereas neurons in V2 fire in response to more simple stimulus properties, such as color, orientation, and spatial frequency, by the level of the inferotemporal (IT) cortex, cells are tuned to much more complex features (e.g., Bruce, Desimone, & Gross, 1981; Perrett, Rolls, & Caan, 1982; Tanaka, 1996). These findings have led to the formulation of several models of visual object recognition (Fukushima, 1980; Perrett & Oram, 1993; Riesenhuber & Poggio, 1999; Wallis & Rolls, 1997). Much of this work has focused on area TE in the anterior portion of the IT cortex, which is traditionally considered the highest area in the ventral visual stream thought to have a role in object recognition (e.g., Ungerleider & Haxby, 1994). The critical features required for the activation of area

TE are said to be "moderately complex" (e.g., Tanaka, 1997), but they are not specific enough to represent the complete objects through the activity of a single cell; a combination of several TE cells is needed.

Area TE is a purely visual structure, but more recently, it has been proposed that the ventral visual stream may extend farther anteriorly to encompass multimodal regions, such as the perirhinal cortex. The perirhinal cortex is a cortical region at the ventral surface of the medial temporal lobe (MTL) that has historically been thought to operate exclusively in the service of declarative memory, with no role in object perception (e.g., Squire & Wixted, 2011). However, recent studies have suggested that the perirhinal cortex may process and represent information at a level of complexity greater than in area TE—perhaps at the level of the whole object (e.g., Bussey & Saksida, 2002; Cowell, Bussey, & Saksida, 2010). These studies have shown that the perirhinal cortex is involved in perceptual tasks involving complex everyday objects (e.g., perceptually similar cars, faces, radios, etc.) when these cannot be solved on the basis of individual object features alone but, instead, require using a complex conjunction of features (e.g., Barense et al., 2005; Barense, Ngo, Hung, & Peterson, 2012; Barense, Rogers, Bussey, Saksida, & Graham, 2010; Bussey, Saksida, & Murray, 2002; Erez, Lee, & Barense, 2013; Lee, Buckley, et al., 2005; O'Neil, Cate, & Kohler, 2009). This requirement to process conjunctions of complex features—as opposed to using a single feature alone—seems to be critical in engaging the perirhinal cortex. For example, task difficulty alone does not activate the perirhinal region: Even when the size of the visual differences was reduced to make the task very difficult, detecting targets that were defined by unambiguous features neither activated nor required the perirhinal cortex (Barense, Gaffan, & Graham, 2007; Barense, Henson, Lee, & Graham, 2010; Devlin & Price, 2007). Thus, the perirhinal region is required not for all tasks requiring the detection of visually similar stimuli, but only for those that stress processing conjunctions of complex visual features (Bussey, Saksida, & Murray, 2003; Lee, Bussey, et al., 2005; Lee, Scahill, & Graham, 2008; Tyler et al., 2004).

In addition to being seen, many real-world objects can be heard. The auditory system is also thought to be hierarchical (Kaas & Hackett, 2000; Wessinger et al., 2001), with cells in the cochlea tuned to simple features such as frequency, feeding into many stages of processing in the ascending pathway comprising the brainstem, thalamus, and cortex. Regions beyond the primary auditory cortex have been shown to represent fairly complex features that could be considered analogous to the visual representations in TE, such as pitch (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002), modulation (Hall et al., 2000), and bandwidth (Rauschecker & Tian, 2004). The aim of the present work is to understand the process that brings together these complex features to allow the recognition of auditory objects and to investigate its relationship to the processes that conjoin complex visual features.

The perirhinal cortex is anatomically well placed to combine inputs from different sensory modalities to create meaningful representations of objects. Although inputs to the perirhinal cortex are primarily visual, it is one of the first cortical fields within the ventral visual stream to have convergence of information from different sensory modalities, including superior temporal regions devoted to auditory sensory processing (Carmichael & Price, 1995; Friedman, Murray, O'Neill, & Mishkin, 1986; Suzuki & Amaral, 1994). This pattern of connectivity is unlike any other in the MTL and suggests that this structure might have a more general role in the integration of auditory and tactile features to create a meaningful representation of an object (Barense, Henson, & Graham, 2011). In support of this idea, there is evidence to suggest that the perirhinal cortex is critical for tactile–visual delayed non-matching-to-sample (Goulet & Murray, 2001), tactile–visual object matching (Holdstock, Hocking, Notley, Devlin, & Price, 2009; Parker & Gaffan, 1998), and auditory–visual integration (Naci, Taylor, Cusack, & Tyler, 2012; Taylor, Moss, Stamatakis, & Tyler, 2006; Taylor, Stamatakis, & Tyler, 2009). While there is evidence of audiovisual interactions even in primary sensory cortices (Calvert et al., 1999; Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003), audiovisual responses to meaningful, complex, multisensory object stimuli have been consistently reported in regions higher up in the object-processing hierarchy, including the lateral temporal (Beauchamp, Lee, Argall, & Martin, 2004; Hein et al., 2007) and perirhinal (Naci et al., 2012; Taylor et al., 2006; Taylor et al., 2009) cortices. It is also known that damage to the rat perirhinal cortex impairs fear conditioning to complex sounds with internal temporal structure, but not to simpler continuous tones (Bang & Brown, 2009; Kholodar-Smith, Allen, & Brown, 2008; Lindquist, Jarrard, & Brown, 2004). However, it has not been established whether the perirhinal cortex has a direct role in processing conjunctions of complex features within nonvisual modalities, such as the features that make up auditory objects. To investigate this idea, here we used a task that taxed the complex visual processes previously demonstrated to recruit the perirhinal cortex (Barense et al., 2007; Barense, Ngo, et al., 2012) and, critically, also introduced an auditory version of the paradigm.

In detection experiments involving simple stimuli (e.g., a single 500-Hz tone or a single lighted circle), it was shown that targets can be detected in concurrent auditory and visual streams with similar performance to detection in one stream alone (Alais, Morrone, & Burr, 2006) and that changing the relative importance of each modality does not lead to a trade-

off in performance between them (Bonnel & Hafter, 1998). These findings suggest modality-specific processing for simple stimuli that do not require processing conjunctions of complex features. The present study used a concurrent dual-modal detection paradigm to assess whether the resources that process conjunctions of complex novel auditory and visual features are amodal or modality specific. We presented concurrent auditory and visual streams of sequential stimuli in which a target had to be detected. Each stream was unimodal. Some streams were of "low ambiguity," meaning that the target shared no features with the distracting stimuli. Other streams were of "high ambiguity," in that each distractor shared many of the features of the target. This manipulation of the distractor–target relationships altered the degree to which the task taxed processing conjunctions of complex features, in a similar way to previous studies that have demonstrated a critical role for the perirhinal cortex in complex object processing (Barense et al., 2005; Barense et al., 2007; Barense, Groen, et al., 2012; Bartko, Winters, Cowell, Saksida, & Bussey, 2007; Bussey et al., 2002). To investigate the degree of modality specificity of these processes, we used a dual-task paradigm in which participants completed two target detection tasks simultaneously. The two tasks were of either the same or different modalities and either did or did not stress processing conjunctions of complex features. We also investigated performance on each task when it was performed alone without a concurrent task. From the pattern of interference among concurrent tasks, we were then able to assess the specificity of the processes (i.e., if performance on one task is not impaired by simultaneously performing a second task, one can surmise that the two tasks are unlikely to be using the same cognitive resource). In particular, we investigated whether processing conjunctions of features in one domain (e.g., audition) interfered with processing conjunctions of features in another domain (e.g., vision), over and above any nonspecific concurrent task interference. In two experiments, we found that concurrent visual tasks that both stressed conjunctive processing strongly interfered with each other but that auditory and visual tasks that both stressed conjunctive processing did not, suggesting that these resources are modality specific.

## Experiment 1

### Method

#### *Participants*

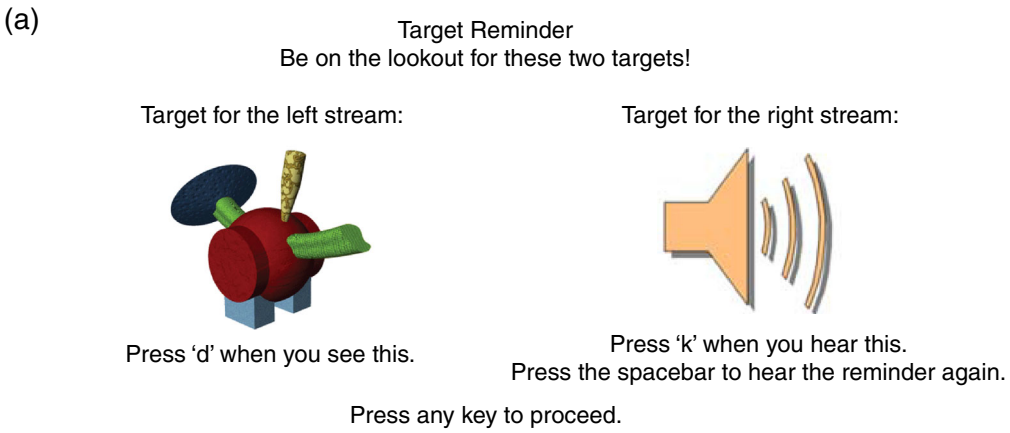
Twenty University of Toronto undergraduate students (11 female; average age = 22.3 years, *SD* = 3.1 years) participated in this study in exchange for $15. All participants gave informed written consent after the nature of the study and its possible consequences were explained to them. This work received ethical approval from the Ethics Review Office at the University of Toronto.

#### *Task and stimuli*

Participants completed a series of rapid serial target detection tasks, during which they monitored either one individual stream or two concurrent streams of sequentially presented targets and distractors (Fig. 1). These streams could be either visual or auditory. Across the different streams, the emphasis on conjunctive processing was manipulated, which was defined as the degree of feature ambiguity (Barense et al., 2005; Bussey et al., 2002), in either the visual or the auditory modality (Fig. 2). Feature ambiguity refers to the presence of overlapping features across different stimuli within a given discrimination. For the high-ambiguity discriminations, the targets differed from the distractors by only a single feature. For the low-ambiguity discriminations, the targets did not share any features with the distractors (except for the "body" and one fixed feature in the visual stream; see Fig. 2 and the text below).

For the visual streams, the stimuli were "fribbles" (Williams & Simons, 2000), novel visual objects composed of a main body and four appendages (Fig. 2a). There were 12 categories (or "species") of fribbles in total. Within a species, all fribbles consisted of the same main body, but each of the four appendages had three possible values, which were manipulated across fribbles to create varying levels of feature ambiguity (Fig. 2). Across different species, the fribbles had a completely different body and set of features. To maintain consistency with the auditory stimuli, which had only three features (see below), we kept one feature constant in any given stream. For example, in the low-ambiguity stream of Fig. 2d, the blue "legs" were fixed across the different fribbles. This effectively meant that the fixed feature could be considered to be part of the body and could be ignored because it was not relevant to target detection. The fribble subtended a horizontal visual angle of 12.35° and a vertical visual angle of 8.42°.

For the auditory streams, the stimuli were novel sound objects composed of sequential parts, or features. The three features were rapid and contiguous and were perceived as a single object. The features were presented sequentially, because pilot work showed that it was not clear how to manipulate orthogonal dimensions of a single sound. Each feature could be one of three types: a pure tone, a warbling tone (with frequency modulation at 5 Hz and a depth of 4 semitones), and a narrowband noise burst (bandwidth of 2 semitones). Analogous to the visual



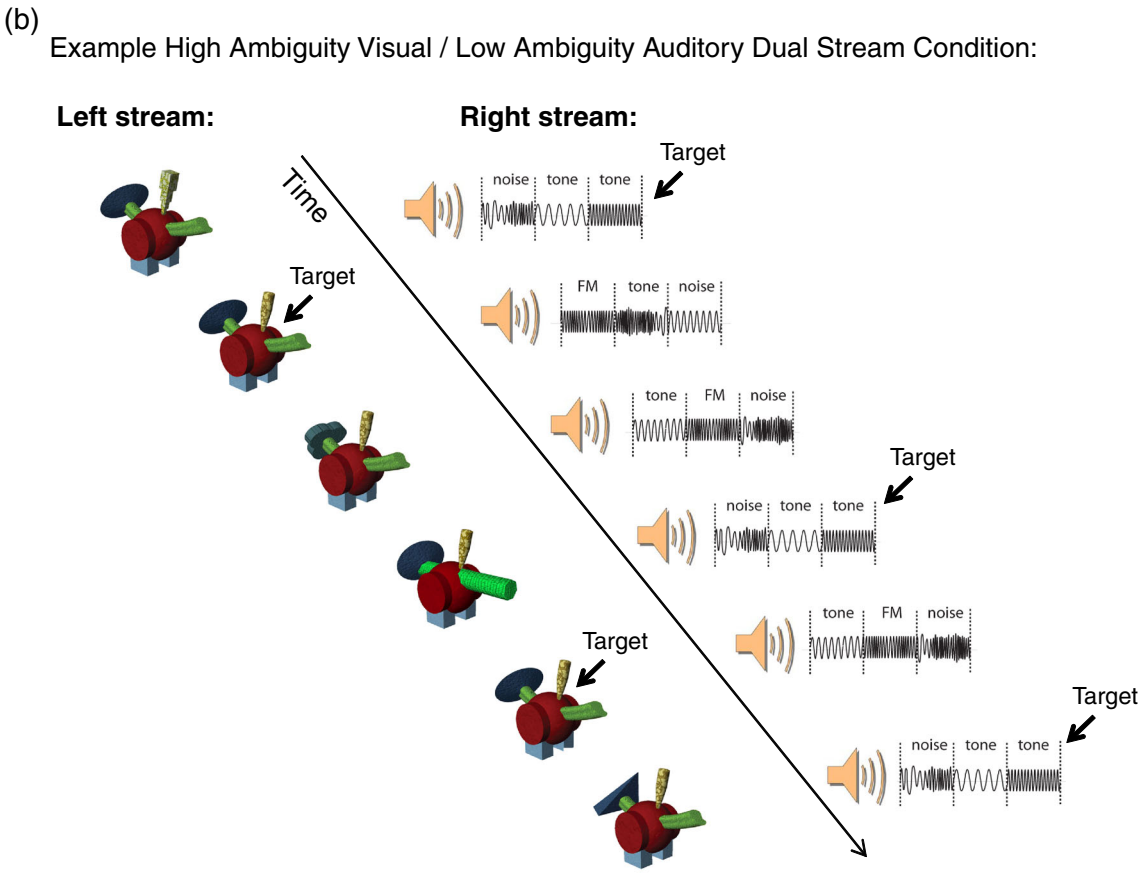
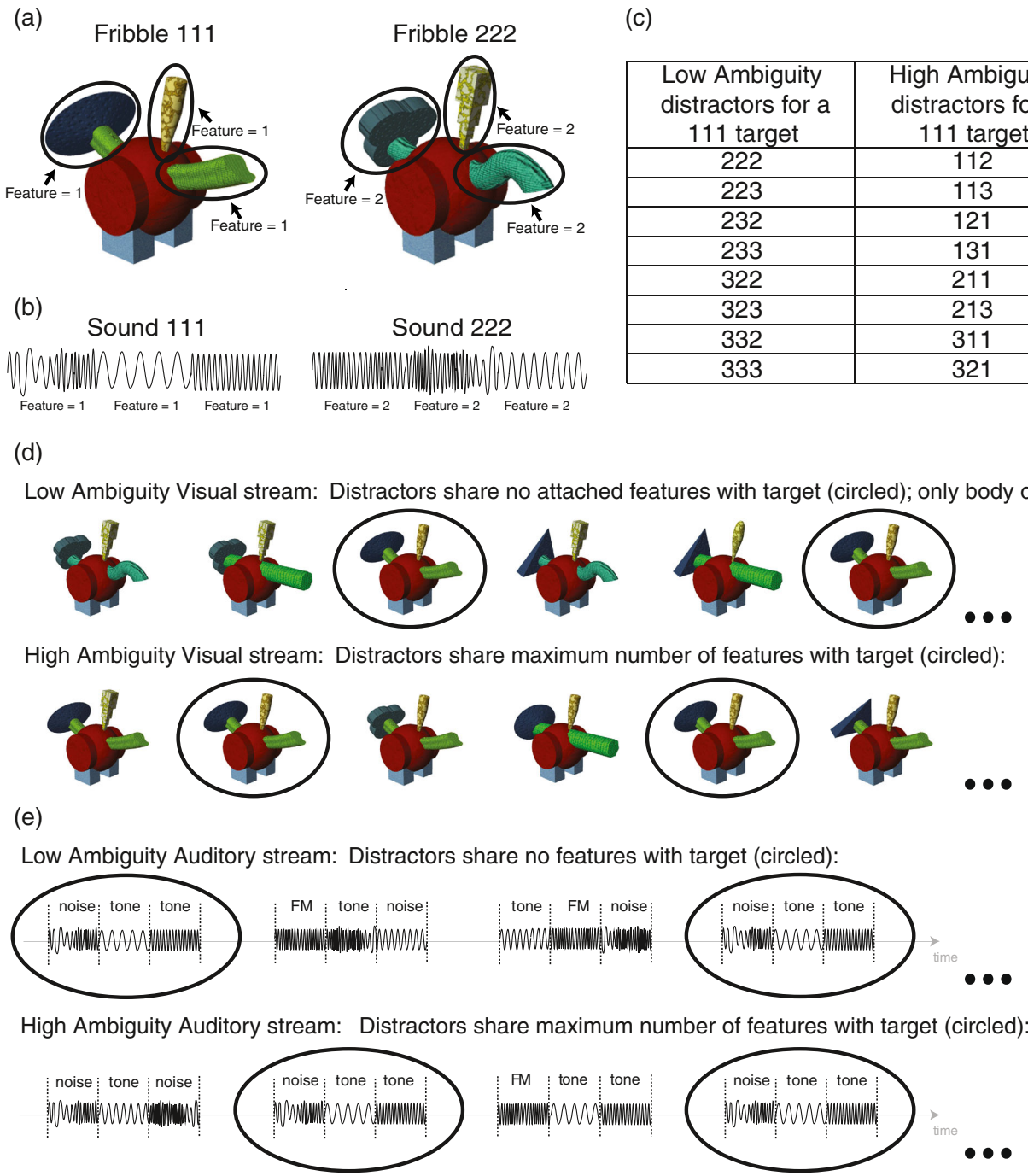




**Fig. 1** Target detection task. **a** Before every block of trials, participants viewed a reminder screen containing the targets for the given condition (here, a reminder screen for a dual-task condition containing visual and auditory streams is shown). The reminder screen remained until participants advanced to the next block of trials. Participants could hear the auditory target as many times as they wished. **b** An example of the high-ambiguity visual/low-ambiguity auditory condition. In the dual-stream conditions, participants simultaneously monitored for the target in each stream. In the high-ambiguity conditions, the individual features overlapped between the target and foils, and thus, participants could not identify the target by a single feature (e.g., the round disk) but, instead, had to identify the target on the basis of the conjunction of features. By contrast, the low-ambiguity stream placed less of a requirement on conjunctive processing, because the target was readily identifiable by a single feature (e.g., a particular tone)

stimuli, 12 different species were created by randomly selecting the frequency of each of the features from the possible values of −12, −10, −8, −4, −2, 0, 4, 6, 8, 12, 14, or 16 semitones, relative to the musical note A above




| Low Ambiguity distractors for a 111 target | High Ambiguity distractors for a 111 target |
|---|---|
| 222 | 112 |
| 223 | 113 |
| 232 | 121 |
| 233 | 131 |
| 322 | 211 |
| 323 | 213 |
| 332 | 311 |
| 333 | 321 |

**Fig. 2** The novel visual objects ("fribbles") were composed of a central body and attached features. **a** Examples of fribbles from the same species "named" 111 and 222. **b** Examples of two sounds (111 and 222) that shared no features. **c** The overlap of features across visual objects and sounds within a stream was varied according to the table. **d** Example of a low- and high-ambiguity visual stream (across different streams, the fribbles were from different species; i.e., they had a different body and a different set of appendages). **e** Example of a low- and high-ambiguity auditory stream

middle-C (440 Hz). Each feature was 200 ms in duration, leading to total sounds of 600 ms. The sounds were presented diotically through headphones, and the volume was adjusted to a comfortable level (approximately 70 dB SPL). There were never two auditory streams concurrently, because it is difficult to design stimuli in which one can

be confident that separate perceptual streams have been achieved.[1]

Within a stream, the stimuli were always from the same species. Any given species was used for one stream only and was never repeated across different conditions. To ensure that any observed effects were not merely due to one target being harder/easier than another, we created four different lists of targets and distractors for every condition. These lists were administered across participants (i.e., 5 participants allocated to each list). Across the different stimulus lists, each target stimulus was counterbalanced so that it appeared equally often as a low-ambiguity or high-ambiguity target, and each condition appeared equally often on the left or the right side of the screen. There was no difference in difficulty across the different stimulus lists, $F(3, 16) = 0.51$, $p = .68$, $\eta_p^2 = .09$, and thus, for all subsequent analyses, we collapsed across stimulus list.

### *Behavioral procedure*

Every participant completed 11 conditions in total (L = low ambiguity; H = high ambiguity; V = visual modality; A = auditory modality): LV, HV, LA, HA, LVLV, LVHV, HVHV, LVLA, LVHA, HVLA, and HVHA. Each stream was associated with its own target. For example, in the LVHV condition, there was one low-ambiguity target and a set of eight distractors associated with the LV stream and a separate high-ambiguity target and a set of eight distractors associated with the HV stream. Each stream contained one target and eight distractors, which were repeatedly presented throughout the condition. Each stream was unimodal (i.e., either visual or auditory) and had its own individual target. Neither the target nor the distractors overlapped across the streams. For each stream, participants were asked to indicate when the target appeared by pressing a button on the keyboard ("k" for the right stream and "d" for the left stream; see Fig. 1a). The two streams were completely separate: For example, if two visual streams were presented concurrently, one stream would always be shown on the left and the other on the right. The left and right streams were separated by an average of 7 cm (10º visual angle) of white space on the screen. Each stream was centered with respect to the vertical axis of the screen. Whether a stream was shown on the left or the right was fixed for that condition but was counterbalanced across participants.

Each condition commenced with a series of practice trials for which participants were given feedback if they identified the target ("Got it!"), missed the target ("Missed it"), or falsely identified a distractor as the target ("False alarm"). For the single-stream conditions, there were 24 practice trials. For the dual-stream conditions, each of the two streams was initially introduced individually for 16 trials. Following familiarization with each stream, there were 8 practice trials of the two streams presented concurrently (i.e., dual streams). After every 8 trials in the practice phase, there was a reminder screen containing the relevant targets for each stream (Fig. 1a). The reminder slide contained an image of the visual target, and participants could press the space bar to hear the auditory target as many times as they wished. The test trials for each condition started immediately after the practice for that condition, and there was no feedback for the test trials. Each test condition comprised 88 trials, which were split into eight blocks of 11 trials. Within each block of 11 trials, the target was repeated 3 times, and each of the eight distractors was shown only once. A given test condition concluded when participants had completed eight such 11-trial blocks. Thus, at the conclusion of each condition, the target had been shown 24 times, and each of the eight distractors had been shown 8 times (i.e., 88 trials per condition). The order of the targets and distractors within each block was randomly determined. After each 11-trial block, there was a target reminder screen (Fig. 1a). After each condition, there was a short break prior to commencing the next condition. The order of the conditions was randomized across participants, but every participant completed each of the 11 conditions.

For all trials regardless of condition or modality, participants had 1,200 ms to respond, during which the visual fribble or an image of the speaker was displayed (Fig. 1b). To discourage participants from engaging in a serial strategy (e.g., attend to left stream, then attend to right stream, then attend to left stream, etc.), the onset of a trial in each stream was slightly jittered in time by 100–600 ms. That is, a trial in one stream would appear, and then, after a randomly determined interval of 100–600 ms, a trial from the second stream would appear. Whether the left or the right stream would appear first was randomly determined on each trial, with 50 % of trials beginning with the left stream. The interstimulus interval randomly varied between 300 and 800 ms for each stream of every condition. The experiments were programmed using E-Prime software (Psychology Software Tools Inc., Pittsburgh, PA) and were administered using a laptop computer (1,024 × 768 resolution) that was placed at a comfortable viewing distance. The entire experiment lasted between 80 and 90 min.

### *Data analysis*

Accuracy data were analyzed using signal detection (for a review, see Sporer, 2001). Correct classifications of the targets were counted as hits, and incorrect classifications of nontargets as

[1] Sounds presented at quite different spatial locations are sometimes perceived as part of the same perceptual stream (Bregman, 1990; Deutsch, 1974) (for a demonstration, see http://deutsch.ucsd.edu/psychology/pages.php?i=203). More effective in evoking auditory stream segregation are differences in frequency, but even then at the start of sequences, with slow presentation rates (such as those required for the present task), or when attention is shifted, concurrent sounds are often heard as a single perceptual stream, and selectively attending to a subset of them is difficult (Cusack, Deeks, Aikman, & Carlyon, 2004; Cusack & Roberts, 2000). We therefore always presented only a single stream of sounds.

targets were counted as false alarms. Categorization accuracy (i.e., the ability to distinguish a target from a nontarget) was then calculated using the formula for $A'$ provided by Rae (1976), which yields scores functionally equivalent to bias-corrected measures of percent correct (Rule & Ambady, 2008). To assess whether the resources for processing conjunctions of features were modality specific, we performed a series of repeated measures ANOVAs on the $A'$ data described below. Depending on the question at hand, the within-subjects factors were *target modality*, *target ambiguity*, *interfering modality*, or *interfering ambiguity*. Because we were interested in how different kinds of interference would affect performance, we restricted subsequent pairwise comparisons to tests between conditions in which the target was of the same modality and ambiguity, allowing us to assess cleanly the effect of interfering modality/ambiguity. There is one exception: For the single streams, there was only one condition of each target modality and ambiguity, and thus, in this case, our pairwise comparisons were within a target modality to assess the effect of ambiguity on visual and auditory target detection. For some conditions, we observed ceiling performance, and thus, the distribution of accuracy data was not always normal. Unfortunately, we are not aware of a nonparametric test that fully parallels a repeated measures ANOVA. However, for the more critical follow-up assessments of our simple effects, we were able to conduct nonparametric related-samples Wilcoxon signed rank tests, which did not qualitatively change our results from those obtained with the parametric analogues. Thus, our findings appear to be robust against the limitations concerning normality among the variables.

## Results

### *Detecting targets in a single visual or auditory stream*

Target detection performance when only one stream was present is shown in Fig. 3 and as percent correct in Table 1. Reaction time (RT) data are provided in Supplemental Table S1. The results demonstrated that detection of high-ambiguity targets was harder than detection of low-ambiguity targets in both modalities, but to an even greater degree for the auditory streams. A repeated measures ANOVA with within-subjects factors of *target modality* (visual vs. auditory) and *target ambiguity* (high vs. low) revealed better performance for visual than for auditory streams overall, $F(1, 19) = 5.72$, $p < .05$, $\eta_p^2 = .23$, and better performance for low-ambiguity than for high-ambiguity streams, $F(1, 19) = 38.42$, $p < .001$, $\eta_p^2 = .67$. There was a marginally significant interaction between target modality and target ambiguity, $F(1, 19) = 4.19$, $p = .06$, $\eta_p^2 = .18$, indicating that the effect of ambiguity was marginally greater for auditory than for visual stimuli. Paired-samples $t$-tests within each modality indicated that there was a significant effect of ambiguity in both the visual, $t(19) = 2.79$, $p < .001$, $d = 0.62$, and auditory, $t(19) = 6.31$, $p < .001$, $d = 1.41$, conditions, indicating that for both visual and auditory streams, performance was better when the target could be identified from single features considered independently (low ambiguity) than when a target could be detected only through a combination of multiple feature dimensions (high ambiguity).

### *Detecting visual targets in dual streams*

We performed a 2 × 2 × 2 repeated measures ANOVA with within-subjects factors of *target ambiguity* (high vs. low), *interfering modality* (visual vs. auditory), and *interfering ambiguity* (high vs. low). This revealed a significant three-way interaction between target ambiguity, interfering modality, and interfering ambiguity, $F(1, 19) = 8.76$, $p < .01$, $\eta_p^2 = .32$ (Fig. 4). Each of the 3 two-way interactions was also significant [target ambiguity × interfering modality, $F(1, 19) = 80.47$,

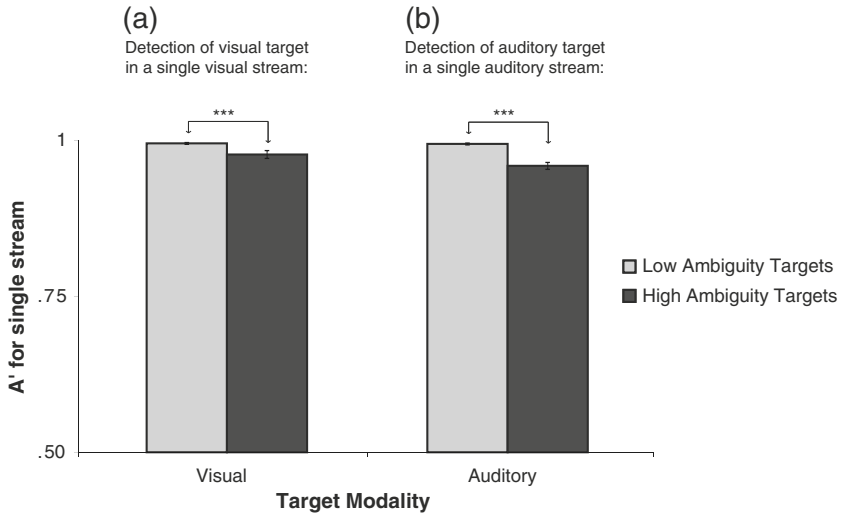



**Fig. 3** Experiment 1 $A'$ scores when only a single visual (**a**) or auditory (**b**) stream was present. These results indicated that detection of high-ambiguity targets was harder than detection of low-ambiguity targets in both modalities. *Error bars* represent *SEM*s. ***$p < .001$

**Table 1** Mean percent correct for the single-stream tasks

| | Experiment 1 | | | | Experiment 2 | | | |
|---|---|---|---|---|---|---|---|---|
| Condition | LV | HV | LA | HA | LV | HV | LA | HA |
| Mean | 99.4 | 95.6 | 99.2 | 91.5 | 99.2 | 93.8 | 96.8 | 91.6 |
| *SD* | 0.8 | 5.4 | 0.8 | 5.3 | 1.4 | 4.2 | 6.1 | 5.2 |

*Note*. *LV* low-ambiguity visual, *HV* high-ambiguity visual, *LA* low-ambiguity auditory, *HA* high-ambiguity auditory

$p < .001$, $\eta_p^2 = .81$; target ambiguity × interfering ambiguity, $F(1, 19) = 6.11$, $p < .05$, $\eta_p^2 = .24$; interfering modality × interfering ambiguity, $F(1, 19) = 48.89$, $p < .001$, $\eta_p^2 = .72$], as were each of the three main effects [target ambiguity, $F(1, 19) = 80.44$, $p < .001$, $\eta_p^2 = .81$; interfering modality, $F(1, 19) = 28.34$, $p < .001$, $\eta_p^2 = .60$; interfering ambiguity, $F(1, 19) = 48.11$, $p < .001$, $\eta_p^2 = .72$]. To further investigate what was driving the significant three-way interaction, we performed a series of 2 × 2 repeated measures ANOVAs at each of the three factors. These are described in turn below.

*Low-ambiguity visual target detection: Visual versus auditory interferers* Low-ambiguity visual targets were easily detected in the presence of either auditory or visual interfering streams (Fig. 4a). A repeated measures ANOVA with within-subjects factors of *interfering modality* (visual vs. auditory) and *interfering ambiguity* (high vs. low) revealed marginally worse performance on low-ambiguity visual target detection if the interfering stream was visual than if it was auditory, $F(1, 19) = 3.86$, $p = .06$, $\eta_p^2 = .17$. There was no effect of interferer ambiguity, $F(1, 19) = 2.61$, $p = .12$, $\eta_p^2 = .12$, and no interaction between the modality of the interferers and the ambiguity of the interferers, $F(1, 19) = 2.83$, $p = .11$, $\eta_p^2 = .13$.

*High-ambiguity visual target detection: Visual versus auditory interferers* The detection of a high-ambiguity visual target showed a more specific pattern of interference consistent with a bottleneck within a modality-specific resource (Fig. 4b). There was greater interference from concurrent visual than from auditory streams, and particularly from high-ambiguity visual interferers. A repeated measures ANOVA with within-subjects factors of *interfering modality* (visual vs. auditory) and *interfering ambiguity* (high vs. low) revealed worse performance on high-ambiguity visual target detection if the interfering stream was visual than if it was auditory, $F(1, 19) = 52.96$, $p < .001$, $\eta_p^2 = .74$, worse performance if the interfering stimulus was high ambiguity than if it was low ambiguity, $F(1, 19) = 28.73$, $p < .001$, $\eta_p^2 = .60$, and a significant interaction between these two factors, $F(1, 19) = 32.95$, $p < .001$, $\eta_p^2 = .63$. Paired-samples $t$-tests to investigate this interaction further revealed that high-ambiguity visual distractors interfered significantly more than low-ambiguity visual distractors, $t(19) = 7.08$, $p < .001$, $d = 1.58$. By contrast, the ambiguity of the auditory distractor did not have a significant effect, $t(19) = -0.54$, $p = .60$, $d = 0.12$. Importantly, this effect was not just a general task difficulty effect, because the high-ambiguity auditory condition was significantly harder than the high-ambiguity visual condition, $t(1, 19) = 2.28$, $p < .05$, $d = 0.51$ (Fig. 3), but interfered less with high-ambiguity visual target detection.

*High-ambiguity versus low-ambiguity visual target detection with high-ambiguity versus low-ambiguity interferers* In the dual-task results described thus far, we found more specific interference effects for high- than for low-ambiguity visual targets. Here, we explicitly tested whether the interference effects from visual interferers significantly differed as a

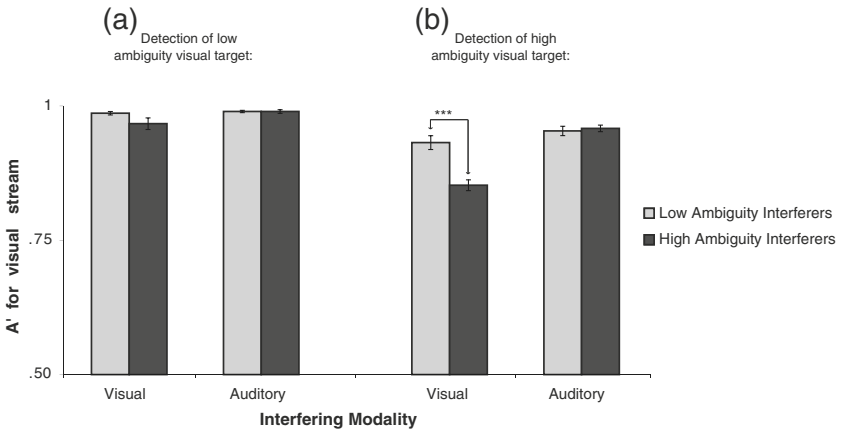



**Fig. 4** Experiment 1 *A′* scores for low-ambiguity visual (**a**) and high-ambiguity visual (**b**) target detection in the presence of low-ambiguity (*light bars*) or high-ambiguity (*dark bars*) interferers in both modalities. These results show that low-ambiguity visual targets were easily detected in the presence of either auditory or visual interfering streams. By contrast, the detection of a high-ambiguity visual target showed greatest interference from high-ambiguity visual interferers and was not affected by the degree of conjunctive processing in the auditory stream. This pattern of interference is consistent with a modality-specific resource for processing conjunctions of features. Note that this is not a generic task difficulty effect, because the high-ambiguity auditory detection was more difficult than the high-ambiguity visual detection (see Fig. 3). *Error bars* represent *SEM*s. ***$p < .001$

function of visual target ambiguity. We performed a repeated measures ANOVA with two within-subjects factors of *target ambiguity* (high-ambiguity visual vs. low-ambiguity visual) and *interfering ambiguity* (high-ambiguity visual vs. low-ambiguity visual) (Fig. 4a and b). This revealed worse performance for high-ambiguity visual targets than for low-ambiguity visual targets, $F(1, 19) = 124.22$, $p < .001$, $\eta_p^2 = .87$, and worse performance when the interferers were high-ambiguity visual than if they were low-ambiguity visual, $F(1, 19) = 84.59$, $p < .001$, $\eta_p^2 = .82$. There was also a significant interaction between these two factors, $F(1, 19) = 9.27$, $p < .01$, $\eta_p^2 = .33$, indicating that a high-ambiguity visual interferer had a greater detrimental effect on detection of high-ambiguity visual target than did a low-ambiguity visual target.

We then conducted a similar analysis to investigate the interference on visual target detection from concurrent auditory stimuli. To investigate the effect of high- versus low-ambiguity auditory interferers on the detection of high- versus low-ambiguity visual targets, we performed a repeated measures ANOVA with two within-subjects factors of *target ambiguity* (high-ambiguity visual vs. low-ambiguity visual) and *interfering ambiguity* (high-ambiguity auditory vs. low-ambiguity auditory) (Fig. 4a and b). This revealed worse performance for high-ambiguity visual targets than for low-ambiguity visual targets, $F(1, 19) = 25.02$, $p < .001$, $\eta_p^2 = .57$. By contrast, the ambiguity of the auditory interferer had no effect on performance, $F(1, 19) = 0.26$, $p = .62$, $\eta_p^2 = .01$, and there was no interaction between the ambiguity of the visual target and the ambiguity of the auditory interferer, $F(1, 19) = 0.23$, $p = .64$, $\eta_p^2 = .01$.

Performance as indexed by percent correct is shown in Table 2, and RT data are shown in Supplemental Table S2.

### *Detecting auditory targets in dual streams*

#### *High-ambiguity versus low-ambiguity auditory target detection with high-ambiguity versus low-ambiguity visual interferers*

We then considered the effect of visual interferers on auditory detection (Fig. 5). To investigate the effect of high- versus low-ambiguity visual interferers on the detection of high- versus low-ambiguity auditory targets, we performed a repeated measures ANOVA with two within-subjects factors of *target ambiguity* (high-ambiguity auditory vs. low-ambiguity auditory) and *interfering ambiguity* (high-ambiguity visual vs. low-ambiguity visual). This revealed worse detection of high-ambiguity auditory targets than of low-ambiguity auditory targets, $F(1, 19) = 48.82$, $p < .001$, $\eta_p^2 = .72$, but no main effect of the ambiguity of the interfering visual stream, $F(1, 19) = 1.26$, $p = .28$, $\eta_p^2 = .06$. There was a marginally significant interaction between these two factors, $F(1, 25) = 3.53$, $p = .08$, $\eta_p^2 = .16$, suggesting that a high-ambiguity visual distractor had a greater detrimental effect on a low-ambiguity auditory target than on a high-ambiguity auditory target. A paired-samples $t$-test revealed that the ambiguity of the visual distractor had a significant effect on low-ambiguity auditory target detection, $t(1, 19) = 3.36$, $p < .001$, $d = 0.75$ (Fig. 5a), but not on high-ambiguity auditory target detection, $t(1, 19) = -0.08$, $p = .94$, $d = 0.17$ (Fig. 5b). Performance as indexed by percent correct and RT is shown in Table 3 and Supplemental Table S3, respectively.

This pattern of results was unexpected and is not consistent with a bottleneck in processing complex conjunctions of features; a shared cross-modal resource for processing conjunctions of complex features would have caused the opposite pattern of results. That is, if the resource were shared, tasks taxing conjunctive processing in vision should have interfered disproportionally more with tasks taxing conjunctive processing in auditon (i.e., the high-ambiguity auditory tasks), rather than the low-ambiguity auditory tasks, as we observed. We consider two alternative explanations for this result. One is that there was modulation of feature segregation processes. The low-ambiguity condition can be resolved through analysis of just one of the three features, but these features must be individuated before use. In pilot work, we found that it was

**Table 2** Mean percent correct for visual streams on the dual-stream tasks in Experiments 1 and 2, split by the eight different conditions

| | Experiment 1 | | | | | | | | Experiment 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Performance on Low-Ambiguity Visual Stream | | | | Performance on High-Ambiguity Visual Stream | | | | Performance on Low-Ambiguity Visual Stream | | | | Performance on High-Ambiguity Visual Stream | | | |
| Condition | LVLV | LVHV | LVLA | LVHA | LVHV | HVHV | HVLA | HVHA | LVLV | LVHV | LVLA | LVHA | LVHV | HVHV | HVLA | HVHA |
| Mean | 98.1 | 96.1 | 98.8 | 98.8 | 90.1 | 80.9 | 92.4 | 93.7 | 97.9 | 95.6 | 98.1 | 97.9 | 88.9 | 81.8 | 88.8 | 93.2 |
| *SD* | 1.5 | 5.3 | 1.1 | 1.8 | 6.3 | 4.2 | 6.8 | 3.9 | 2.2 | 4.1 | 1.6 | 1.9 | 5.9 | 5.5 | 4.8 | 4.3 |

*Note*. *LV* low-ambiguity visual, *HV* high-ambiguity visual, *LA* low-ambiguity auditory, *HA* high-ambiguity auditory. For example, the second column under "Performance on Low-Ambiguity Visual stream" (LVHV, score of 96.1), indicates performance on the low-ambiguity visual stream when the other stream was high-ambiguity visual. Similarly, the first column under "Performance on High-Ambiguity Visual Stream" (LVHV, 90.1) indicates performance on the high-ambiguity visual stream of that same condition (i.e., when the other stream was low-ambiguity visual). When both streams are of the same type (i.e., LVLV and HVHV), we report average performance across the two streams

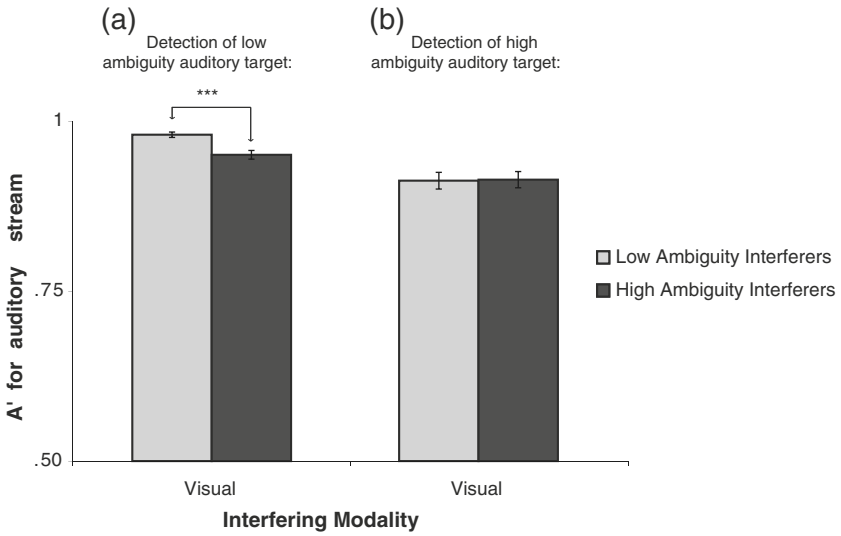



**Fig. 5** Experiment 1 *A′* scores for low-ambiguity auditory (**a**) and high-ambiguity auditory (**b**) target detection in the presence of low-ambiguity (*light bars*) or high-ambiguity (*dark bars*) interferers in the visual modality. (Note that there were never two dual auditory streams.) High-ambiguity auditory targets were harder to detect than low-ambiguity targets, but the ambiguity of the interfering visual stream did not affect high-ambiguity auditory target detection. By contrast, for low-ambiguity auditory target detection, the ambiguity of the interfering visual stream did influence performance *Error bars* represent *SEM*s. ***$p < .001$

hard to decompose a single sound into its three independent features, and so auditory objects comprising three sequentially presented parts were created for this experiment. Even though they were sequential, perhaps some effort was still required to segregate them. We have shown before that segregating sequences of sounds into separate streams requires attention (Carlyon, Cusack, Foxton, & Robertson, 2001; Cusack et al., 2004) and that attending to visual stimuli reduces auditory stream segregation (Carlyon, Plack, Fantini, & Cusack, 2003). Therefore, in the low-ambiguity auditory condition, perhaps the level of concurrent visual load (low vs. high visual ambiguity) modulated segregation and, hence, performance. In contrast, the high-ambiguity auditory task required analysis of all three features together, and so there will be little or no benefit to their segregation—hence, the lack of modulation by the level of concurrent visual load.

A second possible explanation for the lack of an effect of the level of concurrent visual ambiguity on the high-ambiguity auditory task (the similarity of the two bars in Fig. 5b) is that even in the low-ambiguity visual condition in Experiment 1, the distractors shared a single feature with the target (see the Method section). Perhaps even this small demand on conjunctive processing led to maximal interference on the high-ambiguity auditory task, and so no greater interference could be observed with a concurrent high-ambiguity visual task. Although both of these explanations are parsimonious, they are post hoc, and so we conducted another experiment to test whether the unexpected finding is replicable. Furthermore, we made an important modification to try to tease apart the two explanations. We modified the low-ambiguity visual condition so that the distractors did not share any features with the target, to remove even this minimal amount of feature overlap. We conducted a power analysis to ensure that we used an adequate

**Table 3** Mean percent correct for auditory streams on the dual-stream tasks in Experiments 1 and 2, split by the four different conditions

| | Experiment 1ed | | | | Experiment 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | Performance on Low-Ambiguity Auditory Stream | | Performance on High-Ambiguity Auditory Stream | | Performance on Low-Ambiguity Auditory Stream | | Performance on High-Ambiguity Auditory Stream | |
| Condition | LVLA | HVLA | LVHA | HVHA | LVLA | HVLA | LVHA | HVHA |
| Mean | 97.7 | 94.1 | 87.0 | 88.0 | 95.4 | 91.5 | 87.3 | 87.0 |
| *SD* | 2.1 | 3.1 | 5.6 | 6.2 | 4.3 | 6.5 | 6.1 | 8.8 |

*Note*. *LV* low-ambiguity visual, *HV* high-ambiguity visual, *LA* low-ambiguity auditory, *HA* high ambiguity auditory. For example, the first column under "Performance on Low-Ambiguity Auditory Stream" (LVLA, score of 97.7), indicates performance on the low-ambiguity auditory stream when the other stream was low-ambiguity visual

sample size to detect the presence of an effect. On the basis of the effect size from the critical interaction between the ambiguity of the auditory target and the ambiguity of the visual interferer ($\eta_p^2 = .16$; Fig. 5), we determined that a sample of 26 participants would achieve 95 % power (Faul, Erdfelder, Lang, & Buchner, 2007).

## Experiment 2

### Method

#### *Participants*

Twenty-six people (15 female; mean age = 22.2 years, *SD* = 4.1 years) participated in this study in exchange for $15. All participants gave informed written consent after the nature of the study and its possible consequences had been explained to them.

#### *Behavioral procedure*

The behavioral procedure was identical to that described in Experiment 1, except that the ambiguity of the low-ambiguity visual condition was reduced. In Experiment 1, one of the fribble features was fixed, meaning that the low-ambiguity visual condition did contain some feature ambiguity (see the blue "legs" on the fribbles in Fig. 2). Thus, to make the low-ambiguity visual condition even less ambiguous than in Experiment 1, we no longer kept one of the features fixed. This meant that detection of the low-ambiguity visual target could be performed using any one of four unique features and placed minimal demands on conjunctive processing. As in Experiment 1, there was no difference in difficulty across the different stimulus lists, $F(3, 22) = 0.78$, $p = .52$, $\eta_p^2 = .10$, and thus, for all subsequent analyses, we collapsed across stimulus list.

#### *Data analysis*

Data were analyzed in a manner identical to that in Experiment 1. In addition, we also sought to compare performance across the two experiments to determine whether the manipulation of the low-ambiguity targets had any effect. To this end, in addition to the primary ANOVAs within each experiment, we also performed the same series of ANOVAs using an additional between-subjects factor of *experiment* (i.e., Experiment 1 vs. Experiment 2).

### Results

#### *Detecting targets in a single visual or auditory stream*

Target detection performance when only one stream was present is shown in Fig. 6 and as percent correct in Table 1. RT data are provided in Supplemental Table S1. The primary change in this experiment was to remove any requirement for conjunctive processing in the low-ambiguity visual condition by ensuring that the distractors did not share any features with the target. A repeated measures ANOVA with within-subjects factors of *target modality* (visual vs. auditory) and *target ambiguity* (high vs. low) revealed better performance for visual than for auditory streams overall, $F(1, 25) = 6.4$, $p < .05$, $\eta_p^2 = .21$, and better performance for low-ambiguity than for high-ambiguity streams, $F(1, 25) = 18.31$, $p < .001$, $\eta_p^2 = .42$. However, the interaction between target modality and target ambiguity was not significant, $F(1, 25) = 0.12$, $p = .73$, $\eta_p^2 = .005$, indicating that the effect of ambiguity was not significantly different across visual and auditory target detection. Paired-samples *t*-tests within each modality indicated that there was a significant effect of ambiguity in the visual, $t(25) = 5.46$, $p < .001$, $d = 1.07$, and auditory, $t(25) = 2.12$, $p < .05$, $d = 0.42$, conditions.

To compare performance across experiments, we also ran the same ANOVA as that described above with a between-

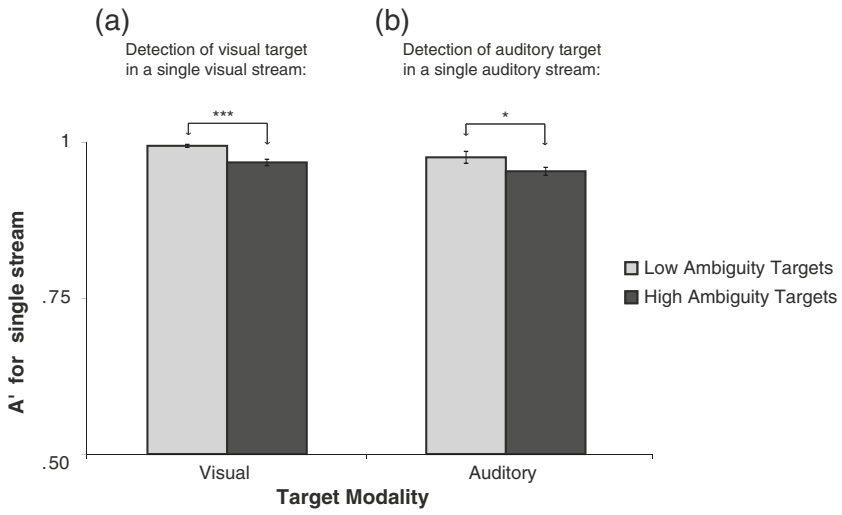



**Fig. 6** Experiment 2 *A′* scores when only a single visual (**a**) or auditory (**b**) stream was present. As in Experiment 1, detection of high-ambiguity targets was harder than detection of low-ambiguity targets in both modalities. *Error bars* represent *SEM*s. ***$p < .001$, *$p < .05$

subjects factor of *experiment* (Experiment 1 vs. Experiment 2). The factor of experiment did not interact with any of the above main effects or interactions (all $F$s < 1.9, $p$s > .2), indicating that the pattern across conditions in Experiment 2 replicated those in Experiment 1. There was a marginal main effect of experiment, driven by the fact that, overall, participants in Experiment 2 performed worse than those in Experiment 1, $F(1, 44) = 3.76$, $p = .06$, $\eta_p^2 = .08$.

*Detecting visual targets in dual streams*

We performed a 2 × 2 × 2 repeated measures ANOVA with within-subjects factors of *target ambiguity* (high vs. low), *interfering modality* (visual vs. auditory), and *interfering ambiguity* (high vs. low). As in Experiment 1, this revealed a significant three-way interaction between target ambiguity, interfering modality, and interfering ambiguity, $F(1, 25) = 25.97$, $p < .001$, $\eta_p^2 = .51$ (Fig. 7). Each of the 3 two-way interactions was also significant [target ambiguity × interfering modality, $F(1, 25) = 50.72$, $p < .001$, $\eta_p^2 = .67$; target ambiguity × interfering ambiguity, $F(1, 25) = 4.80$, $p < .05$, $\eta_p^2 = .16$; interfering modality × interfering ambiguity, $F(1, 25) = 55.38$, $p < .001$, $\eta_p^2 = .69$], as were each of the three main effects [target ambiguity, $F(1, 25) = 138.69$, $p < .001$, $\eta_p^2 = .85$; interfering modality, $F(1, 25) = 36.31$, $p < .001$, $\eta_p^2 = .59$; interfering ambiguity, $F(1, 25) = 25.44$, $p < .001$, $\eta_p^2 = .50$]. To further investigate what was driving the significant three-way interaction, we performed a series of 2 × 2 repeated measures ANOVAs at each of the three factors (described below).

To compare performance across experiments, we also ran the same 2 × 2 × 2 ANOVA as that described above, with an additional between-subjects factor of *experiment* (Experiment 1 vs. Experiment 2). The factor of experiment did not interact with any of the above main effects or interactions (all $F$s < 1.9, $p$s > .2), indicating that the manipulation of the degree of the ambiguity of the low-ambiguity stimuli did not have a significant effect. There was no main effect of experiment on overall performance, $F(1, 44) = 0.17$, $p = .69$, $\eta_p^2 = .004$.

*Low-ambiguity visual target detection: Visual versus auditory interferers* Low-ambiguity visual targets were easily detected in the presence of either auditory or visual interfering streams of both high and low ambiguity (Fig. 7a). A repeated measures ANOVA with within-subjects factors of *interfering modality* (visual vs. auditory) and *interfering ambiguity* (high vs. low) revealed worse performance when interferers were visual, $F(1, 25) = 4.63$, $p < .05$, $\eta_p^2 = .16$, and worse performance when interferers were high ambiguity, $F(1, 25) = 9.52$, $p < .01$, $\eta_p^2 = .28$. There was an interaction between these two factors, $F(1, 25) = 7.91$, $p < .01$, $\eta_p^2 = .24$, which was driven by an effect of the ambiguity of the visual, $t(25) = 3.46$, $p < .01$, $d = 0.68$, but not auditory, $t(25) = -0.06$, $p = .95$, $d = 0.01$, distractors. However, as we will demonstrate below, this effect was stronger for high-ambiguity visual targets than for the low-ambiguity visual targets (see high-ambiguity vs. low-ambiguity visual target detection with high-ambiguity vs. low-ambiguity visual interferers).

*High-ambiguity visual target detection: Visual versus auditory interferers* We observed greater interference from concurrent visual than from auditory streams, and particularly from high-ambiguity visual interferers (Fig. 7b). A repeated measures ANOVA with within-subjects factors of *interfering modality* (visual vs. auditory) and *interfering ambiguity* (high vs. low) revealed worse performance on high-ambiguity visual target detection if the interfering stream was visual than if it was auditory, $F(1, 25) = 63.65$, $p < .001$, $\eta_p^2 = .72$. The effect of the ambiguity of the interferer was also significant, $F(1, 25) =$

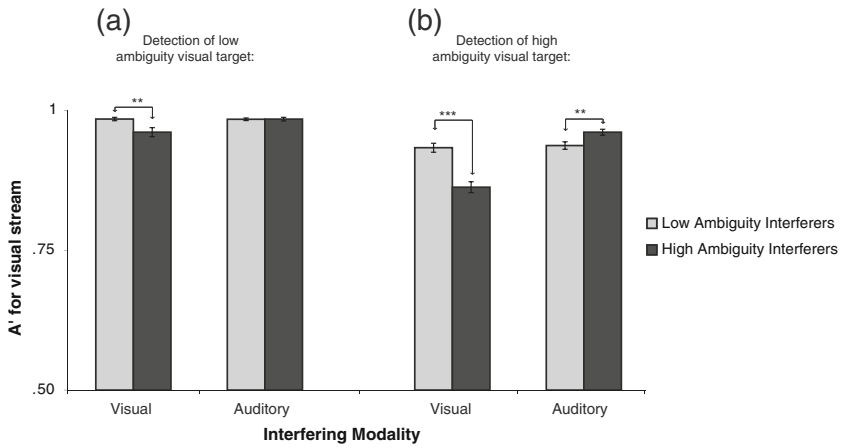



**Fig. 7** Experiment 2 *A′* scores for low-ambiguity visual (**a**) and high-ambiguity visual (**b**) target detection in the presence of low-ambiguity (*light bars*) or high-ambiguity (*dark bars*) interferers in both modalities. We observed the greatest interference when participants completed two high-ambiguity visual tasks. By contrast, the ambiguity of the auditory interferers had the opposite effect than did visual interferers, suggesting that the resources for processing conjunctions of features in vision and audition are not shared. *Error bars* represent *SEM*s. ***$p < .001$, **$p < .01$

22.63, $p < .001$, $\eta_p^2 = .48$, and there was a significant interaction between the interfering modality and the interfering ambiguity, $F(1,25) = 59.33$, $p < .001$, $\eta_p^2 = .70$. Paired samples $t$-tests to investigate this interaction further revealed that high-ambiguity visual distractors interfered significantly more than low-ambiguity visual distractors, $t(25) = 8.18$, $p < .001$, $d = 1.60$. By contrast, the high-ambiguity auditory distractors interfered significantly less than the low-ambiguity auditory distractors, $t(25) = -3.4$, $p < .01$, $d = 0.67$. As with Experiment 1, this was not just a general task difficulty effect, because the single-stream high-ambiguity auditory condition was significantly harder than the single-stream high-ambiguity visual condition, $t(25) = 2.07$, $p < .05$, $d = 0.41$ (Fig. 6), but interfered less with high-ambiguity visual target detection. This crossover interaction provides strong evidence that processing conjunctions of features in vision and audition are qualitatively different in that they tax different cognitive mechanisms. Not only did we fail to observe interference from processing conjunctions of auditory features (high-ambiguity auditory interferers) on processing conjunctions of visual features (high-ambiguity visual targets), in Experiment 2 the effect of auditory interference was in the opposite direction to what we observed from visual interference: Auditory tasks that placed less emphasis on processing auditory conjunctions interfered more with processing visual conjunctions.

*High-ambiguity versus low-ambiguity visual target detection with high-ambiguity versus low–ambiguity interferers* Similar to Experiment 1, we found more specific interference effects for high-ambiguity than for low-ambiguity visual targets. We also tested whether these interference effects from visual stimuli significantly differed as a function of target ambiguity. We performed a repeated measures ANOVA with two within-subjects factors of *target ambiguity* (high-ambiguity visual vs. low-ambiguity visual) and *interfering ambiguity* (high-ambiguity visual vs. low-ambiguity visual) (Fig. 7a and b). This revealed worse performance for high-ambiguity targets than for low-ambiguity targets, $F(1, 25) = 158.02$, $p < .001$, $\eta_p^2 = .86$, and worse performance when the interferers were high ambiguity than if they were low ambiguity, $F(1, 25) = 55.86$, $p < .001$, $\eta_p^2 = .69$. There was also a significant interaction between these two factors, $F(1, 25) = 25.60$, $p < .001$, $\eta_p^2 = .52$, indicating that a high-ambiguity visual interferer had a greater detrimental effect on detection of a high-ambiguity visual target than of a low-ambiguity visual target.

We then conducted a similar analysis to investigate the interference on visual target detection from concurrent auditory stimuli. To investigate the effect of high- versus low-ambiguity auditory interferers on the detection of high- versus low-ambiguity visual targets, we performed a repeated measures ANOVA with two within-subjects factors of *target ambiguity* (high-ambiguity visual vs. low-ambiguity visual) and *interfering ambiguity* (high-ambiguity auditory vs. low-ambiguity auditory) (Fig. 7a and b). This revealed an effect of visual target ambiguity, $F(1, 25) = 52.19$, $p < .001$, $\eta_p^2 = .68$, an effect of the auditory interferer ambiguity, $F(1, 25) = 9.09$, $p < .01$, $\eta_p^2 = .27$, and an interaction between the ambiguity of the visual target and the ambiguity of the auditory interferer, $F(1, 25) = 7.95$, $p < .01$, $\eta_p^2 = .24$. However, this interaction was driven by a different pattern of results than that observed for visual interferers: *Low*-ambiguity auditory interferers negatively impacted high-ambiguity visual target detection more than did high-ambiguity auditory interferers. By contrast, in vision, *high*-ambiguity visual interferers were the most detrimental to high-ambiguity visual target detection.

Performance as indexed by percent correct is shown in Table 2, and RT data are shown in Supplemental Table S2.

### *Detecting auditory targets in dual streams*

*High-ambiguity versus low-ambiguity auditory target detection with high-ambiguity versus low-ambiguity visual interferers* We then considered the effect of visual interferers on auditory detection (Fig. 8). To investigate the effect of high- versus low-ambiguity visual interferers on the detection of high- versus low-ambiguity auditory targets, we performed a repeated measures ANOVA with two within-subjects factors of *target ambiguity* (high-ambiguity auditory vs. low-ambiguity auditory) and *interfering ambiguity* (high-ambiguity visual vs. low-ambiguity visual). This revealed worse detection of high-ambiguity auditory targets than of low-ambiguity auditory targets, $F(1, 25) = 12.95$, $p < .001$, $\eta_p^2 = .34$, and worse detection when the visual interferers were high ambiguity, $F(1, 25) = 5.53$, $p < .05$, $\eta_p^2 = .18$. However, in contrast to Experiment 1, we observed no interaction between these two factors, $F(1, 25) = 0.57$, $p = .46$, $\eta_p^2 = .02$. Paired-samples $t$-tests indicated that the ambiguity of the visual distractor had a significant effect on low-ambiguity auditory target detection, $t(25) = 2.65$, $p < .05$, $d = 0.52$, but not on high-ambiguity auditory target detection, $t(25) = 1.33$, $p = .20$, $d = 0.26$. Performance as indexed by percent correct and RT is shown in Table 3 and Supplemental Table S3, respectively

To compare performance across experiments, we also ran the same ANOVA as that described above with a between-subjects factor of *experiment* (Experiment 1 vs. Experiment 2). There was a marginal main effect of experiment, $F(1, 44) = 3.41$, $p = .07$, $\eta_p^2 = .07$, driven by the fact that overall performance in Experiment 2 was lower than in Experiment 1. However, the factor of experiment did not interact with any of the above main effects or interactions (all $F$s $< 0.85$, $p$s $> .4$), providing no evidence of an effect of using fully unambiguous visual stimuli in Experiment 2 on our pattern of results.

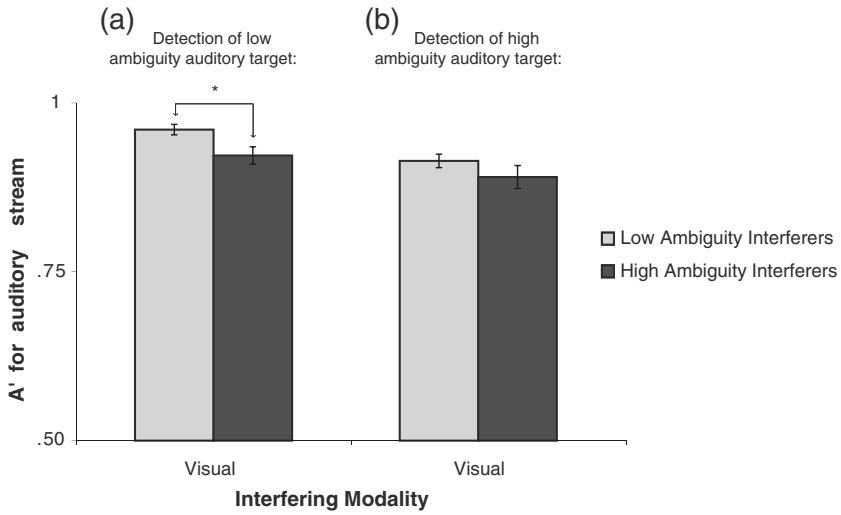



**Fig. 8** Experiment 2 *A′* scores for low-ambiguity auditory (**a**) and high-ambiguity auditory (**b**) target detection in the presence of low-ambiguity (*light bars*) or high-ambiguity (*dark bars*) interferers in the visual modality. (Note that there were never two dual auditory streams.) *Error bars* represent *SEM*s. *$p < .05$

## General discussion

In two experiments, we found that the detection of targets in streams of novel visual stimuli was harder when they were of high feature ambiguity and taxed processing conjunctions of complex features than when they had low feature ambiguity and could be readily solved on the basis of a single complex feature (Figs. 3a and 6a). These results add further support to growing evidence of a capacity-limited mechanism that represents conjunctions of complex visual features to resolve stimuli with high feature overlap. A growing body of work from neuroimaging, human neuropsychology, and lesion studies in animals, using similar paradigms and objects, has suggested that this mechanism involves the perirhinal cortex (e.g., Barense et al., 2005; Barense, Groen, et al., 2012; Bartko et al., 2007; Bussey et al., 2002). Critically, this perirhinal involvement is not a mere function of the perceptual similarity of target and distractors (Duncan & Humphreys, 1989) but, instead, is a function of the requirement to process conjunctions of features. Many experiments have demonstrated that the perirhinal cortex is not required, nor is it recruited, for very difficult detection tasks involving discrimination along a single feature dimension (e.g., size or color; Barense, Groen, et al., 2012; Devlin & Price, 2007; Lee et al., 2008; O'Neil et al., 2009). Another factor that can influence visual search experiments in a way that might be mistaken for an effect of stimulus complexity is the degree of homogeneity of the distractors, with increasing difficulty as the similarity between distractors decreases (Duncan & Humphreys, 1989). However, there was no evidence that this modulated performance in our experiments, because, if anything, the distractors were marginally more homogeneous in the high-ambiguity condition (i.e., they shared more features with each other than in the low-ambiguity conditions). A second key element of the present experiments was the creation of a set of novel auditory stimuli that could be presented at a similar rate to the visual stimuli, in a matched target detection task. As for visual stimuli, we found that auditory targets with high feature ambiguity were harder to detect than low feature ambiguity targets (Figs. 3b and 6b). This suggests that processing conjunctions of auditory features also forms a bottleneck in audition (although see the discussion below on the automaticity of segregation).

To investigate whether the capacity-limited resources for processing conjunctions of features in audition and vision are modality specific or amodal, we also presented two streams of stimuli concurrently. In particular, we investigated whether processing of conjunctions of features in the auditory domain interfered with processing of conjunctions of feature in the visual domain (and vice versa), over and above any nonspecific concurrent task interference. The results showed that concurrent demands on processing conjunctions of features interfered only within a modality: In both Experiments 1 and 2, performance of concurrent visual streams was modulated by the degree to which they taxed the processing of conjunctions of features, but this degree of conjunctive processing in a concurrent auditory task had no effect on its interference with a visual task (Figs. 4 and 7). More specifically, performance on high-ambiguity visual target detection was negatively affected by simultaneously having to process conjunctions of visual features in the concurrent stream (i.e., worse performance on the HV stream of the HVHV relative to HVLV conditions). By contrast, when the concurrent stream was auditory, processing conjunctions of visual features was not affected by simultaneously having to process conjunctions of auditory features (i.e., equivalent performance on the HV stream of the HVHA and HVLA conditions). Similarly, for audition, performance in the high-ambiguity auditory condition was not affected by the requirement to process conjunctions of features in the visual domain (i.e., equivalent

performance on the HA stream of the HAHV and HALV conditions; Figs. 5b and 8b).

Across the two experiments, there was interference when a low-ambiguity auditory task was performed concurrently with a high-ambiguity visual task. In Experiment 1, low-ambiguity auditory target detection was pushed down by high-ambiguity visual interferers, whereas in Experiment 2, high-ambiguity visual target detection was pushed down by low-ambiguity auditory interferers. This suggests a bottleneck in a resource that is shared across low-ambiguity auditory and high-ambiguity visual. We speculate that it might be, for example, that attention is required to segregate sounds (cf. Carlyon et al., 2001) but to integrate visual components (cf. Treisman & Gelade, 1980). However, what is clear is that these results are qualitatively inconsistent with a bottleneck resulting from a shared conjunctive process recruited by high-ambiguity stimuli across vision and audition, providing a clear answer to our initial question. That is, if the resource for processing conjunctions of features were shared across vision and audition, tasks taxing conjunctive processing of auditory features should have interfered *more* with processing conjunctions of visual features; by contrast, we found that tasks taxing conjunctive processing of auditory features interfered *less* with processing conjunctions of visual features, suggesting that this binding resource is not shared across modalities. Further future work is need to understand binding and segregation processes in audition. The difficulty we encountered in pilot work in making single auditory objects with separable features suggests that perhaps the default percept in audition is an integrated representation. This is consistent with the highly integrative, nonlinear responses observed in auditory temporal representations (Machens, Wehr, & Zador, 2004).

One potential criticism of this work is that we cannot be completely sure that the high-ambiguity conditions taxed processing for conjunctions of complex features and were not performed by a serial comparison of the individual features. Although the present experiment does not allow us to unequivocally rule out serial comparison, there is converging evidence from many sources that conjunctions of complex features are rapidly generated when complex visual objects are viewed. The phenomenology of recognizing an object is not of a serial feature-by-feature process of elimination, and concurring with this, electrophysiological and behavioral measures show that object recognition is remarkably rapid (Naci et al., 2012; Thorpe, Fize, & Marlot, 1996). In search tasks, when a single object must be found in a naturalistic scene, search is efficient, which has be taken as evidence of the derivation of a high-level sparse representation in which single objects "pop out" (Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). In one study that used stimuli that were even more similar to those in the present experiments, eye movements were measured while participants compared two visual objects under conditions of high and low ambiguity (Barense, Groen, et al., 2012). Participants made more eye movement transitions *within* an individual object, as compared with *across* objects, in the high-ambiguity condition, relative to the low-ambiguity condition, suggesting that participants were creating a representation of all of the features in one object before encoding the other one, rather than serially comparing single features between the objects in turn. A further line of evidence, which was presented in the introduction and not reiterated here, is from neurophysiological recording, which shows the coding of conjunctions of complex features in the anterior temporal regions. In sum, there is robust evidence from many sources that conjunctive representations of complex features are formed.

There have been far fewer attempts to investigate processing conjunctions of features in audition. In vision, the existence of distinct neural maps that encode different features is established. In audition, there is much less consensus. Even in the early auditory cortex, neurons respond to complex spectrotemporal patterns (Depireux, Simon, Klein, & Shamma, 2001). Although there do appear to be multiple regions in the auditory cortex that are tuned to different kinds of information (Kaas & Hackett, 2000), there is still no consensus on how many brain regions there are or what auditory features they encode. For this reason, in the present experiment, we did not present simultaneous auditory streams, and to maximize separation sound, we distributed auditory features across time. Although this did achieve phenomenological separation of the features, perhaps neurally it achieved only partial separation, because there are nonlinear temporally extended responses even in early auditory regions (Machens et al., 2004).

Our results suggest that the resources required for processing conjunctions of intramodal visual and auditory features are modality specific. If, however, these auditory and visual resources are independent, what processes underlie binding information *across* modalities (i.e., cross-modal integration)? Our experiment did not assess cross-modal integration: There was no requirement that participants integrate information across the two streams of information; sounds and objects were always temporally jittered, randomly presented, and spatially distinct. As such, we cannot address this hypothesis directly. However, integrating information within a given modality likely reflects a distinct process from cross-modal integration. Nearly 100 years ago, it was observed that RTs were shorter for bimodal stimuli (e.g., a light and a tone), as compared with unimodal stimuli (Todd, 1912), an effect that has been replicated many times (e.g., Diederich & Colonius, 1987; Giray & Ulrich, 1993; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; Miller, 1982, 1986; Plat, Praamstra, & Horstink, 2000). Moreover, there is evidence to suggest that auditory–visual integration of speech is more tolerant of auditory and visual temporal asynchrony than when auditory information alone is presented (Grant, van Wassenhove, & Poeppel, 2004), further suggesting that

cross-modal integration reflects a different process than does unimodal integration.

There is also a wealth of evidence that semantic integration of auditory and visual information facilitates object detection, and had the two streams of information in the present experiment been semantically related (e.g., the sound of a bark and a visual image of a dog), the results might have been different. For example, Laurienti, Kraft, Maldjian, Burdette, and Wallace (2004) examined the influence of cross-modal versus intramodal pairings on stimulus detection speed. In cross-modal pairings (visual–sound), colored circles were presented with an auditory verbalization of a color (e.g., "blue"). In intramodal (visual–visual) pairings, the written word "blue" was presented with a colored circle. Both types of pairings could be congruent (e.g., blue with "blue") or incongruent (blue with "green"). Shorter RTs were observed for semantically congruent audiovisual pairs, but not for semantically congruent visual–visual pairs. In contrast to the multisensory facilitation observed for congruent audiovisual pairs, significantly longer RTs were observed when the audiovisual pairs were incongruent. Thus, cross-modal, but not intramodal, semantic congruency affects stimulus detection. Similarly striking effects of audiovisual semantic congruency have been found on more complex tasks, such as object identification, categorization, or recognition (e.g., Chen & Spence, 2010; Murray et al., 2004; Yuval-Greenberg & Deouell, 2009). In one demonstration, RTs were significantly shorter for semantically congruent bimodal stimuli than would be predicted by independent processing of information about the auditory and visual targets (Suied, Bonneel, & Viaud-Delmon, 2009).

A candidate neural structure for this integration of cross-modal feature information into higher-level semantic representations is the perirhinal cortex in the anterior temporal lobe. The perirhinal cortex is one of the first cortical fields within the ventral visual stream to have convergence of information from different sensory modalities (Carmichael & Price, 1995; Friedman et al., 1986; Suzuki & Amaral, 1994), suggesting that it might play a role in the integration of tactile and auditory features to create a meaningful representation of an object. In nonhuman primates, removal of the rhinal cortex impairs tactile–visual delayed nonmatching-to-sample (Goulet & Murray, 2001) and flavor–visual association learning (Parker & Gaffan, 1998). Moreover, recent functional neuroimaging from humans has provided evidence that the perirhinal cortex is sensitive to the degree of congruency between tactile and visual information (Holdstock et al., 2009) and between auditory and visual information (Taylor et al., 2006; Taylor et al., 2009) for complex objects. Whereas it seems highly likely that the perirhinal cortex is critical for high-level cross-modal feature integration, the present research suggests that different systems are involved in processing conjunctions of features intramodally *within* vision and audition.

In summary, results from a series of single-stream and dual-stream target detection tasks indicated that processing conjunctions of features in vision is capacity limited but modality specific, since processing conjunctions of features in the auditory domain did not impact performance on processing conjunctions of features in the visual domain. The same pattern was observed for audition: Performance on the auditory task that taxed feature conjunctions was not affected by the requirement to process conjunctions of features in the visual domain. Interestingly, however, we observed interference between tasks that stressed conjunctive processing in the visual domain and tasks that did *not* stress conjunctive processing in the auditory domain.

**Acknowledgments** We thank Nicholas Rule for statistical assistance and Felicia Yue Zhang for help with data collection. This work was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to M.D.B.

## References

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Biological Sciences, 273*(1592), 1339–1345.

Bang, S. J., & Brown, T. H. (2009). Perirhinal cortex supports acquired fear of auditory objects. *Neurobiology of Learning and Memory, 92*(1), 53–62.

Barense, M. D., Bussey, T. J., Lee, A. C., Rogers, T. T., Davies, R. R., Saksida, L. M., Murray E. A., & Graham, K. S. (2005). Functional specialization in the human medial temporal lobe. *Journal of Neuroscience, 25*(44), 10239–10246.

Barense, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia, 45*(13), 2963–2974.

Barense, M. D., Groen, I. I., Lee, A. C., Yeung, L. K., Brady, S. M., Gregori, M., Kapur, N., Bussey, T. J., Saksida, L. M., & Henson, R. N. (2012). Intact memory for irrelevant information impairs perception in amnesia. *Neuron, 75*(1), 157–167.

Barense, M. D., Henson, R. N., & Graham, K. S. (2011). Perception and conception: Temporal lobe activity during complex discriminations of familiar and novel faces and objects. *Journal of Cognitive Neuroscience, 23*(10), 3052–3067.

Barense, M. D., Henson, R. N. A., Lee, A. C., & Graham, K. S. (2010). Medial temporal lobe activity during complex discrimination of faces, objects, and scenes: Effects of viewpoint. *Hippocampus, 20*(3), 389–401.

Barense, M. D., Ngo, J. K., Hung, L. H., & Peterson, M. A. (2012). Interactions of memory and perception in amnesia: The figure-ground perspective. *Cerebral Cortex, 22*(11), 2680–2691.

Barense, M. D., Rogers, T. T., Bussey, T. J., Saksida, L. M., & Graham, K. S. (2010). Influence of conceptual knowledge on visual object discrimination: Insights from semantic dementia and MTL amnesia. *Cerebral Cortex, 20*(11), 2568–2582.

Bartko, S. J., Winters, B. D., Cowell, R. A., Saksida, L. M., & Bussey, T. J. (2007). Perirhinal cortex resolves feature ambiguity in configural object recognition and perceptual oddity tasks. *Learning and Memory, 14*(12), 821–832.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809–823.

Bonnel, A. M., & Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics, 60*(2), 179–190.

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

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46*(2), 369–384.

Bussey, T. J., & Saksida, L. M. (2002). The organization of visual object representations: A connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience, 15*(2), 355–364.

Bussey, T. J., Saksida, L. M., & Murray, E. A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *European Journal of Neuroscience, 15*(2), 365–374.

Bussey, T. J., Saksida, L. M., & Murray, E. A. (2003). Impairments in visual discrimination after perirhinal cortex lesions: Testing 'declarative' vs. 'perceptual-mnemonic' views of perirhinal cortex function. *European Journal of Neuroscience, 17*(3), 649–660.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport, 10*(12), 2619–2623.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology. Human Perception and Performance, 27*(1), 115–127.

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception, 32*(11), 1393–1402.

Carmichael, S. T., & Price, J. L. (1995). Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology, 363*(4), 642–664.

Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114*(3), 389–404.

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Components of recognition memory: Dissociable cognitive processes or just differences in representational complexity? *Hippocampus, 20*(11), 1245–1262.

Cusack, R., & Carlyon, R. P. (2003). Perceptual asymmetries in audition. *Journal of Experimental Psychology. Human Perception and Performance, 29*(3), 713–725.

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology. Human Perception and Performance, 30*(4), 643–656.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics, 62*(5), 1112–1120.

Depireux, D. A., Simon, J. Z., Klein, D. J., & Shamma, S. A. (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *Journal of Neurophysiology, 85*(3), 1220–1234.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 267–299). New York: Elsevier Science.

Deutsch, D. (1974). An auditory illusion. *Nature, 251*(5473), 307–309.

Devlin, J. T., & Price, C. J. (2007). Perirhinal contributions to human visual perception. *Current Biology, 17*(17), 1484–1488.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component. *Psychological Research, 49,* 23–29.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

Erez, J., Lee, A. C., & Barense, M. D. (2013). It does not look odd to me: Perceptual impairments and eye movements in amnesic patients with medial temporal lobe damage. *Neuropsychologia, 51*(1), 168–180.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22*(13), 5749–5759.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Friedman, D. P., Murray, E. A., O'Neill, J. B., & Mishkin, M. (1986). Cortical connections of the somatosensory fields of the lateral sulcus of macaques: Evidence for a corticolimbic pathway for touch. *Journal of Comparative Neurology, 252*(3), 323–347.

Fukushima, K. (1980). Neocognitron - a Self-Organizing Neural Network Model for a Mechanism of Pattern-Recognition Unaffected by Shift in Position. *Biological Cybernetics, 36*(4), 193–202.

Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology. Human Perception and Performance, 19*(6), 1278–1291.

Goulet, S., & Murray, E. A. (2001). Neural substrates of crossmodal association memory in monkeys: The amygdala versus the anterior rhinal cortex. *Behavioral Neuroscience, 115*(2), 271–284.

Grant, K. W., van Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech Communication, 44,* 43–53.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Summerfield, A. Q., Palmer, A. R., Elliott, M. R., & Bowtell, R. W. (2000). Modulation and task effects in auditory processing measured using fMRI. *Human Brain Mapping, 10*(3), 107–119.

Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience, 27*(30), 7881–7887.

Holdstock, J. S., Hocking, J., Notley, P., Devlin, J. T., & Price, C. J. (2009). Integrating visual and tactile information in the perirhinal cortex. *Cerebral Cortex, 19*(12), 2993–3000.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology. Human Perception and Performance, 20*(1), 131–153.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 11793–11799.

Kholodar-Smith, D. B., Allen, T. A., & Brown, T. H. (2008). Fear conditioning to discontinuous auditory cues requires perirhinal cortical function. *Behavioral Neuroscience, 122*(5), 1178–1185.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158*(4), 405–414.

Lee, A. C., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., Bussey, T. J., Davies, R. R., Kapur, N., Hodges, J. R., & Graham, K. S. (2005). Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus, 15*(6), 782–797.

Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., Hodges, J. R., & Graham, K. S. (2005). Perceptual deficits in amnesia: Challenging the medial temporal lobe 'mnemonic' view. *Neuropsychologia, 43*(1), 1–11.

Lee, A. C., Scahill, V. L., & Graham, K. S. (2008). Activating the medial temporal lobe during oddity judgment for faces and scenes. *Cerebral Cortex, 18*(3), 683–696.

Lindquist, D. H., Jarrard, L. E., & Brown, T. H. (2004). Perirhinal cortex supports delay fear conditioning to rat ultrasonic social signals. *Journal of Neuroscience, 24*(14), 3610–3617.

Machens, C. K., Wehr, M. S., & Zador, A. M. (2004). Linearity of cortical receptive fields measured with natural sounds. *Journal of Neuroscience, 24*(5), 1089–1100.

Martinovic, J., Gruber, T., & Muller, M. M. (2008). Coding of visual object features and feature conjunctions in the human brain. *PLoS One, 3*(11), e3781.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247–279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40*(5), 331–343.

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage, 21*(1), 125–135.

Naci, L., Taylor, K. I., Cusack, R., & Tyler, L. K. (2012). Are the senses enough for sense? Early high-level feedback shapes our comprehension of multisensory objects. *Frontiers in Integrative Neuroscience, 6,* 82.

O'Neil, E. B., Cate, A. D., & Kohler, S. (2009). Perirhinal cortex contributes to accuracy in recognition memory and perceptual discriminations. *Journal of Neuroscience, 29*(26), 8329–8334.

Parker, A., & Gaffan, D. (1998). Interaction of frontal and perirhinal cortices in visual object recognition memory in monkeys. *European Journal of Neuroscience, 10*(10), 3044–3057.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*(4), 767–776.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of Shape Processing. *Image and Vision Computing, 11*(6), 317–333.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47*(3), 329–342.

Plat, F. M., Praamstra, P., & Horstink, M. W. (2000). Redundant-signals effects on reaction time, response force, and movement-related potentials in Parkinson's disease. *Experimental Brain Research, 130*(4), 533–539.

Rae, G. (1976). Table of A'. *Perceptual and Motor Skills, 42,* 98.

Rauschecker, J. P., & Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology, 91*(6), 2578–2589.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology, 50*(1–2), 19–26.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50 ms. *Journal of Experimental Social Psychology, 44,* 1100–1105.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups - An integration of theories. *Psychology, Public Policy, and Law, 7*(1), 36–97.

Squire, L. R., & Wixted, J. T. (2011). The Cognitive Neuroscience of Human Memory Since H.M. *Annual Review of Neuroscience, 34,* 259–288.

Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research, 194*(1), 91–102.

Suzuki, W. A., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *The Journal of Comparative Neurology, 350*(4), 497–533.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19,* 109–139.

Tanaka, K. (1997). Mechanisms of visual object recognition: Monkey and human studies. *Current Opinion in Neurobiology, 7*(4), 523–529.

Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 103*(21), 8239–8244.

Taylor, K. I., Stamatakis, E. A., & Tyler, L. K. (2009). Crossmodal integration of object features: Voxel-based correlations in brain-damaged patients. *Brain, 132*(Pt 3), 671–683.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.

Todd, J. W. (1912). Reaction to multiple stimuli. In R. S. Woodsworth (Ed.), *Archives of Psychology, No. 25* (Vol. XXI). New York: The Science Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. E. (2004). Processing objects at different levels of specificity. *Journal of Cognitive Neuroscience, 16*(3), 351–362.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4*(2), 157–165.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51*(2), 167–194.

Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 13*(1), 1–7.

Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition, 7*(1/2/3), 297–322.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics, 73*(6), 1650–1671.

Yuval-Greenberg, S., & Deouell, L. Y. (2009). The dog's meow: Asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research, 193*(4), 603–614.