Vision and audition do not share attentional resources in sustained tasks

Roberto Arrighi[1,2,3]*, Roy Lunardi[1] and David Burr[1,2,3]

[1] Facoltà di Psicologia, Università degli Studi di Firenze, Firenze, Italy
[2] Istituto Scientifico per la Neuropsichiatria dell'Infanzia e dell'Adolescenza Fondazione Stella Maris, Pisa, Italy
[3] Neuroscience Institute, National Research Council, Pisa, Italy

Our perceptual capacities are limited by attentional resources. One important question is whether these resources are allocated separately to each sense or shared between them. We addressed this issue by asking subjects to perform a double task, either in the same modality or in different modalities (vision and audition). The primary task was a multiple object-tracking task (Pylyshyn and Storm, 1988), in which observers were required to track between 2 and 5 dots for 4 s. Concurrently, they were required to identify either which out of three gratings spaced over the interval differed in contrast or, in the auditory version of the same task, which tone differed in frequency relative to the two reference tones. The results show that while the concurrent visual contrast discrimination reduced tracking ability by about 0.7 d′, the concurrent auditory task had virtually no effect. This confirms previous reports that vision and audition use separate attentional resources, consistent with fMRI findings of attentional effects as early as V1 and A1. The results have clear implications for effective design of instrumentation and forms of audio–visual communication devices.

Keywords: sustained attention, audio–visual integration, cross-modal perception

INTRODUCTION

To successfully interact with the stimuli of our environment, we need to process selectively the information most relevant for our tasks. This process is usually termed "attention" (James, 1890/1950). When stimuli are attended to their processing become more rapid, more accurate, and more detailed (Posner et al., 1980; Desimone and Duncan, 1995; Carrasco and McElree, 2001; Carrasco et al., 2004; Liu et al., 2005, 2009). Attention improves performance on several visual tasks, such as contrast sensitivity, speed and orientation discrimination as well as spatial resolution (Lee et al., 1999; Morrone et al., 2002; Carrasco et al., 2004; Alais et al., 2006a). As attentive resources are limited, when the stimuli demanding attention for a perceptual task exceed system capacity, performance decreases. For example, in a visual search task in which an object (target) has to be detected amongst irrelevant items (distractors), reaction times increase directly with distractor number (unless the difference between the stimuli is so striking to make the target pop out from the cluttered scene). This correlation reflects the limited capacity of selective attention that prevents the observer from monitoring all items at the same time.

Similarly, when more than one perceptual task is performed at the same time, overall performance decreases because of the underlying processing limitations. This occurs even for simple tasks, such as naming a word or identifying the pitch of a tone (Pashler, 1992; Pashler and O'Brien, 1993; Huang et al., 2004). Interference between concurrent perceptual tasks of the same sensory modality has been consistently reported in many psychological and psychophysical studies (Navon et al., 1984; Pashler, 1994; Bonnel and Prinzmetal, 1998; Alais et al., 2006b). However, the evidence for audiovisual cross-modal interference is conflicting (Duncan et al., 1997;

Jolicoeur, 1999; Arnell and Duncan, 2002). Bonnel and Hafter (1998) found that in a identification task in which the sign of a change (luminance in vision and intensity in audition) had to be detected, performance in dual-task conditions were lower than in the single-task conditions regardless the interference was in the same or different modalities. Spence et al. (2000) found that selecting an auditory stream of words presented concurrently with a second (distractor) stream, it is more difficult if a video of moving lips mimicking the distracting sounds it is also displayed. These psychophysical findings are not only congruent with some of the cognitive literature of the 1970s and 1980s (Taylor et al., 1967; Tulving and Lindsay, 1967; Alais et al., 2006b), but also with recent neurophysiological and imaging results. For example, Joassin et al. (2004) examined the electrophysiological correlates for auditory interference with vision by an identification task of non-ambiguous complex stimuli such as faces and voices. Their results suggest that cross-modal interactions occur at various different stages, involving brain areas such as fusiform gyrus, associative auditory areas (BA 22), and the superior frontal gyri. Hein et al. (2007) showed with a functional magnetic resonance (fMRI) study, that even without competing motor responses, a simple auditory decision interferes with visual processing at neural levels including prefrontal cortex, middle temporal cortex, and other visual regions. Taken together these results imply that limitations on resources for vision and audition operate at a central level of processing, rather than in the auditory and visual peripheral senses.

However, much evidence also supports the notion of independence of attentional resources for vision and audition (Allport et al., 1972; Triesman and Davies, 1973; Shiffrin and Grantham, 1974; Alais et al., 2006b; Santangelo et al., 2010). For example,

Larsen et al. (2003) compared subjects' accuracy for identification of two concurrent stimuli (such as a visual and spoken letter) relative to performance in a single-task. They found that the proportion of correct response was almost the same for all experimental conditions and, furthermore, in the divided-attention condition the probability to correctly report a stimulus in one modality was independent of whether the stimulus was correctly reported in the other modality. Similarly, Bonnel and Hafter (1998) used an audiovisual dual-task paradigm to show that when identification of the direction of a stimulus change is capacity-limited (see above), simple detection of visual and auditory patterns is governed by "capacity-free" processes, as in the detection task there was no performance drop compared with single-task controls. Similar results have been achieved by Alais et al. (2006b) by measuring discrimination thresholds for visual contrast and auditory pitch. Visual thresholds were unaffected by concurrent pitch discrimination of chords and vice versa. However, when two tasks were performed within the same modality, thresholds increased by a factor of around two for visual discrimination and four for auditory discrimination. In line with these psychophysical results, a variety of imaging studies suggests that attention can act unimodally at early levels including the primary cortices such as A1 and V1 (Jancke et al., 1999a,b; Posner and Gilbert, 1999; Somers et al., 1999).

Most of the studies mentioned deal with dual-task conditions where both tasks are brief (hundreds of milliseconds) stimuli to be detected or discriminated. Very few consider conditions in which one of the tasks must be performed by continuously monitoring a specific pattern over a temporal scale of seconds, even though this is a typical requirement for many everyday activities, such as reading or driving. These tasks require *sustained* rather than *transient* attention. Here we investigate whether sustained attentional resources are independent for vision and audition. We measure performance on the multiple object-tracking (MOT) task of Pylyshyn and Storm (1988), while asking subjects to perform simultaneously either a visual contrast discrimination task or an auditory pitch discrimination task. The results show strong within-modality interference, but very little cross-modality interference, strongly supporting the idea that in sustained tasks each modality has access to a separate pool of attentional resources.

MATERIALS AND METHODS
SUBJECTS
Four naive subjects (two males and two females, mean age 26 years), all with normal hearing and normal or corrected-to-normal visual acuity, served as subjects. All gave informed consent to participate to the study that was conducted in accordance with the guidelines of the University of Florence. The tasks were performed in a dimly lit, sound-attenuated room.

STIMULI AND PROCEDURE
All visual stimuli were presented on a Sony Trinitron CRT monitor (screen resolution of 1024 × 768 pixels, 32 bit color depth, refresh rate of 60 Hz, and mean luminance 68.5 cd/m²) subtending (40° × 30°) at the subjects view distance of 57 cm. To create visual stimuli we used Psychophysics toolbox (version 2) for MATLAB (Brainard, 1997; Pelli, 1997) on a Mac G4 running Mac OSX 9. Auditory stimuli were digitized at a rate of 65 kHz, and presented

through two high quality loudspeakers (Creative MMS 30) flanking the computer screen and lying in the same plane 60 cm from the subject. Speaker separation was around 80 cm and stimuli intensity was 75 dB at the sound source.

Subjects were tested on two different kinds of perceptual tasks. The primary task was visual tracking of multiple moving objects (MOT; Pylyshyn and Storm, 1988). The MOT task consisted of 12 disks (diameter 0.9°) moving across a gray background at 5°/s. They moved in straight lines, and when colliding with other dots or the sides bounced appropriately (obeying the laws of physics). At the start of each trial 3–5 disks were displayed in green (*xyY* coordinates = 0.25,0.69, 39.5) for 2 s to indicate that those were the targets whilst the remaining were displayed an isoluminant red (*xyY* coordinates = 0.61, 0.33, 39.5). The trial continued for 4 s (tracking period), then disks stopped and four became orange (*xyY* coordinates = 0.52,0.44, 39.4; see **Movie S1** in Supplementary Material). The subjects' task was to choose which of these was the target (only one valid target turned orange on each trial). Subjects were familiarized with the task during a training session of 50 trials before starting the experimental protocol. Each experimental session had five trials per condition (varying in the number of dots to track) for a total of 15 trials per session. All subjects were tested for five sessions for a total of 75 trials. No feedback was provided, but subjects could check their overall performance at the end of each session.

Stimuli for the secondary visual task were luminance-modulated gratings of 0.5 s duration with a spatial frequency of 3 c/deg covering the entire screen. On each trial (4 s duration, during the dot tracking) subjects were presented with a sequence of three gratings, ramped in and out within a raised cosine envelope (over 20 ms), with an inter stimulus interval randomly chosen between 0.5 and 1.3 s. Two out of three gratings had the same contrast (50%) while the target grating (that subjects had to detect), randomly first, second or third in the sequence, had more or less contrast. The size of the contrast difference (Δ) was chosen from trial to trial by means of an adaptive staircase QUEST (Watson and Pelli, 1983) that homed in on threshold (67% of correct responses). The auditory secondary stimulus was of a sequence of three tones with the same presentation duration and temporal spacing as the visual version, two reference stimuli of 880 Hz with the target frequency differing from trial to trial by ±Δ Hz. In the dual-task condition subjects performed both the contrast or frequency discrimination task, and the MOT task. To avoid possible biases for response order we counterbalanced subjects responding first to the MOT task with those that responded for to the secondary task.

RESULTS
To evaluate the costs of dividing attention between sensory modalities, we measured subject performance for visual tracking alone, or with either an auditory or a visual secondary task. **Figure 1** shows the individual results for the three experimental conditions, plotting performance (d′) in the dual-task conditions against single-task performance. Each small symbol indicates individual subject performance in a given condition defined by the number of dots to track whilst large circles indicate the data averaged across subjects and conditions. It is quite clear that the concurrent visual task greatly reduced performance, shown by the average decrease in d′ from 2.48 to 1.50, and also by the fact that all individual data
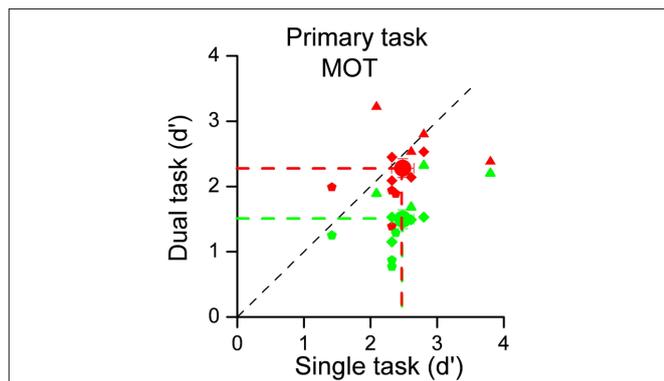
FIGURE 1 | Sensitivities for the MOT task performed alone (on the abscissa) plotted against sensitivities for dual-task conditions (on the ordinate). The 12 data points represent 4 subjects in 3 experimental conditions, defined by the number of dots to track (from 3 to 5). Green symbols refer to the intra-modal condition (secondary task contrast discrimination), red to the cross-modal condition (auditory secondary task). Small symbols refer to individual data (different symbols shape indicates different number of dots to track: three dots → triangles, four dots → diamond, and five dots → pentagon) whilst large symbols to averages. There is a clear effect for intra-modal interference, but not for cross-modal interference.
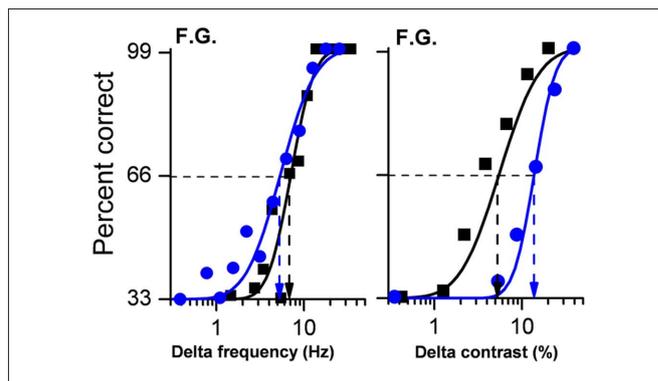

FIGURE 2 | Psychometric functions for auditory frequency discrimination (left panel) and visual contrast discrimination (right panel) for subjects F.G. Performance in the auditory task was almost identical when frequency discrimination was performed alone (black data points and lines) or together with a visual MOT task (blue data points and lines) as shown by the almost overlapping curves. However, when the two concurrent tasks were of the same sensory modality (vision), subject performance was dramatically reduced by around a factor of 3.
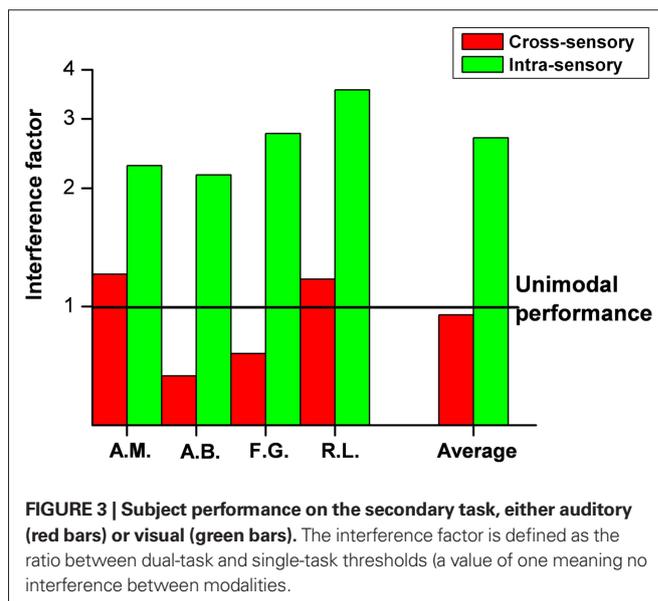

FIGURE 3 | Subject performance on the secondary task, either auditory (red bars) or visual (green bars). The interference factor is defined as the ratio between dual-task and single-task thresholds (a value of one meaning no interference between modalities).

lie below the equality line. The difference was highly significant (one-tailed paired t-test: $t_{11} = 6.98$, $p < 0.001$). However, when the competing task was auditory rather than visual, there was no effect on tracking performance. Average d′ was virtually unchanged (2.48 vs 2.28), certainly not significant (one-tailed paired t-test; $t_{11} = 1.07$, $p = 0.30$).

We also measured sensitivity for both the visual and auditory secondary tasks when performed alone and matched these results with those achieved in the dual-task condition. Examples of psychometric functions for subject F.G, are shown in **Figure 2**.

Auditory frequency discrimination is shown on the left, visual contrast discrimination on the right. It is obvious that the auditory discrimination was little affected by the concurrent visual tracking task. The two psychometric functions (best fitting cumulative Gaussian functions) are virtually identical, yielding thresholds (Δ frequency yielding 66% correct target identification) close to 6–7 Hz in both conditions. However, visual contrast discrimination thresholds were much higher in the dual than in the single-task condition, 5.1 compared with 14.5 (a factor of nearly three).

Figure 3 plots for all subjects the interference factor (ratio of dual- to single-task thresholds) for the within and between modality conditions. It is clear that the auditory task is relatively immune to interference (average factor −0.05), while performance for luminance contrast discrimination thresholds increased by a factor of more than 2.5.

DISCUSSION

In this paper we asked whether vision and audition share cognitive attentional resources in performing sustained tasks, particularly relevant for everyday functioning. As most previous research has been restricted to tasks spanning only a few milliseconds (Larsen et al., 2003; Alais et al., 2006b), or conditions with fast streams of simple auditory or visual patterns (Duncan et al., 1997), our study provides new knowledge about attentional

mechanisms in ecological situations, where prolonged monitoring of information is necessary. The results clearly indicate that under these conditions, vision and audition have access to separate cognitive resources. Performance on a sustained task, typical of everyday requirement, was completely unaffected by a concurrent auditory discrimination task. The lack of interference did not reflect a bias in deploying attention on the visual primary task more than on the auditory task, as both tasks were performed as well as when they were presented alone. On the other hand, sharing attention between two tasks of the same sensory modality produces a robust decrease of performance for both primary and secondary tasks.

That vision and audition have access to separate cognitive resources is consistent with imaging studies showing that attention can modulate responses in primary and secondary visual

and auditory cortexes (Gandhi et al., 1999; Jancke et al., 1999a,b; Somers et al., 1999). If in both modalities attentional effects modulate neural responses at these early of sensory information processing, when the visual and auditory signals are relatively independent, it is reasonable that few interactions are seen between these two senses.

Our results are important not only for the psychophysical data on the role of sustained attention between modalities, but also because they establish guidelines in designing audio–visual instrumentation. Information should be divided as much as possible between modalities, to maximize on the attentional resources available. This becomes increasingly more important as more virtual-reality applications are developed and are used routinely in everyday life.

SUPPLEMENTARY MATERIAL
The Movie S1 for this article can be found online at http://www.frontiersin.org/Perception_Science/10.3389/fpsyg.2011.00056/abstract

REFERENCES

Alais, D., Lorenceau, J., Arrighi, R., and Cass, J. (2006a). Contour interactions between pairs of Gabors engaged in binocular rivalry reveal a map of the association field. *Vision Res.* 46, 1473–1487.

Alais, D., Morrone, C., and Burr, D. (2006b). Separate attentional resources for vision and audition. *Proc. Biol. Sci.* 273, 1339–1345.

Allport, D. A., Antonis, B., and Reynolds, P. (1972). On the division of attention: a disproof of the single channel hypothesis. *Q. J. Exp. Psychol.* 24, 225–235.

Arnell, K. M., and Duncan, J. (2002). Separate and shared sources of dual-task cost in stimulus identification and response selection. *Cogn. Psychol.* 44, 105–147.

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

Bonnel, A. M., and Prinzmetal, W. (1998). Dividing attention between the color and the shape of objects. *Percept. Psychophys.* 60, 113–124.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat. Vis.* 10, 433–436.

Carrasco, M., Ling, S., and Read, S. (2004). Attention alters appearance. *Nat. Neurosci.* 7, 308–313.

Carrasco, M., and McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proc. Natl. Acad. Sci. U.S.A.* 98, 5363–5367.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Duncan, J., Martens, S., and Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature* 387, 808–810.

Gandhi, S. P., Heeger, D. J., and Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 96, 3314–3319.

Hein, G., Alink, A., Kleinschmidt, A., and Muller, N. G. (2007). Competing neural responses for auditory and visual decisions. *PLoS ONE* 2, e320. doi: 10.1371/journal.pone.0000320

Huang, L., Holcombe, A. O., and Pashler, H. (2004). Repetition priming in visual search: episodic retrieval, not feature priming. *Mem. Cognit.* 32, 12–20.

James, W. (1890/1950). *The Principles of Psychology*, Vol. 1. New York: Dover.

Jancke, L., Mirzazade, S., and Shah, N. J. (1999a). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci. Lett.* 266, 125–128.

Jancke, L., Mirzazade, S., and Shah, N. J. (1999b). Attention modulates the blood oxygen level dependent response in the primary visual cortex measured with functional magnetic resonance imaging. *Naturwissenschaften* 86, 79–81.

Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M., and Campanella, S. (2004). When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. *Neurosci. Lett.* 369, 132–137.

Jolicoeur, P. (1999). Restricted attentional capacity between sensory modalities. *Psychon. Bull. Rev.* 6, 87–92.

Larsen, A., McIlhagga, W., Baert, J., and Bundesen, C. (2003). Seeing or hearing? Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention. *Percept. Psychophys.* 65, 568–574.

Lee, D. K., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nat. Neurosci.* 2, 375–381.

Liu, T., Abrams, J., and Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychol. Sci.* 20, 354–362.

Liu, T., Pestilli, F., and Carrasco, M. (2005). Transient attention enhances perceptual performance and FMRI response in human visual cortex. *Neuron* 45, 469–477.

Morrone, M. C., Denti, V., and Spinelli, D. (2002). Color and luminance contrasts attract independent attention. *Curr. Biol.* 12, 1134–1137.

Navon, D., Gopher, D., Chillag, N., and Spitz, G. (1984). On separability of and interference between tracking dimensions in dual-axis tracking. *J. Mot. Behav.* 16, 364–391.

Pashler, H. (1992). Attentional limitations in doing two tasks at the same time. *Curr. Dir. Psychol. Sci.* 1, 44–48.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychol. Bull.* 116, 220–244.

Pashler, H., and O'Brien, S. (1993). Dual-task interference and the cerebral hemispheres. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 315–330.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.

Posner, M. I., and Gilbert, C. D. (1999). Attention and primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2585–2587.

Posner, M. I., Snyder, C. R., and Davidson, B. J. (1980). Attention and the detection of signals. *J. Exp. Psychol.* 109, 160–174.

Pylyshyn, Z. W., and Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197.

Santangelo, V., Fagioli, S., and Macaluso, E. (2010). The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space. *Neuroimage* 49, 2717–2727.

Shiffrin, R. M., and Grantham, D. W. (1974). Can attention be allocated to sensory modalities? *Percept. Psychophys.* 15, 460–474.

Somers, D. C., Dale, A. M., Seiffert, A. E., and Tootell, R. B. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1663–1668.

Spence, C., Ranson, J., and Driver, J. (2000). Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. *Percept. Psychophys.* 62, 410–424.

Taylor, M. M., Lindsay, P. H., and Forbes, S. M. (1967). Quantification of shared capacity processing in auditory and visual discrimination. *Acta Psychol. (Amst)* 27, 223–229.

Triesman, A. M., and Davies, A. (1973). *Divided Attention to Ear and Eye.* New York, NY: Academic Press.

Tulving, E., and Lindsay, P. H. (1967). Identification of simultaneously presented simple visual and auditory stimuli. *Acta Psychol. (Amst)* 27, 101–109.

Watson, A. B., and Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Percept. Psychophys.* 33, 113–120.