Exp Brain Res (2008) 184:533–546
DOI 10.1007/s00221-007-1120-z

RESEARCH ARTICLE

Semantic congruency and the Colavita visual dominance effect

Camille Koppen · Agnès Alsius · Charles Spence

Received: 27 April 2007 / Accepted: 23 August 2007 / Published online: 19 September 2007

Abstract Participants presented with auditory, visual, or bimodal audiovisual stimuli in a speeded discrimination task, fail to respond to the auditory component of bimodal targets significantly more often than to the visual component, a phenomenon known as the Colavita visual dominance effect. Given that spatial and temporal factors have recently been shown to modulate the Colavita effect, the aim of the present study, was to investigate whether semantic congruency also modulates the effect. In the three experiments reported here, participants were presented with a version of the Colavita task in which the stimulus congruency between the auditory and visual components of the bimodal targets was manipulated. That is, the auditory and visual stimuli could refer to the same or different object (in Experiments 1 and 2) or audiovisual speech event (Experiment 3). Surprisingly, semantic/stimulus congruency had no effect on the magnitude of the Colavita effect in any of the experiments, although it exerted a significant effect on certain other aspects of participants' performance. This finding contrasts with the results of other recent studies showing that semantic/stimulus congruency can affect certain multisensory interactions.

Keywords Colavita effect · Semantic congruence · Visual dominance · Audiovisual · Speech

C. Koppen (✉) · C. Spence
Crossmodal Research Laboratory,
Department of Experimental Psychology,
University of Oxford, Oxford, OX1 3UD, UK
e-mail: camille.koppen@psy.ox.ac.uk

A. Alsius
Cognitive Neuroscience Group, Parc Científic de Barcelona,
Departament de Psicologia Bàsica, Universitat de Barcelona,
Passeig de la Vall d'Hebron, 171, 08035 Barcelona, Spain

Introduction

As we interact with the rich and varied multisensory environments that surround us, our brains constantly integrate the information impinging on the different sensory receptors in order to generate the unified multisensory perceptual experiences that fill our daily lives (Driver and Spence 2000). The phenomenon of visual dominance provides a particularly fascinating example of multisensory integration (see Bertelson and de Gelder 2004; Partan and Marler 1999, for reviews). Vision appears to be the dominant sense when people make spatial judgments, as demonstrated, for example, by the fact that visual stimuli frequently bias our judgments of the apparent location of auditory stimuli (e.g. Hay et al. 1965; Heron et al. 2004; Howard and Templeton 1966), and also modify our perception of auditory stimuli, as demonstrated by illusions such as the McGurk effect (McGurk and MacDonald 1976). Under the appropriate conditions, however, such as in the temporal domain, audition can also influence people's perception of visual events (e.g. Morein-Zamir et al. 2003; Recanzone 2003; Shams et al. 2000; Shimojo and Shams 2001; Vroomen and Keetels 2006; Watanabe and Shimojo 1998; Welch et al. 1986).

One of the most striking instances of visual dominance has been provided by research on the Colavita effect. In a typical study, participants are presented with unimodal auditory, unimodal visual, or bimodal audiovisual stimuli in a speeded response task setting, in which they have to identify the modality of presentation (auditory, visual, or bimodal) of the target presented on each trial. Participants typically fail to respond to the auditory component of the bimodal targets significantly more often than they fail to respond to the visual component, often responding to the bimodal stimulus as if only a unimodal visual stimulus had been presented (Colavita 1974; Colavita et al. 1976;

Colavita and Weisberg 1979; Egeth and Sager 1977; Johnson and Shapiro 1989; Koppen and Spence 2007a, b, c; Quinlan 2000; Sinnett et al., 2007).

The Colavita effect has been shown to be modulated by both the spatial and temporal coincidence between the auditory and visual stimuli; specifically, a larger Colavita effect has been reported when the auditory and visual stimuli are presented closer together in time than when they are presented asynchronously (Koppen and Spence 2006; C. Koppen and C. Spence, submitted data). A larger Colavita effect has also been reported when the auditory and visual stimuli are presented from the same spatial location rather than from different locations (see Koppen and Spence 2007b). Koppen and Spence explained these results in terms of the ‘unity effect’ (Vatakis and Spence, 2007a; Welch and Warren 1980), according to which, the stronger an observer’s belief that two sensory events refer to the same unimodal object rather than to separate events, the greater the intersensory bias, or intersensory interaction, between them (Spence 2007). Indeed, an extensive body of research has highlighted the fact that structural factors such as the spatial and temporal coincidence of auditory and visual events can modulate audiovisual integration (see Calvert et al. 2004; Slutsky and Recanzone 2001; Spence 2007; Stein and Meredith 1993; Wallace et al. 2004; Welch 1999; Welch and Warren 1986; Zampini et al. 2005). It would therefore seem plausible to hypothesize that the Colavita visual dominance effect would also be modulated by factors other than spatial and temporal coincidence that modulate the strength of the assumption of unity that people have concerning specific pairs of auditory and visual stimuli. One such factor that has been considered in this context recently is semantic congruency (e.g. Heron et al. 2004; Vatakis and Spence, 2007a).

Multisensory cues that originate from a single object (or event) will typically share not only their temporal and spatial attributes, but may also share certain semantic features, learnt from prior experience (Laurienti et al. 2004). Thus, the semantic congruency between the cues originating from different sensory modalities may also help to facilitate the cross-modal binding of sensory information. However, the majority of the multisensory studies that have been published to date, including those on the Colavita effect, have tended to use simple stimuli with little (if any) semantic content (e.g. bursts of noise or flashes of light have been used most frequently). Furthermore, researchers have tended to combine such auditory and visual stimuli in an arbitrary manner. Consequently, the contribution of semantic congruency to the audiovisual integration observed in the Colavita effect has not received any attention by researchers thus far.

The few recent studies that have examined the role of semantic congruency on multisensory integration have shown that the behavioural aspects of audiovisual object recognition can be affected by the semantic congruence between the component unisensory stimuli under certain conditions. In particular, participants tend to respond more rapidly and accurately to semantically congruent auditory and visual stimuli than to semantically incongruent stimulus pairings (e.g. Laurienti et al. 2004; Molholm et al. 2004). It is important to note, however, that while certain studies have shown that the speed and accuracy of object recognition can be affected by audiovisual semantic congruence, other studies have failed to demonstrate any such effects on human behaviour (see Taylor et al. 2006).

To date, only one study has attempted to extend the Colavita effect to the processing of complex and meaningful stimuli. In particular, Sinnett et al. (2007, Experiment 1) conducted a Colavita study using complex sounds (such as a cat meowing) and line drawings (such as a light bulb) of familiar objects as stimuli. In Sinnett et al.’s study, the participants had to search for predefined auditory, visual, or bimodal targets (i.e. the sound of a cat meowing, a picture of a stoplight, or both stimuli presented simultaneously) amongst continuous streams of irrelevant distractors. However, while Sinnett et al. demonstrated a significant Colavita effect when using meaningful stimuli, they did not investigate the influence of semantic congruency on the magnitude of the Colavita effect. That is, they made no attempt to address the question of whether the Colavita effect would be larger for semantically congruent audiovisual pairings than for semantically incongruent stimulus pairings. Instead, the stimuli that were presented on the bimodal trials were always semantically incongruent (e.g. a picture of traffic lights paired with the sound of a bird).

Hence, while the Colavita effect has been demonstrated using complex stimuli, the effect of semantic congruency on the magnitude of the visual dominance effect has yet to be explored. Therefore, the aim of the present study was to investigate whether the Colavita effect would be modulated by the semantic congruency between the auditory and visual stimuli, using complex and semantically meaningful stimuli. In Experiment 1, the participants were presented with auditory and visual stimuli from two different classes of animal (cats and dogs), as well as bimodal stimuli, which could either be composed of congruent or incongruent pairings of audiovisual stimuli. If the Colavita effect reflects a robust empirical phenomenon, one should expect the Colavita effect to emerge even though complex, meaningful stimuli were now being used (cf. Sinnett et al., 2007). If the Colavita effect is modulated by the semantic congruency between auditory and visual stimuli, then a larger Colavita effect would be expected to occur for bimodal targets consisting of semantically congruent than for semantically incongruent pairs of stimuli. Furthermore, the use of semantically rich stimuli should in fact strengthen the assumption of unity between them (as the stimuli would be more overlapping features in common with each other). Finally, one might also expect the Colavita effect to be larger, using complex stimuli than in similar paradigms that have used simple stimuli (cf. Koppen and Spence 2007a, b, c; Koppen and Spence, submitted), given that the presentation of more complex stimuli should presumably increase the perceptual load of the participants' task (e.g. Lavie 2005), which would be expected to increase the overall error rate.

Experiment 1

Methods

Participants

About 12 naïve participants (mean age of 21 years, age range from 18 to 28 years; 6 males and 6 females) took part in Experiment 1. All except one of the participants were right-handed by self-report, with normal or corrected-to-normal sight and normal hearing. The experimental session lasted for ~25 min. The participants were given a £5 gift voucher in return for taking part in the study.

Apparatus and materials

The participants sat ~60 cm from the light and sound sources in a dimly illuminated testing booth. A 17 in. FD Trinitron CRT monitor (60 Hz refresh rate) positioned directly in front of the participant at eye-level was used to present the visual stimuli. The visual stimuli consisted of full-colour photographs ($15 \times 25\ \text{cm}^2$) of either a cat or dog, subtending ~8° of visual angle on a white background for 350 ms. The auditory stimuli (the sound of a cat meowing or of a dog barking) were presented for 350 ms from two loudspeakers; one positioned 24 cm to either side of the centre of the monitor, such that the auditory and visual stimuli appeared to emanate from the same position. The sounds were presented at 65 dB(A), as measured from the participant's ear position. Amplitude enveloping was applied to the first and last 5 ms of the auditory stimulus, using the Adobe Audition 1.5 audio editing software. There was one exemplar of each type of sound or image. There were four possible unimodal stimuli which could be presented together in any combination to give rise to four possible combinations of bimodal stimuli, consisting of both semantically congruent and semantically incongruent stimulus pairings (see Table 1).

Responses were collected from a computer keyboard placed on the table directly in front of the participant. The participants were instructed to press one key in response to auditory stimuli and another key in response to visual stimuli, with the allocation of the stimuli to the response keys (the 'n' and 'm' keys) counterbalanced across participants. The participants were instructed to press both response keys whenever a bimodal target was presented (i.e. when an auditory and a visual stimulus were presented at the same time). No specific instructions were provided to the participants as to whether they should press the two response keys simultaneously or not. The experiment was controlled using the E-Prime software (Schneider et al. 2002a, b).

Design

In order to ensure that the participants were able to discriminate the identity (cat or dog) of the auditory and visual target stimuli, they were presented with two control blocks (one auditory and the other visual) consisting of 12 trials each, in which they had to specify whether the target stimuli represented cats or dogs. None of the participants made any errors in these control blocks.

Next, the participants were presented with 6 blocks of 100 trials, each consisting of 40 visual trials, 40 auditory trials and 20 bimodal trials (the stimulus probabilities matched those used in the majority of previous studies, e.g. Colavita 1974; Egeth and Sager 1977; Koppen and Spence 2007a, b, c; Sinnett et al., 2007). See Table 1 for the relative frequencies of presentation of the four types of unimodal stimuli and the four types of bimodal stimuli. The order of stimulus

Table 1 Table showing the semantic category (cat vs. dog), semantic congruency (congruent vs. incongruent) and the number of auditory and visual stimuli presented in each block in Experiment 1

Target type	Semantic category		Semantic congruency	Number of trials per block
	Auditory stimulus	Visual stimulus		
Unimodal	Cat	–	–	20
	Dog	–	–	20
	–	Cat	–	20
	–	Dog	–	20
Bimodal	Cat	Cat	Congruent	5
	Cat	Dog	Incongruent	5
	Dog	Cat	Incongruent	5
	Dog	Dog	Congruent	5

presentation was randomized within each block of trials. A block of 30 practice trials that were identical to the main experimental trials (but which were not analysed) was presented before the main experimental session.

Procedure

On each trial, the participants were presented with an auditory, visual, or bimodal target. The target was presented at the start of each trial and was followed by a white screen which was presented for between 1,450 and 1,700 ms (ranging randomly with a rectangular distribution). On unimodal auditory trials the auditory stimulus was presented together with a white screen. On unimodal visual trials the visual stimulus was presented on its own (i.e. silently). The auditory and visual responses were collected from the onset of the presentation of the target stimuli. On bimodal trials in which participants correctly pressed both the auditory and visual response keys, both the auditory and visual RTs were collected independently. The next trial began immediately after the end of the preceding trial. The participants were instructed to respond to the targets as rapidly and accurately as possible. Note that response speed was emphasized over response accuracy. No feedback regarding the correctness of a participant's responses was provided.

Results

The participants failed to make any response on 0.6% of the trials overall, and these trials were not included in the data analyses. The Colavita effect is defined as occurring when participants make significantly more *visual-only responses* (i.e. when the participant only pressed the visual response key) than *auditory-only responses* on the bimodal trials. The magnitude of the Colavita effect therefore reflects the difference between the percentage of visual-only and auditory-only responses on the bimodal trials. The results of Experiment 1 are highlighted in Table 2, and in Figs. 1 and 2.

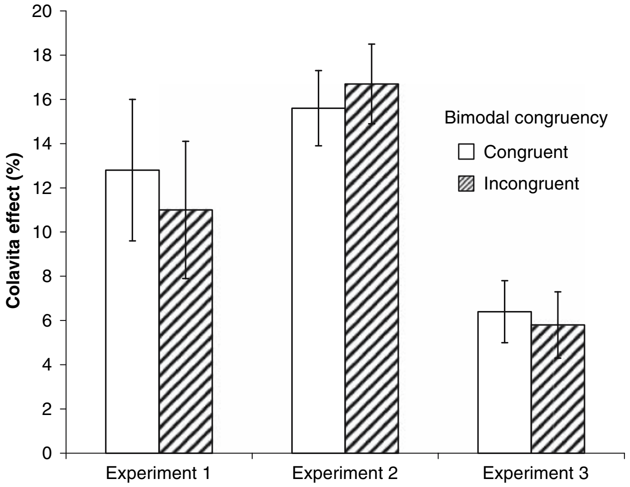

Fig. 1 Mean magnitude of the Colavita effect (percentage of visual-only responses minus percentage of auditory-only responses) for the congruent and incongruent conditions in Experiments 1–3. The *error bars* indicate the standard error of the mean

Table 2 Mean error rates for the unimodal auditory, unimodal visual and bimodal target stimuli (congruent and incongruent) in Experiments 1–3

	Experiment 1			Experiment 2			Experiment 3		
	Unimodal	Bimodal		Unimodal	Bimodal		Unimodal	Bimodal	
		Congruent	Incongruent		Congruent	Incongruent		Congruent	Incongruent
Error rates (%)									
Unimodal auditory	8.6 (2.1)	–	–	7.3 (1.0)	–	–	3.5 (0.5)	–	–
Unimodal visual	11.4 (2.3)	–	–	14.1 (2.8)	–	–	3.8 (1.0)	–	–
Bimodal									
Auditory-only responses	–	7.1 (1.7)	5.4 (2.3)	–	9.8 (1.3)	7.5 (1.4)	–	4.3 (0.9)	4.8 (1.1)
Visual-only responses	–	19.9 (4.5)	16.4 (3.9)	–	25.4 (2.7)	24.2 (2.9)	–	10.7 (2.3)	10.6 (2.6)
RTs (ms)									
Unimodal auditory	577 (30)	–	–	525 (15)	–	–	559 (16)	–	–
Unimodal visual	522 (30)	–	–	477 (14)	–	–	584 (16)	–	–
Bimodal					567 (16)	579 (17)		624 (16)	631 (13)
Auditory responses	–	607 (24)	574 (26)	–	–	–	–	–	–
Visual responses	–	602 (26)	579 (29)	–	–	–	–	–	–

There were two types of error that participants could make on bimodal trials: they could either make an auditory-only or a visual-only response. Mean reaction times (RTs; ms) for correct responses to unimodal auditory, unimodal visual, and congruent and incongruent bimodal target stimuli. Bimodal auditory and visual responses refer to the RTs to the auditory and visual components of the bimodal target stimuli (on trials where participants correctly made both responses). Standard error are shown in parentheses

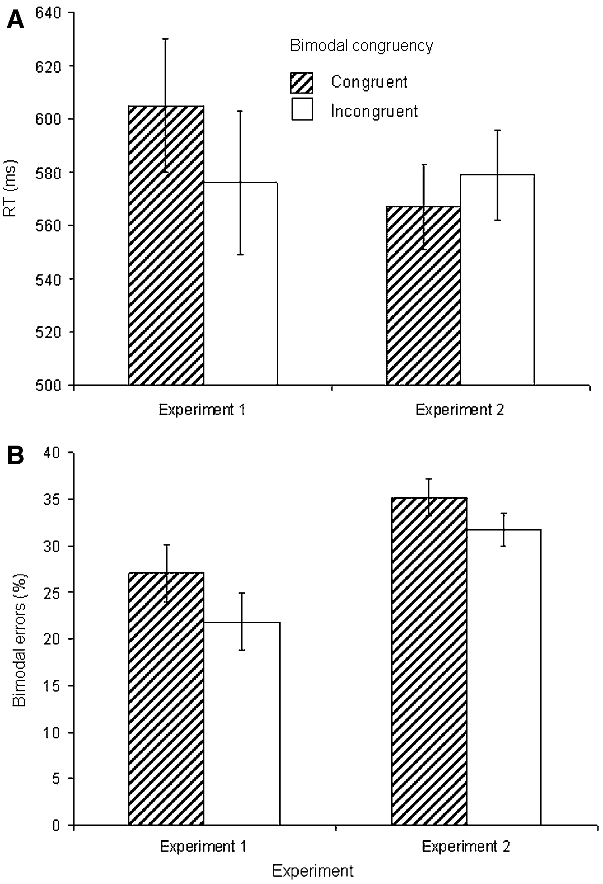

Fig. 2 Summary of the **a** mean RT (ms), and **b** errors (%) values for bimodal congruent and incongruent targets in Experiments 1 and 2. The *error bars* indicate the standard error of the mean

Error data

A preliminary analysis of the data revealed no significant main effect of the identity of the targets (i.e. cats vs. dogs), and so the data were combined across this factor to simplify the analysis. The data from the bimodal trials in which the participants failed to respond to one of the two stimuli were analysed using an analysis of variance (ANOVA) with the factors of Response (Auditory-only or Visual-only) and Semantic Congruency (Congruent or Incongruent). In this and all subsequent analyses, Greenhouse-Geisser corrections were made whenever the assumption of sphericity was violated. The analysis of the error data revealed a significant main effect of Response [$F(1, 11) = 10.34$, $P = 0.008$], as predicted. In particular, the participants made significantly more visual-only than auditory-only responses (18.2% vs. 6.2% of all bimodal trials, respectively); thus demonstrating a robust Colavita visual dominance effect (Koppen and Spence 2007a). The analysis of the data also revealed a significant main effect of Semantic Congruency [$F(1, 11) = 6.91$, $P = 0.023$], attributable to participants making significantly more errors when they were presented with semantically congruent bimodal stimuli (13.5% errors) than when they were presented with semantically incongruent stimuli (10.9% errors). Crucially, however, there was no significant interaction between Response and Semantic Congruency [$F < 1$, n.s.], thus suggesting that the magnitude of the Colavita effect itself was not modulated by the semantic congruency between the auditory and visual components of the bimodal stimulus.

Next, the congruent and incongruent bimodal target data were combined in order to compare the unimodal with the bimodal error data. An ANOVA revealed a significant main effect of Stimulus (Auditory, Bimodal, or Visual) [$F(1.06, 11.62) = 9.13$, $P = 0.010$], attributable to participants making significantly more errors on the bimodal trials (24.4% errors) than on either the unimodal auditory (8.6% errors; $t(11) = 3.15$, $P = 0.009$) or unimodal visual trials (11.4% errors; $t(11) = 2.87$, $P = 0.015$), and more errors on unimodal visual trials than on the unimodal auditory trials ($t(11) = 2.71$, $P = 0.020$).

RT data

The RT data from those trials in which the participants responded correctly were analysed in an ANOVA with the factors of Target Modality (Auditory or Visual) and Target Type (Unimodal, Congruent Bimodal, or Incongruent Bimodal). The analysis revealed a significant main effect of Target Modality [$F(1, 11) = 5.64$, $P = 0.037$], with the participants responding more rapidly to visual (567 ms) than to auditory targets (586 ms) overall. The analysis of the RT data also revealed a significant main effect of Target Type [$F(2, 22) = 36.97$, $P < 0.001$], with participants responding more rapidly to unimodal (549 ms) than to either congruent or incongruent bimodal targets (604 and 576 ms, respectively; $t(11) = 7.34$, $P < 0.001$; $t(11) = 4.38$, $P = 0.001$), and more rapidly to incongruent bimodal than to congruent bimodal targets ($t(11) = 5.27$, $P < 0.001$). There was also a significant interaction between Target Modality and Target Type [$F(1.27, 13.98) = 7.79$, $P = 0.011$], attributable to there being a significant effect of Target Modality for the unimodal targets but not for either the congruent or the incongruent bimodal targets (mean difference between auditory and visual response latencies = 55, 5 and 5 ms, respectively; $t(11) = 3.08$, $P = 0.010$; $t(11) = 0.64$, $P = 0.530$; $t(11) = 0.63$, $P = 0.542$, respectively). Note that there was a speed accuracy trade-off in participants' responses to the unimodal auditory and unimodal visual targets.

Discussion

The results of Experiment 1 revealed a robust Colavita visual dominance effect; that is, when participants failed to respond correctly on the bimodal target trials (which they did on 24.4% of all bimodal trials), they made significantly more visual-only than auditory-only responses (18.2 vs. 6.2% of all bimodal trials, respectively). Hence, the results of Experiment 1 clearly show that the Colavita effect can be extended to the processing of complex stimuli, in addition to the simple lights and sounds that have been used in the majority of previous research (though see Sinnett et al., 2007, for a recent exception). Note that the magnitude of the Colavita visual dominance effect (the percentage of visual-only responses minus the percentage of auditory-only responses) in this experiment using congruent (12.8%) and incongruent (11.0%) complex stimuli (colour photographs and animal vocalizations) was larger than that reported in many other recent studies of the Colavita effect that have used simple light and sound stimuli (e.g. Koppen and Spence 2007a, where a mean Colavita effect of 4.2% was reported). This difference may reflect the increased perceptual load attributable to the use of complex auditory and visual stimuli (i.e. meaningful sounds and pictures of actual objects, rather than simple lights and flashes) in the present study. Presumably, the additional information contained in the presentation of the complex, meaningful stimuli (e.g. including details such as the colour, average frequency, average volume, how the stimulus changes over time, semantic category, etc.) would add to the processing load of the participants, which would, in turn, perhaps result in more errors being made (see Lavie 2005; Sinnett et al., 2007).

Contrary to our predictions, however, the magnitude of the Colavita visual dominance effect was not affected by the semantic congruency between the auditory and visual components of the bimodal target stimuli. The comparison yielded a *P* value[1] of 0.552, with a correspondingly low-observed power[2] of 0.129, which suggests that one would be justified in accepting the null hypothesis (see Frick 1995, for the criteria that should be met before one accepts the null hypothesis) that the Colavita visual dominance effect is simply not modulated by the semantic congruency between the auditory and visual stimuli on bimodal target trials.

It is important to note, however, that stimulus congruency did influence performance. In particular, stimulus congruency influenced the difficulty of processing of the bimodal stimuli; with significantly slower RTs and higher error rates being reported for congruent bimodal targets (604 ms; 13.5% errors) than for incongruent bimodal targets (576 ms; 10.9% errors). The slower RTs and higher error rates observed for congruent bimodal targets suggest that participants found it harder to separate the auditory and visual components of the congruent stimuli (i.e. it took them longer to realize that a bimodal stimulus had been presented) than to separate the components of the incongruent stimuli. This would have resulted in participants taking longer to respond to both components of the bimodal target and making more errors (i.e. responding to only one component of the bimodal stimulus) on the bimodal trials.

One factor that may have contributed to participants responding more rapidly and accurately on the incongruent bimodal target trials (as opposed to on the congruent target trials) may have been that the mismatch between the stimuli for the incongruent bimodal targets could have provided participants with an extra cue to inform them that a bimodal target had, in fact, been presented. Another explanation for participants' difficulty in responding to congruent (versus incongruent) bimodal targets could be a 'failure of binding' (Baylis et al. 2002). That is, when participants bound the auditory and visual stimuli together, the perception of the visual stimulus may have hindered their processing of the auditory stimulus because it adequately described (i.e. it provided sufficient information concerning) the unitary audiovisual event. If participants had experienced such a failure to represent all aspects of the stimuli appropriately, it would have been more difficult for them to realise that two stimuli had been presented (which would in turn be expected to lead to an increase in their RTs and error rates). In sum, the semantic congruency between the auditory and visual components of the bimodal targets influenced participants' performance on the Colavita task (in terms of their RTs and error rates on bimodal trials). It did not, however, affect the magnitude of the Colavita effect that was observed.

It has been argued elsewhere that animals have both more features per se, and also more features in common, than non-living objects (e.g. McRae et al. 1997; Tyler and Moss 2001). This makes it more difficult to distinguish two animals from each other, than to distinguish an animal from a non-living object. It is therefore possible, if rather unlikely, that one reason why no effect of semantic congruency on the Colavita effect was observed in Experiment 1 may have been because of the large overlap in the number of object features for the particular stimuli used. In order to address this possibility, a control study was run (using an experimental procedure and design that was identical to that used in Experiment 1) in which the participants ($N = 14$) were now presented with stimuli in different

---

[1] Frick (1995) has argued that one cannot confidently accept the null hypothesis when the *P* value is in the range from 0.200 to 0.500, but if the *P* value is greater than 0.500, he argues that this provides one important criterion for accepting the null hypothesis.

[2] Observed power is the probability of correctly rejecting a false statistical null hypothesis (Type II error; the probability of a Type II error is referred to as β) and is equal to $1 - \beta$. Thus, while a low *P*-value and a high observed power would provide support for the H1 hypothesis, a *P*-value above 0.500 and a low observed power would be evidence supporting the null hypothesis (Frick 1995).

semantic categories (animals and non-living objects; e.g. cats and phones). However, once again no significant main effect of semantic congruency on the magnitude of the Colavita effect was found (mean Colavita effect of 7.2 and 5.9% in the semantically congruent and semantically incongruent conditions, respectively).[3] This result shows that the null effect of semantic congruency on the Colavita visual dominance effect reported in Experiment 1 cannot simply be attributable to the specific stimuli that were presented.

Another reason why no effect of semantic congruency on the Colavita effect was observed in either Experiment 1, or in the control experiment (see Footnote 3), may have been because there were only four different stimulus pairings on bimodal trials (auditory-cat auditory-dog, auditory-cat visual-dog, etc.) and only one exemplar of each stimulus. It may therefore have been that these particular stimulus pairings may simply have become over-learned by the participants (even for the incongruent pairings), and consequently, over-represented in long-term episodic memory. In order to rule out this possible explanation of the null effect of semantic congruency on the magnitude of the Colavita effect reported in Experiment 1, the size of the stimulus set in the next experiment was increased from 4 to 80 stimuli. This modification to the design made it possible for us to present participants with many more different semantically congruent and semantically incongruent stimulus pairings than had been the case in Experiment 1. In addition, a greater number of participants were tested, thus increasing the statistical power of the experimental design, and hence making it more likely that an effect of semantic congruency would be found, should one exist. Finally, in order to make the experiment more similar to the one reported by Sinnett et al. (2007), an extra response key was introduced (the three-key response condition). Thus, there were now three separate response keys, one for each type of stimulus (an auditory response key, a visual response key and a bimodal response key). Adding the third response key made it possible to evaluate the possibility that the errors participants made on bimodal trials were due to a difficulty in the processing of the stimuli, rather than due to any difficulties in making two keypresses (cf. Koppen and Spence 2007a).

Experiment 2

Methods

Participants

About 30 naïve participants (mean age of 21 years, age range from 19 to 27 years; 11 males and 19 females) took part in Experiment 2. All except three of the participants were right-handed by self-report, with normal or corrected-to-normal sight and normal hearing. The experimental session lasted for ~30 min.

Apparatus, materials and procedure

These were exactly the same as in Experiment 1 with the exception of the particular stimuli used and the response requirements of the task. The auditory stimuli consisted of animal sounds (8 bit; mono-channel; 11,500 Hz digitization), some of which were obtained from an online library (http://www.cofc.edu/~marcellm; downloaded 9 May 2005; for normative data concerning these stimuli, see Marcell et al. 2000) and the rest from http://www.a1freesound-effects.com/animal.html. The visual stimuli were comprised of 40 line-drawing pictures of different animals which were chosen from the Snodgrass and Vanderwart picture database (see Snodgrass and Vanderwart 1980, for standardization statistics) and edited using Microsoft Paint Version 5.1. On the congruent bimodal trials, the auditory and visual stimuli depicted the same animal, while on the incongruent bimodal trials they depicted different animals. There were now three response keys (the three-key response condition), one for auditory, one for visual and one for bimodal targets. The allocation of the stimuli to the response keys (the ‘b’, ‘n’ and ‘m’ keys) was counterbalanced across participants.

Design

The design was exactly the same as in Experiment 1 with the exception that the participants were presented with 4 blocks of 200 trials, each consisting of 80 visual trials, 80 auditory trials and 40 bimodal trials. The participants were presented with an equal number (20) of bimodal congruent and bimodal incongruent trials per block.

---

[3] Fourteen naïve participants took part in this control study. The apparatus, materials, design, and procedure were exactly the same as in Experiment 1 with the sole exception that the experimental stimuli now consisted of pictures and sounds of cats and phones (rather than of cats and dogs as used in Experiment 1). The data from the bimodal trials in which the participants failed to respond to one of the stimuli were analysed using an ANOVA with the factors of Response (Auditory-only or Visual-only) and Target Congruency (Congruent or Incongruent). The analysis of the error data revealed a significant main effect of Response [$F(1, 13) = 12.00$, $P = 0.004$], attributable to participants making significantly more visual-only than auditory-only responses (21.3 vs. 8.2% of all bimodal trials, respectively); once again demonstrating a Colavita effect. The analysis of the error data also revealed a significant main effect of Target Congruency [$F(1, 13) = 10.64$, $P = 0.006$], attributable to participants making significantly more errors when the auditory and visual components of the bimodal stimuli were congruent (16.4% errors) than when they were incongruent (13.1% errors). Crucially, however, there was no interaction between Response and Target Congruency [$F(1, 13) = 1.47$, $P = 0.247$]. Hence, once again, no effect of semantic congruence on the magnitude of the Colavita effect was observed.

Results

Error data

The results of Experiment 2 are highlighted in Table 2. The data from the bimodal trials in which the participants failed to respond to one of the stimuli were analysed using an ANOVA with the factors of Response (Auditory-only or Visual-only) and Semantic Congruency (Congruent or Incongruent). The analysis of the error data revealed a significant main effect of Response [$F(1, 29) = 52.07$, $P < 0.001$], attributable to participants making significantly more visual-only than auditory-only responses (24.8% vs. 8.6% of all bimodal trials, respectively); thus, a large and highly significant Colavita visual dominance effect was once again observed. The analysis of the data also revealed a significant main effect of Semantic Congruency [$F(1, 29) = 6.61$, $P = 0.016$], attributable to participants making significantly more errors on the congruent bimodal trials (17.6% errors) than on the incongruent bimodal trials (15.8% errors). Crucially, however, there was no significant interaction between Response and Semantic Congruency [$F < 1$, n.s.].

An ANOVA performed on the unimodal and bimodal error data revealed a significant main effect of Stimulus (Auditory, Bimodal, or Visual) [$F(1.67, 48.43) = 45.53$, $P < 0.001$], attributable to participants making significantly more errors on the bimodal trials (33.4% errors) than on either unimodal auditory (7.3% errors; $t(29) = 8.50$, $P < 0.001$) or unimodal visual trials (14.1% errors; $t(29) = 6.05$, $P < 0.001$), and more errors on unimodal visual than on the unimodal auditory trials ($t(29) = 3.12$, $P = 0.003$).

RT data

The RT data from the congruent and incongruent bimodal target trials were combined in order to compare the unimodal RT data with the bimodal RT data. An ANOVA performed on this data with the factor of Stimulus (Auditory, Bimodal, or Visual) revealed a significant main effect [$F(2, 58) = 61.74$, $P < 0.001$]. This term was attributable to participants responding significantly more rapidly to unimodal visual stimuli (477 ms) than to the unimodal auditory stimuli (525 ms; $t(29) = 5.76$, $P < 0.001$) or to the bimodal stimuli (573 ms; $t(29) = 11.47$, $P < 0.001$), and significantly more rapidly to the unimodal auditory stimuli than to the bimodal stimuli ($t(29) = 5.23$, $P < 0.001$).

Next, the congruent and incongruent bimodal target data were analysed in an ANOVA with the factor of Semantic Congruency (Congruent or Incongruent). This analysis revealed a significant main effect of Semantic Congruency [$F(1, 29) = 9.78$, $P = 0.004$], with the participants responding significantly more rapidly to congruent bimodal stimuli (567 ms) than to incongruent bimodal stimuli (579 ms). Thus, there was a speed-accuracy trade-off in the data in terms of participants' responses to bimodal congruent and incongruent stimuli.

Discussion

Once again, the results of Experiment 2 revealed a robust Colavita visual dominance effect, with participants making significantly more visual-only than auditory-only responses (24.8 vs. 8.6% of all bimodal trials). In addition, no effect of semantic congruence was observed on the magnitude of the Colavita effect. The comparison yielded a $P = 0.664$, with a correspondingly low-observed power of 0.124, which once again supports the null hypothesis that the Colavita visual dominance effect is simply not modulated by the semantic congruency between the auditory and visual stimuli on the bimodal target trials. This null effect was observed despite the fact that a greater variety of stimuli were now presented to the participants, so that there was less chance of the stimuli becoming over-learned and over-represented in the participants' long-term episodic memory. Note that the magnitude of the Colavita effect, and the overall error rate, was larger in Experiment 2 (Colavita effect = 16.2%, bimodal error rate = 33.4%) than in Experiment 1 (Colavita effect = 12.0%, bimodal error rate = 24.4%). This difference may have been due to the greater variability in the stimuli that were presented to participants, which would have been expected to increase the perceptual load of their task (Lavie 2005), thus increasing the processing demands of the task somewhat. Indeed, Sinnett et al. (2007) recently reported a similar increase of error rates as a result of increasing the size of the stimulus set that they presented participants with.

Once again, semantic congruency had a significant effect on participants' performance (faster RTs but more errors in the congruent than in the incongruent conditions), showing that the manipulation of semantic congruency was effective in modulating certain aspects of participants' performance (see Fig. 2). However, one question to arise from the comparison of the results of our manipulation of semantic congruency in Experiments 1 and 2 concerns why semantic congruency resulted in a decrease of the RTs in Experiment 2, while apparently having the opposite effect in Experiment 1.

One explanation for this difference is in terms of the speed-accuracy trade-off that was present in Experiment 2 but not in Experiment 1 (for bimodal congruent and incongruent stimuli). The speed-accuracy trade-off observed in Experiment 2 meant that participants responded more rapidly to congruent stimuli at the expense of response accuracy, thus resulting in faster RTs but higher error rates for congruent bimodal targets (than for incongruent bimodal targets). This speed-accuracy trade-off could have been caused by participants experiencing time pressure, possibly caused by the higher perceptual load (due to the greater variety of stimuli) in Experiment 2.

A second explanation for the reversal of RTs for congruent and incongruent trials between Experiments 1 and 2 may be related to the different responses requirements used in the two experiments (i.e. the participants responded with two responses keys in Experiment 1, but with three response-keys in Experiment 2) that makes any simple comparison of the results of the two experiments somewhat difficult. In Experiment 1, the participants responded to bimodal targets using the auditory and visual response keys and thus did not have to suppress their responses to the individual components of the bimodal stimuli (i.e. the participants could respond to the bimodal target in the same way as if they were responding separately to the auditory and visual components of the stimulus). In contrast, in Experiment 2, participants responded to the bimodal targets using a dedicated bimodal response key, which meant that they had to suppress their tendency to respond to the auditory and visual stimuli individually (i.e. they had to refrain from pressing either the auditory or visual response keys).

Therefore, in Experiment 2, when an incongruent bimodal stimulus was presented, the participants may have found it easier to distinguish the two stimuli (this may be reflected in the higher accuracy of their responses to the incongruent stimuli), but may have taken longer to respond due to having to suppress their responses to the individual auditory and visual components of the stimulus. Whereas, when a congruent bimodal stimulus was presented, the participants may have found it harder to distinguish the two stimuli (as reflected in the higher error rate when they responded to congruent stimuli), and hence may have tended to perceive the stimulus array as consisting of a the presentation of a single stimulus (despite the fact that it actually consisted of stimuli in two different sensory modalities). The participants may therefore have responded more rapidly because they did not have to suppress any tendency to respond to the auditory and visual stimuli individually.

It should be noted here that the analysis of the RT data in both Experiments 1 and 2 revealed that participants responded significantly more rapidly to visual than to auditory stimuli overall, and more rapidly to unimodal visual than to unimodal auditory stimuli. It could therefore be argued that the occurrence of the Colavita effect may have been due to participants responding more rapidly to the visual than to the auditory component of the bimodal target stimuli. However, this explanation can be ruled out given that other researchers have demonstrated significant Colavita effects even when participants respond significantly more rapidly to the auditory stimuli than to the visual stimuli (e.g. Colavita 1974; Colavita et al. 1976; Colavita and Weisberg 1979; Koppen and Spence 2007a; Sinnett et al., 2007). Furthermore, responses to the auditory and visual components of the bimodal targets tend to be coupled (see Koppen and Spence 2007a on these points), further arguing against this possibility. Finally, as we shall see in Experiment 3, the Colavita effect still occurs even under conditions where participants respond more rapidly to unimodal auditory targets than to unimodal visual targets. Hence, the Colavita effect cannot simply occur due to participants responding more rapidly to the visual component of bimodal target stimuli.

The finding that the participants in Experiment 2 responded significantly more rapidly and accurately to the unimodal visual and unimodal auditory targets than to the bimodal targets, suggests that the relatively high-error rates and response latencies seen for the bimodal targets in Experiment 1 were not simply caused by the specific response requirements of the task (i.e. having to make two responses rather than just one). Instead, these results suggest that the poorer performance towards bimodal targets may have been an outcome of participants' difficulty in perceptually processing the bimodal stimuli.

In the two experiments reported thus far, the auditory and visual stimuli only represented token attributes of the particular stimulus categories (e.g. cats) concerned. For example, the actual cat sounds that were presented did not directly correspond with the particular image of the cat shown on the screen. A further difference between the auditory and visual stimuli in Experiments 1 and 2 was that the auditory signals were dynamic (which, by definition, auditory signals must be) whilst the visual images consisted of statically presented images. It could therefore be argued that one reason why no modulation of the Colavita effect by the semantic congruency of the stimuli was observed in Experiments 1 and 2 was because there was no dynamic relationship between the auditory and visual stimuli. One important question that arises here regards whether using dynamic audiovisual stimuli belonging to the same underlying perceptual event (so that there would be a direct correspondence, rather than a token correspondence, between the auditory and visual stimuli) would result in a significant modulation of the Colavita effect as a function of the congruency of the stimuli.

The time-varying correlation between auditory and visual stimuli has been shown to be a critical factor contributing to multisensory integration (Calvert et al. 2004; Rosenblum et al. 1996); the time-varying relationship between the auditory and visual events contributes to the structure, or the physical relationship, between the auditory and visual stimuli. Therefore, a dynamic stimulus pairing would provide an even more rigorous test of the impact of stimulus congruency on the Colavita effect, because the likelihood of the auditory and visual stimuli being 'bound' into a singular object or event would be much greater (presumably because there would be more automaticity to the binding). For these reasons, the effects of stimulus congruency on the Colavita visual dominance effect was explored in Experiment 3 using time-varying dynamic stimuli (i.e. audiovisual speech). One of the reasons why speech stimuli were used (rather than non-speech stimuli) was because it has recently been shown that participants are more sensitive to the match versus mismatch between dynamic auditory and visual stimuli for speech stimuli than they are for non-speech stimuli (see Vatakis and Spence, 2007a, b). Therefore, any effects of stimulus congruency would be more likely to emerge with speech stimuli.

Experiment 3

Methods

Participants

About 15 participants (mean age of 23 years, age range from 19 to 36 years; 11 females) took part in Experiment 3. All of the participants were right-handed by self-report, with normal or corrected-to-normal sight and normal hearing. The experimental session lasted for ~30 min.

Apparatus and materials

These were exactly the same as in Experiment 2 with the exception of the particular stimuli used. The auditory and visual stimuli were created from video clips (DV/PAL, 720 × 576 pixels) which consisted of digital recordings of an English speaker (only the lower half of the face was shown) pronouncing the syllables 'mo' and 'da' (i.e. there were two video clips). The sound and video clips were edited using Adobe Premiere 6.0 software. Each clip was 360-ms long (nine frames, at 40 ms/frame). The visual stimuli were presented at a rate of 25 frames/s. For the visual stimuli, the speech event started at the point of articulation of the first consonant (which lasted for 200 ms) and ended after the production of the vowel (which lasted for 160 ms). Note that, as both /m/ and /d/ are voiced consonants, the acoustic information concerning the pronunciation of the syllable was already available at the beginning of the stimuli. The auditory stimuli were digitized at 48,000 Hz, and presented at 65 dB(A) from the participants' ear position.

There was one example of each type of sound or video clip. There were therefore four possible unimodal stimuli which could be presented together in any combination to give rise to four possible combinations of bimodal stimuli, consisting of both congruent and incongruent stimulus pairings (as in Experiment 1). The unimodal auditory stimuli (/mo/, /da/) were presented together with a blank black screen, whilst the unimodal visual ([mo], [da]) stimuli consisted of faces which articulated silently. Incongruent bimodal stimuli were created by dubbing the acoustic syllable /mo/ onto the visual syllable [da], or the syllable /da/ onto the visual syllable [mo]. On the incongruent bimodal trials, the participants clearly perceived a conflict between the acoustic and visual components of the stimulus (i.e. the features of these two syllables did not lead to experience a McGurk illusion), whereas on the congruent bimodal trials, the visual and the acoustic components matched. The participants were presented with the three-key response condition (as in Experiment 2).

Design and procedure

The design was exactly the same as in Experiment 2 with the sole exception that the participants were presented with three blocks of trials (rather than four blocks). The procedure was exactly the same as in Experiment 1 with the exception that the target was presented followed by a silent black screen which was presented for 800 ms.

Results

The participants failed to make any response on 9.3% of the trials overall, and these trials were not included in the data analyses. The results of Experiment 3 are highlighted in Table 2.

Error data

The data from the bimodal trials in which the participants failed to respond to one of the stimuli were analysed using an ANOVA with the factors of Response (Auditory-only or Visual-only) and Congruency (Congruent or Incongruent). There was a significant main effect of Response [$F(1, 14) = 9.24$, $P = 0.009$], indicating that participants made significantly more visual-only than auditory-only responses (10.6% vs. 4.6% of all bimodal trials, respectively), demonstrating a significant Colavita visual dominance effect. The main effect of Congruency failed to reach significance, and crucially, there was no significant interaction between Response and Congruency, for both terms [$F < 1$, n.s.].

An ANOVA performed on the unimodal and bimodal error data revealed a significant main effect of Stimulus (Auditory, Bimodal, or Visual) [$F(1.08, 15.13) = 19.10$, $P < 0.001$], attributable to participants making significantly more errors on the bimodal trials (15.2% errors) than on either unimodal auditory (3.5% errors; $t(29) = 4.35$, $P = 0.001$) or unimodal visual trials (3.8% errors; $t(29) = 4.53$, $P < 0.001$), but no more errors on unimodal visual than on the unimodal auditory trials ($t(29) = 0.51$, $P = 0.620$).

RT data

The RT data from the congruent and incongruent bimodal target trials was combined in order to compare the unimodal RT data with the bimodal RT data. An ANOVA performed on this data with the factor of Stimulus (Auditory, Bimodal, or Visual) revealed a significant main effect [$F(2, 28) = 19.40$, $P < 0.001$]. This term was attributable to participants responding significantly more slowly to bimodal stimuli (627 ms) than to unimodal visual stimuli (584 ms; $t(14) = 4.01$, $P = 0.001$) or to the unimodal auditory stimuli (559 ms; $t(14) = 5.95$, $P < 0.001$), and more rapidly to the unimodal auditory than to the unimodal visual stimuli ($t(14) = 2.24$, $P = 0.042$).

Next, the congruent and incongruent bimodal target data were analysed in an ANOVA with the factor of Congruency (Congruent or Incongruent). However, this analysis did not reveal a significant main effect of Congruency [$F < 1$, n.s.].

Discussion

The results of Experiment 3 highlighted a robust Colavita visual dominance effect, with participants making significantly more visual-only than auditory-only responses (10.6% vs. 4.6% of all bimodal trials). Once again, no effect of semantic congruency was observed on the magnitude of the Colavita effect. The comparison yielded a $P = 0.741$, with a correspondingly low-observed power of 0.061, once again supporting the null hypothesis that the Colavita visual dominance effect is not modulated by the congruency between the auditory and visual stimuli on bimodal target trials. This finding was observed despite the fact that dynamic audiovisual stimuli that belonged to the same underlying perceptual event (i.e. so that there was a time-varying correspondence between the auditory and visual stimuli) were used (cf. Vatakis and Spence, 2007a).

In contrast to the results of Experiments 1 and 2, the congruency of the stimuli had no significant main effect on participants' performance (in terms of their response latencies and error rates). This does not mean, however, that the manipulation of stimulus congruency was ineffective; indeed, other researchers have found effects of stimulus congruency which did not emerge in the behavioural data (e.g. see Taylor et al. 2006; where the effects of stimulus congruency emerged in the fMRI data instead). One reason why behavioural effects of congruency may not have emerged in Experiment 3 is because the effects of stimulus congruency on the integration of audiovisual speech signals is weaker when the auditory information is clear (i.e. when it is not degraded; Callan et al. 2001, 2003; Sekiyama et al. 2003). The auditory speech stimuli presented in Experiment 3 were clear, which could have meant that the effects of stimulus congruency on participants' behavioural performance may have been too weak to have resulted in a significant effect on performance.

The analysis of the RT data in Experiment 3 revealed that participants responded significantly more rapidly to the unimodal auditory than to the unimodal visual targets. The fact that the Colavita effect still emerged despite this pattern of response latencies, strengthens our argument (outlined earlier) that the Colavita effect cannot simply be attributable to participants responding to the visual stimuli more rapidly than to auditory stimuli.

As dynamic audiovisual stimuli were presented in Experiment 3, which would have been expected to increase the perceptual load of the participants' task (relative to the more static stimuli that were presented in Experiments 1 and 2), one would predict that the magnitude of the Colavita effect, and the overall error rate would be higher in Experiment 3 than in Experiments 1 and 2. However, the magnitude of the Colavita effect, and the overall error rate, were in fact smaller in Experiment 3 (Colavita effect = 6.2%, bimodal error rate = 15.2%) than in Experiments 1 and 2 (Colavita effects = 12.0 and 16.2%, respectively; bimodal error rates = 24.4 and 33.4%, respectively). This pattern of error rates can be explained in terms of the speed-accuracy trade-off between the experiments: that is, participants responded more accurately and less rapidly to stimuli in Experiment 3 than they did in Experiments 1 and 2. One reason for this trade-off might have been because different experimenters ran the experiments (C. Koppen conducted Experiments 1 and 2, whereas A. Alsius conducted Experiment 3), which may have resulted in a difference in the emphasis on participants' error rates and speed of responding (i.e. it is possible that speed of response was emphasised more strongly in Experiments 1 and 2 than in Experiment 3).

General discussion

The three experiments reported in the present study represent one of the very few occasions in which the Colavita effect has been investigated using stimuli that are more complex than the simple auditory brief beeps and visual flashes that have been used in so much of the previous research on multisensory information processing (and, in particular, in previous research on the Colavita effect). Furthermore, Experiment 3 represents the first time that speech stimuli have been used to investigate the Colavita effect. The primary aim of the experiments reported in this study was to investigate whether manipulating the stimulus congruency between the auditory and visual stimuli would influence the magnitude of the Colavita visual dominance effect. A null effect of stimulus congruency on the magnitude of the Colavita effect was found in all three of the experiments. The fact that this null result was replicated in three separate experiments fulfils the good effort criterion for accepting the null hypothesis (see Frick 1995).

It should, however, be noted that the participants were not simply insensitive to the manipulations of semantic congruency used in the present study. Stimulus congruency had a significant effect on certain aspects of participants' behavioural performance in Experiments 1 and 2; in particular, the congruency of the bimodal targets significantly affected the speed and accuracy of participants' responses to the bimodal targets. This pattern of results suggests that participants found it harder to separate the auditory and visual components of a bimodal target when they were congruent than when they were incongruent. There was, however, no effect of stimulus congruency on the magnitude of the Colavita visual dominance effect itself.

It is interesting to contrast the null results of stimulus congruency on the Colavita effect reported here with the results of previous studies where significant effects of semantic congruency on behavioural performance have been observed (Laurienti et al. 2004; Molholm et al. 2004). In the studies of both Laurienti et al. and Molholm et al., participants had to identify pre-specified targets (e.g. an auditory or visual red or blue target in Laurienti et al.'s study, or a sound or a static image of a cow in Molholm et al.'s study) which could be presented in either a congruent audiovisual pairing (e.g. the same target would be presented in both sensory modalities) or an incongruent audiovisual pairing (both a target and non-target would be presented). In both studies, the participants responded significantly faster and more accurately to targets in congruent pairings than to targets in incongruent pairings. The authors argued that the redundant target effect (Miller 1982; the phenomenon whereby participants respond more rapidly to single versus multiple targets, which has typically been explained by neural co-activation) may have contributed to the improved performance observed for congruent target pairings because participants had two targets pertaining to a single target response. In contrast, the participants in the present study had to make one response to each stimulus, thus any redundant target effects that may have been the basis of the stimulus congruency effects observed in the aforementioned studies would not have affected the Colavita effects reported in this study.

One final contrast between Laurienti et al.'s (2004) and Molholm et al.'s (2004) studies and the experiments reported in this study is that they used an identification task (in which a redundant target could help to speed up the identification process), whereas participants in the experiments reported in this study were presented with a modality detection task[4] (in which the identification of the stimuli may not necessarily have contributed to the judgment of which modality the target appeared in).

In order to investigate whether the RT benefits for semantically congruent stimuli observed in the aforementioned studies could also be observed in a Colavita-type paradigm, a follow-up study was conducted in which participants had to respond to the semantic attributes of the stimuli (cf. Laurienti et al. 2004; Molholm et al. 2004; Sinnett et al., 2007). The new participants ($N = 10$) were presented with the same stimuli as in Experiment 1, but now they had to respond to the semantic category of the target (cat or dog) rather than to its modality of presentation. Thus, for the congruent bimodal trials, the participants only had to press one key (e.g. the 'cat' response key). Whenever an incongruent bimodal stimulus (containing both a cat and a dog target) was presented, the participants were explicitly instructed to press both response keys. Despite the fact that participants now had to respond to the semantic category of the targets, the Colavita effect was observed once again (mean Colavita effect of 10.1%). Furthermore, the magnitude of the Colavita visual dominance effect reported in Experiment 2 (10.1%) was just as large as that

---

[4] It should be noted that the task that participants had to perform could also be considered to be an identification task (where participants discriminate which modality a target is presented in). Quite what is the most appropriate description for the task is ambiguous, as participants were both detecting stimuli in different modalities, and discriminating which modality a target was presented in.

observed in Experiment 1 (10.1%; $t(20) = 0.34$, $P = 0.740$).[5]

Importantly, the participants in our follow-up study responded more rapidly to congruent bimodal targets (where only a single manual response was required; 512 ms) than to either of the unimodal auditory (570 ms) or unimodal visual targets (539 ms). They also responded more rapidly to congruent than to incongruent bimodal targets (539 ms vs. 712 ms). This result might be explained (in the same way as the results of Laurienti et al. 2004 and Molholm et al. 2004) in terms of a redundancy gain effect (cf. Miller 1982). Thus, it is possible to demonstrate an effect of semantic congruency, in terms of RT benefits to congruent targets, whilst using a Colavita task design in which participants have to identify the semantic category (but not the sensory modality) of the target. Stimulus congruency does not, however, appear to modulate the Colavita effect itself when participants have to respond to the sensory modalities of the targets.

While no evidence of an effect of semantic congruency on the Colavita visual dominance effect was found, spatial and temporal coincidence, factors that may contribute to the 'unity effect' (and therefore to the multisensory binding of auditory and visual stimuli; Spence 2007), have recently been shown to modulate the Colavita effect (Koppen and Spence 2006, 2007b; C. Koppen and C. Spence, submitted data). Although both semantic congruency and spatiotemporal factors contribute to multisensory binding, the important distinction between them is that the latter can be considered as structural (i.e. bottom-up) factors that determine the actual physical relationship between auditory and visual stimuli (see Radeau and Bertelson 1977, for a discussion of structural/bottom-up and cognitive/top-down factors influencing multisensory perception; Spence 2007; Welch 1999). In contrast, the semantic congruency of audiovisual stimuli is learnt through experience, and is thought to occur later in object processing than, for example, the calculation of an object's spatial position (see Laurienti et al. 2004). Therefore, it could be argued that the Colavita visual dominance effect is affected by factors that contribute to the structural binding of audiovisual stimuli, but not by those factors, such as semantic congruency, that presumably emerge later in stimulus processing.

Acknowledgements C. K. was supported by a Departmental Studentship from the Department of Experimental Psychology, University of Oxford. Correspondence regarding this article should be addressed to C. K., at the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD. E-mail: camille.koppen@psy.ox.ac.uk.

References

Baylis GC, Simon SL, Baylis LL, Rorden C (2002) Visual extinction with double simultaneous stimulation: what is simultaneous? Neuropsychologia 40:1027–1034

Bertelson P, de Gelder B (2004) The psychology of multimodal perception. In: Spence C, Driver J (eds) Crossmodal space and crossmodal attention. Oxford University Press, Oxford, pp 141–177

Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001) Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Cogn Brain Res 10:349–353

Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E (2003) Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport 14:2213–2218

Calvert GA, Spence C, Stein BE (eds) (2004) The handbook of multisensory processes. MIT, Cambridge, MA

Colavita FB (1974) Human sensory dominance. Percept Psychophys 16:409–412

Colavita FB, Tomko R, Weisberg D (1976) Visual prepotency and eye orientation. Bull Psychon Soc 8:25–26

Colavita FB, Weisberg D (1979) A further investigation of visual dominance. Percept Psychophys 25:345–347

Driver J, Spence C (2000) Multisensory perception: beyond modularity and convergence. Curr Biol 10:R731–R735

Egeth HE, Sager LC (1977) On the locus of visual dominance. Percept Psychophys 22:77–86

Frick RW (1995) Accepting the null hypothesis. Mem Cognit 23:132–138

Hay JC, Pick HLJ, Ikeda K (1965) Visual capture produced by prism spectacles. Psychon Sci 2:215–216

Heron J, Whitaker D, McGraw PV (2004) Sensory uncertainty governs the extent of audio-visual interaction. Vision Res 44:2875–2884

Howard IP, Templeton WB (1966) Human spatial orientation. Wiley, New York

---

[5] Ten naïve participants took part in this control study. The apparatus, materials, design, and procedure were exactly the same as in Experiment 1 with the sole exception that the participants were now instructed to press one key in response to cats and another key in response to dogs. Thus, for the congruent bimodal trials, the participants only had to press one key (a single-response trial). Whenever an incongruent bimodal stimulus (containing both a cat and a dog target) was presented, the participants were explicitly instructed to press both response keys (a two-response trial). The data from the incongruent bimodal trials in which the participants failed to respond to one of the two stimuli were analysed using an ANOVA with the factors of Response (Auditory-only or Visual-only) and Auditory Stimulus (Cat or Dog). The analysis revealed a significant main effect of Response [$F(1, 9) = 8.27$, $P = 0.010$], with participants making significantly more visual-only than auditory-only responses (19.2 vs. 9.1% of all bimodal trials, respectively), thus demonstrating a robust Colavita visual dominance effect. None of the other terms in this analysis of the error data reached significance [both Fs < 1, n.s.]. The RT data from those trials in which only a single response was required were analysed using an ANOVA with the factor of Stimulus Type (Auditory, Visual, or Bimodal) and revealed a significant main effect [$F(2, 18) = 17.95$, $P < 0.001$]. Participants responded more rapidly to bimodal targets (512 ms) than to either auditory (570 ms; $t(9) = 5.59$, $P < 0.001$) or visual targets (539 ms; $t(9) = 3.84$, $P = 0.004$), and more rapidly to visual than to auditory targets ($t(9) = 2.79$, $P = 0.021$). Finally, an ANOVA performed on the bimodal congruent and incongruent RT data with the factors of Semantic Congruency (Congruent or Incongruent) revealed a significant main effect [$F(1, 9) = 38.92$, $P < 0.001$], with participants responding significantly more rapidly to congruent (539 ms) than to incongruent targets (712 ms).

Johnson TL, Shapiro KL (1989) Attention to auditory and peripheral visual stimuli: effects of arousal and predictability. Acta Psychol 72:233–245

Koppen C, Spence C (2006) Prior entry and the Colavita effect. Poster presented at the 7th meeting of the IMRF. Trinity College, Dublin, 18–21 June [Abstract no. 111]

Koppen C, Spence C (2007a) Seeing the light: exploring the Colavita visual dominance effect. Exp Brain Res 180:737–754

Koppen C, Spence C (2007b) Spatial coincidence modulates the Colavita visual dominance effect. Neurosci Lett 417:107–111

Koppen C, Spence C (2007c) Assessing the role of stimulus probability on the Colavita visual dominance effect. Neurosci Lett 418:266–271

Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004) Semantic congruence is a critical factor in multisensory behavioral performance. Exp Brain Res 158:405–414

Lavie N (2005) Distracted and confused? Selective attention under load. Trends Cogn Sci 9:75–82

Marcell ME, Borella D, Greene M, Kerr E, Rogers S (2000) Confrontation naming of environmental sounds. J Clin Exp Neuropsychol 22:830–864

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748

McRae K, de Sa VR, Seidenberg MS (1997) On the nature and scope of featural representations of word meaning. J Exp Psychol Gen 126:99–130

Miller JO (1982) Divided attention: evidence for coactivation with redundant signals. Cogn Psychol 14:247–279

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452–465

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn Brain Res 14:121–134

Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. Cogn Brain Res 17:154–163

Partan S, Marler P (1999) Communication goes multimodal. Science 283:1272–1273

Quinlan P (2000) The 'late' locus of visual dominance. Abstr Psychon Soc 5:64

Radeau M, Bertelson P (1977) Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. Percept Psychophys 22:137–146

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078–1093

Rosenblum LD, Wuestefeld AP, Anderson KL (1996) Auditory reachability: an affordance approach to the perception of sound source distance. Ecol Psychol 8:1–24

Schneider W, Eschman A, Zuccolotto A (2002a) E-prime user's guide. Psychology Software Tools, Pittsburgh

Schneider W, Eschman A, Zuccolotto A (2002b) E-prime reference guide. Psychology Software Tools, Pittsburgh

Sekiyama K, Kanno I, Miura S, Sugita Y (2003) Auditory-visual speech perception examined by fMRI and PET. Neurosci Res 47:277–287

Shams L, Kamitani Y, Shimojo S (2000) What you see is what you hear: sound induced visual flashing. Nature 408:788

Shimojo S, Shams L (2001) Sensory modalities are not separate modalities: plasticity and interactions. Curr Opin Neurobiol 11:505–509

Sinnett S, Spence C, Soto-Faraco S (2007) Visual dominance and attention: the Colavita effect revisited. Percept Psychophys

Slutsky DA, Recanzone GH (2001) Temporal and spatial dependency of the ventriloquism effect. Neuroreport 12:7–10

Snodgrass JG, Vanderwart M (1980) A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn Mem 6:174–215

Spence C (2007) Audiovisual multisensory integration. Acoust Sci Technol 28:61–70

Stein BE, Meredith MA (1993) The merging of the senses. MIT, Cambridge, MA

Taylor KI, Moss HE, Stamatakis EA, Tyler LK (2006) Binding crossmodal object features in perirhinal cortex. Proc Natl Acad Sci USA 103:8239–8244

Tyler LK, Moss HE (2001) Towards a distributed account of conceptual knowledge. Trends Cogn Sci 5:244–252

Vatakis A, Spence C (2007a) Crossmodal binding: evaluating the 'unity assumption' using audiovisual speech. Percept Psychophys 69:744–756

Vatakis A, Spence C (2007b) Evaluating the influence of the 'unity assumption' for the temporal perception of realistic audiovisual stimuli. Acta Psychol doi:10.1016/j.actpsy.2006.12.002

Vroomen J, Keetels M (2006) The spatial constraint in intersensory pairing: no role in temporal ventriloquism. J Exp Psychol Hum Percept Perform 32:1063–1071

Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004) Unifying multisensory signals across time and space. Exp Brain Res 158:252–258

Watanabe K, Shimojo S (1998) Attentional modulation in perception of visual motion events. Perception 27:1041–1054

Welch RB (1999) Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In: Aschersleben G, Bachmann T, Müsseler J (eds) Cognitive contributions to the perception of spatial and temporal events. Elsevier Science, B.V., Amsterdam, pp 371–387

Welch RB, DuttonHurt LD, Warren DH (1986) Contributions of audition and vision to temporal rate perception. Percept Psychophys 39:294–300

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychol Bull 3:638–667

Welch RB, Warren DH (1986) Intersensory interactions. In: Boff KR, Kaufman L, Thomas JP (eds) Handbook of perception and performance, vol 1. Sensory processes and perception. Wiley, New York, pp 25-1–25-36

Zampini M, Guest S, Shore DI, Spence C (2005) Audio-visual simultaneity judgments. Percept Psychophys 67:531–544