

The Sensory-Dependent Nature of Audio-Visual Interactions for Semantic Knowledge

Guillaume Vallet (guillaume.vallet@univ-lyon2.fr)

Université Lumière Lyon 2. Laboratoire EMC, 5 avenue Pierre Mendès France,
69676, Bron cedex, France &

Laval University, School of Psychology, 2325 rue des Bibliothèques
Quebec City (Quebec), G1V 0A6 Canada

Benoit Riou (benoit.riou@univ-lyon2.fr), Rémy Versace (remy.versace@univ-lyon2.fr)

Université Lumière Lyon 2. Laboratoire EMC, 5 avenue Pierre Mendès France,
69676, Bron cedex, France

Martine Simard (martine.simard@psy.ulaval.ca)

Laval University, School of Psychology, 2325 rue des Bibliothèques
Quebec City (Quebec), G1V 0A6 Canada

Abstract

The nature of audio-visual interactions is poorly understood for meaningful objects. These interactions would be indirect through semantic memory according to the amodal nature of knowledge, whereas these interactions would be direct according to the modal nature of knowledge. This question, central for both memory and multisensory frameworks, was assessed using a cross-modal priming paradigm from auditory to visual modalities tested on familiar objects. For half of the sound primes, a visual abstract mask was simultaneously presented to the participants. The results showed a cross-modal priming effect for semantically congruent objects compared to semantically incongruent objects presented without the mask. The mask interfered in the semantically congruent condition, but had no effect in the semantically incongruent condition. The semantic specificity of the mask effect demonstrates a memory-related effect. The results suggest that audio-visual interactions are direct. The data support the modal approach of knowledge and the grounded cognition theory.

Keywords: Memory; Perception; Audio-visual; Masking; Priming; Grounded Cognition.

Introduction

Our environment is filled with meaningful objects representing semantic knowledge. These objects are perceptually processed using several sensory channels in which the auditory and visual modalities dominate the other senses in Human (for a review see Spence, 2007). The sensory information is mainly integrated on the basis of the temporal and spatial relationships between the stimuli (Calvert & Thesen, 2004), and also on the basis of the semantic relationships existing between them (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004). Yet it remains uncertain how semantic memory aspects are involved in multisensory perception (for a review see Doehrmann & Naumer, 2008). This issue depends on the perceptual or semantic nature of cross-modal interactions, and thus questions the modal or amodal nature of knowledge (Vallet, Brunel, & Versace, 2010). The present study therefore aims at assessing the nature of audio-visual interactions using an innovative masking procedure.

Communication between different modalities is called interaction (or interplay). If this interaction involves a representation of higher level, this interaction is called integration (Driver & Noesselt, 2008). An integrated object is a representation that is more than the sum of its part. Previous research in the multisensory perception theoretical framework principally studied the neural basis of the integration mechanism using meaningless stimuli (for review see Calvert & Thesen, 2004; Koelewijn, Bronkhorst, & Theeuwes, 2010). Fewer studies were conducted with meaningful stimuli, and the goal of these studies was also to determine the brain substrates of multisensory integration (Doehrmann & Naumer, 2008). The semantic constraint is generally assessed by manipulating the semantic congruency. A congruent trial is when the prime and the target refer to the same semantic object (meowing sound - cat's picture). Semantic congruent stimuli usually facilitate information processing (Chen & Spence, 2010), and may enhance memory performances in semantic (Laurienti et al., 2004) and episodic tasks (Lehmann & Murray, 2005).

In the memory theoretical framework, cross-modal interactions tested on meaningful stimuli are generally studied by inserting a delay between the stimuli. The most famous paradigm in this field is the cross-modal priming paradigm. The cross-modal priming effect is the facilitation of the processing of one stimulus in one modality (the target) by the previous presentation of another stimulus in another modality (the prime). The cross-modal priming effect may be observed between different modalities, such as the haptic and visual modalities (Easton, Srinivas, & Greene, 1997), but most of the studies were realized between the auditory and visual modalities (for a review see Schneider, Engel, & Debener, 2008). The increasing number of studies on the audio-visual interactions involving meaningful stimuli are aimed at a better understanding of these effects. Nevertheless, the nature of audio-visual interactions, which is the central issue underlying these effects, remains poorly understood.

The nature of these interactions depends on the nature

of knowledge. This question is much less studied since it was supposed that semantic knowledge is amodal, i.e., context free (e.g., Coccia, Bartolini, Luzzi, Provinciali, & Lambon Ralph, 2004). Semantic knowledge was defined as general knowledge on objects and their properties, words meaning and facts in general (Tulving, 1972). In the amodal knowledge theoretical framework of memory, the interactions between the auditory and visual modalities are supposed to be semantic. The co-activation between modalities is supposed to be indirect through an abstract semantic representation (Chen & Spence, 2010). In other words, the presentation of one component in one modality (e.g., meowing sound) should activate the abstract conceptual representation in semantic memory ("cat") through a bottom-up activation. In a second step, this activation would activate all the associated features through a top-down activation (e.g., visual representation of a cat).

The amodal nature of knowledge is challenged nowadays by the grounded cognition theory (e.g., Brunel, Labeye, Lesourd, & Versace, 2009). In this approach, knowledge is modal and the cognitive system is supposed to simulate the situation to be processed (Barsalou, 2008; Versace, Labeye, Badard, & Rose, 2009), so that processing a familiar sound shall automatically activate the associated representations in the other sensory modalities (e.g., Molholm, Martinez, Shpaner, & Foxe, 2007). Since the simulation is done in the same brain areas than perception (e.g., Slotnick & Schacter, 2006), then the co-activation between modalities should be direct and perceptual (Brunel et al., 2009; Vallet et al., 2010).

As perception remains dominant, the simulation should not occur efficiently if a rival sensory perception is presented at the same time in the simulation's modality. This hypothesis was recently tested in young adults (Vallet et al., 2010). In this study, we developed an innovative long-term cross-modal priming paradigm using familiar bimodal items (sound-picture). In a long-term priming paradigm all the primes are first presented in the study phase, whereas all the targets are presented in a second phase, called the test phase. A mask was presented with half of the primes and it shared the target's modality rather than the prime's modality. For instance, in the auditory to visual modalities direction, a visual abstract mask was presented with half of the auditory primes. A cross-modal priming effect was observed for the targets associated with unmasked primes in the study phase compared to new pictures (no sound heard). The main result was that visual targets associated to auditory masked-primes in the study phase were processed as new pictures. No significant effect was observed in the study phase for the masked primes suggesting that the mask interfered with the simulation of the representations associated to the prime. Nevertheless and coherent with amodal approach of knowledge, attention resources could have been divided between modalities. In this case, the mask might have produced a less efficient processing of the prime and thus of the target (Mulligan, 2003). In addition, the semantic congruency was

not manipulated in this particular study. Consequently, the nature and the specificity of the mask remain unexplored.

The objective of the present study is therefore to assess the nature and the specificity of the mask effect for audiovisual interactions in the processing of meaningful stimuli and then in semantic knowledge. This research topic questions the nature of semantic audio-visual interactions and is thus an attempt to clarify the issue about the amodal or modal nature of knowledge. To this aim, the paradigm used by Vallet et al. (2010) was adapted into a short-term priming paradigm. In this form, the prime is immediately followed by the target in the same trial so that semantic congruency can be manipulated. In each trial, the participants first heard a sound as prime. Half of these primes were presented with a visual abstract mask. Then, they had to categorize the picture target as an animal or as an artefact. Half of the trials were category-congruent, i.e. the sound prime and the picture target belonged to the same category. The other half of the trials were category-incongruent, i.e. the sound prime and the picture target belonged to two different categories. In addition, half of the trials in the category-congruent condition were item-congruent (e.g., meowing sound - cat's picture) and half item-incongruent (e.g., meowing sound - eagle's picture). The item-congruency manipulation permits the precise assessment of the specificity of the mask effect and the avoidance of cognitive interference resulting from the utilization of two different categories (Taylor, Moss, & Tyler, 2007).

Two hypotheses may be contrasted. First, according to the amodal framework, a sensory meaningless mask effect should be explained by attention since no direct link should exist between the modalities. In this case, the mask should modulate the processing of the target regardless of the semantic congruency. On the contrary, according to the modal hypothesis, a sensory mask should alter the processing of the target only in the semantically congruent condition. In this case, the mask should have a perceptual memory effect. A visual mask should interfere with the automatic activation of the visual representation associated to the auditory prime (semantically congruent). The authors of the present study hypothesize that the mask will have a perceptual effect.

Method

Participants

Twenty-four right-handed students (4 men; 20 women; $\bar{x} = 21.71 \pm 3.87$) recruited at Lyon 2 University (France) took part in the experiment. The participants had no history of medical or psychiatric disorder. They were all native French speakers and demonstrated adequate visual and hearing performances.

Stimuli and material

Overall 200 stimuli were used: half of them were sounds and half photographs. Half of the stimuli were familiar animals (e.g., cow, cat, dog, lion), and the other half familiar artefacts

(e.g., piano, guitar, bell, airplane). All the photographs had the same format (393 x 295 pixels, resolution of 72 x 72 dots per inch). All the sounds were edited to last 1,000 ms. Each participant himself adjusted the auditory intensity in order to reach a comfortable level. Ten visual color masks were created using Photoshop CS3 Mac. A ripple effect was applied to 10 color pictures not included in the experimental material. This procedure was meant to make the result impossible to be identified just like an abstract painting. Different masks were created to avoid a systematic association between the stimuli and a specific mask, and to avoid repetition.

Prediction in the categorization task was avoided by defining an equal number of category-congruent trial (in which primes and targets belong to a same category) and of category-incongruent trials (in which primes and targets belong to different semantic categories). This design is the most used to manipulate semantic congruency. Yet some attention effect such as inhibition may be involved when the prime and the target belong to different categories (Taylor et al., 2007). Consequently, we chose to focus on the item-congruency level to assess precisely the specificity of the mask effect. In this case, the prime and the target belong to the same general category, and could either be semantically congruent (e.g., meowing sound then cat's picture) or semantically incongruent (e.g., meowing sound then eagle's picture).

Out of these stimuli, 120 were the same items (60 photographs and 60 sounds) as in our previous study (Vallet et al., 2010). These items were selected in a pre-test experiment to be easily recognizable in each modality, and to be as prototypical and familiar as possible. The pre-test has also assessed the sound-picture association (see Vallet et al., 2010). From these items, 20 bimodal items (20 sounds - 20 pictures) were assigned to the item-congruent condition (e.g., meowing sound-cat's picture). Twenty sounds with 20 different pictures were assigned to the item-incongruent condition (e.g., barking sound-eagle's picture). These two conditions were included in the category-congruent condition in which the sound and the picture belong to the same general semantic category.

Eighty new stimuli (40 sounds, 40 pictures) were included in the category-incongruent condition (e.g., photocopier's sound – ant's picture). However, these new items were not counterbalanced with the others conditions (category-congruent), because it was impossible to find the same exact bimodal, familiar, and recognizable features as those previously chosen. These items were thus excluded from the analyses. The item-congruency level was preferred to the category-congruency level since it allows a more precise evaluation of the mask specificity.

Finally, 16 sounds and 16 pictures were included as practice trials representing all the experimental conditions. They were the same for all the participants.

In summary, the general design was congruency (item-congruent, item-incongruent and category-incongruent con-

ditions) by masking (masked, unmasked primes). All the stimuli of the category-congruent condition (item-congruent and item-incongruent, 10 stimuli per condition) were counterbalanced between subjects into the unmasked item-congruent, masked item-congruent, unmasked item-incongruent and masked item-incongruent conditions according to 4 different lists. The uncontrolled stimuli (20 per condition) were assigned into the 2 following conditions: unmasked category-incongruent and masked category-incongruent conditions.

Procedure and design

The experiment was conducted using a Macintosh MacBook Pro. Psyscope software X B53 (Cohen, MacWhinney, Flatt, & Provost, 1993) was used to set up and manage the experiment. Informed written consent was obtained from each participant. Each participant was tested individually in one session lasting approximately 12 minutes (see Figure 1 for an illustration of the protocol). The participants were informed that they were taking part in a study on reaction speed to visual stimuli. Participants were told that before the presentation of each picture, they will hear a sound which could match or not to the picture. They were also informed that sometimes a color rectangle may appear on the center of the screen as they hear the sound. The participants were instructed to ignore these stimuli (sounds and rectangles) in order to focus only on the pictures and the categorization task.

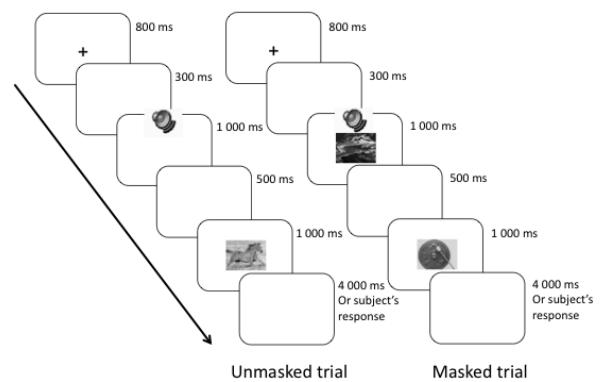


Figure 1: Illustration of the experimental protocol. A sound is presented as the prime. For half of the sound primes, an abstract visual mask is presented. Then, a photograph is categorized as an animal or as an artefact.

The experiment began with 16 practice trials which were followed by the 80-trial test phase. Each trial started with a central fixation point displayed for 800 ms. This was followed, 300 ms later, by a 1,000 ms sound presented bi-aurally through a stereo headset: half of these sounds corresponded to animals and the other half to artefacts. For half of these sounds, a visual mask was presented simultaneously during 1,000 ms. Five hundred ms later, a centrally positioned picture appeared for 1,000 ms.

Each mask was associated with four different sounds. Finally,

a white screen was displayed for 3,000 ms or until the participant responded. The participants were asked to judge, as quickly and as accurately as possible, whether the picture corresponded to an animal or to an artefact. They answered by pressing the appropriate key on the keyboard. Response logging started with the presentation of the picture.

The sounds and the pictures were presented in random order. The response keys were counterbalanced across the participants.

Results

The mean correct reaction times and mean rates of correct responses were calculated across subjects for each experimental condition. The practice trials were excluded from the analyses as were the category-incongruent items¹. Reaction times that differed by more than 2.5 standard deviations from the mean in each condition were treated as outliers (less than 2% of the data). Separate analyses of variance (ANOVA) were performed on percentages of correct responses and correct reaction times. The analyses were performed with subjects as random variable according to a 2 (item-congruency: item-congruent vs. item-incongruent) x 2 (mask: masked vs. unmasked) within-subjects variables. The data were analyzed using PASW for Macintosh (SPSS Inc.).

The analyses performed on correct responses revealed no significant effect of any factor. There might be ceiling effects since the overall correct response rate was 95.1%.

The analysis of reaction times revealed a main effect of the item-congruency, $F(1, 23) = 12.91, p < .05, \eta^2_{partial} = .35$. There was no effect of the mask ($F(1, 23) = 2.82, p = .11$), but a significant interaction between item-congruency and mask, $F(1, 23) = 5.96, p < .05, \eta^2_{partial} = .21$.

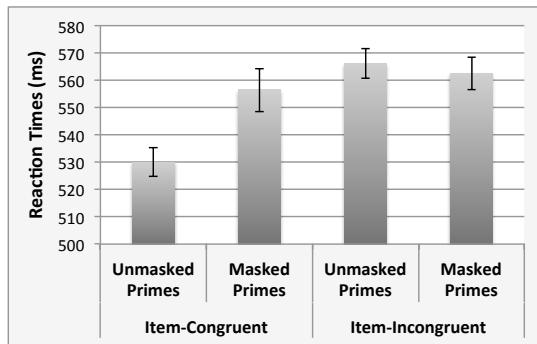


Figure 2: Means and standard errors for reaction times of the item-congruency by mask interaction.

The detailed analysis of this interaction (see Figure 2) demonstrated that the items in the semantically congruent

¹ANOVA with the category-incongruent condition revealed no effect for the correct response rates. For the reaction times, the ANOVA revealed an effect of the item-congruency, $F(1,23) = 6.18, p < .05$ and item-congruency by mask interaction, $F(1,23) = 4.79, p < .05$. No difference in the category-incongruent conditions $t(23) = .30, p = .76$.

unmasked condition were processed faster than masked items, $t(23) = 2.96, p < .05, d = .46$. In contrast, no significant difference was observed in the item-incongruent condition between the unmasked and masked items, $t(23) = .39, p = .70$. The subtraction of the reaction times of the unmasked item-congruent condition from the reaction times of the unmasked item-incongruent condition indicated a priming effect of 36 ms.

In summary, the analyses revealed no effect of any factor for the correct response rates. Regarding the reaction times, the main finding was that the unmasked semantically-congruent were processed faster than the masked semantically-congruent items.

Discussion

This study was designed to assess the nature of audio-visual interactions in semantic knowledge using a masking short-term cross-modal paradigm with familiar bimodal objects. Half of the sound primes were presented simultaneously with a visual abstract mask. The picture targets were categorized into animals or artefacts. The picture targets and the sound primes could be semantically congruent (item-congruent), or semantically incongruent (item-incongruent and category-incongruent).

The reaction times analyses showed that congruent stimuli were processed faster than incongruent stimuli, as typically expected (Laurienti et al., 2004). The results also demonstrated a cross-modal priming effect. The unmasked item-congruent stimuli were processed faster than the unmasked item-incongruent stimuli with a gain of 36 ms. This result replicates the finding of a cross-modal priming for familiar objects (e.g., Schneider et al., 2008). However, the most important finding of this study was the mask by item-congruency interaction. The results demonstrated that the mask interfered with the processing of the target only in the item-congruent condition. In the item-incongruent condition, no significant difference was observed between masked and unmasked items. The mask interference replicated our previous findings in a long-term cross-modal priming paradigm (Vallet et al., 2010).

The mask interference could be explained in an amodal approach of knowledge by an attention effect only since, according to this theory, no direct relation is supposed to exist between the sensory modalities (cf. the SPI model, Tulving, 1995). Should this hypothesis be true, an attention effects should impact both congruent and incongruent semantic items conditions, because attention would be divided into the different modalities (Mulligan, 2003) or because attention would be enhanced by a multisensory stimulation (e.g., Koelewijn et al., 2010; Sperdin, Cappe, Foxe, & Murray, 2009). In the present study, the attention hypothesis can be rejected since the mask effect is specific to the semantically congruent condition. In addition, an attention effect was also insufficient to explain the interference observed in our previous findings (Vallet et al., 2010).

ous study (Vallet et al., 2010). Indeed, in this study, there was no significant difference on correct response rates and reaction times between the prime presented with the mask and the prime presented without the mask in the study phase.

Supporting our hypothesis, the mask interference is specific to the semantically associated features. The masking procedure used here is unusual since the masking procedure is classically explained by a superposition of the same kind of sensory information on the prime (for a review see van den Bussche, van den Noortgate, & Reynvoet, 2009). Yet the mask seems to interfere with the target rather than with the prime in our paradigm. This effect is not a forward masking, i.e. a mask before the stimulus. Indeed, forward masking is limited to 300 ms (Enns & Di Lollo, 2000) whereas an inter-stimuli interval (ISI) of 500 ms was used in the present study. This interference effect thus appears to be related to memory rather than to perception. While perception is supposed to occur at a lowest level than memory recent studies have demonstrated that learned associations or expertise could play a central role in multisensory perception (Mitterer & Jesse, 2010; Petrini, Russell, & Pollick, 2009). These data suggest that memory and perception are closer than previously hypothesized. Results from different studies support this hypothesis with common activations for visual imagery and visual perception (Ishai & Sagi, 1995) and with direct influence of memory features on perceptual tasks (Riou, Lesourd, Brunel, & Versace, in press). These relationships between memory and perception are supposed to exist in the grounded cognition theory (Barsalou, 2008). The presentation of a visual mask during the perception of a sound prime would interfere with the simulation of the visual associated representation of the object in memory. This hypothesis could explain why, in our study, the mask's interference is specific to the semantically congruent condition.

Our interpretation of the mask-congruency interaction is therefore that the visual mask has interfered with the automatic and direct activation of the visual representation of the object associated with the sound prime. The visual mask might then overlap with the activation (simulation) of the visual associated representation of the sound prime. These data support a perceptual (or sensory-dependent) nature of the audio-visual interactions and thus support the grounded cognition theory.

However, the present study has some limitations. For instance, the time window chosen might be surprising. An ISI of 500 ms is unusual for a study on multisensory interaction (e.g., Chen & Spence, 2010). Yet multisensory interaction and integration could occur with an ISI of 500 ms as in the present study (Wallace et al., 2004). This ISI was chosen based on a study demonstrating that shorter ISI (100 ms) produced an additive effect compared to longer ISI (300) leading to an integration of the activations (Labeye, Oker, Badard, & Versace, 2008); and because the masking effect was observed if the mask was presented until 250 ms before the presentation

of the prime, and until 300 ms after (Enns & Di Lollo, 2000). Consequently, an ISI of 500 ms should be long enough to allow an integration of the features and long enough to avoid forward masking (i.e. a perceptual interference of a mask on a stimulus presented after the mask).

In conclusion, this study showed that cognition could be multimodal as supposed by the grounded cognition theory. Knowledge would be sensory-dependent so that the co-activation between sensory modalities should be automatic and direct. The masking effect observed in the present study seems to refer to both memory and multisensory perception. This effect is an additional argument in favour of studies combining multiple sub-domains of cognition. The modal hypothesis has important repercussions on the understanding of cognition and eventually has an impact on clinical practice. Sensory-dependent knowledge has also recently been demonstrated in healthy aging (Vallet, Simard, & Versace, in press). Consequently, memory disorders and memory rehabilitation programs in the elderly might find some new perspectives based on multisensory knowledge. Some cognitive rehabilitation programs focusing on the link between perception and memory may eventually be developed, that may improve memory functioning by enhancing multimodal presentation and mental imagery.

Acknowledgments

Guillaume Vallet and Rémy Versace are supported by a grant from the Rhône-Alpes Region through the cluster "Handicap – Aging – Neurosciences".

References

- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Brunel, L., Labeye, E., Lesourd, M., & Versace, R. (2009). The sensory nature of episodic memory: Sensory priming effects due to memory trace activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1081–1088.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology*, 98, 191–205.
- Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, 114, 389–404.
- Coccia, M., Bartolini, M., Luzzi, S., Provinciali, L., & Lambon Ralph, M. (2004). Semantic memory is an amodal, dynamic system: Evidence from the interaction of naming and object use in semantic dementia. *Cognitive Neuropsychology*, 21, 513–527.
- Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods*, 25, 257–271.
- Doehrmann, O., & Naumer, M. (2008). Semantics and the

multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, 1242, 136–150.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*, 57, 11–23.

Easton, R., Srinivas, K., & Greene, A. (1997). Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 153–163.

Enns, J., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, 4, 345–352.

Ishai, A., & Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science*, 268, 1772–1774.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134, 372–384.

Labeyre, E., Oker, A., Badard, G., & Versace, R. (2008). Activation and integration of motor components in a short-term priming paradigm. *Acta Psychologica*, 129, 108–111.

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

Lehmann, S., & Murray, M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, 24, 326–334.

Mitterer, H., & Jesse, A. (2010). Correlation versus causation in multisensory perception. *Psychonomic Bulletin & Review*, 17, 329–334.

Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. (2007). Object-based attention is multisensory: Co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*, 26, 499–509.

Mulligan, N. W. (2003). Effects of cross-modal and intramodal division of attention on perceptual implicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 262–276.

Petrini, K., Russell, M., & Pollick, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, 110, 432–439.

Riou, B., Lesourd, M., Brunel, L., & Versace, R. (in press). Visual memory and visual perception: When memory improves visual search. *Memory & Cognition*, 1–9.

Schneider, T., Engel, A., & Debener, S. (2008). Multisensory identification of natural objects in a two-way cross-modal priming paradigm. *Experimental Psychology*, 55, 121–132.

Slotnick, S. D., & Schacter, D. L. (2006). The nature of memory related activity in early visual areas. *Neuropsychologia*, 44, 2874–2886.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, 28, 61–70.

Sperdin, H., Cappe, C., Foxe, J., & Murray, M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Frontiers in Integrative Neuroscience*, 3, 1–10.

Taylor, K., Moss, H., & Tyler, L. (2007). The conceptual structure account: A cognitive model of semantic memory and its neural instantiation. In J. Hart & M. Kraut (Eds.), *The neural basis of semantic memory* (pp. 265–301). Cambridge University Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (1995). Organization of memory: Quo Vadis? In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 839–847). Cambridge, Mass: MIT Press.

Vallet, G., Brunel, L., & Versace, R. (2010). The perceptual nature of the cross-modal priming effect: Arguments in favor of a sensory-based conception of memory. *Experimental Psychology*, 57, 376–382.

Vallet, G., Simard, M., & Versace, R. (in press). Sensory-dependent knowledge in young and elderly adults: Arguments from the cross-modal priming effect. *Current Aging Science*.

van den Bussche, E., van den Noortgate, W., & Reynvoet, B. (2009). Mechanisms of masked priming: A meta-analysis. *Psychological Bulletin*, 135, 452–477.

Versace, R., Labeyre, E., Badard, G., & Rose, M. (2009). The contents of long-term memory and the emergence of knowledge. *European Journal of Cognitive Psychology*, 21, 522–560.

Wallace, M., Roberson, G., Hairston, W., Stein, B., Vaughan, J., & Schirillo, J. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, 158, 252–258.