Cogn Process (2008) 9:69–76
DOI 10.1007/s10339-007-0195-6

REVIEW

How the bimodal format of presentation affects working memory: an overview

Serena Mastroberardino · Valerio Santangelo · Fabiano Botta · Francesco S. Marucci · Marta Olivetti Belardinelli

Received: 16 March 2007 / Revised: 24 September 2007 / Accepted: 26 September 2007 / Published online: 12 October 2007

Abstract The best format in which information that has to be recalled is presented has been investigated in several studies, which focused on the impact of bimodal stimulation on working memory performance. An enhancement of participant's performance in terms of correct recall has been repeatedly found, when bimodal formats of presentation (i.e., audiovisual) were compared to unimodal formats (i.e, either visual or auditory), in providing implications for multimedial learning. Several theoretical frameworks have been suggested in order to account for the bimodal advantage, ranging from those emphasizing early stages of processing (such as automatic alerting effects or multisensory integration processes) to those centred on late stages of processing (as postulated by the dual coding theory). The aim of this paper is to review previous contributions to this topic, providing a comprehensive theoretical framework, which is updated by the latest empirical studies.

Keywords Working memory · Bimodal · Unimodal · Dual code theory · Multisensory interaction · Multimedia learning

S. Mastroberardino (✉) · V. Santangelo · F. Botta · F. S. Marucci · M. Olivetti Belardinelli
Department of Psychology, University of Rome "La Sapienza", via dei Marsi 78, 00185 Rome, Italy
e-mail: serena.mastroberardino@uniroma1.it

F. S. Marucci · M. Olivetti Belardinelli
ECONA, Inter-university Centre for the Research on Cognitive Processing in Natural and Artificial Systems, Rome, Italy

Introduction

In recent years, novel technologies have introduced new ways of transmitting information and facilitating communication processes, in particular as a suitable basis for learning and teaching. As a consequence, many universities have started e-learning programmes in order to allow more students to enter higher education. This development represents a challenge for psychologists and neuroscientists, especially to those interested in the processes of learning and memory. In this context, two issues appear to be of: the format of the information that needs to be recalled and the relation between the stimulus presentation format and working memory (WM). The aim of this paper is to review studies relevant to these questions in order to provide a comprehensive theoretical framework, updated by recent empirical studies.

The modality effect

A key issue that has been widely investigated in the context of the relation between stimulus presentation and WM performance, concerns the format of memory representations thus focusing on the sensory modality of the input (e.g., visual or auditory) rather than the content of what is remembered (e.g., speech or non-speech). More specifically, several studies have examined whether different sensory modalities determine different types of memory processing. A step forward in the understanding of this phenomenon has been the finding of the *modality effect* (Crowder and Morton 1969; Penney 1975, 1989). This effect highlights the advantage in terms of recency of auditory vs. visual material in immediate serial recall tasks. The modality effect was originally attributed to a pre-categorical acoustic stage, through which auditory speech was supposedly retained in the form of an auditory-based code for a couple of seconds. Subsequent studies (Campbell and Dodd 1980; Spoehr and Corin 1978) suggested that the critical aspect was not the auditory form as such but whether the material is coded as speech or in other forms.

The results obtained using lip-reading material (which was remembered in the same way as speech) supported this *speech processing hypothesis*, which was also strengthened by the finding that adding an irrelevant speech sound at the end of a list of to-be-remembered words, overruled the advantage of the last item (see Surprenant et al. 2000; Jones 1994; Jones and Macken 1993, 1995). This phenomenon, called the ‘suffix effect’ (see e.g., Greene and Samuel 1986), was originally considered to be strictly auditory, and then attributed to speech processes, since lip-read information could interfere with auditory recency (see Spoehr and Corin 1978). As a consequence of the lip-reading studies, various authors began to investigate the content of the stimuli (speech or non-speech) instead of the input modality (visual or auditory). We will now briefly focus on the main findings of these studies, before addressing our main question: is there a relation between the format of input and WM performance?

Interesting findings, derived from the investigation of the content of stimuli, have been provided by De Gelder and Vroomen (1997). In this study, recency effects derived from four different formats of stimulus presentations: two with speech items (written and spoken words) and two with non-speech items (environmental sounds or drawings). The findings indicated clear recency advantage for spoken words; no advantage for printed words and drawings, and an intermediate performance was observed for environmental sounds. The authors proposed that the intermediate recency effect for environmental sounds was due to sounds and spoken words sharing auditory features although sounds do not involve the use of a phonological code. Moreover, the intermediate effect of environmental sounds seemed to favour the views that consider auditory explanations of the recency advantage as a combination of the auditory format and the linguistic content code (see also De Gelder and Vroomen 1992, 1994). In fact, the representation of a stored item may consist of a bundle of features. The exact composition of this bundle of features may vary as a function of the content of the items and the presentation modality. As a consequence, spoken words may therefore have two sets of features: one related to an auditory-based code, and the other related to a speech-based code. Conversely, environmental sounds may have auditory features, but not those resulting from speech coding, while written words and drawings would not be coded in an auditory or a speech-like code, thus overriding the recency effect.

The dual code account

Although the ‘modality effect’ studies indicated a shift of interest from the input modality (auditory or non-auditory) to the content of WM processes (speech or non-speech), at the same time other authors focused their investigation on the sensory modality of the input. For instance, in order to assess the independence of auditory and visual codes, Thompson and Paivio (1994) used a task in which participants had to memorize environmental items presented in three different formats: visual (i.e., pictures), auditory (i.e., sounds), or audiovisual (i.e., pictures and sounds). In order to minimize the role of verbal encoding, they used three verbal conditions: one included a verbal distractor task, one an incidental learning procedure, and a third intentional recall instructions without distractors. Thompson and Paivio’s results showed that participant recall in the dual modality condition (i.e., when pictures and sounds were presented together) was additive, when compared to the single modality conditions under incidental learning instructions. This result supported the hypothesis that auditory and visual components of audiovisual objects are functionally independent in memory, and it was also one of the first studies to imply a short-term memory advantage of the bimodal format of presentation with respect to the unimodal one.

Thompson and Paivio (1994) discussed their results in the light of the dual code theory, according to which memory traces are a conglomerate of modalities and components that can be retrieved independently. In terms of this theory therefore, bimodal items are more likely to be recalled than unimodal items because the latter are encoded only once, while the former are encoded twice. Incidentally, the dual code theory assumes that bimodal items affect not only WM performance, but also object identification. According to Paivio (1971, 1986), nameable objects can be retrieved via the activation of their names, their non-verbal images, or both. Object identification therefore appears to be affected by the input modality, given that it comprises memory traces corresponding to the various sensory modalities (e.g., vision, touch, and audition) in which an object is experienced (see Johnson et al. 1996). These modality-specific traces, often integrated in experience, may function independently, additively contributing to cognitive performance (Thompson and Paivio 1994). In fact, an experience in one modality often reintegrates object memories specific to other modalities. In the case of naming tasks, (as well as WM tasks), *multimodal* object representations might therefore play a key role. Object naming may primarily rely on visual characteristics, but it can also be based on auditory or tactile characteristics (see Buckingham 1981; Goodglass 1980; Thompson and Paivio 1994).

Bimodal effects using non-verbal stimuli

In summary, the bimodal advantage observed in terms of WM performance (as well as in terms of object identification) has been explained by several authors by means of the dual code theory (e.g., Johnson et al. 1996; Thompson and Paivio 1994). Specifically, the dual code theory assumes the existence of two different representation subsystems: A non-verbal subsystem, consisting of non-linguistic object representation or imagens (Paivio 1986), and a verbal subsystem consisting of linguistic word representations or logogens (Morton 1969, 1979). In terms of the dual code theory, therefore, the bimodal advantage is due to the fact that bimodal stimuli is represented by two codes, while unimodal stimuli are represented by a single code. Consequently, the representation of bimodal stimuli would appear to be more robust and would also be easier to recall than unimodal stimuli.

However, it is worth noting that all the studies that have been mentioned in the previous sections have used verbal (i.e., semantic) stimuli to demonstrate the bimodal advantage, while a crucial test for the dual code account might consist of the use of non-semantic (i.e., non-verbal) stimuli, i.e., stimuli that are not necessarily related to a particular meaning. Indeed, evidence for a bimodal advantage, emerging from the use of non-semantic stimuli would not be consistent with the dual code theory: given that it is not possible to assume that non-semantic stimuli are represented verbally (or linguistically) of, both unimodal and bimodal stimuli should be represented only by the non-linguistic representational subsystem (i.e., a single code in both cases). Thus, the bimodal advantage would be accounted for by different (and likely pre-semantic) processes.

Even though bimodal semantic effects have been demonstrated with non-linguistic materials (Olivetti Belardinelli et al. 2004), there is now considerable evidence that bimodal stimulation is more effective than unimodal stimulation regardless of the fact that it conveys semantic information. Non-semantic audio–visual interactions have also been demonstrated by means of neuro-imaging studies in several cognitive domains such as perception (e.g., Fort et al. 2002; Giard and Peronnet 1999) selective attention (e.g., Calvert et al. 2000; Teder-Sälejärvi et al. 2002), and, at a behavioral level, in the exogenous orienting of spatial attention (e.g., Santangelo et al. 2007; Santangelo and Spence 2007).

Santangelo et al. (2006) conducted a study using non-semantic stimuli in order to directly address the issue of whether or not bimodal non-verbal stimuli also provide WM advantage, when compared to unimodal non-verbal stimuli. In this study, non-verbal visual stimuli consisted of Chinese ideograms (note that participants were not familiar with these stimuli, i.e., they were not Chinese speakers or readers); non-verbal auditory stimuli consisted of musical fragments, which were created in order not to match any known song or melody; the bimodal stimuli consisted of the combination of both ideograms and musical fragments. Visual, auditory, and bimodal stimuli were presented in separate blocks of trials. The participants had to perform an *n*-back task, by pressing a response key when the presented stimulus was the same as the stimulus presented two positions before.

Santangelo et al. (2006) conducted two experiments which were identical, except that in Experiment 2 the participants had to perform articulatory suppression (cf. Baddeley and Hitch 1974) while performing the task, in order to avoid (or at least reduce) any spontaneous verbalization. Their results showed an advantage [i.e., faster response times (RTs)] for the bimodal presentation format compared to the unimodal one. This pattern of results was identical in both experiments, showing no effect of the articulatory suppression (see Fig. 1), implying that participants did not use verbalizations to carry out the task. The results are not consistent with the dual code theory, which assumes that the bimodal advantage is due to a dual representation. They do however suggest that non-semantic stages of processing are involved in the advantage observed on WM performance when bimodal stimuli are compared with unimodal stimuli. In principle, the bimodal improvement on WM performance might be attributed to an early multisensory facilitation process (see Stein and Meredith 1993), or to a functional link between modality-specific memory systems (see Engelkamp and Zimmer 1994).

Although further research on this topic seems necessary to assess the influence of these "pre-semantic" stages of elaboration on WM performance, Santangelo et al.'s (2006) results clearly indicated that the dual code account

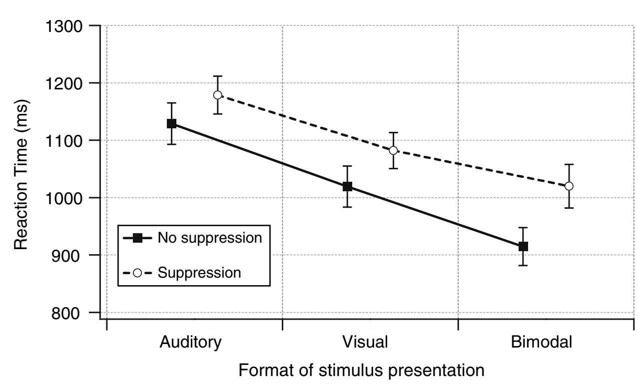

Fig. 1 Mean RTs and standard errors (*error bars*) as a function of the format of stimulus presentation with articulatory suppression or no-articulatory suppression observed by Santangelo et al. (2006) in their n-back task

is not sufficient to explain the bimodal advantage, thus ruling out the need of a verbal representation of the incoming (bimodal) information.

Format and content of input: a possible synthesis?

The previous section highlighted that although several studies investigated the ‘format’ (visual vs. auditory; e.g., Johnson et al. 1996; Thompson and Paivio 1994) and the ‘content’ (speech vs. non-speech; e.g., De Gelder and Vroomen 1992, 1994, 1997) of the to-be-remembered items separately, more fruitful results might be obtained from the simultaneous investigation of both these variables. For this very reason, this section will focus on those studies that specifically addressed the relation between the format and the content of the input. If bimodal formats of presentation enhance memory performance when compared to unimodal formats, a relevant issue may be whether the content of the bimodal items is crucial.

A study that aims to answer this question has been recently reported by F. Delogu and M. Olivetti Belardinelli (submitted). In their experiment, their participants had to memorize a list of verbal (spoken and written words) and non-verbal (environmental sounds and pictures) stimuli, presented either bimodally (e.g., the picture of a cow and the sound of the bellow) or unimodally (e.g., the picture of the cow or the bellow). The authors found that participants’ performance in a subsequent free recall task were superior for bimodal presentations, but *only* when non-verbal materials (environmental sounds and pictures) rather than verbal materials (spoken and written words) were employed. Although spoken words are different from written words in real life situations, Delogu and Olivetti Belardinelli argued that when these are used as experimental materials during free recall tasks, the information provided is quite similar. In fact, the perceptual information provided by the speaker’s voice (mood, characteristics, etc) is carefully controlled and kept constant across participants. In other words, according to Delogu and Olivetti Belardinelli, spoken words do not provide the same amount of affective and perceptual information as real life situations, and is thus comparable with written words. These results appear to indicate the existence of a clear interaction between the effect of the bimodal formats of presentation and the content of the presented items on WM performance, as previously reported by other studies as well.

For instance, Goolkasian and Foos (2005) asked their participants to recall three or six items while performing a concurrent task (i.e., verifying the accuracy of mathematical sentences). Items were presented either in a unimodal or bimodal format. Unimodal formats of presentation consisted of presenting the item either as a picture, a printed word, or a spoken word. Bimodal formats of presentation consisted of presenting the same items combined in three different formats: picture and printed words (i.e., visual formats), picture and spoken words (i.e., audiovisual formats), spoken words and printed word (i.e., audiovisual formats). An advantage was found for all *bimodal* formats of presentation, as compared to the unimodal format of presentation. Moreover, participants’ performances were superior when the bimodal presentation format was *audiovisual* rather than visual. That is, spoken words, when combined with either pictures or printed words (i.e., audiovisual formats), produced superior performances than pictures with printed words (i.e., visual formats).

The nature of the interaction between format and content of item presentation was further investigated by Foos and Goolkasian (2005). In four different experiments, participants tried to remember three or six concrete nouns presented as pictures, as spoken or printed words while performing a concurrent task. The concurrent task difficulty varied according to the difficulty of the math equation that had to be verified, [i.e., “Is $(8 \times 2) + 4 = 20$?”] vs. [“Is $(2 \times 3) + (4/1) - 5 = 5$?”], in order to control for central executive effort levels. According to the WM model, in fact, limiting central executive resources can affect WM performance (Baddeley and Hitch 1974; see also Hasher and Zacks 1979). An increase of the resources devoted to processing a more difficult task, should decrease the performance on the storage task, and as a consequence the effect of presentation should be more pronounced. By contrast, if independent resources are used in order to store and process inputs, no effect on the presentation format should be found (e.g., Oberauer 2002; Oberauer, et al. 2001). In the most recent formulation of the WM model, Baddeley (2000) included a temporary and limited capacity storage system, the episodic buffer, capable of integrating information from different sources. The episodic buffer is controlled by the central executive which is responsible for binding information from different sources into an episode, which can then be temporarily stored in the episodic buffer. Moreover, the buffer is supposed to provide a temporary interface between the phonological loop and the visuospatial scratchpad and LTM. If, as in the case of Foos and Goolkasian (2005), the central executive is limited by increasing the difficulty of a concurrent task, it is conceivable that the episodic buffer would be affected, since the central executive would be allocating more cognitive resources to perform the concurrent task and could thus not bind information from different sources in order to create an episode which can be stored in the buffer.

Foos and Goolkasian (2005) results were consistent with this view. In fact, increasing the difficulty of the concurrent task produced a decrease on recall performance for pictures and spoken words, but not for printed words. The authors explained this finding in terms of written words requiring fewer processing resources than spoken words and pictures (see also Goolkasian and Foos 2002). However, it is worth noting that in this experiment the authors could not rule out the role of the verbal buffer, since the concurrent task involved an arithmetic task. Therefore, in a further experiment, Foos and Goolkasian (2005) used a visuo/spatial reasoning task (sentence verification), which was believed to interfere with the storage of pictures (likely to be stored in the visual buffer), but not with words. In particular, they used a problem-solving task in which the difficulty of the sentence that had to be verified varied from an easy level (e.g., “Is the circle below the arrow?” when showing a circle above an arrow) to a difficult level (e.g., “Are red squares least likely?”, when showing four rows of different shapes—stars, triangles, etc.—presented in different numbers with different colours). The authors assumed that if the advantage of spoken words and pictures was due to different mechanisms, long term modality effects might explain the advantage of spoken words, while specific processing tasks used in the experiments (e.g., math verification sentences or visual/spatial reasoning) might underlie the advantages of pictures found. Compared to their previous experiments, the only different result, consisted of a general decrease of participants’ recall performance for all presentation formats, ruling out the type of processing task as an explanation of the superior performance of pictures and spoken words over printed words. Importantly, the existence of a long term modality effect showed a recall advantage for spoken words when compared to printed words.

Taken together, these results support the conclusion that the interaction between format and content of the to-be-remembered items is primarily based on shared resources between storage and processing in WM. This appears to be well accounted for by the existence of an episodic buffer (e.g., Baddeley 2000), which requires both storage and processing resources. As mentioned above, the episodic buffer is a temporary storage based on a multimodal code, controlled by the central executive that is responsible of binding information from different sources into an episode temporarily stored in the episodic buffer. If the central executive is involved in a concurrent task, performance will decrease as a consequence of cognitive resources being allocated to performing the concurrent task. Thus, it will be more difficult for this system to integrate information from different sources in order to create the episode stored in the buffer. The ‘episodic buffer’, seems to clarify why written words are less affected by the increasing difficulty of the concurrent task, compared to spoken words and pictures. In fact, written words require fewer cognitive resources than spoken words and pictures, thus reducing the central executive’s effect on the episodic buffer.

Beyond memory performance: a bimodal effect on reasoning and problem solving?

So far, evidence demonstrate the advantage, in terms of recall, of bimodally-presented items over unimodal formats of presentation. In order to shed more light on the mechanisms underlying this bimodal advantage, the discussion should briefly extend to reasoning and problem solving domains. In fact, it may be hypothesized that the format of stimuli presentations are also relevant to these other topics.

Several studies on multimedia learning (e.g., Sweller et al. 1990; Tindall-Ford et al. 1997) outlined the conditions under which problem solving? may benefit from bimodal rather than a unimodal formats of presentation. For instance, focusing on multiple sources of information requires participants to mentally integrate different information prior to problem solving, which may result in the splitting of attentional resources, thus interfering with problem solving itself. Conversely, if the material is presented in different formats, physically integrated problem solving is facilitated because the load on WM is reduced. In fact, according to Baddeley’s WM model (1990, 1992) there are two independent processors, the visuo-spatial sketch-pad and the phonological loop for verbal materials, and if materials are presented in more than one sensory modality, the WM capacity increases, when compared to a single format of presentation.

It is surprising that despite these theories to date no evidence seems to exist, which supports that bimodal formats of presentation aid participants in reasoning or problem solving tasks. By contrast, there is evidence which shows no interaction between format of presentation and performance of these tasks (see Goolkasian 1996, 2000). For instance, Goolkasian (2000) evaluated the effect of presentation format on problem solving in a test during which the participants were presented a series of grouped items in different format and asked to make true/false judgements to test sentences. Two different problem solving items (i.e., probability judgements on colours and category inclusion) were used in two different experiments. In the category inclusion task, the participants were presented with three pictures on the same slide (e.g., three cows, three of horses, and three of roses) and were subsequently asked to evaluate one of two sentences: a true one (“*there are fewer flowers than animals*”), and a false one (“*there are fewer animals than flowers*”). In the probability judgements with colours tasks, for example, seven red, five blue, and three yellow squares were presented and participants were required to make a judgment of the truth of sentences: “Yellow squares are least likely to appear” (true) or “Red squares are least likely to appear” (false). According to Goolkasian, probability judgements with colours are more dependent on visual representation, while category inclusions are more related to membership in categories. The author hypothesised an advantage of multiple presentations across format (picture/printed word) or across modalities (picture/spoken or printed/spoken), resulting in shortened RTs, when compared to a single-format condition.

Regardless of the type of problem solving task (i.e., probability judgements with colours or category inclusions), the data showed the same pattern or results. Consistently with previous studies (e.g., Bauer and Johnson-Laird 1993; Goolkasian 1996), Goolkasian (2000) found an advantage for pictures presentation in terms of RTs, while the spoken materials took longer to be processed and produced more response errors. This suggests that the presentation format may influence the way information is acquired and processed by participants. This result can be explained by hypothesising that pictures benefit from a more direct access to information than printed words (Larkin and Simon 1987). However, this *direct access hypothesis* does not explain why the picture advantage decreases with the increase of the delay between the presentation of items. In fact, when both study materials and test sentence appeared simultaneously, picture and printed words differences decreased. Crucially, however, in the bimodal format condition the observed RTs were longer than those of the unimodal format conditions. Indeed, the redundancy of the information provided by the bimodally-presented items did not help participants to make inferences, especially when the combined condition included all three formats. However, it is worth noting here that the studies that reported a bimodal advantage in memory tasks typically used the number of reported items as a dependent measure, while Goolkasian used RT measures. One may therefore argue that the differential effect (i.e., bimodal advantage in memory but not in reasoning and problem solving tasks) is just accounted for by the different variables measured. However, given that a bimodal advantage in WM performance has been observed also when measuring RTs (see Santangelo et al. 2006), we believe that this differential effect is based on more complex differences between these types of tasks. Future research with the non-trivial task has to investigate why the bimodal advantage, which seems so effective on memory performance, disappears in reasoning and problem solving tasks, although these are clearly memory-based tasks.

Conclusion

This paper reviewed studies assessing whether or not a bimodal formats of stimulus presentation increase performance in terms of WM, when compared to unimodal format of presentation. Although the bimodal format of presentation appeared not to fulfil a special role when compared to the unimodal formats of presentation on cognitive tasks such as problem solving (e.g., Goolkasian 2000), a clear bimodal advantage has been found in a number of studies on WM performance (e.g., Foos and Goolkasian 2005; Goolkasian and Foos 2002; Santangelo et al. 2006; Thompson and Paivio 1994), and occasionally on reasoning tasks (e.g., Goolkasian 1996). Taken together these results support the view that a bimodal (i.e., audio-visual) format of presentation is more effective when information needs to be remembered, with respect to unimodal (either visual or auditory) formats of presentation. It is worth mentioning that the notion of a bimodal advantage may have significant implications for everyday learning situations (such as e-learning), thus strengthening the multimedia approach to teaching. As mentioned, the influence of innovative technology on learning has recently created new interest in researchers. E-learning has been viewed as a potential source for teaching and the spreading of culture. Many universities are now offering on-line courses to allow people to attend classes and enter higher education. The possibility of an interaction between students and computers, (i.e., e-learning or multimedia encyclopaedias and textbooks), stimulates new theories and further research on how to design computer based instructions using words and pictures. For instance, Mayer and Sims (1994), extended Paivio's dual coding theory to multimedia learning.

According to the dual coding theory (Paivio 1986), when visual and verbal information is presented separately or simultaneously, two different information processing system are activated, one for the representation of the verbal and one for the representation of the visual information. In a multimodal learning condition, when animation and narration are presented, they must be integrated during learning within the learner's WM. However, the dual code theory does not sufficiently explain how this integration takes place. The modified version of the dual code theory for multimedia learning proposed by Mayer and Sims (1994) hypothesizes the existence of three different processes. First, the learner constructs a mental representation of the verbal presentation (i.e., verbal encoding). If the presentation is visual (e.g., an animation) the learner constructs a mental representation of it (i.e., visual encoding). These two mental representations are then connected by means of referential connections, which create links between the information presented either in visual or verbal modalities. The learner must then create connections between causal relations as stated in words and their corresponding visual presentation, as well as referential connections between visual and verbal representation of essential parts in the system. Finally, connections must be made between underlying casual principles in formal and visual forms.

The extension of the dual code theory to problem solving, could provide a theoretical basis for analysing methods that facilitate all types of connections. In this case, when visual and verbal materials are contiguous, the creation of referential connections is enhanced, thus aiding the learning process. Multimodal presentations of learning material thus seem to aid learning by automatically creating connections between verbal and visual information. In contrast, separate presentations, require more effort since the learner needs to create the connections.

Acknowledgment We wish to thank Angela Tagini for the language revision.

References

Baddeley A (1990) Human memory: theory and practice. Lawrence Erlbaum Associates Inc, Hillsdale

Baddeley A (1992) Working memory. Science 255:556–559

Baddeley AD (2000) The episodic buffer: a new component of working memory? Trends Cogn Sci 4:417–423

Baddeley AD, Hitch GJ (1974) Working memory. In: Bower G (ed) The psychology of learning and motivation. Academic, New York, pp 47–89

Bauer MI, Johnson-Laird PN (1993) How diagrams can improve reasoning. Psychol Sci 6:372–378

Buckingham HW (1981) Lexical and semantic aspects of aphasia. In: Sarno MT (ed) Acquired aphasia. Academic, New York, pp 183–214

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657

Campbell R, Dodd B (1980) Hearing by eye. Q J Exp Psychol 32:85–99

Crowder RG, Morton J (1969) Precategorical acoustic storage (PAS). Percept Psychophys 5:363–373

De Gelder B, Vroomen J (1992) Abstract versus modality-specific memory representations. Mem Cognit 20:533–538

De Gelder B, Vroomen J (1994) Memory for consonants versus vowels in heard and lipread speech. J Mem Lang 31:737–756

De Gelder B, Vroomen J (1997) Modality effects in immediate recall of verbal and non-verbal information. Eur J Cogn Psychol 9:97–110

Engelkamp J, Zimmer HD (1994) The human memory: a multi-modal approach. Hogrefe, Toronto

Foos PW, Goolkasian P (2005) Presentation format effects in working memory: the role of attention. Mem Cognit 33:499–513

Fort A, Delpuech C, Pernier J, Giard MH (2002) Early auditory–visual interactions in human cortex during nonredundant target identification. Cogn Brain Res 14:20–30

Giard MH, Peronnet F (1999) Auditory–visual integration during multimodal object recognition in humans: a behavioural and electrophysiological study. J Cogn Neurosci 11:473–490

Goodglass H (1980) Disorders of naming following brain injury. Am Sci 68:647–655

Goolkasian H (1996) Picture-word difference in a sentence verification task. Mem Cognit 24:584–594

Goolkasian P (2000) Pictures, words, and sounds: from which format are we best able to reason? J Gen Psychol 127:439–459

Goolkasian P, Foos PW (2002) Presentation format and its effect on working memory. Mem Cognit 30:1096–1105

Goolkasian P, Foos PW (2005) Bimodal format effects in working memory. Am J Psychol 118:61–77

Greene RL, Samuel AG (1986) Recency and suffix effects in serial recall of musical stimuli. J Exp Psychol Learn Mem Cogn 12:517–524

Hasher L, Zacks RT (1979) Automatic and effortful processes in memory. J Exp Psychol Gen 108:356–388

Johnson CJ, Paivio A, Clark JM (1996) Cognitive components of picture naming. Psychol Bull 120:113–139

Jones DM (1994) Disruption of memory for lip read lists by irrelevant speech: further support for the changing state hypothesis. Q J Exp Psychol 47A:143–160

Jones DM, Macken WJ (1993) Irrelevant tones produce an irrelevant speech effect: implications for phonological coding in working memory. J Exp Psychol Learn Mem Cogn 19:369–381

Jones DM, Macken WJ (1995) Phonological similarity in the irrelevant speech effect: within- or between-stream similarity? J Exp Psychol Learn Mem Cogn 21:103–133

Larkin JH, Simon HA (1987) Why an illustration is (sometimes) worth ten thousand words. Cogn Sci 11:65–99

Mayer RE, Sims VK (1994) Why a diagram is (sometimes) worth ten thousand words. Cogn Sci 11:65–69

Morton J (1969) Interaction of information in word recognition. Psychol Rev 76:165–178

Morton J (1979) Facilitation in word recognition: experiments causing change in the logogen model. In: Kolers PA, Wrolstead M, Bouma H (eds) Processing of visible language, vol 1. Plenum Press, New York, pp 259–268

Oberauer K (2002) Access to information in working memory: exploring the focus of attention. J Exp Psychol Learn Mem Cogn 28:411–421

Oberauer K, Demmruch A, Mayr U, Kilegl R (2001) Dissociating retention and access in working memory: an age-comparative study of mental arithmetic. Mem Cognit 29:18–33

Olivetti Belardinelli M, Sestieri C, Di Matteo R, Delogu F, Del Gratta C, Ferretti A, Caulo M, Tartaro M, Romani GL (2004) Audio–visual crossmodal interactions in environmental perception: An fMRI investigation. Cogn Process 5:167–174

Paivio A (1971) Imagery and verbal processes. Holt, Rinehart and Winston, New York

Paivio A (1986) Mental representations: a dual coding approach. Oxford University Press, Oxford

Penney CG (1975) Modality effects in short-term verbal memory. Psychol Bull 82:68–84

Penney CG (1989) Modality effects and the structure of short-term verbal memory. Mem Cognit 17:398–422

Santangelo V, Spence C (2007) Multisensory cues capture spatial attention regardless of perceptual load. J Exp Psychol Hum Percept Perform (in press)

Santangelo V, Mastroberardino S, Botta F, Marucci FS, Olivetti Belardinelli M (2006) On the influence of audio–visual interactions on working memory performance: a study with non-semantic stimuli. Cogn Process 7:187

Santangelo V, Olivetti Belardinelli M, Spence C (2007) The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. J Exp Psychol Hum Percept Perform 33:137–148

Spoehr KT, Corin WJ (1978) The stimulus suffix effect as a memory coding phenomenon. Mem Cognit 6:583–589

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

Surprenant AM, LeCompte DC, Neath I (2000) Manipulations of irrelevant information: suffix effects with articulatory suppression and irrelevant speech. Q J Exp Psychol A 53:325–348

Sweller J, Chandler P, Tierney P, Cooper M (1990) Cognitive load as a factor in the structuring of technical material. J Exp Psychol Gen 119:176–192

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio–visual crossmodal integration by means of event-related potential (ERP) recordings. Cogn Brain Res 14:106–114

Thompson V, Paivio A (1994) Memory for pictures and sounds: independence of auditory and visual codes. Can J Exp Psychol 48:380–396

Tindall-Ford S, Chandler P, Sweller J (1997) When two sensory modes are better than one. J Exp Psychol Appl 3:257–287