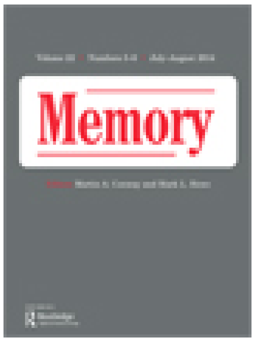

MEMORY

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Semantic encoding in working memory: Is there a (multi)modality effect?

Franco Delogu , Antonino Raffone & Marta Olivetti Belardinelli

To cite this article: Franco Delogu , Antonino Raffone & Marta Olivetti Belardinelli (2009) Semantic encoding in working memory: Is there a (multi)modality effect?, MEMORY, 17:6, 655-663, DOI: 10.1080/09658210902998054

To link to this article: https://doi.org/10.1080/09658210902998054

MEMORY, 2009, 17 (6), 655–663

Psychology Press
Taylor & Francis Group

Semantic encoding in working memory: Is there a (multi)modality effect?

Franco Delogu
Sapienza University of Rome, and ECONA, Rome, Italy, and RIKEN BSI, Wako, Japan

Antonino Raffone
Sapienza University of Rome, and ECONA, Rome, Italy, RIKEN BSI, Wako, Japan, and University of Sunderland, UK

Marta Olivetti Belardinelli
Sapienza University of Rome, and ECONA, Rome, Italy

In spite of a large body of empirical research demonstrating the importance of multisensory integration in cognition, there is still little research about multimodal encoding and maintenance effects in working memory. In this study we investigated multimodal encoding in working memory by means of an immediate serial recall task with different modality and format conditions. In a first non-verbal condition participants were presented with sequences of non-verbal inputs representing familiar (concrete) objects, either in visual, auditory or audio-visual formats. In a second verbal condition participants were presented with written, spoken, or bimodally presented words denoting the same objects represented by pictures or sounds in the non-verbal condition. The effects of articulatory suppression were assessed in both conditions. We found a bimodal superiority effect on memory span with non-verbal material, and a larger span with auditory (or bimodal) versus visual presentation with verbal material, with a significant effect of articulatory suppression in the two conditions.

Keywords: Working memory; Encoding; Multimodal; Episodic Buffer; Semantics; Span.

Most events in everyday environments produce a nearly simultaneous multimodal stimulation in different perceptual systems. For example, when a falling bottle reaches the ground we perceive the glass-breaking sound at almost the same moment in which we are seeing the impact. While seeing the bottle falling we expect the sound to come, and we can even foresee the auditory features of the incoming crash. This synchronisation, called *valid co-occurence* by Bertelson and de Gelder (2004), is very powerful in creating strong crossmodal associations. Behavioural and neurophysiological research has repeatedly demonstrated that inputs in different modalities are not perceived in isolation (Calvert, 2001; Olivetti Belardinelli et al., 2004; Spence & Driver, 2004). Our perceptual systems continuously integrate information in order to provide a unified representation of our surroundings.

Researchers' attention in multisensory integration is mostly focused on spatio-temporal input concomitance. In the experiments on this topic,

Address correspondence to: Franco Delogu, Department of Psychology, Sapienza University of Rome, Rome, Italy. E-mail: franco.delogu@uniroma1.it

We are grateful to Perla Amodio and Chiara Bozzacchi for their help in the data collection, and to an anonymous reviewer for his/her constructive comments and helpful suggestions.

http://www.psypress.com/memory DOI:10.1080/09658210902998054

very simple stimuli like brief noise bursts, tones, and flashes are generally preferred to stimuli deriving from meaningful multimodal events (for exceptions see Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004, Moholm, Ritter, Javitt, & Foxe, 2004; Olivetti Belardinelli et al., 2004). However, the use of meaningless stimuli does not allow the analysis of semantic associations generally occurring during the encoding of multisensory stimuli in ecological situations.

A similar lack of attention for semantic associations can also be found in working memory research. In particular, no studies have been conducted on serial working memory encoding of multimodal non-verbal information. Modality and multimodality effects in working memory have been mostly investigated with phonological materials. Several studies consistently reported that in typical forward serial recall tasks, the serial recall of auditory–verbal items produces a greater recency effect than that of visual–verbal items. This is referred to as the *modality effect* (Crowder & Morton, 1969). Given the efficiency of the phonological store in serial recall, adult participants typically name and subvocally rehearse visually presented items, thereby transferring the information from a visual to an auditory code (Baddeley, 1986). Articulatory suppression prevents this by removing the effect of phonological similarity for visually presented items, but not for the auditory presented ones, as they are automatically registered in the phonological store (Murray, 1968). As regards maintenance, the first working memory models (Baddeley, 1986; Baddeley & Hitch, 1974), entailing a strict separation between stores, cannot explain the integrated maintenance of multimodal events and audio-visual codes in working memory.

In spite of the increasing number of converging evidences of a more integrated functioning of the working memory subsystems (Baddeley, 2000; Prabhakaran, Narayanan, Zhao, & Gabrieli, 2000; Repovš & Baddeley, 2006), the research on multisensory integration in working memory is still rather limited. In particular, no previous research has tried to investigate semantic activation using environmental bimodal items.

In the present study we investigate the influence of semantic encoding in serial working memory by comparing the efficiency of verbal and non-verbal encoding in both unimodal and multimodal conditions.

In the non-verbal condition participants memorised concrete concept items. These items were either presented in a bimodal audio-visual format (a picture and the semantically associated sound) or in separate auditory-only and visual-only conditions. In the verbal condition participants memorised words presented in either a bimodal audio-visual format or in separate auditory and visual formats.

As regards semantic encoding, in the non-verbal condition the simultaneous presentation of non-redundant visual and auditory information ensures an augmented semantic encoding while the bimodal verbal presentation does not. As regards the format of maintenance, in all of three non-verbal conditions participants need to convert stimuli into their verbal labels. Differently, in the verbal encoding condition, a direct access of words to the phonological loop does occur in the case of spoken- and spoken–written materials, and a grapheme–phoneme conversion is needed for the written stimuli.

Therefore, in spite of the phonological nature of the task (verbal recall), we are expecting to find a better span performance for bimodally presented items in the non-verbal encoding condition. We assume that the enriched semantic encoding should help the rehearsal and recall of the phonological elements during the span task. This result should be in accordance with evidences about the interactive activation of phonological, lexical and semantic nodes in verbal short-term memory (Martin & Saffran, 1997).

In the verbal condition, since it is well established that verbal short-term memory draws heavily on a phonological code (Conrad & Hull, 1964), we expect to find an advantage of verbal auditory presented items (both in isolation and with written words) on written words items in isolation, but no a multimodal benefit per se.

In order to assess the relative reliance on the articulatory loop and phonological store of the different formats and modalities of presentation, we will test the effects of articulatory suppression in all the experimental conditions.

METHOD

Participants

A total of 80 university students participated in the study. Their age ranged from 19 to 35 years, with a mean age of 23.05 years. No participants presented auditory or visual sensory deficits.

Apparatus

A personal computer with 1 GHz Intel processor and 256 MB RAM running Windows XP OS was used for data collection. E-prime software was used for to program stimulus presentation and data collection. Images and sounds were presented by means of a 14-inch monitor and two loudspeakers displaced on the left and the right of the computer monitor.

Stimuli

A total of 40 non-verbal pairs and 40 verbal pairs of auditory and visual stimuli were used in the experiment. The non-verbal auditory stimuli were 40 environmental sounds and the non-verbal visual stimuli were 40 colour pictures of easily identifiable objects which were semantically linked to the environmental sounds (the complete list of the stimuli is shown in the Appendix). All the non-verbal stimuli were selected from a previous study (Delogu, Auricchio, & Olivetti Belardinelli, 2006) according to the criterion that all sounds and pictures had been correctly recognised and named with an appropriate verbal label in the totality of the cases.

All materials belonged to three semantic macro-categories: human beings, animals, and non-living things. Auditory and visual stimuli were organised in semantically matched bimodal pairs, in which pictures and sounds referred to the same environmental source, e.g., the image of a cat and its meowing. The items were arranged in three different kinds of memory sequences, depending on different experimental conditions: auditory sequences, consisting of environmental sounds only; visual sequences consisting of images (pictures) only; audio-visual (bimodal) sequences consisting in simultaneously presented sounds and pictures, referring to the same object. In all sequences items were displaced in a pseudo-random order to avoid easy semantic associations (e.g., cat following dog), and to generate sequences of comparable difficulty.

Verbal stimuli consisted in the labels of the objects presented in the non-verbal condition. Participants were presented with 40 spoken and written words, which corresponded semantically to the 40 environmental sounds and pictures used in the non-verbal condition.

Design and procedure

Participants were randomly divided into two experimental groups with 40 participants in each group. The two groups performed the same task with different experimental materials. The non-verbal group was presented with non-verbal item sequences, while the verbal group with sequences of words. Words and non-verbal items were semantically matched. Each group was divided into two sub-groups: the first one was asked to perform concurrent articulatory suppression by repeatedly subvocalising the syllables ‘‘cola-cola’’ while the other sub-group did not. For all conditions, memory performance was assessed by means of a span task adapted from the digit span task of the WAIS-III (Wechsler, 1997).

Each participant performed three blocks of trials. Each block consisted in the presentation of either auditory, or visual, or bimodal audio-visual item sequences of increasing length in terms of number of items (see Figure 1). The order of the blocks was randomised across subjects. Each item was presented for 2500 ms, and was preceded by a fixation point centred on the monitor for 500 ms (see Figure 1).

Participants were individually tested in a silent and dimly lit room. For each sequence, participants were instructed to memorise all the items for a subsequent verbal recall in the correct serial order of presentation. Each of the three presentation modality blocks, started from a sequence length of three items. If participants were able to recall correctly two sequences of a specific length within three attempts, they were allowed to try to recall sequences including one more item. By contrast, when participants made two errors out of three attempts, the presentation ended for that condition. The participant's span for each of the three presentation conditions was given by the longest sequence in which the participant performed correctly twice. A pause of 10 seconds was included between the end of a trial and the start of the next sequence in order to avoid interference effects. Recalls with wrong item order, missing items, and insertion of non-presented items were classified as wrong responses. Vice versa, in the non-verbal condition, those trials in which participants could not name the item using the same labels of the verbal condition, but in which they used synonyms, were considered correct.

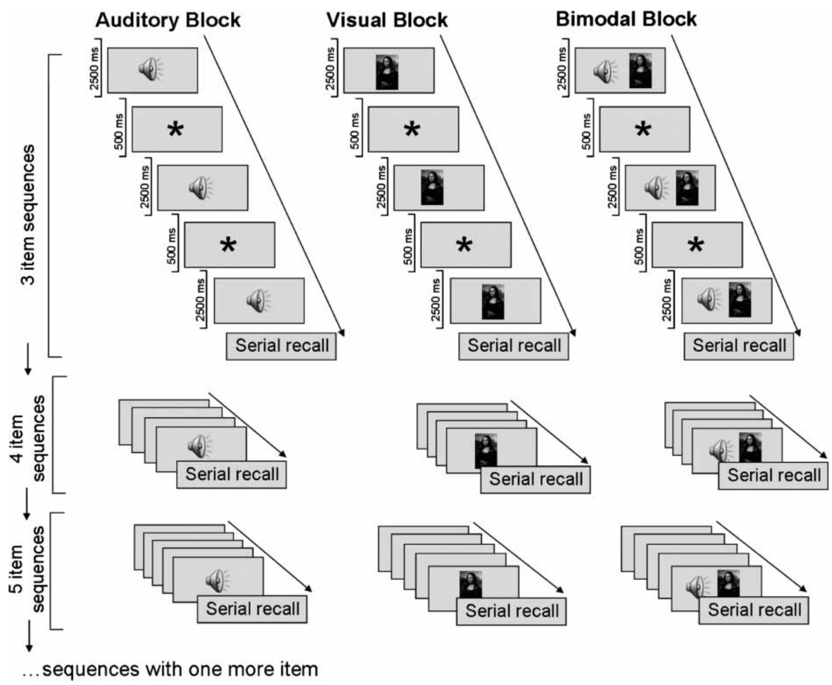

Figure 1. Structure and time course of each block in the three modality presentation conditions: auditory, visual, and bimodal (audio-visual). In this illustration visual and auditory inputs are represented by icons. Participants perform recall of one item longer sequences if they have provided correct recall of two out of three sequences with a given number of items (starting with three items). See text for more explanation.

RESULTS

Individual span scores were averaged for each presentation condition, with and without articulatory suppression. A mixed design ANOVA was performed, with Modality (with three levels: bimodal, visual and auditory presentation) as within-participants variable and Format (two levels: non-verbal, Verbal) and Articulatory suppression (two levels: presence vs. absence) as between-participants variables.

The analysis reveals a significant main effect of Modality, $F(2, 152) = 13.31$ $p < .01$ in which the bimodal span is superior to the span for both the unimodal conditions and the auditory span was superior to the visual one.

There is also a significant main effect of Articulatory suppression, $F(1, 76) = 43.3$, $p < .01$. The interaction between Modality and Articulatory suppression is not significant, $F(2, 152) = 0.604$, $p > .05$.

The main factor Format has no significant influence on span performance, $F(1, 76) = 0.246$, $p > .05$.

The most important results for our theoretical purposes come from the analysis of the interaction between Format and Modality factors. This interaction is significant, $F(2, 152) = 3.16$, $p < .05$. Specifically, in the non-verbal condition (and not in the verbal one), the post-hoc analysis (Fisher's LSD Test) indicates a better span in the bimodal presentation than in the unimodal visual ($p < .01$) and auditory, ($p < .01$) presentations. This multimodal effect, which is present only in the non-verbal format condition, suggests that a non-redundant multimodal encoding plays an important role in serial verbal recall (Figure 2).

Conversely, the post-hoc analysis shows a higher span for auditory than for visually presented items, only in the case of verbal material, ($p < .01$). In fact, while the span for spoken words is significantly larger than the span for corresponding written words, the spans for environmental sounds and for pictures are of comparable extent ($p > .05$), see Figure 3.

Finally, a non-significant interaction between Format and Articulatory suppression was found: $F(1, 76) = 2.17$, $p > .05$.

DISCUSSION

The effect of the main factor Modality is significant, indicating that the bimodal span is

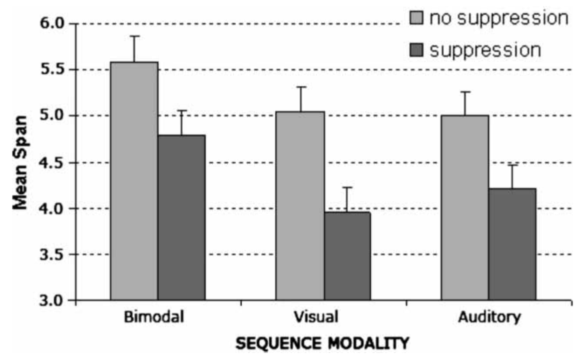

Figure 2. Memory span for non-verbal items with auditory, visual, and bimodal presentations. Note the bimodal superiority effect, and the global span reduction with articulatory suppression.

superior to both the unimodal spans and also that the auditory span is superior to the visual span. This result derives from collapsing together verbal and non-verbal conditions, and seems to summarise the multimodal benefit in the non-verbal condition and the auditory benefit in the verbal condition.

The effect of the main factor Format is not significant, indicating that verbal and non-verbal encoding do not lead to different span performances. Considering that only in the non-verbal condition participants have to link the presented stimulus (picture or sound) to an appropriate lexical label in order to use it in the phonological rehearsal, we can draw to different explanations. The first one is that the costs for lexical selection are not high enough to influence span performance. The second account is that the costs of lexical selection are counterbalanced by an advantage of the analogical (non-verbal) encoding.

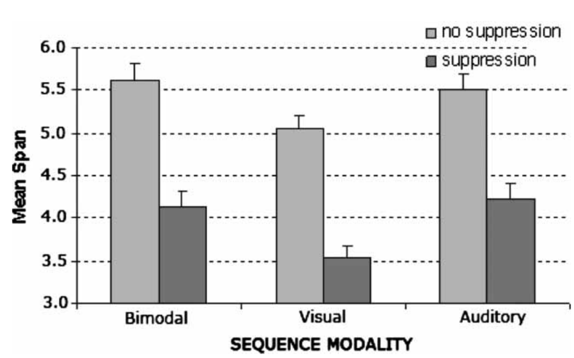

Figure 3. Memory span for words with auditory, visual, and bimodal presentations. Note that the span is lower with visual presentation as compared to both auditory and bimodal presentation conditions. Also note the global reduction of span with articulatory suppression.

The effect of the main factor Articulatory Suppression is significant and does not interact with the other main factors Format and Modality. This result indicates that the phonological loop was critical for rehearsal and recall in both verbal and in non-verbal conditions.

Our results showed interesting differences between verbal and non-verbal span performances (Figure 4). In the non-verbal condition we found a larger span performance for the bimodal presentation, which was not found in the verbal condition. Such an effect is likely to be due to the presence of simultaneous sources of non-redundant auditory and visual information during encoding. This semantic bimodal superiority is not attenuated by the articulatory suppression, thus suggesting that it is not linked to an enhanced access to a phonological representation of the items. The absence of differences between the two non-verbal unimodal conditions, and a comparable attenuation effect due to the articulatory suppression in all conditions of encoding, is a proof that participants convert non-verbal stimuli in verbal labels in a comparable way regardless of the modality of presentation.

By contrast, in the verbal condition we observed a better span performance for the auditory sequences than for the visual ones. This modality effect is consistent with several findings showing a better performance with auditory verbal materials than with visual verbal materials in immediate serial recall tasks (see, for example, Crowder & Morton, 1969). In our case, the poorer span performance for written word sequences is likely due to the costs of the grapheme–phoneme

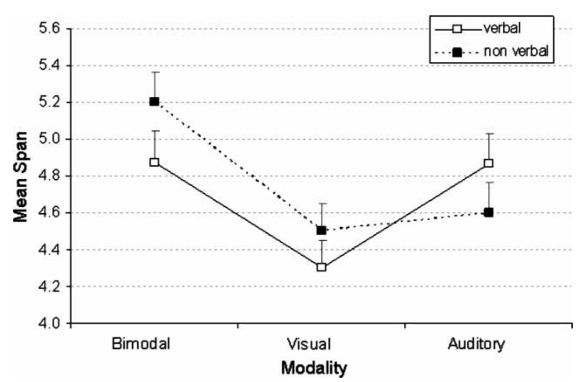

Figure 4. Comparison between spans in non-verbal and verbal conditions. Note the different effects of bimodal and auditory presentations with verbal and non-verbal material. See text for more explanations. Since articulatory suppression does not show significant interactions, we collapsed the two articulatory suppression conditions in order to facilitate the reading of the figure.

conversion. Moreover, as in the non-verbal condition, we found that articulatory suppression affects the span in all verbal presentation modality conditions. The absence of interactions with the factor Modality is counterintuitive, as we would expect articulatory suppression to affect the span more for written than for spoken words. This lack of interaction may be due to the relatively long presentation time of auditory and visual verbal items in this experiment, which is likely to eliminate the detrimental effect of the articulatory control process on the encoding of visually presented words (Baddeley, 1986). However, as shown by the significant overall effects of articulatory suppression, the articulatory control loop would indeed play a key role in maintenance.

Finally, as expected, we did not observe a multimodality effect in the verbal condition. In this case, the presented auditory and written words do not serve to ensure a richer degree of semantic activation as environmental sounds and pictures do in the non-verbal condition. Consequently, since the auditory verbal input has a direct access to the phonological store (Baddeley, 1986), and since the concomitant presentation of written words brings no additional semantic content, the bimodal presentation does not produce any benefits.

CONCLUSION

Previous research on multisensory encoding mostly focused on spatiotemporal contingencies, generally avoiding the involvement of semantics in investigating and modelling multisensory integration. On the other hand, working memory studies have generally paid little attention to multisensory interaction of cross-modal environmental information coming into an integrated episode (see Baddeley, 2000).

The results of our experiment lead to some interesting considerations about multimodal encoding, semantic activation, and phonological recoding in working memory. First of all, the fact that serial verbal recall is superior after non-verbal bimodal encoding is consistent with an interactive model of working memory. In particular, the multimodal benefit we have shown indicates that visual and auditory congruent but non-redundant information are inclined to interact at one or more stages of processing in working memory.

However, in our experiment multimodal presentation per se does not lead to any memory benefit. It seems that such an effect can only be possible with an involvement of semantic memory, which is able to establish a meaningful link between the otherwise independent auditory and visual concurrent stimulus presentations. Semantic coherence between stimuli encoded through different modalities ensures an enriched multimodal encoding and a consequent enhancement at the level of recall, even though it is mediated by a phonological recoding. Consistently with this interpretation, the multimodal effect is limited to the non-verbal domain.

For verbal materials we actually found that the auditory format is dominant, and that there is no additional benefit from a bimodal (audio-visual) encoding of words. It could be argued that we are not dealing with a semantically coherent multimodal encoding, but only with an independent activation of the auditory and visual storage systems. In further research, by adding to the design a condition in which auditory and visual stimuli are semantically mismatched, it could be possible to dissociate the effect of the spatiotemporal contingency from the effect of semantic coherence. In our view, however, it seems very unlikely that by mismatching auditory and visual stimuli we would observe a multimodal effect similar to the one we observed in the present non-verbal condition.

The significant main effect of articulatory suppression without any interactions with the format (verbal, non-verbal) and the presentation modality (auditory, visual, bimodal) suggests that the phonological store is the primary component for maintenance in our task, with performance based on serial recall. A possible explanation to the fact that articulatory suppression affects performances with verbal and non-verbal material in the same way is that there are no other working memory systems, but only the phonological store, involved in the serial maintenance of items regardless of the format of encoding. In favour of this account there are several findings showing the difficulty to maintain the serial order of visuo-spatial information (e.g., Phillips & Christie, 1977). It is arguable that both the verbal and non-verbal stimuli were simply recoded verbally and held in the phonological store. Following this interpretation, in case of difficulty in the identification or in the verbally recoding of sounds and pictures, two routes of naming can lead to a better memory span. We can disconfirm this alternative explanation for two reasons. First, all the stimuli were verified and selected for their ease of recognition and naming. Second, a difficulty in the naming of non-verbal items would result in differences between the non-verbal and the verbal span, in which the names to recall are already provided. This is not the case. In fact if we exclude the verbal auditory span, which is superior to all the other unimodal spans because of the modality effect (Crowder & Morton, 1969), no other verbal span appears to be higher than the non-verbal ones.

We can conclude that the most straightforward explanation for our effect is in terms of multimodal encoding of episodic sequences, mediated by active semantic representations, interacting with phonological codes. A strong role of semantics in serial recall tasks has been previously shown in several studies (Bourassa & Besner, 1994; Jefferies, Frankish, & Lambon-Ralph, 2006; Walker & Hulme, 1999).

A likely candidate for this role of interface between the working memory sub-systems and long-term memory is the *episodic buffer*. In fact, our evidence of an advantage in recall provided by the meaningful relation between auditory and visual concurrent items seems to fit with a buffer described as ‘‘an interface between a range of systems, each involving a different set of codes’’ (Baddeley, 2000, p. 421). In particular, as pointed out by Repovš and Baddeley (2006), one of the aspects of working memory that could be satisfactorily accounted for by the episodic buffer mechanisms in future research is how the subsidiary systems interact with each other and how they are related to semantics. In our experiment, multimodal encoding leads to an enhanced memory performance that is not explainable without taking into account an interaction between a phonological, a visual, and an auditory-but-not-phonological system. Therefore, a memory buffer that allows the temporary interaction of integrated codes is a likely explanation for our data. However, the understanding of the mechanisms that mediate such multimodal interactions is not completely inferable from our result and needs to be specifically addressed by future investigations. Our data are also interpretable within the theoretical framework of the dual-coding theory (Paivio, 1986). Paivio states that ‘‘human cognition is unique in that it has become specialized for dealing simultaneously with language and with non-verbal objects and events’’ (p. 53). His classic theory claims that both non-verbal and verbal codes are used to organise incoming information into knowledge that can be stored and retrieved for subsequent use. In our experiment, participants deal with two non-verbal codes, one deriving from the environmental sounds and one from the pictures. Both of these codes retain the main perceptual features of what is being observed and heard, and both of them can activate the phonological code. When these two different, but semantically related, analogical codes are present we have an advantage in memory recall. Further research should be conducted to investigate how the dual coding theory deals with the simultaneous activation of verbal and various non-verbal codes and, more in general, how ‘‘modal’’ and linguistic representations interact (Barsalou, 1999).

Finally, this research stresses some aspects of environmental sound encoding in short-term memory, an almost unexplored field of research. Auditory short-term memory models commonly do not comprise cognitive mechanisms devoted to non-phonological auditory processing (for a few exceptions, see de Gelder, 1997; Greene & Samuel, 1986; Rowe & Rowe, 1976). Consequently it is currently unclear where and how non-phonological auditory information is stored and processed in working memory. In further research, a dissociation between different auditory formats (phonological materials, environmental sounds, music) has to be more extensively investigated in order to provide working memory models that takes into account short term processing of auditory non-phonological materials and their relation with the other, more investigated, components.

Manuscript received 21 December 2008
Manuscript accepted 24 April 2009
First published online 17 June 2009

REFERENCES

Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). New York: Academic Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.

Bertelson, P., & de Gelder, B. (2004). The psychology of multi-modal perception. In J. Driver & C. Spence

(Eds.), *Cross modal space and cross modal attention*. London: Oxford University Press.
Bourassa, D., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomics Bulletin & Review*, *1*, 122–125.
Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110–1123.
Conrad, R., & Hull, AJ. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 429–432.
Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage. *Perception & Psychophysics*, *5*, 365–373.
de Gelder, B. (1997). Modality effects in immediate recall of verbal and non-verbal information. *European Journal of Cognitive Psychology*, *9*, 97–110.
Delogu, F., Auricchio, L., & Olivetti Belardinelli, M. (2006). Interazioni crossmodali semantiche nell'elaborazione del suono ambientale. *Ricerche di Psicologia*, *2*, 130–156.
Greene, R. L., & Samuel, A. G. (1986). Recency and suffix effects in serial recall of musical stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 517–524.
Jefferies, E., Frankish, C., & Lambon-Ralph, M. (2006). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *Quarterly Journal of Experimental Psychology*, *59*, 949–964.
Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioural performance. *Experimental Brain Research*, *158*, 405–414.
Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641–682.
Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, *14*, 452–465.
Murray, D. J. (1968). Articulation and acoustic confusability in short term memory. *Journal of Experimental Psychology*, *78*, 679–684.
Olivetti Belardinelli, M., Sestieri, C., Di Matteo, R., Delogu, F., Del Gratta, C., Ferretti, A., et al. (2004). Audio-visual crossmodal interactions in environmental perception: An fMRI investigation. *Cognitive Processing*, *5*, 167–174.
Paivio, A. (1986). *Mental representation: A dual coding approach*. Oxford, UK: Oxford University Press.
Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, *29*, 117–133.
Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. (2000). Integration of diverse information in working memory within the frontal lobes. *Nature Neuroscience*, *3*, 85–90.
Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, *139*, 5–21.
Rowe, E. J., & Rowe, W. G. (1976). Stimulus suffix effects with speech and nonspeech sounds. *Memory and Cognition*, *4*, 128–131.
Spence, C., & Driver, J. (2004). *Crossmodal space & crossmodal attention*. Oxford, UK: Oxford University Press.
Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *25*, 1256–1271.
Wechsler, D. (1997). *Weschsler Adult Intelligence Scale-III*. San Antonio, TX: The Psychological Corporation.

APPENDIX

The non-verbal and verbal stimuli used in the experiment

SEMANTIC CATEGORY		TYPE OF STIMULUS		
		Picture	Sound	Verbal label
	Human	hands clapping	clapping	applauso
		shivering person	shivering	brivido
		singing person	singing	canto
		teeth brushing	teeth brushing	denti
		whistling person	whistling	fischiare
		nose blowing	nose blowing	naso
		crying baby	baby crying	pianto
		yawning person	yawn	sbadiglio
		laughing person	laugh	risata
		fingers snapping	snapping fingers	schiocco
		screaming person	scream	urlo
		coughing person	coughing	tosse
		sneezing person	sneeze	starnuto
	Animal	dog	dog barking	cane
		horse	horse neigh	cavallo
		elephant	elephant trumpet	elefante
		seagull	seagull cry	gabbiano
		hen	hen chuckling	gallina
		cat	cat meowing	gatto
		wolf	wolf howling	lupo
		cow	cow mooing	mucca
		bird	bird chirping	uccello
		sheep	sheep bleating	pecora
		frog	croaking	rana
		pig	grunt	maiale
		monkey	monkey cry	scimmia
	Non living things	ambulance	ambulance siren	ambulanza
		bell	bell toll	campana
		helicopter	helicopter sound	elicottero
		chalk on blackboard	chalk on blackboard	gesso
		hammer	hammering	martello
		clock	clock ticking	orologio
		coin	spinning coin	moneta
		motorbike	motorbike roar	moto
		door	door squeak	porta
		drill	drill sound	trapano
		hand saw	hand saw sound	sega
		toilet	flushing toilet	water
		telephone	telephone ring	telefono
		train	train sound	treno