School-aged children can benefit from audiovisual semantic congruency during memory encoding

Jenni Heikkilä & Kaisa Tiippana

Volume 234 · Number 5 · May 2016

special

Experimental Brain Research

Special Issue on Multisensory Processing

Guest Editors: Durk Talsma · Jean Vroomen

Springer

Springer

Springer

CrossMark

RESEARCH ARTICLE

School-aged children can benefit from audiovisual semantic congruency during memory encoding
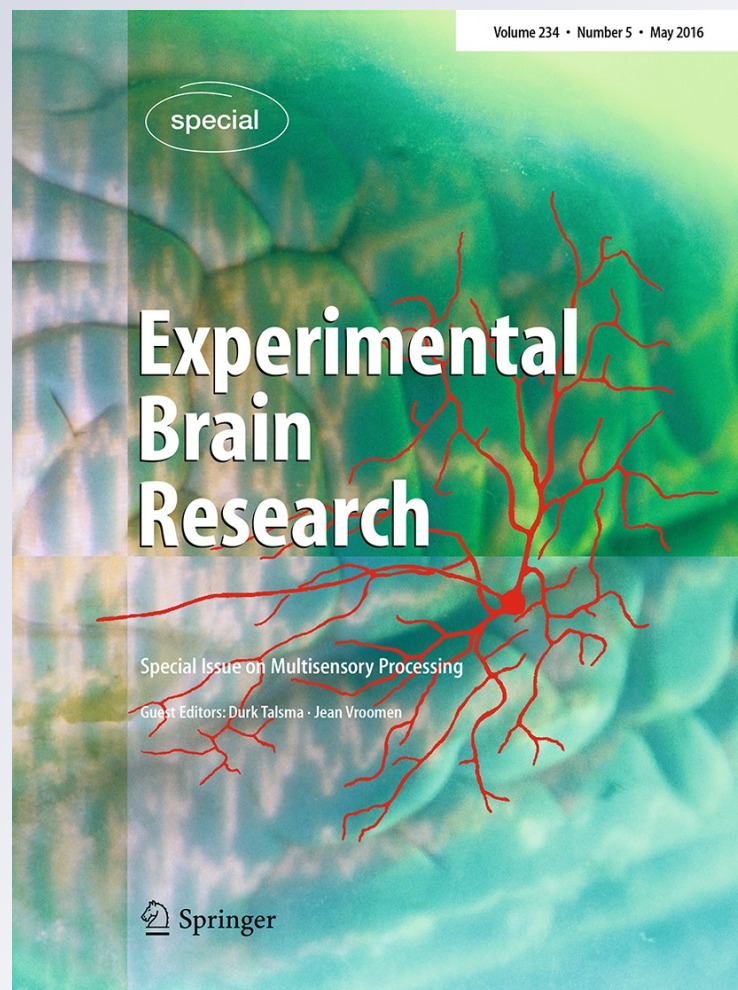
Jenni Heikkilä[1] · Kaisa Tiippana[1]

Abstract Although we live in a multisensory world, children's memory has been usually studied concentrating on only one sensory modality at a time. In this study, we investigated how audiovisual encoding affects recognition memory. Children ($n = 114$) from three age groups (8, 10 and 12 years) memorized auditory or visual stimuli presented with a semantically congruent, incongruent or non-semantic stimulus in the other modality during encoding. Subsequent recognition memory performance was better for auditory or visual stimuli initially presented together with a semantically congruent stimulus in the other modality than for stimuli accompanied by a non-semantic stimulus in the other modality. This congruency effect was observed for pictures presented with sounds, for sounds presented with pictures, for spoken words presented with pictures and for written words presented with spoken words. The present results show that semantically congruent multisensory experiences during encoding can improve memory performance in school-aged children.

Keywords Audiovisual · Children · Development · Memory · Multisensory · Semantic congruency

Introduction

Using different senses is a natural way for children to explore their environment and learn from it. For example,

a child can learn to recognize a ball by seeing it, hearing its bounce and feeling its shape. Benefits of multisensory experiences are present from very early in development, directing attention toward congruent multisensory events in the environment and guiding perceptual development (Bahrick and Lickliter 2012; Lewkowicz and Ghazanfar 2009). Almost all previous studies of memory and learning in children have focused only on a single sensory modality, yet the human brain has evolved to learn and operate in multisensory environments. The positive effects of multisensory learning have been demonstrated in adults (for a review, see Shams and Seitz 2008), but children's ability to benefit from multisensory cues in memory and learning is still a largely unexplored topic.

Integration of multisensory information is necessary in constructing meaningful representations of the environment. Adults integrate multisensory information that is temporally, spatially and semantically coherent, and such congruent multisensory information facilitates perception (Calvert et al. 2004). The capability to integrate multisensory information develops from infancy through childhood and adolescence (Bahrick and Lickliter 2012; Ernst 2008; Brandwein et al. 2011; Hillock et al. 2011; Gori et al. 2008). Different multisensory capabilities mature at different rates. Integration of visual and haptic cues matures between 8 and 10 years of age (Gori et al. 2008). Brandwein et al. (2011) demonstrated that children from 7 years of age exhibit faster reaction times for simultaneously presented audiovisual stimuli compared with unisensory stimuli. This multisensory facilitation was still immature in 8-year-olds, but had reached the mature level in 15-year-olds. In line with these behavioral findings, their electrophysiological recordings of brain activity showed systematic relationships between age and multisensory processing. They concluded that multisensory integration continues to mature through middle childhood and early adolescence.

✉ Jenni Heikkilä
jenni.heikkila@helsinki.fi

[1] Division of Cognitive Psychology and Neuropsychology, Institute of Behavioural Sciences, University of Helsinki, P.O. Box 9, 00014 Helsinki, Finland

Congruent multisensory information enables more efficient interaction with the environment. For example, when semantically congruent auditory and visual stimuli (such as a sound and a picture of an animal) are perceived simultaneously, adults recognize the stimuli more accurately and more efficiently than semantically incongruent multisensory stimuli or unisensory stimuli (Chen and Spence 2010; Molholm et al. 2004; Laurienti et al. 2004). The effect of multisensory semantic congruency on the perception of numerosity in children has been studied by Jordan and Baker (2011), who found that redundant audiovisual information improves preschoolers' numerical matching performance compared with unisensory information. When congruent numerical information was presented simultaneously both visually (series of squares) and auditorily (series of tones), children were more accurate in matching numerosities compared with auditory or visual unisensory presentation.

In addition to perceptual benefits, multisensory information can facilitate memory performance. Previously, Lehmann and Murray (2005) and Murray et al. (2004) demonstrated that in adults, memory performance of drawings of common objects was improved when the drawings were presented together with semantically congruent sounds. Conversely, Moran et al. (2013) and Thelen et al. (2015) showed similar improvement in memory performance of natural sounds presented together with semantically congruent pictures. We have previously investigated the effects of audiovisual encoding on recognition memory performance with several different stimulus combinations (Heikkilä et al. 2015). In our study, adult participants memorized auditory or visual stimuli (sounds, pictures, spoken words or written words), each of which co-occurred with either a semantically congruent or incongruent stimulus or a non-semantic stimulus in the other modality during encoding. The encoding phase was followed by a recognition memory task, in which the memorized stimuli were presented again, but now alone (unisensory presentation) and randomized among new stimuli. The participants indicated whether each stimulus had been presented in the encoding phase. The results showed that recognition memory performance was better when the memorized auditory or visual stimulus was initially presented together with a semantically congruent stimulus in the other modality than when it was accompanied by a non-semantic stimulus in the other modality. This congruency effect was observed for sounds presented together with pictures, for written words presented together with spoken words and for spoken words presented together with written words. Memory performance was slightly better when pictures were presented together with semantically congruent sounds than with non-semantic sounds, but the effect did not reach statistical significance. An interference effect

was not observed, that is, memory performance was not worse when the stimulus was initially accompanied by an incongruent stimulus in the other modality compared with a non-semantic stimulus in the other modality. There was no effect of congruency in a purely visual task (pictures presented together with written words). These results suggest that semantically congruent multisensory experiences enhance encoding of both nonverbal and verbal materials, resulting in an improvement in later recognition memory. These audiovisual congruency effects on memory may be due to the fact that congruent multisensory stimuli receive more efficient encoding (Murray et al. 2004). Multisensory presentation of a congruent stimulus during encoding leads to a more elaborate memory trace. Later, the whole multisensory memory trace can be activated by the presentation of a unisensory component of the stimulus, leading to better memory performance.

There is very scarce prior knowledge on whether semantically congruent multisensory information can benefit memory performance in children. Multisensory integration develops throughout childhood, and this may affect the way children use multisensory information in memory encoding. In adults, semantically congruent multisensory stimuli improve memory performance. However, to which extent children benefit from multisensory memory cues is an open question. In the only memory study using multisensory information we are aware of, Constantidou et al. (2011) investigated modality preferences in working memory in 7- to 13-year-old children using a free-recall paradigm. Participants studied 12-item lists of spoken words, pictures or simultaneous presentations of spoken words and pictures, followed by recall tasks. Constantidou and colleagues found that working memory performance for simultaneous presentation of spoken words and pictures was better than performance for spoken words presented alone, but not better than performance for pictures presented alone. This finding may be interpreted to mean that memory of spoken words was enhanced by the presentation of semantically congruent pictures during encoding. Alternatively, it may mean that vision was superior in this task and thus determined performance, as in the visual dominance effect (Colavita 1974; Ngo et al. 2010), which points to the dominant role of vision over audition in perceiving audiovisual events. Visual dominance effect emerges already in childhood after 7 years of age (Nava and Pavani 2013). In the extreme, in the audiovisual situation, children may have ignored the auditory information and only memorized the pictures. In contrast, in a recognition memory paradigm with a unisensory recall phase, as in Heikkilä et al. (2015), audition cannot be ignored when it is the to-be-memorized modality. That is, in the recognition memory paradigm, the task is to remember spoken words, which are presented with semantically congruent,

incongruent or non-semantic pictures during the encoding phase. During the memory phase, only spoken words are presented. If memory performance is better for those spoken words that were initially presented with semantically congruent pictures during encoding, there is a congruency effect. This result would confirm that semantically congruent pictures improve memory of spoken words. We set out to address this question as one part of the present study, in which we investigated a more extensive set of different audiovisual stimulus combinations and their effect on children's memory performance.

The aim of the present study was to investigate the following questions: (1) Do school-aged children benefit from semantically congruent audiovisual information during memory encoding? (2) Which combinations of semantically congruent information improve recognition memory performance in children? and (3) Are there developmental differences in the utilization of semantically congruent audiovisual information between the ages of 8 and 12 years?

We investigated whether recognition memory performance is better in school-aged children when a stimulus in one modality is initially presented together with a semantically congruent stimulus in the other modality than when it is initially presented together with a stimulus without semantic content in the other modality. We administered six experiments: Pictures presented with sounds (Experiment 1), sounds presented with pictures (Experiment 2), spoken words presented with pictures (Experiment 3), written words presented with spoken words (Experiment 4), spoken words presented with written words (Experiment 5) and a visual experiment in which pictures were presented with written words (Experiment 6). Experiments 1, 2, 4, 5 and 6 were identical to the experiments conducted with adults by Heikkilä et al. (2015). Experiment 3 was added to investigate the effect of pictures on memorizing spoken words. Constantidou et al. (2011) found that working memory performance was better when spoken words were presented together with pictures than when spoken words were presented alone. Our aim was to study whether this finding can be attributed to the audiovisual semantic congruency effect, as described above.

Participants were typically developing children from elementary school grades 2, 4 and 6, with average ages of 8, 10 and 12 years, respectively. We chose children from these age groups in order to investigate the developmental course of multisensory memory during middle childhood since previous research has shown multisensory integration and perceptual benefits at these ages, but sometimes not at mature levels. We expected to find congruency effect for audiovisual experiments, i.e., for pictures presented with sounds (Experiment 1), sounds presented with pictures (Experiment 2), spoken words presented with pictures

(Experiment 3), written words presented with spoken words (Experiment 4) and spoken words presented with written words (Experiment 5), but not for the unisensory situation where pictures were presented with written words (Experiment 6). However, we hypothesized that the ability to utilize semantically congruent audiovisual information for memory encoding develops together with multisensory integration and other cognitive skills, and therefore, performance might not be at mature for all tasks. In particular, if literacy skills are still not fully developed in middle childhood (Dekker et al. 2014), children may not be very efficient in utilizing written words to enhance memory encoding. In fact, the youngest, 8-year-old children of the present study, did not participate in Experiments 4–6 because reading skills are generally not yet fluent at that age in Finnish-speaking children (Torppa et al. 2010). So, an alternative hypothesis was that even though children may benefit from audiovisual semantic congruency in other experiments, congruent written words may not enhance recognition memory performance of spoken words (Experiment 5) due to developing literacy skills.

Methods

In the present study, we utilized otherwise the same experimental design as Heikkilä et al. (2015), except that in the present study an additional condition was included where spoken words were initially presented with congruent and incongruent pictures (Experiment 3).

Participants

A total of 114 children (48 boys) aged from 7 years 8 months to 13 years 3 months (mean 10 years 4 months) participated in the experiment. Forty-one of the participants were from school grade 2 and aged from 7 years 8 months to 9 years 1 month (mean 8 years 4 months). Thirty-five of the participants were from school grade 4 and aged from 9 years 9 months to 10 years 1 month (mean 10 years 4 months). Thirty-eight of the participants were from school grade 6 and aged from 11 years 10 months to 13 years 3 months (mean 12 years 5 months). All participants had Finnish as their mother tongue. According to the guardians of the participants, all participants had normal or corrected to normal vision, normal hearing, and no dyslexia or other learning disabilities or neurological disorders. The guardians were informed about the study, and they all signed permission for their child's participation. The children were informed that participation was voluntary and that they could stop their participation at any time without any consequences. The research had ethical approval from the University of Helsinki Review Board in the Humanities

and Social and Behavioural Sciences, and a research permit from the Educational Board of Espoo, Finland.

Stimuli

Pictures of natural objects, sounds of natural objects, written words, spoken words, visual noise and auditory noise were used as stimulus material.

The pictures were 272 photographs obtained either from the Multimodal Stimulus Set (Schneider et al. 2008) or from Internet and modified to resemble those in the Multimodal Stimulus Set. The photographs presented common objects from several semantic categories, for example, animals, tools, vehicles, musical instruments and household items. They were converted into grayscale images and presented on a black background in the center of a laptop screen. The images were presented for 400 ms in Experiments 1 and 2 and 800 ms in Experiments 3 and 6. The sounds were 136 recordings of complex sounds from several semantic categories, for example, animals, tools, vehicles, musical instruments and household items. The sounds were either obtained from the Multimodal Stimulus Set (Schneider et al. 2008) or selected from Internet and modified to resemble those from the Multimodal Stimulus Set. The sound duration was 400 ms. The written words were 170 common two-syllable Finnish nouns from several categories. They were presented for 800 ms in white 40-point Times New Roman font in the center of the laptop monitor. The spoken words were 238 common Finnish nouns from several categories. The duration of the spoken words was between 350 ms and 780 ms. The auditory non-semantic stimulus was a burst of white noise. Its duration was 400 ms in Experiment 1 and 800 ms in Experiment 4. The visual non-semantic stimulus was a frame of white noise presented for 400 or 800 ms (Exps. 2 and 3) or a row of Xs presented for 800 ms (Exps. 5 and 6). The auditory stimuli were presented with headphones at an intensity of approximately 55 dB(A). All items were presented in random order with simultaneous onsets of the auditory and visual components. All stimuli were different across all experiments; for example, "dog" was presented only once (either as a word, sound or picture) during the series of experiments. For further details, see Heikkilä et al. (2015).

Design and procedure

The study included six experiments presented to the 10-year-old and 12-year-old participants. For the 8-year-old participants, only the three experiments that did not require reading skills (Experiments 1, 2 and 3) were presented.

All experiments consisted of two parts: (1) an encoding phase consisting of audiovisual items, followed by (2) a recognition memory task consisting of unisensory items.

During encoding, the children were instructed to memorize the stimuli in one modality while ignoring the stimuli in the other modality. There were three congruency conditions for the audiovisual items: congruent, incongruent and non-semantic. The congruent items were semantically matching: for example, a picture of a horse and a horse's hee-haw, or the same word (e.g., "cheese") written and spoken. The incongruent items were semantically mismatching: for example, a picture of a wolf and the sound of a violin. The non-semantic items consisted of the semantic stimulus to be memorized in one modality presented together with a non-semantic stimulus in the other modality. A total of 51 items were memorized in each experiment, with 17 items in each congruency condition.

The recognition memory task was presented immediately after the encoding phase. Here, the memorized stimuli were presented again, but now alone and intermingled with an equal number of new stimuli. The task was to decide whether or not each stimulus had been presented in the encoding phase or not by pressing one of two keys on the keyboard.

Experiment 1: Pictures with sounds

In the encoding phase, pictures were paired with congruent or incongruent natural sounds, or bursts of noise. The task was to memorize the pictures, while ignoring the sounds. In the visual recognition memory phase, the child decided whether each picture had been presented in the encoding phase.

Experiment 2: Sounds with pictures

In the encoding phase, sounds of natural objects were paired with congruent or incongruent pictures or visual noise. The task was to memorize the sounds, while ignoring the pictures. In the auditory recognition memory phase, the child decided whether each sound had been presented in the encoding phase.

Experiment 3: Spoken words with pictures

In the encoding phase, spoken words were paired with semantically congruent or incongruent pictures or visual noise. The task was to memorize the spoken words, while ignoring the pictures. In the auditory recognition memory phase, the child decided whether each spoken word had been presented in the encoding phase.

Experiment 4: Written words with spoken words

In the encoding phase, written words were paired with congruent or incongruent spoken words, or noise bursts. The

task was to memorize the written words, while ignoring the auditory stimuli. In the visual recognition memory phase, the child decided whether each written word had been presented in the encoding phase.

Experiment 5: Spoken words with written words

In the encoding phase, spoken words were paired with congruent, incongruent or non-semantic written words. The children's task was to memorize the spoken words, while ignoring the written words. In the auditory recognition memory phase, the child decided whether each spoken word had been presented in the encoding phase.

Experiment 6: Pictures with written words

In the encoding phase of this unisensory visual experiment, pictures were paired with congruent, incongruent or non-semantic written words. The task was to memorize the pictures, while ignoring the written words. In the recognition memory phase, the child decided whether each picture had been presented in the encoding phase.

The experiments were conducted in two elementary schools in Espoo, Finland. Each child was tested individually in a private room during school days. The experimenter (the first author) sat next to the participant to ensure proper concentration on the task and that the participant maintained the gaze on the screen of the laptop computer. The participants sat in a chair about 50 cm from the laptop. They responded by pressing two marked keys on the keyboard. Before the experiment, they received oral instructions and performed a short practice session. The six experiments (lasting about 7 min each) were presented to 10- and 12-year-old participants in random order. Most of the 8-year-old participants started with Experiment 1, and then, Experiments 2 and 3 were presented in random order. This procedure was used because Experiment 1 was the easiest to understand for young children.

The percentages of correct responses[1] were analyzed with analyses of variance (ANOVA) including Congruency and Experiment as repeated-measures factors and Age as a between-subjects factor. The significance level was set to $p < .05$. Greenhouse–Geisser correction was applied for p

---

[1] The sensitivity index (d') was not used here since the results would remain exactly the same with this transformation and d' would bring no additional information compared with the percentage correct. This is because there was a single recall phase consisting of the memorized items for all congruency categories and new items. Consequently, the false alarm rates (calculated on the basis of yes responses to new items) were equal for all congruency categories, and therefore, the d' values would be in direct proportion to the percentage correct (hit) rates.

Table 1 Mean percentages of correct responses (and standard deviations) in Experiments 1–6 separately for each age group

|	8-year-olds	10-year-olds	12-year-olds
Exp. 1 Congruent	75.88 (13.76)	73.41 (14.59)	75.06 (16.50)
Exp. 1 Incongruent	74.41 (12.82)	74.94 (12.82)	74.00 (15.32)
Exp. 1 Non-semantic	67.24 (16.05)	73.76 (15.59)	71.06 (15.82)
Exp. 2 Congruent	65.12 (14.70)	61.64 (13.23)	57.74 (13.60)
Exp. 2 Incongruent	53.94 (18.29)	57.47 (15.47)	50.92 (11.52)
Exp. 2 Non-semantic	56.24 (19.47)	51.76 (15.00)	50.15 (14.23)
Exp. 3 Congruent	79.76 (13.34)	78.82 (15.82)	73.82 (14.58)
Exp. 3 Incongruent	67.12 (16.64)	70.23 (15.70)	62.54 (18.70)
Exp. 3 Non-semantic	67.15 (16.88)	61.65 (15.47)	54.02 (15.23)
Exp. 4 Congruent		71.41 (13.35)	66.25 (17.22)
Exp. 4 Incongruent		64.00 (13.70)	58.17 (15.47)
Exp. 4 Non-semantic		62.18 (17.18)	55.88 (12.74)
Exp. 5 Congruent		63.00 (13.82)	58.94 (12.76)
Exp. 5 Incongruent		60.00 (15.23)	56.64 (15.60)
Exp. 5 Non-semantic		57.64 (19.35)	56.80 (15.64)
Exp. 6 Congruent		81.47 (13.11)	82.35 (15.22)
Exp. 6 Incongruent		76.12 (15.29)	80.29 (12.76)
Exp. 6 Non-semantic		78.84 (14.41)	79.41 (13.31)

values when appropriate. The original degrees of freedom are reported with F values. Bonferroni-corrected significance levels were used in pairwise comparisons.

Results

The recognition memory performance scores (percentages of correct responses) for stimuli that were initially presented with a semantically congruent, incongruent or non-semantic stimulus during encoding in Experiments 1–6 are presented in Table 1.

The effect of audiovisual semantic congruency was statistically tested first with an ANOVA including all six experiments. We expected to find a main effect of congruency, and interaction between congruency and experiment, suggesting that the congruency effect occurs in some of the experiments, but not in all of them. We expected to see the congruency effect in audiovisual experiments (1–5) but not in the unisensory visual experiment (6). The ANOVA with Congruency (congruent, incongruent, non-semantic), Experiment (1–6) and Age (10, 12) as factors revealed significant main effects of Congruency [$F(2,140) = 50.2$, $p < .001$, $\eta^2 = .418$] and Experiment [$F(5,350) = 67.4$, $p < .001$, $\eta^2 = .490$], and interaction between Congruency and Experiment [$F(10,700) = 6.18$, $p < .001$, $\eta^2 = .081$], suggesting that there were differences between the congruency conditions and experiments but not between the age groups. In the second ANOVA, the focus was more on the

Fig. 1 Mean percentages of correct responses in the recognition memory task for stimuli presented with congruent, incongruent or non-semantic stimuli during encoding, calculated across age groups. Experiment 1 (*Exp. 1*): recognition memory of pictures presented together with sounds during encoding (all age groups). Experiment 2 (*Exp. 2*): recognition memory of sounds presented together with pictures during encoding (all age groups). Experiment 3 (*Exp. 3*): recognition memory of spoken words presented together with pictures during encoding (all age groups). Experiment 4 (*Exp. 4*): recognition memory of written words presented together with spoken words during encoding (10- and 12-year-olds). Experiment 5 (*Exp. 5*): recognition memory of spoken words presented together with written words during encoding (10- and 12-year-olds). Experiment 6 (*Exp. 6*): recognition memory of pictures presented together with written words during encoding (10- and 12-year-olds). *Asterisks* indicate statistically significant congruency effects (*p* < .05). *Error bars* depict the standard error of the mean
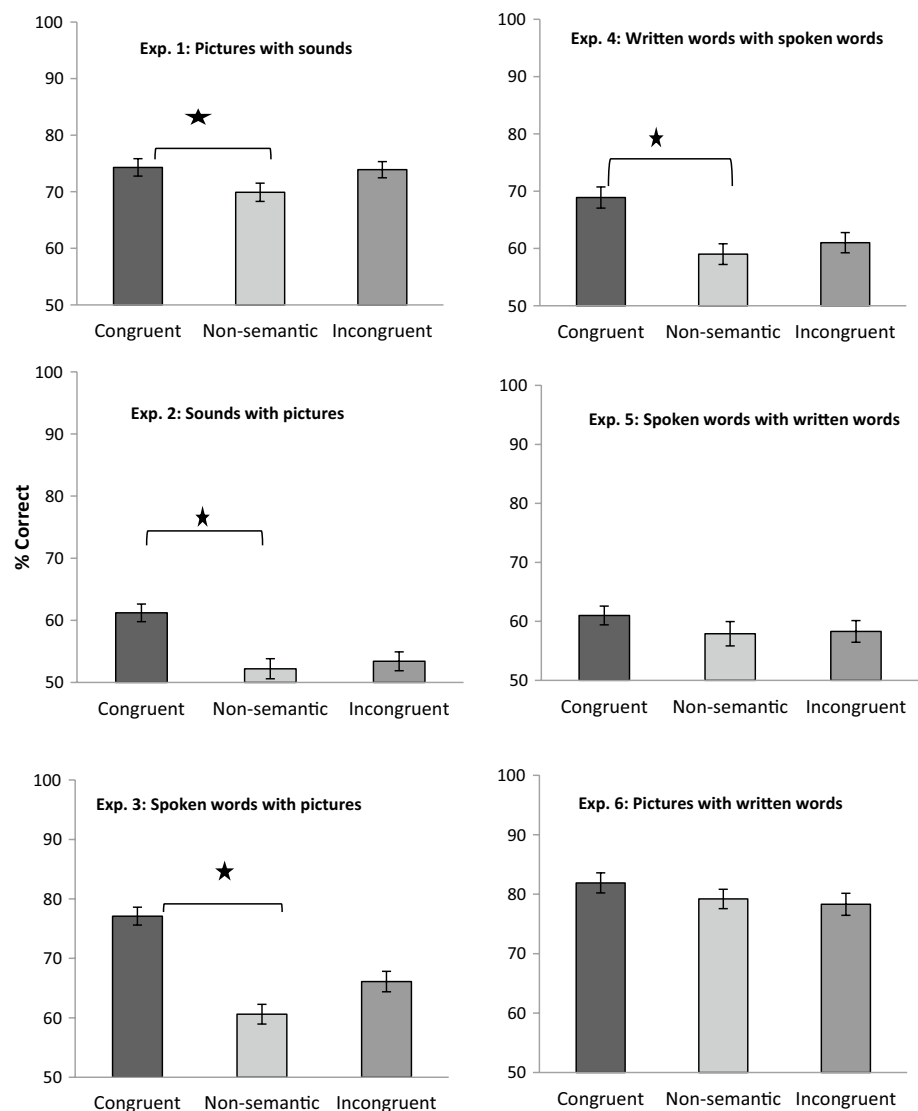
effect of age, so that all three age groups were included for those experiments that were done by all children. The second ANOVA with Congruency (congruent, incongruent, non-semantic), Experiment (1, 2, 3) and Age (8, 10, 12) as factors showed the same significant main effects of Congruency [$F(2,222) = 70.1$, $p < .001$, $\eta^2 = .387$] and Experiment [$F(2,222) = 85.6$, $p < .001$, $\eta^2 = .435$], and interaction between Congruency and Experiment [$F(4,444) = 12.2$, $p < .001$, $\eta^2 = .099$]. In addition, there was a three-way interaction between Congruency, Experiment and Age [$F(8, 444) = 2.16$, $p = .032$, $\eta^2 = .038$]. This was the only significant effect where age was involved. It was investigated further by running separate ANOVAs for Experiments 1, 2 and 3 with Age and Congruency as factors. No significant interactions of Age and Congruency were found in any of the experiments [$p > .05$ in each case], suggesting that the children's age did not affect the way semantic congruency influenced memory performance.

Since the factor Age had no significant effect, the data from all age groups were combined to investigate the effect of congruency in more detail. Figure 1 shows the averaged data for the three congruency conditions in each experiment. Post hoc tests for the Congruency × Experiment interaction revealed that Experiments 1 (pictures with sounds), 2 (sounds with pictures), 3 (spoken words with pictures) and 4 (written words with spoken words) showed congruency effects, while Experiments 5 (spoken words with written words) and 6 (pictures with written words) did not. That is, when the stimuli were presented together with semantically congruent stimuli in the other modality in the encoding phase, memory performance improved compared with a non-semantic stimulus presentation for pictures presented together with sounds [$t(113) = 3.32$, $p = .012$], sounds presented together with pictures [$t(113) = 5.50$, $p < .001$], written words presented together with spoken words [$t(71) = 4.87$, $p < .001$] and spoken words

presented together with pictures [$t(113) = 11.8\ p < .001$]. On the other hand, the congruency effect was not observed for spoken words presented together with written words and pictures presented together with written words in the encoding phase [$p > .05$ for both comparisons]. There was no interference effect in any of the experiments, that is, when the stimuli were presented together with semantically incongruent stimuli in the encoding phase, memory performance was not poorer compared with a non-semantic stimulus presentation in any of the experiments [$p > .05$ in each case]. However, in Experiment 3 memory performance for incongruent stimulus presentation was better compared with a non-semantic stimulus presentation [$t(113) = 3.94$, $p = .001$].

Discussion

The present study shows that school-aged children can benefit from audiovisual semantic congruency during memory encoding. Children's recognition memory was better for pictures that were initially presented with semantically congruent sounds compared with sounds that were presented with non-semantic auditory stimuli, as well as sounds that were presented with semantically congruent pictures compared with sounds that were presented with non-semantic visual stimuli (Experiments 1 and 2). Similarly, when spoken words were presented together with semantically congruent pictures during encoding, children's memory of these words was better compared with words that had been initially presented with non-semantic visual stimuli (Experiment 3). These effects were found in children of all age groups. A congruency effect was also seen in 10- and 12-year-olds for written words, which were remembered better when initially presented together with a congruent spoken word than when initially presented together with a non-semantic auditory stimulus (Experiment 4). In contrast, written words did not facilitate the memory of spoken words (Experiment 5). Finally, in the unisensory visual experiment (Experiment 6), written words did not facilitate the recognition memory of pictures.

When the stimuli were accompanied by incongruent stimuli during encoding, there was no interference effect, that is, memory performance was not poorer than for non-semantic trials in any of the experiments. In the non-semantic trials, meaningless auditory or visual noise was presented in the other modality, so that these were not purely unisensory trials. The non-semantic stimuli might cause some interference compared with purely unisensory stimuli. However, some studies have shown that memory performance is similar for semantically incongruent and unisensory stimuli (Lehmann and Murray 2005; Moran et al. 2013), suggesting that unisensory and non-semantic

noise stimuli may be equivalent in that respect. Furthermore, in real learning situations, for example, in a classroom, background noise is usually present, so that the non-semantic noise trials resemble real learning conditions more than purely unisensory trials.

There were no age-related differences in the utilization of audiovisual information when comparing different age groups in the current study. Children aged 8, 10 and 12 years showed congruency effects in Experiments 1, 2 and 3 without any differences between age groups. Similarly, there were no age differences between 10- and 12-year-old children in the congruency effect in Experiment 4, either. On the other hand, neither 10- nor 12-year-olds benefitted from the congruent presentation of written words during the encoding of spoken words (Experiment 5). This might be due to the fact that in 10–12-year-olds literacy skills have not yet been automatized, which may weaken the ability to integrate congruent written information in order to form a stronger memory trace. Children may rely more on the spoken word, resulting in no benefit from the written word to encoding. Usually Finnish-speaking children acquire literacy skills and understand grapheme–phoneme correspondence during the first few school years, around 7–9 years of age (Torppa et al. 2010), but the automatic use of written language may not be fully developed during the primary school years.

Note that 8-year-old children did not participate in all experiments, and it may be expected that age differences could be found with written stimuli. Furthermore, a wider age range may have revealed some developmental differences. For example, preschool children may have a weaker ability to use multisensory information; meanwhile, adolescents may be more adult-like in their performance, especially for spoken words accompanied by written words. Further studies are needed to address this issue.

Taken together, the present findings provide the first evidence that audiovisual semantic congruency can facilitate recognition memory performance not only in adults but also in school-aged children. It has been shown previously that multisensory perceptual integration occurs in children (e.g., Ernst 2008). The current study indicates that multisensory presentation of semantically congruent information can benefit memory encoding in middle childhood. This suggests that children can establish the semantic correspondence between auditory and visual signals and integrate the information in order to encode it as a meaningful whole. Heikkilä et al. (2015) discussed the congruency effect in the light of Potter's (1993, 2012) conceptual short-term memory (CSTM), a memory buffer that rapidly integrates new perceptual information with information held in long-term memory and creates new representations that can be encoded into long-term memory. Instead of being only unisensory construct, the CSTM may integrate also

audiovisual information (Chen and Spence 2010; Heikkilä et al. 2015). Here, we suggest that also in children, semantically congruent audiovisual information leads to more effective encoding of the semantic meaning of the stimulus, causing greater signal strength by binding the congruent information and leading to a richer memory trace. The whole memory trace can be later activated by one component of the stimulus, resulting in improved memory performance.

Our results show that semantically congruent pictures enhanced the recognition memory of spoken words in children. In a similar vein, Constantidou et al. (2011) found that pictures improved the working memory performance of spoken words compared with words-only presentation, but spoken words did not improve the working memory performance for pictures compared with pictures-only presentation. Visual dominance emerges in multisensory perceptual tasks after about 7 years of age (Nava and Pavani 2013). This visual dominance could account for the findings in Experiments 2 and 3, i.e., that pictures facilitate the memory encoding of sounds and spoken words, as well as for the findings of Constantidou and colleagues. In contrast, the finding that written words did not improve memory of spoken words in children is contrary to a general visual dominance account. Instead, the more informative and reliable modality generally dominates (Ernst and Bülthoff 2004), and often it is vision. In children, however, processing of written words is still inefficient compared with spoken language, and therefore, visual influence does not emerge in this case.

We have shown that pictures facilitate the recognition memory of both verbal and nonverbal auditory information, sounds facilitate the memory of pictures and spoken words facilitate the memory of written words in school-aged children. It could be speculated that in educational settings, it might be beneficial to use multisensory semantically congruent information to strengthen the learning results, especially in situations that require rote memory learning (such as foreign languages). Multisensory information might also be useful in education and rehabilitation of children with learning disabilities. This topic should be addressed in future studies.

In conclusion, the current study shows for the first time that semantically congruent audiovisual information can enhance memory performance in school-aged children. When a stimulus is presented together with its congruent counterpart from the other sensory modality during encoding, recognition memory performance is improved.

References

Bahrick L, Lickliter R (2012) The role of intersensory redundancy in early perceptual, cognitive and social development. In: Bremner A, Lewkowich D, Spence C (eds) Multisensory development. Oxford University Press, UK, pp 183–206

Brandwein A, Foxe J, Russo N, Altschuler T, Gomes H, Molholm S (2011) The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. Cereb Cortex 21:1042–1055

Calvert G, Spence C, Stein B (2004) The Handbook of Multisensory Processes. The MIT Press, London

Chen Y, 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

Colavita F (1974) Human sensory dominance. Percept Psychophys 25:345–347

Constantidou F, Danos M, Nelson D, Baker S (2011) Effects of modality presentation of working memory in school-age children: evidence for the pictorial superiority hypothesis. Child Neuropsychol 17:173–196

Dekker TM, Mareschal D, Johnson MH, Sereno MI (2014) Picturing words? Sensorimotor cortex activation for printed words in child and adult readers. Brain Lang 139:58–67

Ernst M (2008) Multisensory integration: a late bloomer. Curr Biol 18:519–521

Ernst MO, Bülthoff HH (2004) Merging the senses into a robust percept. Trends Cogn Sci 8:162–169

Gori M, Del Viva M, Sandini G, Burr D (2008) Young children do not integrate visual and haptic information. Curr Biol 18:689–693

Heikkilä J, Alho K, Hyvönen H, Tiippana K (2015) Audiovisual semantic congruency during encoding enhances memory performance. Exp Psychol 62:123–130

Hillock A, Powers A, Wallace M (2011) Binding sights and sounds: age-related changes in multisensory temporal processing. Neuropsychologia 49:461–467

Jordan K, Baker J (2011) Multisensory information boosts numerical matching abilities in young children. Dev Sci 14:205–213

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

Lehmann S, Murray MM (2005) The role of multisensory memories in unisensory object discrimination. Cogn Brain Res 24:326–334

Lewkowicz DJ, Ghazanfar AA (2009) The emergence of multisensory systems through perceptual narrowing. Trends Cogn Sci 13:470–478

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

Moran Z, Bachman P, Pham P, Cho S, Cannon T, Shams L (2013) Multisensory encoding improves auditory recognition. Multisens Res 26:581–592

Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schnider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage 21:125–135

Nava E, Pavani F (2013) Changes in sensory dominance during childhood: converging evidence from the Colavita effect and the sound-induced flash illusion. Child Dev 84:604–616

Ngo M, Sinnet S, Soto-Faraco S, Spence C (2010) Repetition blindness and the Colavita effect. Neurosci Lett 480:186–190

Potter MC (1993) Very short-term conceptual memory. Mem Cogn 21:156–161

Potter MC (2012) Conceptual short term memory in perception and thought. Front Psychol 3:1–11

Schneider TR, Engel AK, Debener S (2008) Multisensory identification of natural objects in a two-way crossmodal priming paradigm. Exp Psychol 55:121–132

Shams L, Seitz AR (2008) Benefits of multisensory learning. Trends Cogn Sci 12:411–417

Thelen A, Talsma D, Murray M (2015) Single-trial multisensory memories affect later auditory and visual object discrimination. Cognition 138:148–160

Torppa M, Lyytinen P, Erskine J, Eklund K, Lyytinen H (2010) Language development, literacy skills and predictive connections to reading in Finnish children with and without familiar risk for dyslexia. J Learn Disabil 43:308–321