Department of Psychology and Logopedics
Faculty of Medicine
University of Helsinki
Finland

Benefits of audiovisual memory encoding across the life span

Jenni Heikkilä

ACADEMIC DISSERTATION

Academic dissertation to be publicly discussed,
by due permission of the Faculty of Medicine
at the University of Helsinki in Auditorium XII at the Main building, Fabianinkatu 33
on the 17th of September, 2018, at 12 o'clock

ISBN 978-951-51-4443-0 (pbk.)

ISBN 978-951-51-4444-7 (PDF)

Unigrafia

Helsinki 2018

Supervisors

Docent Kaisa Tiippana
Department of Psychology and Logopedics
Faculty of Medicine
University of Helsinki
Finland

Professor Kimmo Alho
Department of Psychology and Logopedics
Faculty of Medicine
University of Helsinki
Finland

Reviewers

Professor Jari Hietanen
Department of Psychology
University of Tampere
Finland

Associate Professor Antonino Raffone
Department of Psychology
Sapienza University of Rome
Italy

Opponent

Professor Fiona Newell
School of Psychology
Trinity College Dublin
Ireland

CONTENTS

Abstract.....6
Tiivistelmä.....7
Acknowledgements.....8
List of original publications.....10
1. Introduction.....11
1.1. Benefits of audiovisual perception.....12
1.2. The development of audiovisual perception in childhood.....13
1.3. Changes in audiovisual processing in older age.....15
1.4. Semantically congruent audiovisual contexts can enhance memory performance.....16
1.4.1. Semantically congruent audiovisual stimuli improve recognition memory.....16
1.4.2. Audiovisual context and working memory.....17
1.4.3. Audiovisual memory across the life span.....18
2. The aim of the present thesis.....20
3. General methods.....24
3.1. Stimuli.....24
3.2. Recognition memory tasks.....25
3.3. Statistical methods.....26
3.3.1. Signal detection theory.....26
3.3.2. Statistical analyses.....27
4. Specific methods and results of Studies I – IV.....28
4.1. Study I: Audiovisual semantic congruency enhances memory performance in young adults.....28
4.1.1. Introduction.....28
4.1.2. Specific methods.....28
4.1.3. Results.....30
4.2. Study II: Semantically congruent visual stimuli improve auditory memory performance in young adults.....32
4.2.1. Introduction.....32

4.2.2. Specific methods....32
4.2.3. Results....33
4.3. Study III: Audiovisual semantic congruency enhances memory performance in children....36
4.3.1. Introduction....36
4.3.2. Specific methods....36
4.3.3. Results....37
4.4. Study IV: Semantically congruent visual stimuli improve auditory memory performance in elderly people....40
4.4.1. Introduction....40
4.4.2. Specific methods....41
4.4.3. Results....41
5. General discussion....44
5.1. The benefits of audiovisual memory encoding....44
5.2. Conceptual Short Term Memory model....46
5.3. Age-related differences in the congruency effect....47
5.4. Future directions....49
6. Conclusions....51
7. References....52

ABSTRACT

Although we live in a multisensory world, human memory has been traditionally studied concentrating on just one sensory modality, for example, either audition or vision. Yet, some previous studies have shown better memory performance for audiovisual information than for unisensory information. However, such studies are scarce and they have mainly focused on young adults. In the present series of studies, the effects of audiovisual encoding on later unisensory recognition memory performance were studied in children, young adults, and elderly people. In Studies I and II, these effects were studied in younger adults using both verbal and non-verbal audiovisual stimuli. Study III, in turn, investigated how audiovisual encoding affects recognition memory in school-aged children (mean age 10 years 4 months). Finally, Study IV compared how audiovisual encoding affects auditory recognition memory in elderly people (mean age 71 years) and young adults (mean age 23 years). Overall, recognition memory performance was better in all age groups when the stimulus to be memorized was initially accompanied by a semantically congruent stimulus in the other modality than when it was accompanied by a non-semantic stimulus in the other modality or by no stimulus. Altogether, the results of the present series of studies suggest that semantically congruent audiovisual experiences enhance memory encoding not only in young adults but also in children and elderly people. These results might be useful when developing educational practices for children and young adults, as well as when designing practical applications to alleviate mild memory problems due to normal aging.

TIIVISTELMÄ

Muistia on perinteisesti tutkittu yhden aistin, joko näön tai kuulon, avulla, vaikka elämme moniaistisessa ympäristössä. On jonkin verran aiempaa näyttöä siitä, että muistaminen on parempaa, kun mieleen painettava tieto on yksiaistisen sijaan audiovisuaalista. Tutkimukset aiheesta ovat kuitenkin harvalukuisia, ja keskittyneet pääosin vain nuoriin aikuisiin. Tässä väitöskirjassa tutkittiin audiovisuaalisen mieleen painamisen vaikutusta yksiaistiseen tunnistusmuistiin lapsilla, nuorilla aikuisilla sekä ikääntyvillä. Väitöskirjan ensimmäisessä ja toisessa osatyössä ilmiötä tutkittiin nuorilla aikuisilla käyttäen sekä kielellisiä että ei-kielellisiä ärsykkeitä. Väitöskirjan kolmannessa osatyössä tutkittiin, miten audiovisuaalinen mieleen painaminen vaikuttaa tunnistusmuistiin alakouluikäisillä lapsilla (keski-ikä 10 v 4 kk). Väitöskirjan neljännessä osatyössä verrattiin audiovisuaalisen mieleen painamisen vaikutusta auditoriseen tunnistusmuistiin ikääntyvillä (keski-ikä 71 vuotta) ja nuorilla aikuisilla (keski-ikä 23 vuotta). Suoritus tunnistusmuistitehtävissä oli kaikilla ikäryhmillä parempi, kun mieleen painettavan ärsykkeen kanssa esitettiin merkitykseltään yhtenevä ärsyke toiselta aistikanavalta verrattuna merkityksettömän ärsykkeen esittämiseen tai vain yksiaistiseen esittämiseen. Väitöskirjatutkimuksen tulokset osoittavat, että merkitykseltään yhtenevät audiovisuaaliset ärsykkeet parantavat mieleen painamista nuorten aikuisten lisäksi myös lapsilla ja ikääntyvillä. Tutkimuksista saadun tiedon avulla voidaan mahdollisesti kehittää uusia menetelmiä lasten kouluoppimisen ja nuorten opiskelun tukemiseen, ja myös ikääntymisen aiheuttamaa muistin heikkenemistä voidaan mahdollisesti kompensoida audiovisuaalisen tiedon avulla.

ACKNOWLEDGEMENTS

I am grateful to many people who made this thesis possible. First and foremost, I want to express my deepest gratitude to my supervisors Docent Kaisa Tiippana and Professor Kimmo Alho for their excellent supervision and support throughout my postgraduate studies. I am grateful that I have had the opportunity to work in and be funded by the Multisensory Learning-project (led by Kaisa Tiippana) at the University of Helsinki. I am deeply grateful for the guidance, encouragement and never-ending support that I have received, while working in the project.

I thank my official reviewers Professor Jari Hietanen and Dr. Antonino Raffone for their valuable comments on the manuscript of this thesis. I am grateful to Professor Fiona Newell for agreeing to act as the opponent at the defence of this dissertation.

I am grateful to the funding agencies that made this thesis possible: the University of Helsinki, Arvo and Lea Ylppö Foundation, Avohoidon Tutkimussäätiö Foundation, Emil Aaltonen Foundation, Finnish Brain Foundation, Finnish Concordia Fund and Otto Malm Foundation.

I thank Heidi Hyvönen for her help in the data collection of Study I. I want to thank the children and teachers of Iivisniemi elementary school and Tähtiniitty elementary school, Espoo, Finland, for participating in the Study III. I also thank Petra Fagerlund for collaboration and data collection for the Study IV. I am indebted to Miika Leminen, Tommi Makkonen and Kalevi Reinikainen for their help in technical matters.

It has been a pleasure to work as a part of the Perception Action Cognition-research group and I want to thank Markku Kilpeläinen, Tarja Peromaa, Lari Vainio, Viljami Salmela, Mikko Tiainen and Ilmari Kurki for their collegial support. I also want to thank Kaisa Kanerva and Kaisa Kaseva for their collegial support and friendship. I am grateful to the members of the NuTu

(the Young Researchers' Division of the Finnish Psychological Society) for their peer support.

I thank Leena and Pekka Heikkilä, Antti Heikkilä, Anna-Liisa Malinen, and my extended family for encouragement and support. Finally, thank you Tuomas Malinen for encouraging me, believing in me, and being by my side.

LIST OF ORIGINAL PUBLICATIONS

I Heikkilä, J., Alho, K., Hyvönen, H. & Tiippana, K. (2015). Audiovisual semantic congruency during encoding enhances memory performance. *Experimental Psychology,* 62, 123–30.

II Heikkilä, J., Alho, K. & Tiippana, K. (2017). Semantically congruent visual stimuli can improve auditory memory. *Multisensory Research,* 30, 639–51.

III Heikkilä, J. & Tiippana, K. (2016). School-aged children can benefit from audiovisual semantic congruency during memory encoding. *Experimental Brain Research*, 234, 1199–207.

IV Heikkilä, J., Fagerlund, P. & Tiippana, K. (2018). Semantically congruent visual information can improve auditory recognition memory in older adults. *Multisensory Research* 31, 213–25.

The articles are re-printed according to the guidelines of the copyright holders. The original published versions can be found from the respective journals.

1. INTRODUCTION

In everyday life, most situations contain both auditory and visual elements. People integrate auditory and visual information that is spatially, temporally and semantically coherent. For example, when we are talking with someone, we usually hear the speaker's voice and see the lip movements simultaneously. Brain mechanisms have evolved to receive audiovisual environmental cues and combine them into coherent percepts. Coherent audiovisual information facilitates our ability to perceive the world around us (Calvert, Spence, & Stein, 2004). In memory studies, however, the focus has been mostly on unisensory stimuli. Moreover, although it has been demonstrated that audiovisual information presented during encoding can enhance memory in young adults (Murray et al. 2004, 2005; Lehmann & Murray, 2005; Moran et al. 2013; Thelen, Talsma, & Murray, 2015; Ueno et al., 2015), little is known about benefits of audiovisual encoding in children or elderly people. Yet the topic is relevant both in children, who are in the middle of their basic education, and in elderly people, whose memory processes are often declining due to the aging (Old & Naveh-Benjamin, 2008).

The idea that encoding affects retrieval from memory has been popular throughout the history of cognitive psychology. In as early as 1890, it was suggested that remembering involved the same sensory and motor components and the same brain areas as encoding (James, 1890, pp. 653–689). Later, it was proposed that memory representations are stored in the same neural ensembles where activity occurred during perception (Damasio, 1998). More recently, it has been suggested that memories of past episodes rely on perceptual representations and that multisensory neural simulations of previous episodes are essential to their retrieval (Barsalou, 2008). Evidence from brain imaging studies supports the idea that many brain regions involved in memory encoding are also activated during subsequent retrieval (Wheeler, Petersen & Buckner, 2000; Nyberg et al., 2000, Vaidya, Zhao, Desmond & Gabrieli, 2001), yet there is asymmetry between left and right frontal brain

areas activated in memory encoding and memory retrieval (for a review, see Habib, Nyberg & Tulving, 2003).

Nyberg and collegues (2000) suggested that the sensory properties of audiovisual events are stored in the cortical regions that are activated during encoding and that these regions are reactivated during retrieval. They discussed their findings in terms of redintegration, to mean reactivation of the whole representation of an event upon the occurrence of its encoded component. On this account, the auditory and visual components of an audiovisual stimulus are combined into a single representation in the encoding process, and this representation is later activated by either unisensory component. Consequently, many brain regions involved in audiovisual encoding are also activated during unisensory retrieval.

1.1. Benefits of audiovisual perception

Audiovisual interactions are often automatic and can occur in our nervous system at early stages of perceptual processing (Ghazanfar & Schroeder, 2006). Several factors, such as temporal and spatial correspondence, facilitate the integration between auditory and visual inputs. This integration can lead to facilitation, for example when two stimuli of different sensory modalities coincide at the same location, they can be detected faster and more accurately than either stimulus occurring alone (e.g., Morrell, 1968; Simon & Craft, 1970; Soto-Faraco, Kingstone, & Spence, 2003). Perceptual facilitation caused by audiovisual inputs is also influenced by stimulus congruency, which is related to the similarity of information from the modalities. Shams and Seitz (2008) define congruency as "the relationship between the stimuli that are consistent with the prior experience of the individual or the relationship between the senses found in nature." For example, speech perception is improved when the acoustic speech signal is observed together with the speaker's facial movements (Sumby & Pollack, 1954; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007).

Congruency can occur between higher-level features, for example between the semantic content of meaningful auditory and visual stimuli, such as a sound and picture of an animal. There has been interest in the effect of semantic congruency on audiovisual perception. This is typically studied by contrasting congruent (a visual stimulus matched with its auditory counterpart), incongruent (semantically unrelated auditory and visual stimuli), and unisensory (auditory or visual) stimulus conditions. When semantically congruent auditory and visual stimuli are presented simultaneously, participants usually recognize the stimuli faster (Molholm et al., 2004, Laurienti et al., 2004; Suied, Bonneel, & Viaud-Delmon, 2009) and more accurately (Molholm et al., 2004, Laurienti et al., 2004; Chen & Spence, 2010) than incongruent stimulus pairs or just unisensory stimuli. However, when two congruent features of the same stimulus are presented simultaneously only in the visual modality, facilitation is not observed (Laurienti et al., 2004). This suggests that perceptual facilitation caused by stimulus congruency resulted from audiovisual interaction rather than from additional sensory information.

1.2. The development of audiovisual perception in childhood

The ability to integrate information from different senses develops from infancy through childhood and adolescence (Ernst, 2008; Gori, Del Viva, Sandini & Burr, 2008; Brandvein et al., 2010; Hillock, Powers & Wallace, 2011; Bahrick & Lickliter, 2012). The ability to integrate auditory and visual speech signals (speech-related facial movements) starts to develop in infancy. Already infants can detect temporal synchrony of lip movements and speech sounds (Dodd, 1979) and match auditory and visual vowels (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999). Still, children are less sensitive to audiovisual speech than adults, and the development of audiovisual speech integration continues to adolescence and adulthood (McGurk & MacDonald, 1976; Massaro, 1984; Hockley & Polka, 1994; Tremblay, Champoux, Voss, Bacon,

Lapore & Theoret, 2007; Sekiyama & Burnham, 2008; Ross, Molholm, Blanco, Gomez-Ramirez, Saint-Amour & Foxe, 2011).

Audiovisual stimuli can facilitate reaction times also in children. 7-year-old children have faster reaction times to simultaneously presented auditory and visual stimuli compared to unisensory presentation (Brandwein et al., 2010). This facilitation is still immature in 8-year-olds, but matures to adult-like level at 15 years of age. This suggests that audiovisual integration continues to mature through middle childhood and early adolescence.

Semantically congruent audiovisual associations have been shown to influence perception also in childhood. The ability to perceive numerosity in children has been shown to improve by audiovisual semantic congruency (Jordan & Baker, 2011). Audiovisual information improves numerical matching performance compared to unisensory information in preschool children. When congruent numerical information was presented simultaneously in both visual (series of squares) and auditory modality (series of tones), the children matched numerosities more accurately than with auditory or visual unisensory presentation. Recently, the effect of semantically congruent, incongruent and non-semantic pictures on auditory object recognition has been studied in 5–9-year-old children and adults using a paradigm that required identification of sounds while ignoring the pictures that were either congruent or incongruent with the sounds, or non-semantic (Thomas, Nardini and Mareschal, 2017). Adults and also children from 8 years onwards were faster at responding to sounds when coinciding semantically congruent pictures were presented, and slower at responding to sounds when semantically incongruent pictures were presented. Younger children were faster at responding to sounds presented together with congruent pictures, but slowing of responses was not observed when sounds were presented together with incongruent pictures.

Taken together, the previous studies show that audiovisual processing changes from infancy to childhood and adulthood. Adult-like audiovisual integration develops from middle childhood to adolescence. However,

children can benefit from semantically congruent audiovisual stimuli in perceptual tasks, even though the audiovisual integration is still developing.

1.3. Changes in audiovisual processing in older age

Aging affects audiovisual processing. Some studies have shown that the gain associated with audiovisual integration increases in older age, suggesting enhanced audiovisual integration (Laurienti et al., 2006; Peiffer et al., 2007; Diedrich et al., 2008). Elderly people often show greater audiovisual facilitation than young adults in response time tasks (Peiffer, 2007) and discrimination tasks (Laurienti et al., 2006), but opposite results have also been found (Stephen et al., 2010). Prolonged time windows for audiovisual integration have been reported in the elderly (Laurienti et al., 2006; Diedrich et al., 2008), which means that they can integrate two stimuli presented further apart in time than young adults. However, some studies suggest that aging might decrease the probability of audiovisual integration. Elderly people are less likely to integrate sensory cues from auditory and visual modalities than young adults because of general slowing in sensory processing (Diedrich et al., 2008). This results in a lower probability of audiovisual integration in spite of the prolonged integration window which can only partly compensate for the slowing of sensory processing.

The visual dominance effect (Colavita, 1974; Ngo, Sinnet, Soto-Faraco & Spence, 2010) refers to the dominant role of vision over audition in perceiving audiovisual events so that visual stimulus can aid processing of auditory information more than vice versa. It is also affected by aging (Diaconescu et al., 2013). Both young adults and elderly people are faster at responding to sounds when they are presented together with pictures than vice versa, but they also show more pronounced gain in audiovisual than unisensory conditions suggesting that they have a larger visual dominance effect than young adults. Elderly people show more visual influence also in speech perception than young adults, suggesting that they utilize visual speech cues more efficiently to support auditory speech comprehension (Cienkowski and Carney, 2002; Winneke and Phillips, 2011; Setti et al., 2013; Sekiyama, Soshi

& Sakamoto, 2014; Stevenson et al., 2015;), even though some studies have not found age differences (e.g., Sommers, Tye-Murray & Spehar, 2005). The content of the speech can affect audiovisual speech perception both in elderly people and young adults (Tye-Murray et al., 2008; Gordon and Allen, 2009; Maguinness et al., 2011). Visual speech cues may be particularly beneficial to elderly adults when the semantic content of the speech is unpredictable (Maguinness et al., 2011).

Taken together, previous studies show that audiovisual integration changes during normal aging. Elderly people can benefit from audiovisual semantic congruency in perceptual tasks more than young adults, and audiovisual enhancement may be greater in elderly people.

1.4. Semantically congruent audiovisual contexts can enhance memory performance

1.4.1. Semantically congruent audiovisual stimuli improve recognition memory

Previous studies have investigated the effects of audiovisual semantic congruency on recognition memory, and have found a congruency effect, i.e. that semantic congruency between simultaneous auditory and visual stimuli (e.g., an animal call and a picture of the animal) during encoding facilitates memory performance. In previous studies, the participants have been almost exclusively young adults.

Several studies have found that semantically congruent sounds facilitate the recognition memory of pictures. Lehmann and Murray (2005) and Murray and colleagues (2004, 2005) paired drawings with semantically congruent and incongruent sounds, and presented these audiovisual stimulus pairs among unisensory visual stimuli (drawings with no sound). In subsequent repeated presentations, only drawings were shown. The participants' task was to perform a continuous recognition memory task in which they decided for each drawing whether it had appeared previously. Recognition memory performance was better for drawings initially presented with semantically

congruent sounds than for drawings presented with incongruent sounds or without a sound. Moran and colleagues (2013) and Thelen and colleagues (2015) later replicated this finding.

The congruency effect is also observed for auditory stimuli (sounds) presented together with semantically congruent visual stimuli (pictures). Moran and colleagues (2013), Matusz and colleagues (2015), and Thelen and colleagues (2015) used a similar experimental design as Murray and colleagues (2004, 2005) and demonstrated that performance was better for sounds initially presented with semantically congruent pictures than for sounds presented with incongruent pictures, with non-semantic pictures or without pictures. In contrast, Cohen, Horowitz and Wolfe (2009) found that auditory recognition memory performance for sounds was not affected by semantically congruent pictures presented during memory encoding.

Very few previous studies investigated whether the congruency effect can be observed also by using verbal material, that is, spoken or written words. Only Ueno and colleagues (2015) have shown that semantically congruent natural sounds enhance recognition memory for written words compared with unisensory written word presentation. However, Cohen and colleagues (2009) found that there was no congruency effect in the converse situation, that is, recognition memory performance for sounds was not affected by congruent written words presented during encoding.

1.4.2. Audiovisual context and working memory

In studies on working memory, the effect of audiovisual stimulus presentation on immediate free recall has been studied with non-verbal and verbal stimuli (for a review, see Quak, London & Talsma, 2015). The free recall paradigm requires a shorter maintenance of memorized stimuli than the recognition memory paradigm. The free recall paradigm also requires retrieval without memory cues, whereas in the recognition memory paradigm the participants gets cues since he decides for each stimulus whether it was presented before. In a pioneering study using simultaneous auditory and visual objects as stimuli, Thompson and Paivio (1994) found that free recall was better for congruent picture-sound pairs than for unisensory stimuli. However, when

two pictures of the same object were shown, there was no enhancement in performance.

In audiovisual working memory research, verbal material has been used more often than pictures and sounds. Lewandowski and Kobus (1993) pioneered the use of written and spoken words to study the effect of audiovisual stimulus presentation on recall. They found superior recall for congruent spoken and written words over unisensory auditory and visual words, as well as over incongruent spoken and written words. Goolkasian and Foos (2005) observed a similar effect. They also found that written words presented together with pictures are not recalled any better than either visual component presented alone.

Delogu, Raffone and Belardinelli (2009) studied the memory for both verbal (words) and non-verbal stimuli (pictures of objects) using immediate serial recall task with auditory, visual and congruent audiovisual stimulus presentations. They found a distinct congruency effect for objects, confirming Thompson and Paivio's (1994) earlier finding. However, the memory for written and spoken words was better for audiovisual and auditory presentation than for visual presentation, while audiovisual performance did not exceed unisensory auditory performance.

Summarizing the working memory studies, it can be concluded that verbal recall can be enhanced for congruent audiovisual stimuli.

1.4.3. Audiovisual memory across the life span

Most studies on audiovisual memory have been conducted in young adults, and less is known about audiovisual memory in children or elderly people, as the studies reviewed above only used young adults as participants.

Constantinidou, Danos, Nelson and Baker (2011) used a free-recall paradigm to investigate children's working memory for spoken words, pictures, and simultaneous presentations of congruent spoken words and pictures. Working memory performance was better for simultaneous presentation of spoken words and pictures than for spoken words presented alone, but not better than performance for pictures presented alone.

Luo, Hendriks and Craik (2007), in turn, studied audiovisual recognition memory in elderly people. They found that written words presented with congruent sounds during encoding did not facilitate memory performance in elderly people, as it did in young adults. This suggests that memory for written words presented together with sounds during encoding does not offer the same benefit to the elderly as it does to young adults.

In sum, there is some evidence that semantically congruent audiovisual stimuli may facilitate later memory performance also in children and in elderly people. However, the evidence is extremely scarce, and it is not known which kind of audiovisual stimulus combinations can improve memory performance in these age groups.

2. THE AIM OF THE PRESENT THESIS

The aim of the present thesis was to study the effects of audiovisual encoding on later unisensory recognition memory performance in school-aged children, young adults and elderly people, and investigate how the different age groups utilize audiovisual information in memory encoding. The effects were studied with verbal and non-verbal materials. The aim was to expand the investigation of audiovisual memory to address the question of whether semantic congruency of audiovisual stimuli during memory encoding can improve also recognition memory for verbal material (spoken and written words), in addition to sounds and pictures of objects or animals used in previous studies (Murray et al., 2004; Lehmann & Murray, 2005; Cohen et al., 2009; Moran et al., 2013; Matusz et al., 2015; Thelen et al., 2015). To our knowledge, this was the first series of studies where both verbal and non-verbal materials were used in the same participants to study audiovisual memory with a recognition memory paradigm.

The benefits of audiovisual encoding were investigated using a two-part recognition memory paradigm with an audiovisual or a unisensory encoding phase, followed by a unisensory retrieval phase. Previously, a continuous recognition memory task has often been used, and the evidence for the congruency effect has been obtained mainly using that paradigm (Murray et al., 2004; Lehmann & Murray, 2005; Moran et al., 2013; Matusz et al., 2015; Thelen et al., 2015). In the present studies, the two-part recognition memory paradigm was utilized, because it separates the encoding and retrieval phases and does not require free recall of memorized items. This reduces the cognitive load in the retrieval phase and allows to study the benefits of audiovisual encoding without the high cognitive demands of free recall. This is especially important when studying children and elderly people. The present aim was also to study longer maintenance of memorized items than working memory tasks, or the continuous task.

The present series of studies included two variants of the two-part recognition memory paradigm. In the first paradigm, participants memorized auditory or visual stimuli, each of which co-occurred during encoding with

either a semantically congruent stimulus, incongruent stimulus or a non-semantic stimulus in the other modality. Different congruency conditions were utilized in order to study whether a congruent stimulus facilitates memory performance compared with a non-semantic stimulus, and whether an incongruent stimulus causes interference, that is, worsens memory performance compared with a non-semantic stimulus. Non-semantic audiovisual stimuli were used instead of unisensory stimuli to control for potential differences in alerting effects, which may occur if some to-be-remembered stimuli are audiovisual, while some are not. In this paradigm, stimuli from the three congruency conditions were intermingled in the encoding phase. This intermingled paradigm was used in Studies I and III.

In the second paradigm, participants memorized auditory or visual stimuli which co-occurred during encoding with a semantically congruent stimulus in the other modality or which were presented alone. This design allowed investigation of the congruency effect without possible interference by incongruent or non-semantic stimuli. Possible alerting effects were controlled for by presenting audiovisual and unisensory stimuli in different blocks, not intermingled as in the first paradigm. This blocked recognition memory paradigm was used in Studies II and IV.

The aim of Study I was to investigate whether semantically congruent audiovisual stimuli presented during memory encoding improve recognition memory performance for non-verbal (sounds and pictures of natural objects) and verbal material (spoken and written words) in young adults. Recognition memory was investigated using the intermingled paradigm. The hypothesis was that semantically congruent verbal and non-verbal audiovisual stimuli facilitate encoding, leading to better memory performance compared with stimuli presented with non-semantic stimuli in the other modality.

The aim of Study II was to investigate which kind of semantically congruent audiovisual information facilitates the precision of recognition memory in relation to unisensory stimuli in young adults. The study applied the blocked recognition memory paradigm, which allowed investigation of the congruency effect without possible interference caused by incongruent or non-

semantic stimuli. It also allowed to analyze the recognition memory data using measure of detectability *d'* (Green & Swets, 1966; Murdock, 1982; Stanislav & Todorov, 1999; Macmillan & Creelman, 2005). The hypothesis was that semantically congruent audiovisual presentation facilitates memory encoding of verbal and non-verbal auditory and visual stimuli and leads to better recognition memory performance compared with the auditory and visual component stimuli presented alone during encoding.

The aim of Study III was to investigate whether school-aged children benefit from semantically congruent audiovisual information during memory encoding. The aim was also to investigate whether there are developmental differences in the utilization of semantically congruent audiovisual information between the ages of 8 and 12 years, and between children and adults. Recognition memory was investigated using the intermingled recognition memory paradigm. This paradigm was utilized in order to study whether a congruent stimulus facilitates memory performance and whether an incongruent stimulus causes interference in children. Interference effect was studied because of possible age-related changes in the interference effect in perceptual studies (Thomas et al., 2017). The hypothesis was that congruent audiovisual stimuli can facilitate memory encoding of their auditory and visual components and lead to better recognition memory performance also in children.

The aim of Study IV was to investigate whether semantically congruent audiovisual information enhances auditory memory encoding also in elderly people, and whether young adults and elderly people differ in how they utilize audiovisual information in memory encoding. The effect of congruent visual stimuli on recognition memory performance for auditory stimuli was investigated. The study applied blocked recognition memory paradigm. This paradigm was chosen because this allowed comparison the precision of recognition memory between age groups, and because audiovisual congruency effect was found from young adults in Study II using this paradigm. The hypothesis was that congruent visual stimuli presented together with auditory stimuli facilitate memory encoding of auditory stimuli, and lead to better recognition memory performance also in elderly people. It was also expected

that elderly people might benefit more from audiovisual semantic congruency than young adults, in line with previous findings of greater audiovisual perceptual enhancement in elderly people compared with young adults (Laurienti et al., 2006; Peiffer et al., 2007; Diedrich et al., 2008).

3. GENERAL METHODS

3.1. Stimuli

Photographs of natural objects, sounds of natural objects, written words, spoken words, visual noise, meaningless writing (a row of 6 x-letters), and auditory noise were used as stimulus material in all studies. The photographs were obtained from the Multimodal Stimulus Set (Schneider, Engel & Debner, 2008) and from the internet and modified to resemble those in the Multimodal Stimulus Set. The photographs presented objects from several semantic categories (animals, tools, vehicles, musical instruments, and household items). They were converted into gray scale and presented on a black background. They were positioned centrally on a computer screen. Their sizes varied between 3 cm and 23 cm horizontally and vertically. The sounds were recordings of complex sounds of objects from the same semantic categories as visual stimuli. They were also obtained from the Multimodal Stimulus Set (Schneider et al., 2008) or from the internet and edited to resemble those from the Multimodal Stimulus Set in duration (400 ms) and intensity. The written and spoken words were common two-syllable Finnish nouns from various semantic categories (animals, tools, foods, plants and household items etc.). The visual noise stimulus was a frame of white noise or a row of six letters X (XXXXXX). Written words were presented in white Times New Roman font with a 40-point font size in the center of the computer or laptop screen. Spoken words were spoken by a female voice (the author of the present thesis), recorded, and edited to a constant intensity. Their duration varied between 350 and 780 ms. A burst of white noise was used as a non-semantic auditory stimulus. The auditory stimuli were presented through loudspeakers (in Study I) or binaurally through headphones (in Studies II, III and IV) at an intensity of approximately 55 dB(A) at the ear drums.

3.2. Recognition memory tasks

In Studies I–IV, memory performance was measured using recognition memory paradigms (the intermingled recognition memory paradigm and the blocked recognition memory paradigm), which consists of several recognition memory tasks.

All recognition memory tasks had two parts: (1) an encoding phase consisting of audiovisual, dual visual or unisensory auditory or visual items, followed by (2) a recognition memory phase consisting of only unisensory items. In the encoding phase, the participants were instructed to memorize the stimuli of a designated modality while not memorizing other possibly coinciding stimuli. The audiovisual stimulus pairs were presented in the encoding phase with simultaneous onsets of their auditory and visual components. The dual visual stimulus pairs, a picture and written word or a row of letters X were presented in the encoding phase simultaneously at the monitor the text on the top of the picture. In unisensory recognition memory tasks, only one unisensory visual or auditory stimuli were presented in the encoding phase. The recognition memory phase immediately followed the encoding phase. Here the stimuli in the memorized modality were presented again, intermingled randomly with an equal number of new stimuli. All stimuli were unisensory in the recognition memory phase. The participant's task was to indicate for each stimulus by pressing one of two keys whether or not the stimulus had been presented in the encoding phase. The next item was presented 1000 ms after the response.

The presentation order of stimuli or stimulus pairs was always random in both encoding and recognition memory phase. Presentation durations varied depending on stimulus. For pictures, sounds and picture-sound pairs, the presentation time was 400 ms. For written and spoken words and stimulus pairs that included written words or spoken words, the presentation time was 800 ms. Inter-stimulus-interval was always 800 ms, during which, a black screen with a fixation cross in the center was presented.

The recognition memory tasks were presented and the response data gathered with Presentation software (Neurobehavioral Systems Inc., Berkeley,

California, USA). Before the experiment, the participants read written instructions and performed a short practice session. During the tasks, the experimenter sat next to the participant to ensure that the participant maintained his or her gaze on the screen.

3.3. Statistical methods

3.3.1. Signal detection theory

Signal detection theory (SDT) is a method in analyzing recognition memory data since it allows to estimate the precision of recognition memory using measure or detectability *d'* (Green & Swets, 1966; Murdock, 1982; Stanislav & Todorov, 1999; Macmillan & Creelman, 2005). In Studies I and III, like in previous studies by Murray and colleagues (2004), Lehmann and Murray (2005), Moran and colleagues (2013), and Thelen and colleagues (2015), *d'* was not informative because of the experimental design. In experimental designs where different congruency conditions are intermingled, *d'* does not offer additional information compared with the percentage of correct responses, because there is only one recognition memory phase consisting of the memorized items for all congruency conditions and new items. Therefore, the false alarm (FA) rates are equal for all congruency conditions, and the *d'* values are in direct proportion to the percentage correct rates.

In the analyses of Studies II and IV, *d'* was used as a measure. Experimental designs with congruent audiovisual stimulus pairs and with unisensory stimuli presented in different blocks allow the use of *d'* in a meaningful way because it was possible to calculate the parameters needed. For the analysis, *d'* was calculated using the hit rate and the false alarm rate in each recognition memory task. Hits were the trials in which the participant recognized the memorized items correctly (correct responses; already seen or heard stimuli identified as "old"), whereas FAs were the trials in which the participant recognized new items as previously presented (new stimuli recognized as "old"). These values were normalized to obtain the *z*-score values *z*(*HIT*) and *z*(*FA*), and subtracted from one another to get $d' = z(HIT) - z(FA)$.

False alarm rates of 0 were corrected using $1/(2n)$, where n is the number of trials (Miller, 1996). In our studies, no hit rates of 1 were observed.

3.3.2. Statistical analysis

The percentages of correct responses in Studies I and III, and *d'* values in Studies II and IV were analyzed using analysis of variance (ANOVA). Greenhouse-Geisser correction was applied for p values when appropriate. However, the original degrees of freedom are reported. Bonferroni corrected significance levels ($p < .05$) were used in pairwise comparisons.

4. SPECIFIC METHODS AND RESULTS OF STUDIES I-IV

4.1. Study I: Audiovisual semantic congruency enhances memory performance in young adults

4.1.1. Introduction

Study I investigated whether the semantic congruency of audiovisual stimuli during memory encoding can improve recognition memory performance for non-verbal (sounds and photographs of natural objects) and verbal material (spoken and written words) in young adults. The aim was to investigate, for the first time, whether semantic congruency of audiovisual stimuli during encoding improves recognition memory also for verbal material (spoken and written words) as it does for auditory objects (sounds of natural objects, cf. Moran et al, 2009). The aim was also to test whether the congruency effects for objects (found by Lehmann and Murray (2005) and Murray et al. (2004)) is present in a two-part task with separate encoding and retrieval phases and longer memory delay.

4.1.2. Specific methods

Fourty-two volunteers (10 men; 37 right-handed, mean age 26 years, SD 7.7 years) participated in Study I. All participants were native speakers of Finnish. They reported normal or corrected-to-normal vision and normal hearing and no dyslexia or neurological illnesses.

The study included five recognition memory tasks (four audiovisual tasks and one unisensory dual visual task), each including stimuli to be encoded in three congruency condition: semantically congruent, semantically incongruent and non-semantic. The congruent items were semantically matching: for example, a picture of a sheep and a sheep´s baa, or the same noun (e.g. “car”) written and spoken. The incongruent items were semantically conflicting: for example, a picture of a cow and the sound of a foghorn, or a written word paired with a different spoken word. The non-semantic stimuli consisted of the item to be memorized in one modality presented with either

white noise or a meaningless written stimulus (XXXXXX) in the other modality. There were always 17 items per congruency category, so that a total of 51 items were to be memorized in each task.

In Block 1, in the encoding phase, pictures were paired with congruent or incongruent natural sounds, or noise bursts. The participants` task was to memorize the pictures. In the recognition memory phase, the participants were instructed to decide for each picture whether it had been presented in the encoding phase.

In Block 2, in the encoding phase, the sounds of natural objects were paired with congruent, incongruent or noise pictures. The participants` task was to memorize the sounds. In the recognition memory phase, the participants were instructed to decide for each sound whether it had been presented in the encoding phase.

In Block 3, in encoding phase, written words were paired with congruent or incongruent spoken words, or noise bursts. The participants` task was to memorize the written words. In the recognition memory phase, the participants were instructed to decide for each written word whether it had been presented in the encoding phase.

In Block 4, in the encoding phase, spoken words were paired with congruent or incongruent written words, or a row of letters X. The participants` task was to memorize the spoken words. In the recognition memory phase, the participants were instructed to decide for each spoken word whether it had been presented in the encoding phase.

In Block 5, in the encoding phase, pictures were paired with congruent or incongruent written words, or a row of letters X. The participants` task was to memorize the pictures. In the recognition memory phase, participants decided whether each picture had been presented in the encoding phase.

The hypothesis was that audiovisual congruent stimuli facilitates memory encoding and leads to better recognition memory performance compered to auditory or visual stimuli presented with noise. This facilitation was expected

to occur in audiovisual tasks but not in a unisensory dual visual task, where two visual stimuli were presented together.

4.1.3. Results

The recognition memory performance (percentage of correct responses) for stimuli that were presented with a semantically congruent, incongruent or non-semantic stimulus during encoding in Blocks 1-5 is shown in Figure 1. In a two-way repeated-measures ANOVA, the main effects of Congruency [$F(2,82) = 32.60$, $p < .001$, partial $\eta^2 = .443$] and Block [$F(4,164) = 52.43$, $p < .0001$, partial $\eta^2 = .561$] were significant, and the interaction between Congruency and Block was also significant [$F(8,328) = 4.28$, $p < .001$, partial $\eta^2 = .095$]. Pairwise comparisons between congruent and non-semantic trials showed that there was a congruency effect in Blocks 2–4, i.e., memory performance was better for stimuli that were accompanied by a congruent rather than non-semantic stimulus in the other modality during encoding, for sounds paired with pictures [$t(41) = 5.23$, $p < .001$], for written words paired with spoken words [$t(41) = 3,13$, $p < .05$], and for spoken words paired with written words [$t(41) = 4.06$, $p < .001$]. For the unisensory dual visual Block 5, there was no congruency effect. Performance was the best in Block 1 (pictures) and the poorest in Block 2 (sounds), differing from the other blocks [$p < .05$ in each case]. There was no interference effect in any of the tasks; when the stimuli were presented together with semantically incongruent stimuli in the encoding phase, memory performance was not any worse compared with a non-semantic stimulus presentation ($p > .05$ in each case).

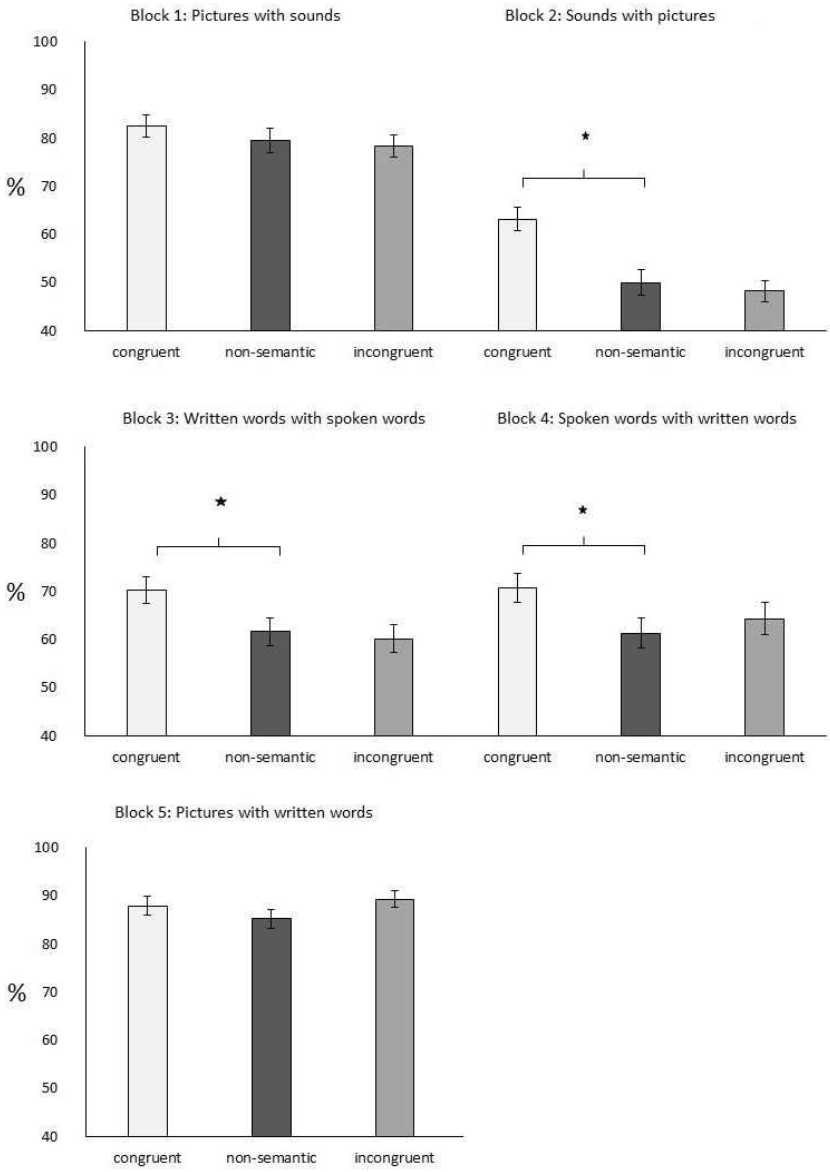


Figure 1. Percentages of correct responses in the recognition memory task for stimuli presented together with congruent, incongruent or non-semantic stimuli during encoding in blocks 1-5. Block 1: Recognition memory of pictures presented together with sounds during encoding phase. Block 2: Recognition memory of sounds presented together with pictures during encoding phase. Block 3: Recognition memory of written words presented together with spoken words during encoding phase. Block 4: Recognition memory of spoken words presented together with written words during encoding phase. Block 5: Recognition memory of pictures presented together with written words during encoding phase. Stars indicate statistically significant differences ($p < .05$). Error bars depict standard errors of the mean.

Thus, the results of Study I show that when pictures and sounds are presented together during encoding, memory for sounds is better for stimuli that were initially presented together with semantically congruent pictures than when than for stimuli that were initially presented together with a non-semantic visual stimulus. Moreover, when written and spoken words were presented together during encoding, the memory for both written and spoken words was better for stimuli that were initially presented together with a congruent word in the other modality than for stimuli that were initially presented together with a non-semantic stimulus. Furthermore, according to the results of Study I, semantically congruent sounds did not facilitate recognition memory for pictures or, nor did semantically congruent written words coinciding with pictures during encoding.

4.2. Study II: Semantically congruent visual stimuli improve auditory memory performance in young adults

4.2.1. Introduction

Study II investigated which combinations of semantically congruent audiovisual information can improve recognition memory performance in relation to only unisensory stimuli, without any possible interference caused by different congruency conditions. The experimental design consisted of three experimental blocks (sounds, spoken words, written words), in which semantically congruent audiovisual stimulus pairs and unisensory stimuli were presented to young adult participants in separate recognition memory tasks. This design allowed to compare the precision of recognition memory (*d'*) between congruent audiovisual and unisensory stimuli. We expected that congruent audiovisual stimuli facilitate memory encoding and lead to better recognition memory performance compared to unisensory stimuli.

4.2.2. Specific methods

Forty-three participants (mean age 23.0 years, SD 3.5 years; 9 males) participated in Blocks 1 and 2 of Study II, and another group of 43 participants (mean age 24.6 years, SD 6 years; 6 males) participated in Block 3. All

participants reported Finnish as their mother tongue. They reported normal or corrected-to-normal vision and normal hearing, and reported no dyslexia or neurological illnesses.

Study II included three experimental blocks: 1. Sound block, 2. Spoken word block and 3. Written word block.

Block 1 consisted of three recognition memory tasks: one unisensory task and two audiovisual tasks. In the encoding phase of the unisensory task, 25 sounds were presented alone. In the encoding phase of one audiovisual task, 25 sounds were presented with congruent pictures, and in the other audiovisual task, 25 sounds were presented with congruent written words. The participants were instructed to memorize the sounds in all tasks.

Block 2 consisted of three recognition memory tasks: one unisensory task and two audiovisual tasks. In the encoding phase of the unisensory task, 51 spoken words were presented alone. In the encoding phase of one audiovisual task, 51 spoken words were presented with congruent pictures, and in the other audiovisual task, 51 spoken words were presented with congruent written words. The participants were instructed to memorize the spoken words in all tasks.

Block 3 consisted of two recognition memory tasks, one unisensory task and one audiovisual task. In the encoding phase of a unisensory task, 51 written words were presented alone. In the encoding phase of audiovisual task, 51 written words were presented with congruent spoken words. The participants were instructed to memorize the written words in both tasks.

4.2.3. Results

The *d'* values for stimuli that were initially presented either with a congruent stimulus in the other modality or presented alone during encoding in Blocks 1–3 are presented in Figure 2.

The *d'* values were analyzed with a repeated-measures analysis of variance (ANOVA) for each block with stimulus type (Blocks 1 and 2: sounds presented with pictures, sounds presented with written words, sounds presented alone;

Block 3: written words presented with spoken words, written words presented alone) as factor.

For sounds (Block 1), significant main effect of stimulus type was found [$F(2,82) = 20.13$, $p < .001$, $\eta^2 = .329$]. Pairwise comparisons showed that memory performance was better when sounds were initially presented with either semantically congruent written words or semantically congruent pictures compared with unisensory presentation ($p < .001$ for both comparisons).

For spoken words (Block 2), the main effect of stimulus type was also significant [$F(2,82) = 18.10$, $p < .001$, $\eta^2 = .306$]. Pairwise comparisons showed that memory performance was better for spoken words presented with semantically congruent pictures than with written words or just alone ($p < .001$ for both comparison). When spoken words were presented alone or with congruent written words, there was no difference in memory performance ($p > .05$).

For written words (Block 3), effect of stimulus type was not observed [$F(1,41) = .530$, $p > .05$, $\eta^2 = .013$]. There was no difference in memory performance when written words were presented with congruent spoken words or alone.

Thus, the results of Study II show that the semantic congruency of audiovisual stimuli improved the precision of recognition memory in younger adults. The *d'* value was higher for sounds and spoken words when they were presented with semantically congruent pictures during encoding than when they were presented alone. It was also higher for sounds presented with semantically congruent written words, but not when spoken words were paired with written words. However, *d'* for written words was unaffected by congruent spoken words compared with unisensory presentation.

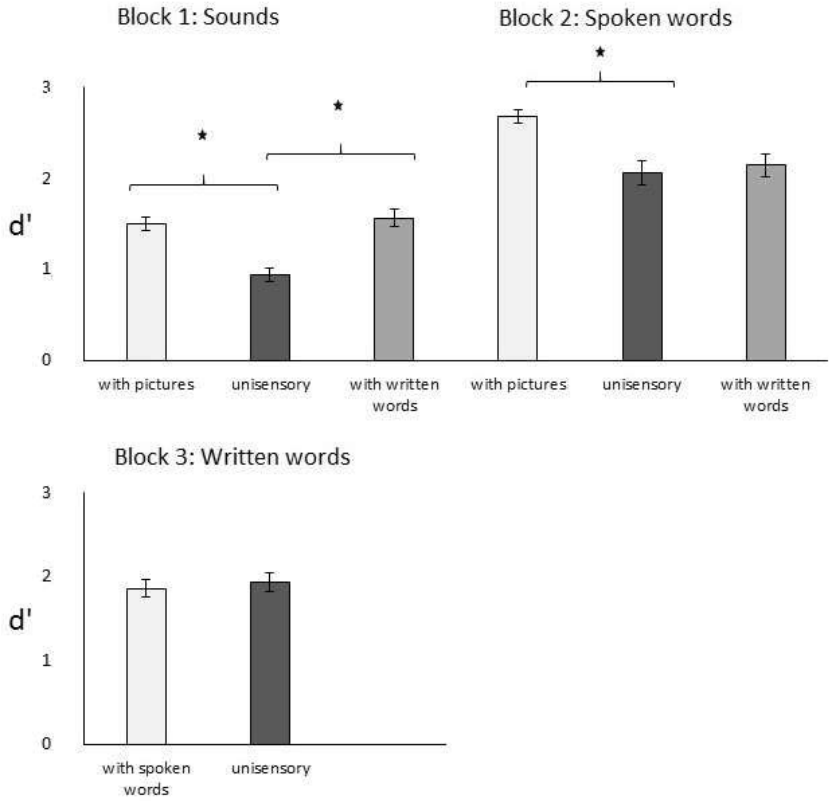


Figure 2. Performance accuracy (*d'*) in the recognition memory task for the memorized stimuli presented alone or together with semantically congruent stimuli during encoding in Blocks 1-3. Block 1: Recognition memory for sounds presented with congruent pictures, presented alone or together with congruent written words during encoding. Block 2: Recognition memory for spoken words presented together with congruent pictures, presented alone or together with congruent written words during encoding. Block 3: Recognition memory for written words presented together with congruent spoken words or presented alone during encoding. Stars indicate statistically significant ($p < .05$) differences. Error bars depict standard errors of the mean.

4.3. Study III: Audiovisual semantic congruency enhances memory performance in children

4.3.1. Introduction

Study III investigated whether the semantic congruency of audiovisual stimuli during memory encoding can improve recognition memory performance in school-aged children, and which combinations of audiovisual information would benefit children's memory encoding. The aim was also to investigate, whether there are developmental differences in the utilization of semantically congruent audiovisual information between the ages of 8 and 12 years, and between children and adults. The study included six recognition memory tasks, and each encoding phase included three congruency conditions: semantically congruent, incongruent and non-sematic. The hypothesis was that audiovisual congruent stimuli can facilitate memory encoding and lead to better recognition memory performance also in children.

4.3.2. Specific methods

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 children were from school grade 2 and aged from 7 years 8 months to 9 years 1 month (mean 8 years 4 months). Thirty-five children were from school grade 4 and aged from 9 years 9 months to 10 years 1 month (mean 10 years 4 months). Thirty-eight children 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 had no dyslexia or other learning disabilities or neurological disorders. The adult participants were the same as in Study I.

We utilized otherwise the same experimental design as in Study I, except that an additional block was included where the effect of pictures on the recognition memory of spoken words was studied (Block 6). Thus, the blocks used in Study I were: pictures memorized with sounds (Block 1), sounds

memorized with pictures (Block 2), written words memorized with spoken words (Block 3), spoken words memorized with written words (Block 4), pictures memorized with written words (Block 5) and spoken words memorized with congruent or incongruent pictures, or visual noise (Block 6). The blocks were presented to the 10-year-old and 12-year-old participants in a random order. The 8-year-old participants were presented only with the Blocks 1, 2 and 6 not requiring reading skills.

4.3.3. Results

The effect of audiovisual semantic congruency was tested first with an ANOVA including all six blocks. The ANOVA with Congruency (to-be-memorized stimulus presented with congruent stimuli, to-be-memorized stimulus presented with incongruent stimuli, to-be-memorized stimulus presented with non-sematic stimuli), Block (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 Block [$F(5,350) = 67.4$, $p < .001$, $\eta^2 = .490$], and interaction between Congruency and Block [$F(10,700) = 6.18$, $p < .001$, $\eta^2 = .081$] suggesting that there were differences between the congruency conditions and tasks but not between the age groups. In the second ANOVA, the focus was on the age-related effects, so that all three age groups were included for the blocks that were done by all children. The second ANOVA with Congruency (congruent, incongruent, non-semantic), Block (1, 2, 6), and Age (8, 10, 12) as factors showed the main effects of Congruency [$F(2,222) = 70.1$, $p < .001$, $\eta^2 = .387$] and Block [$F(2,222) = 85.6$, $p < .001$, $\eta^2 = .435$], and interaction between Congruency and Block [$F(4,444) = 12.2$, $p < .001$, $\eta^2 = .099$]. In addition, there was a three-way interaction between Congruency, Block and Age [$F(8, 444) = 2.16$, $p = .032$, $\eta^2 = .038$]. It was further tested by separate ANOVAs for Blocks 1, 2 and 6 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 how they utilized semantic congruency in memory encoding.

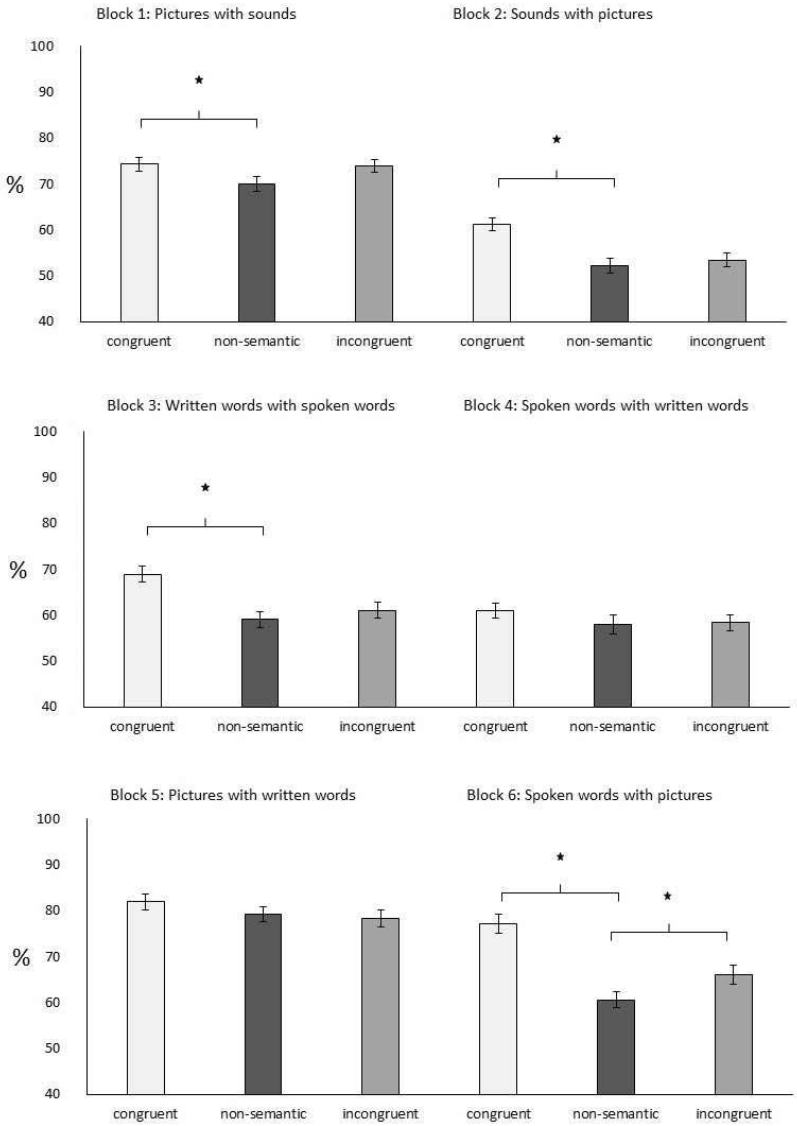


Figure 3. Mean percentages of correct responses in the recognition memory task for stimuli presented together with congruent, incongruent or non-semantic stimuli during encoding. Block 1: Recognition memory for pictures presented together with sounds during encoding (means across all age groups). Block 2: Recognition memory for sounds presented together with pictures during encoding (means across all age groups). Block 3: Recognition memory for written words presented together with spoken words during encoding (means across 10- and 12-year-olds). Block 4: Recognition memory for spoken words presented together with written words during encoding (means across 10- and 12-year olds. Block 5: Recognition memory for pictures presented together with written words during encoding (means across 10- and 12-year olds). Block 6: Recognition memory of spoken words presented together with pictures during encoding (means across all age groups). Stars indicate statistically significant congruency effects ($p < .05$). Error bars depict standard errors of the mean.

The data from all age groups were combined to further investigate the effect of congruency, because factor Age did not have a significant effect. In Figure 3, the averaged data for the three congruency conditions in each block is presented. Post hoc tests for the Congruency × Block interaction revealed that Blocks 1 (pictures with sounds), 2 (sounds with pictures), 3 (spoken words with pictures), and 4 (written words with spoken words) showed congruency effects, while Blocks 5 (spoken words with written words) and 6 (pictures with written words) did not. When the stimuli were presented together with semantically congruent stimuli in the other modality during encoding, 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 [$t(113) = 11.8$ $p < .001$]. However, 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 blocks; when the stimuli were presented together with semantically incongruent stimuli in the encoding phase, memory performance was not any worse compared with a non-semantic stimulus presentation [$p > .05$ in each case]. However, in Block 3 (spoken words with pictures) memory performance for incongruent stimulus presentation was better compared with a non-semantic stimulus presentation [$t(113) = 3.94$, $p < .001$].

Differences in how children and young adults utilize audiovisual information in memory encoding were investigated by comparing the congruency effect of children and young adults. The adult data were from Study I. Block 6 was excluded from the analysis because it was not performed by the adult participants. The differences in congruency effect between children and adults was investigated by studying the interaction between Block and Age with gain indices. The gain indices were calculated as the difference between the memory performance when to-be-memorized stimuli was presented with congruent stimuli and the memory performance when to-

be-memorized stimuli was presented with non-semantic stimuli (congruent minus non-semantic). The gain index was calculated separately for each participant and each block. Gain indices were studied using a mixed models ANOVA with Block (1-5) and Age (adults, children) as factors. The main effect of Block was significant [$F(4,448) = 5.9$, $p < .0001$, $\eta^2 = .046$], which indicates that there were differences in memory performance between blocks. The gain was larger in Block 2 (sounds) compared to Block 1 (pictures) and Block 5 (unisensory). The gain was larger in Block 3 (text) compared to Block 5 ($p < .05$ in every comparison; $p = .01$, $p = .002$ and $p = .032$ respectively). The other comparisons were not significant ($p > .05$). There was no interaction between Block and Age [$F(4,448) = 1.4$, $p = .259$, $\eta^2 = .011$] suggesting that there were no statistically significant differences in the congruency effect between adults and children.

The results of Study III show that children can benefit from audiovisual semantic congruency during memory encoding. There were no age-related differences in the utilization of audiovisual information when comparing different age groups, or comparing adults and children.

4.4. Study IV: Semantically congruent visual stimuli improve auditory memory performance in elderly people

4.4.1. Introduction

Study IV investigated whether semantically congruent visual information can facilitate the recognition memory of auditory information in elderly people with normal aging. Differences in how older and younger participants utilize audiovisual information in memory encoding was also investigated. The effect of congruent visual stimuli in the recognition memory performance of auditory stimuli was investigated. Semantically congruent visual stimuli were expected to facilitate the recognition memory performance in not only younger adults but also in elderly people.

4.4.2. Specific methods

The participants of Study IV were 42 elderly people (mean age 71 years, SD 4.4, range 65 – 85 years; 15 males). Their results were compared with the results from 42 young adults (mean age 23 years, SD 3.5, range 18 – 34 years; 9 males) in Study II. All participants were native speakers of Finnish, reported normal or corrected-to-normal vision and normal hearing. They reported no problems in memory or any neurological illnesses or learning difficulties.

The experimental design of Study IV was similar to Study II, except that Block 3 (written words presented with spoken words) was left out because the spoken words did not facilitate the memory of written words in Study II. The measure or detectability *d'* was used in the statistical analyses.

4.4.3. Results

A mixed models ANOVA was conducted with Modality (to-be-memorized auditory stimulus presented alone vs. with a congruent picture vs. with congruent text), Stimulus type (sounds, spoken words) and Age (young, elderly) as factors. The main effects of Modality [$F(2,164) = 70.9$, $p < .0001$, $\eta^2 = .464$], Stimulus type [$F(1,82) = 271.4$, $p < .0001$, $\eta^2 = .768$] and Age [$F(1,82) = 360.6$, $p < .0001$, $\eta^2 = .425$] were significant, as were the interactions between Modality and Stimulus type [$F(2,164) = 21.1$, $p < .0001$, $\eta^2 = .205$] and Modality, Stimulus type and Age [$F(2,164) = 6.3$, $p = .003$, $\eta^2 = .072$]. Memory performance was better when the memorized stimulus was paired with either a congruent picture or text compared with auditory stimuli presented alone. Performance was better with congruent pictures than with congruent text. Moreover, spoken words were remembered better than sounds. Overall, memory performance was better in younger adults than in elderly people. In Figure 4, the data for the three congruency conditions in each experiment is presented.

The congruency effect was further investigated by studying the interaction between Modality, Stimulus type and Age using gain indices. The gain index was calculated as the difference between memory performance for stimuli presented with congruent visual stimuli and unisensory stimuli (audiovisual *d'* - auditory *d'*). The gain index was calculated for each participant and each

audiovisual condition (sounds with pictures; sounds with text; spoken words with pictures; spoken words with text). They were studied using an ANOVA with Modality (gain due to pictures, gain due to text), Stimulus type (sounds vs. spoken words) and Age (young vs. elderly) as factors. The main effect of Modality was significant [$F(1,82) = 25.6$, $p < .0001$, $\eta^2 = .238$], as were the interactions between Modality and Stimulus type [$F(1,82) = 36.6$, $p < .0001$, $\eta^2 = .309$] and Stimulus type and Age [$F(1,82) = 14.6$, $p < .0001$, $\eta^2 = .151$]. The main effect of Modality suggested that pictures produced a greater memory gain than text. However, post-hoc tests for the Modality and Stimulus type interaction showed that this difference was significant only for spoken words [$t(83) = 7.39$, $p < .0001$], not for sounds [$t(83) = -.512$, $p = .610$].

In order to study the interaction between Stimulus type and Age, the gain indices were combined (gain for sounds with pictures plus gain for sounds with text; gain for spoken words with pictures plus gain for spoken words with text), and the combined gain indices were studied using an ANOVA with Stimulus type (gain for sounds vs. gain for spoken words) and Age (young vs. elderly) as factors. When encoding sounds, young adults benefited more from audiovisual presentation than elderly people [$F(1,82) = 7.19$, $p = .009$, $\eta^2 = .081$]. When encoding spoken words, elderly people benefited more from audiovisual presentation than young adults [$F(1,82) = 6.65$, $p = .012$, $\eta^2 = .075$].

In summary, the results of Study IV show that semantically congruent audiovisual stimuli during encoding improve memory performance for spoken words and sounds in the elderly with normal aging. When comparing audiovisual gains between the age groups, the elderly people benefited more from audiovisual presentation than the young adults when the task was to memorize spoken words. However, when memorizing sounds young adults benefited more from audiovisual presentation than elderly people.

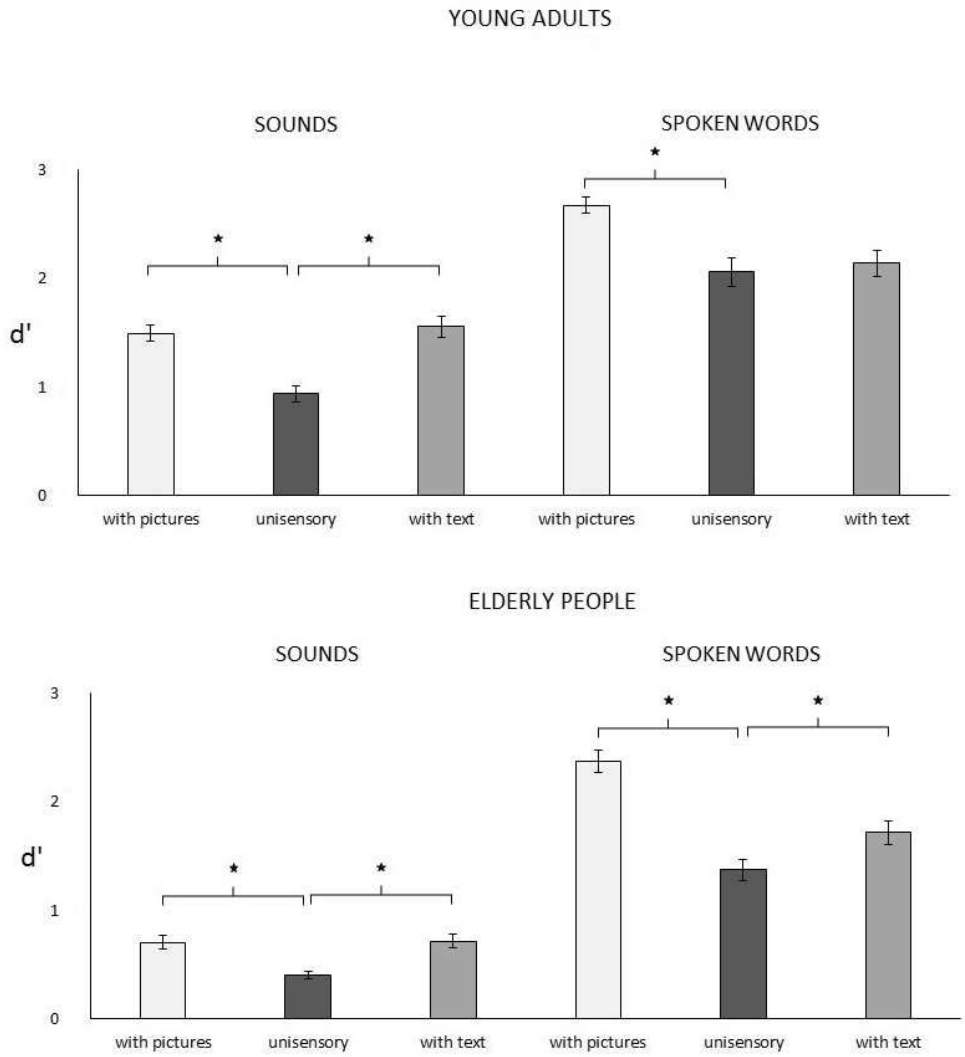


Figure 4. Recognition memory performance accuracy (*d'*) in young adults and elderly people for sounds and spoken words presented alone or together with semantically congruent pictures or semantically congruent written words during encoding. Block 1: Recognition memory for sounds presented together with congruent pictures, presented alone or together with congruent written words during encoding. Block 2: Recognition memory for spoken words presented together with congruent pictures, presented alone or together with congruent written words during encoding. Stars indicate statistically significant congruency effects ($p < .05$). Error bars depict standard errors of the mean.

5. GENERAL DISCUSSION

5.1. The benefits of audiovisual memory encoding

The results of the present series of studies show that the semantic congruency of audiovisual stimuli can improve recognition memory performance in school-aged children, young adults and elderly people. When auditory and visual stimuli were presented together during encoding, the participants' memory was better for auditory and visual stimuli that were initially presented with a semantically congruent stimulus in the other modality compared with stimuli initially presented with non-semantic stimuli (Studies I and III) or compared with unisensory stimuli (Studies II and IV). This congruency effect in memory performance was seen with non-verbal and verbal material. Especially semantically congruent pictures facilitated memory performance of sounds and spoken words in all age groups. These results provide empirical evidence that audiovisual semantic congruency facilitates recognition memory performance across the life span.

The memory for sounds was better when they were initially presented together with semantically congruent pictures compared to sounds presented with visual noise or compared to unisensory sound presentation. This facilitation was seen in younger adults (Studies I and II), in line with previous studies of Moran and colleagues (2013), Matusz and colleagues (2015), and Thelen and colleagues (2015). Importantly, the congruency effect was shown for the first time to be present also in children (Study III) and in elderly people (Study IV). Semantically congruent pictures also facilitated memory encoding of spoken words in young adults, children and elderly people. Studies II, III and IV showed that the memory performance was better for spoken words when they were paired with semantically congruent pictures during encoding. This facilitation was seen in all present age groups. The findings are in line with Constantinidou et al. (2011) who showed that in children, congruent pictures can improve the working memory performance for spoken words compared with spoken words presented alone.

When to-be-memorized pictures were paired with semantically congruent sounds during encoding, the congruency effect was observed in children (Study III). This congruency effect did not reach significance in young adults in Study I, in contrast to previous studies (Lehmann and Murray, 2005; Murray et al., 2004; Moran et al., 2013; Thelen, Talsma and Murray, 2015).

Congruent pictures can aid processing of auditory information more than vice versa, and this facilitation can also be seen in memory performance. Pictorial information seems to be more helpful than other forms of information in memory encoding. Picture superiority effect which refers to the finding that pictures are remembered better than words (Paivio, Rogers & Smythe, 1968; Nelson, Reed & Walling, 1976; Snodgrass & Asiaghi, 1977) might influence the findings that pictures facilitate the memory performance.

Congruent written words facilitated encoding of spoken words also in young adults and elderly people (Studies I and IV), but not in school-aged children (Study III). Also, semantically congruent spoken words facilitated the memory encoding of written words in young adults and in children (Studies I, III). Previously, Delogu and colleagues (2009) found that working memory performance for congruent audiovisual words is better than for written words. However, in Study II with young adults, the memory for written words was not affected by congruent spoken words compared with unisensory presentation of words, and memory for spoken words were not affected by congruent written words compared with unisensory spoken word presentation. This contradiction between studies I and II is most likely due to differences in the experimental designs. In Study I, congruent word pairs were presented intermingled with incongruent word pairs and written words paired with auditory noise. In this kind of design, a non-semantic noise stimuli presented in the other modality might have slightly interfered the encoding of target words, and the congruency effect observed might have been boosted due to that (see Thelen et al., 2015). This might suggest that environmental conditions during encoding may affect the verbal congruency effect. In more "noisy" conditions, audiovisual congruency of verbal material can aid the memory encoding.

In the present studies, no interference effect due to the incongruent stimuli was observed. When the stimuli were paired with incongruent stimuli during encoding, there was no interference, that is, memory performance was not worse than when the stimuli were paired with non-semantic stimuli in the other modality. This suggests that incongruent stimuli from the other modality did not interfere the memory encoding of pictures, sounds or verbal material.

Taken together, the results of the present studies show that especially semantically congruent pictures facilitate auditory memory performance in children young adults, and elderly people. Also congruent written and spoken words can facilitate the memory encoding of their counterpart in the other modality in many conditions, but the effect is not as robust as the facilitation caused by semantically congruent pictures.

5.2. Conceptual Short Term Memory model

Conceptual Short Term Memory (CSTM) model (Potter, 1993; 2012) is a useful framework to explain how encoding is enhanced for congruent stimuli if the model is extended to the audiovisual domain. According to Potter (1993, 2012) CSTM is a memory buffer that integrates perceptual information with information held in long term memory and creates new representations that can be encoded into long term memory. Perceived stimuli are identified at a conceptual level quickly, and the related information in the long term memory is activated. New links modified by information from long term memory, current task goals and perceptual elements of the stimuli, are formed in CSTM. The resulting structure can be encoded into long term memory or held in working memory. Information that is not linked into these structures is forgotten. Chen and Spence (2010) have suggested that CSTM can also utilize audiovisual information. They suggested that when for example a semantically congruent sound is presented with a picture, the sound provides coherent information that enhances the encoding of the representation of the picture into CSTM. In contrast, incoherent information provided by an incongruent sound is not linked to the picture and is thus forgotten.

On the basis of the present studies I and II, it may be proposed that CSTM model may explain how congruent auditory stimuli facilitate the recognition memory of visual stimuli, and vice versa. When memorizing visual stimuli, congruent auditory stimuli might enhance the encoding of these stimuli in CSTM because of the coherent information that congruent stimuli provide. Incongruent auditory stimuli, however, do not provide coherent information within CSTM and therefore do not influence encoding. Therefore, memory performance is improved for visual stimuli presented together with semantically congruent auditory stimuli compared with visual stimuli presented alone. CSTM might operate in audition in a similar way as in vision. That is, when the task is to memorize auditory stimuli, congruent visual stimuli may improve the encoding of the auditory stimuli in CSTM by providing coherent information. During memory retrieval, when only the auditory stimulus is presented, the whole audiovisual representation is activated, causing a congruency effect. CSTM model might thus explain the audiovisual congruency effect.

CSTM model may explain why the semantically congruent pictures improved memory encoding more often than other kinds of stimuli. Pictorial information may offer qualitatively stronger memory cue that enhances the memory encoding within CSTM more than for example sounds. Furthermore, semantic access is faster for pictures than spoken words (Chen & Spence, 2013), which may enhance the effect of pictures in the CSTM.

5.3. Age-related differences in the congruency effect

Developmental differences in audiovisual processes may influence the way different age groups benefit from audiovisual information in memory encoding. Previous studies of audiovisual perception show that there are differences in how children and adults benefit from audiovisual information. Children under seven years of age are less likely to integrate auditory and visual stimuli than adults, and adult-like audiovisual integration develops in middle childhood (Brandwein et al., 2010). Previous studies show that school-aged children can benefit from semantically congruent audiovisual stimuli in

perceptual tasks (Jordan & Baker, 2011; Thomas, Nardini and Mareschal, 2017). This may be reflected also in the way children utilize audiovisual information in memory encoding. Study III shows that children utilized semantically congruent audiovisual information in their memory encoding in most conditions, and statistically significant differences between children and adults were not observed. This suggests that adult-like benefits on audiovisual memory encoding can occur at school age. Murray, Lewkowicz, Amedi and Wallace (2016) have proposed theoretical framework that takes account not only stimulus characteristics but also learned associations in multisensory processes. When children grow and acquire more experience of the world, their ability to perceive more complex multisensory associations grows. Study III suggest that school-aged children can already utilize learned audiovisual associations (picture of a cat and cat's vocalization) in memory encoding in adult-like manner.

In Study III, no age-related differences were found in the utilization of audiovisual information when comparing different age group or comparing children and adults. Children of 8, 10 and 12 years of age showed similar congruency effects in tasks with non-verbal stimuli, and there was no difference between 10- and 12-year-old children in the congruency effect for verbal stimuli either. A wider age range might have revealed developmental differences. For example, preschool children may have a weaker ability to benefit from audiovisual information.

In Study IV, elderly people utilized audiovisual information in memory encoding. However, there were some differences between younger adults and elderly people. The overall memory performance was poorer in the elderly people, in line with previous findings (Old and Naveh-Benjamin, 2008). When this baseline difference was taken in to account using gain indices, the elderly people benefited more from congruent audiovisual presentation than the young adults when encoding spoken words. Previous studies have shown that elderly people can benefit more from the presence of visual speech than younger adults when the task is to recall meaningless sentences (Maguinness et al., 2011). It can be speculated that not only visual speech but also other semantically congruent verbal visually presented information (e.g., written

words in Study IV) can aid elderly people to resolve the phonetic and semantic information in unpredictable speech (in this case a random sequence of unrelated spoken words) and therefore facilitate memory encoding relatively more than they do in younger adults.

However, according to Study IV, when encoding sounds young adults benefited more from audiovisual presentation than elderly participants. The elderly might have more often problems in processing briefly presented stimuli because of the general slowing of information processing (Salthouse, 1996, 2000). Memorizing brief sounds (400 ms) was a difficult task for elderly people, and semantically congruent pictures and written words improved memory encoding of sounds only little. Diedrich and colleagues (2008) proposed that elderly people are less likely than young adults to integrate sensory cues because of the age-related slowing in sensory processing, and that this slowing may result in a lower probability of audiovisual integration. However, further research is needed from this topic.

Taken together, children and young adults benefited similarly from audiovisual semantic congruency in memory encoding. Elderly people, however, benefited more from congruent pictures and written words than the young adults when encoding spoken words but less when encoding sounds. It may be that elderly people who have more experience of the world than young adults, can utilize the learned audiovisual associations to spoken words more efficiently than young adults to compensate the declining memory. However, this does not apply to natural sounds of objects that are difficult to remember for elderly people.

5.4. Future directions

In the future, it would be informative to investigate the benefits of audiovisual memory using a same recognition memory paradigm applied to different age groups from preschoolers to elderly people. Optimally the design would be longitudinal, but also a cross-sectional design would offer information about audiovisual memory across the life span. This would allow to investigate the maturation of audiovisual memory encoding.

Also, it would be informative to study possible practical applications of audiovisual memory cues. In education, the use of audiovisual semantically congruent information might be beneficial for learning especially in situations that require rote memory learning, such as foreign languages. Also audiovisual applications for rehabilitation of children with developmental learning disabilities, such as dyslexia or language impairment, could be an important topic for further research. Combining congruent auditory cue with written text might help poor readers to improve their reading skills. In the field of aging research, it would be important to investigate the benefits of audiovisual information for elderly people with mild memory impairments, and the possibilities for practical applications, for example, audiovisual memory aids.

6. CONCLUSIONS

The current series of studies shows that semantically congruent audiovisual information enhances memory performance in school-aged children, younger adults and elderly people. When an auditory or visual stimulus was presented together with a congruent stimulus of the other sensory modality during encoding, later recognition memory performance was better compared to stimuli accompanied by a non-semantic stimulus, an incongruent stimulus or no stimulus in the other modality. Especially semantically congruent pictures occurring during encoding of spoken words or sounds improved memory performance. In conclusion, semantically congruent audiovisual cues, especially semantically congruent pictures, can help memory encoding throughout the life span.

7. 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 (pp. 183–206). Oxford University Press, Oxford, UK.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617 – 45.

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. *Cerebral Cortex*, 21, 1042–55.

Calvert, G., Spence, C. & Stein, B. (2004). The handbook of multisensory processes. The MIT Press, Cambridge, MA, USA.

Cienkowski, K. & Carney, A. (2002). Audio-visual speech perception and aging. *Ear and Hearing*, 23, 439 –49.

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.

Cohen, M.A., Horowitz, T.S. & Wolfe, J.M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences of the United States of America,* 106, 6008–10.

Colavita, F. (1974). Human sensory dominance. *Perception & Psychophysics*,16,409, 12.

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 Neuropsychology*, 17, 173–96.

Damasio, A. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition,* 33, 25–62.

Delogu, F., Raffone, A. & Belardinelli, M.O. (2009). Semantic encoding in working memory: is there a (multi)modality effect? *Memory,* 17, 655–63.

Diaconescu, A., Hasher, L. & McIntosh, A. (2013). Visual dominance and multisensory integration changes with age. *NeuroImage*, 65, 152–66.

Diederich, A., Colonius, H. & Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia*, 46, 2556–62.

Dodd, B. (1979). Lip reading in infants: attention to speech presented in- and out - of - synchrony. *Cognitive Psychology*, 11, 478–84.

Ernst, M. (2008). Multisensory integration: A late bloomer. *Current Biology,* 18, 519- 21.

Ghazanfar, A. & Schroeder, C. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10, 278-85.

Goolkasian, P. & Foos, P.W. (2005). Bimodal format effects in working memory. *American Journal of Psychology,* 118*,* 61–7.

Gordon, M.S. & Allen, S. (2009). Audiovisual speech in older and younger adults: Integrating a distorted visual signal with speech in noise. *Experimental Aging Research,* 35, 202–19.

Gori, M., Del Viva, M., Sandini, G. & Burr, D. (2008). Young children do not integrate visual and haptic information. *Current Biology*, 18, 689–93.

Green D. M. & Swets J. A. (1966). Signal Detection Theory and Psychophysics. Peninsula Publishing, Los Altos Hills, CA, USA.

Habib, R., Nyberg, L. & Tulving, E. (2003). Hemispheric asymmetries of memory: The HERA model revisited. *Trends in Cognitive Sciences*, 7, 241–5.

Hillock, A., Powers, A. & Wallace, M. (2011). Binding sights and sounds: Age-related changes in multisensory temporal processing. *Neuropsychologia*, 49, 461–7.

Hockley, N. & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America,* 96, 3309.

James, W. (1890). The Principles of Psychology, H. Holt and Company, New York, NY, USA.

Jordan, K. & Baker, J. (2011). Multisensory information boosts numerical matching abilities in young children. *Developmental Science*, 14, 205–13.

Kuhl, P. & Meltzoff, A. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138–41.

Laurienti, P.J., Burdette, J.H., Maldjian, J.A. & Wallace, M.T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging,* 27, 1155–63

Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H. & Wallace, M.T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research,* 158, 405–14.

Lehmann, S. & Murray, M.M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research,* 24, 326–34.

Lewandowski, L.J. & Kobus, D.A. (1993). The effects of redundancy in bimodal word processing. *Human Performance,* 6, 229–39.

Luo, L. Hendriks, T. & Craik, F. (2007). Age differences in recollection: Three patterns of enhanced encoding. *Psychology and Aging*, 22, 269–80.

Macmillan, N. & Creelman, C. (2005). Detection Theory: A User's Guide. Cambridge University Press, New York, NY, USA.

Maguinness, C., Setti, A., Burke, K., Kenny, R. A. & Newell, F. (2011). The effect of combined sensory and semantic components on audio-visual speech perception in older adults. *Frontiers in Aging Neuroscience,* 3, 19.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development,* 55, 1777–88.

Matusz, P., Thelen, A., Amrein, S., Geiser, E., Anken, J. & Murray, M. (2015). The role of auditory cortices in the retrieval of single-trial auditory-visual object memories. *European Journal of Neuroscience*, 41, 699-708.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices, *Nature*, 264, 746–8.

Miller, J. (1996). The sampling distribution of d'. *Perception and Psychophysics*, 58, 65–72.

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, 452–65.

Moran, Z., Bachman, P., Pham, P., Cho, S., Cannon, T. & Shams, L. (2013). Multisensory encoding improves auditory recognition. *Multisensory Research*, 26, 581–92.

Morrell, L. K. (1968). Temporal characteristics of sensory interaction in choice reaction times. *Journal of Experimental Psychology,* 77, 14–18.

Murdock, B. (1982). Recognition memory. In: Puff, C.R. (ed.), Handbook of Research Methods in Human Memory and Cognition. Academic Press Inc., London, UK.

Murray, M.M., Foxe, J.J. & Wylie, G.R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage,* 27, 473–78.

Murray, M., Lewkowicz, D., Amedi, A. & Wallace, M. (2016). Multisensory processes: a balancing act across life span. *Trends in Neurosciences*, 39, 567–79.

Murray, M.M., Michel, C.M., 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*,* 125–35.

Nelson, D., Reed, V. & Walling, J. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 523–8.

Ngo, M.K., Sinnet, S. Soto-Faraco, S. & Spence, C. (2010). Repetition blindness and the Colavita effect. *Neuroscience Letters,* 480*,* 186–90.

Nyberg, L., Habib, R., McIntosh, A. & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America,* 97, 11120–24.

Old, S. & Naveh-Benjamin, M. (2008). Age-related changes in memory. In Hofer, S. & Alwin, D. (eds.) Handbook of Cognitive Aging (pp. 151–67). Sage Publications, Thousand Oaks, CA, USA.

Paivio, A., Rogers, T.B. & Smythe, P. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11, 137–8

Patterson, M. & Werker, J. (1999). Matching phonetic information in lips and voice is robust in 4.5-month old infants. *Infant Behavior and Development*, 22, 237–47.

Peiffer, A., Mozolic, J., Hugenschmidt, C. & Laurienti, P. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *NeuroReport*, 18, 1077–81

Potter, M.C. (1993). Very short-term conceptual memory. *Memory and Cognition,* 21*,* 156–61.

Potter, M.C. (2012). Conceptual short term memory in perception and thought. *Frontiers in Psychology,* 3, 113.

Quak, M., London, R. & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers of Human Neuroscience*, 9, 197.

Ross, L., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D. & Foxe, J. (2011). The development of multisensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, 33, 2329–37.

Ross, L., Saint-Amour, D., Leavitt, V., Javitt, D. & Foxe, J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex,* 17, 1147–53.

Salthouse, T. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–28.

Salthouse, T. (2000). Aging and measures of processing speed. *Biological Psychology*, 54, 35–54.

Schneider, T.R., Engel, A.K. & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology,* 55, 121–32.

Sekiyama, K. & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science,* 11, 306–20.

Sekiyama, K., Soshi, T. & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: a heightened McGurk effect in older adults. *Frontiers in Psychology,* 5, 323.

Setti, A., Burke, K. E., Kenny, R. & Newell, F. N. (2013). Susceptibility to a multisensory speech illusion in older persons is driven by perceptual processes. *Frontiers in Psychology,* 4, 575.

Shams L. & Seitz, A.R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences,* 12*,* 411–7.

Simon, J. R. & Craft, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of Experimental Psychology,* 86, 272–4.

Snodgrass, J. & Asiaghi, A. (1977). The pictorial superiority effect in recognition memory. *Bulletin of Psychonomic Society,* 10, 1–4.

Sommers, M., Tye-Murray, N. & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal hearing younger and older adults. *Ear and Hearing*, 26, 263–75.

Soto-Faraco, S., Kingstone, A. & Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia*, 41, 1847–62.

Stanislav, H. & Torodov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137–49.

Stephen, J., Knoefel, J., Adair, J., Hart, B. & Aine, C. (2010). Aging-related changes in auditory and visual integration measured with MEG. *Neuroscience Letters*, 484, 76–80.

Stevenson, R.A., Nelms, C. E., Baum, S.H., Zurkovsky, L., Barense, M.D., Newhouse, P.A. & Wallace, M.T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of Aging,* 36, 283–91.

Suied, C., Bonneel, N. & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research,* 194, 91–102.

Sumby W. H. & Pollack I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America,* 26, 212–5.

Thelen, A., Talsma, D. & Murray, M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*, 138, 148–60.

Thomas, R., Nardini, M. & Mareschal, D. (2017). The impact of semantically congruent and incongruent visual information on auditory object recognition across development. *Journal of Experimental Child Psychology*, 162, 72–88.

Thompson, V.A. & Paivio, A. (1994). Memory for pictures and sounds: independence of auditory and visual codes. *Canadian Journal of Experimental Psychology,* 48, 380–98.

Tremblay, C., Champoux, F., Voss, P., Bacon, B., Lepore, F. & Theoret, H. (2007). Speech and non-speech audio-visual illusions: a developmental study. *PlosOne*, 8, 742.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S. & Rose, N.S. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology,* 47, 31–7.

Ueno, D., Masumoto, K., Sutani, K. & Iwaki, S. (2015). Latency of modality-specific reactivation of auditory and visual information during episodic memory retrieval. *NeuroReport,* 26, 302–8.

Vaidya, C.J., Zhao, M., Desmond, J.E. & Gabrieli, J. (2002). Evidence for cortical encoding specifity in episodic memory: memory-induced re-activation of picture processing areas. *Neuropsychologia*, 40, 2136–43.

Wheeler, M., Petersen, S. & Buckner R. (2000). Memory´s echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America,* 97, 11125–9.

Winneke, A. H. & Phillips, N.A. (2011). Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. *Psychology of Aging,* 26, 427–38.