EFFECTS OF MULTISENSORY STIMULATION ON INFANTS' VISUAL PROCESSING AND LEARNING

Natasa Ganea

A thesis submitted for the degree of

Doctor of Philosophy (PhD)

University of London

2020

Goldsmiths, University of London, New Cross, SE14 6NW

I confirm that this thesis has been composed solely by myself and that I have not submitted it, in part or whole, for any other degree or diploma at the University of London or any other educational institution. The work presented is entirely my own, except for where it is stated otherwise by reference and acknowledgement. I declare that the intellectual content of this thesis is the product of my work, except for the assistance I received from others in the conceptualization of the research project, the coding of a small part of the data to establish inter-rater reliability, the selection of the presentation style, and the improvement of linguistic expression.

Signed: Natasa Ganea

Dated: 23/12/2020

To the babies who have participated in our studies:

Your curiosity and resilience have inspired this research.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisors, Dr. Andrew Bremner and Dr. Caspar Addyman, for their support and guidance throughout my PhD. Andy, when I started my PhD, I knew little about multisensory development. I am very grateful that you introduced me to this area of research. Without a doubt, you have influenced my understanding of the topic, and our discussions have motivated me to pursue further research in this area. Caspar, you became my supervisor halfway through my PhD, but you have been incredibly supportive, and I thank you so much for that. Your practical solutions to some of the data coding and study design issues I encountered speeded up my PhD progress and have made me see that sometimes complex problems have simple solutions.

A special thank you goes to the students who have volunteered in the Goldsmiths InfantLab and have helped me by coding some of the infant looking behaviour data for reliability purposes. Betül, Aanchal, Emma, Burcu, Nicky, Catherine, Marena, Megan, Sam, thank you very much! I hope that the research experience you gained in the InfantLab has helped you choose a career path.

During my PhD, I have had the support of my amazing colleagues and friends, who have not hesitated to offer me both professional and personal advice. Anna Lena, Bianca, Jiale, Giulia, Zehra, Jack, Maurice, Colin, Teea, you have helped me a lot! Thank you and a big hug!

Throughout the four years of my PhD, I have had the unconditional love and support of my parents. Although my parents live thousands of

kilometres away and have expertise in other areas, they have patiently listened to my research interpretations and advised me on how to write my PhD thesis faster. Mom, Dad, I am forever grateful for your support and for always believing in me! I love you!

Last but not least, I would like to thank the Economic and Social Research Council for funding my PhD, and the families that have participated in the studies. If it had not been for your motivation and enthusiasm, none of this research would have possible. Thank you very much, all!

PREFACE

I came across the Intersensory Redundancy Hypothesis, the theory I investigated in this thesis, at the beginning of my PhD. It drew my attention because it addressed an aspect of multisensory processing in which I was interested. Namely, how does multisensory stimulation affect infants' learning? Back then, the theory sounded intuitive. Plus, it made specific predictions about how infants would respond in different conditions - perfect for a theory-based approach to empirical research. Delving into the theory and the literature on multisensory processing in infants, children, and adults, I realised that there is a rift between the theory and some of the empirical findings. This rift is, in part, because researchers have used different methods to investigate the same topic. Besides, some of the terminology used in multisensory research is inconsistent, sometimes the same problem having two or three labels. Another challenge that I had to overcome during my PhD is that researchers disagree on the cognitive mechanisms underlying multisensory processing in infants. Many multisensory researchers assume that infants only associate cross-modal stimuli. As infants grow up, they gain perceptual experience, and they learn to fuse the different sensory inputs into multisensory representations. At the opposite end, many infant researchers assume that babies integrate multisensory information like adults do, and they perceive correlated audiovisual cues as features of the same object/event. In this thesis, I have tried to cover these different theoretical approaches and to provide comprehensive explanations for the various findings reported.

ABSTRACT

The research reported here investigated how the congruency of audiovisual stimulation affects infants' perceptual processing and learning. The research questions addressed were based on the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012). This theory states that certain kinds of stimulation (i.e., intersensory redundancy) drive infants' attention to specific object/event properties. The first study examined the effect of audiovisual stimulation on 10-month-old infants' encoding of object pattern. Since the results were inconclusive, the following experiment used a different set of visual stimuli. The second study found that the type of audiovisual stimulation that 10-month-old infants received did not affect their learning of object pattern. The third study examined whether younger infants, such as 4- and 6-month-olds, would be affected by the multisensory nature of stimulation. The experiment revealed that only the 6-month-old infants encoded the visual object pattern, and the effect was more robust when they received only visual stimulation. The fourth study extended these findings and showed that, contrary to the predictions of the IRH, 6-month-old infants learn both the pattern and the trajectory of objects when they receive congruent audiovisual stimulation. The final study investigated the effect that audiovisual gender matching on 6-month-old infants' perception of audiovisual speech synchrony. It revealed that, in the gender congruent condition, the infants looked longer at a video of a person who spoke in synchrony with a voice recording. This finding is inconsistent with the IRH and suggests that arbitrary cross-modal relations influence infants' responses to intersensory redundancy. Altogether, the results provide partial support for the IRH in indicating that infants' perception and learning is affected by multisensory stimulation. They also highlight some limitations of the IRH and directions for further research.

TABLE OF CONTENTS

Acknowledgements **4**

Preface **6**

Abstract **7**

Table of Contents **9**

List of Tables and Figures **12**

List of Abbreviations **15**

Chapter 1 .... **16**

Introduction **16**

1.1. Multisensory processing in adults 20

1.1.1. Rules of binding 21

1.1.2. Benefits of congruent stimulation 27

1.2. Multisensory processing in infants 33

1.2.1. Rules of binding 39

1.2.2. Benefits of congruent stimulation 42

1.3. Intersensory Redundancy Hypothesis (IRH) 46

1.3.1. Intersensory facilitation 49

1.3.2. Unimodal facilitation 51

1.3.3. Developmental and episodic changes 51

1.3.4. Task difficulty and processing expertise 53

1.3.5. Empirical findings inconsistent with the IRH 54

1.4. How I planned to test the IRH 57

1.4.1. Testing the unisensory facilitation prediction 57

1.4.2. Testing the developmental prediction 60

1.4.3. Testing the intersensory facilitation prediction 60

1.4.4. Testing the interaction between amodal and arbitrary relations 61

1.5. Thesis overview 64

Chapter 2 .... **67**

Effects of multisensory stimulation on 10-month-old infants' encoding of object pattern. Part 1 .... **67**

2.1. Introduction 68

2.2. Methods 74

2.2.1. Design 74

2.2.2. Participants 74

2.2.3. Apparatus and Stimuli 76

2.2.4. Procedure 80
2.3. Results 83
2.3.1. Exploratory Analysis 87
2.4. Discussion 92
Chapter 3 100
Effects of multisensory stimulation on 10-month-old infants' encoding of object pattern. Part 2 100
3.1. Introduction 101
3.2. Methods 107
3.2.1. Design 107
3.2.2. Participants 108
3.2.3. Apparatus and Stimuli 111
3.2.4. Procedure 114
3.3. Results 117
3.3.1. Exploratory Analysis 122
3.4. Discussion 124
Chapter 4 131
Effects of multisensory stimulation on 4- and 6-month-old infants' encoding of object pattern 131
4.1. Introduction 132
4.2. Methods 138
4.2.1. Design 138
4.2.2. Participants 138
4.2.3. Apparatus and Stimuli 142
4.2.4. Procedure 143
4.3. Results 145
4.3.1. Exploratory Analysis 153
4.4. Discussion 154
Chapter 5 162
Effects of multisensory stimulation on 6-month-old infants' encoding of object pattern and trajectory 162
5.1. Introduction 163
5.2. Methods 168
5.2.1. Design 168
5.2.2. Participants 168
5.2.3. Apparatus and Stimuli 171
5.2.4. Procedure 173

5.3. Results ... 175
5.3.1. Exploratory Analysis ... 182
5.4. Discussion ... 185
Chapter 6 ... 192
Interactions between audiovisual gender correspondences and speech synchrony perception in 6-month-old infants ... 192
6.1. Introduction ... 193
6.2. Methods ... 199
6.2.1. Design ... 199
6.2.2. Participants ... 200
6.2.3. Apparatus and Stimuli ... 202
6.2.4. Procedure ... 207
6.3. Results ... 209
6.3.1. Exploratory Analysis ... 213
6.4. Discussion ... 215
Chapter 7 ... 221
Discussion ... 221
7.1. Summary of findings ... 223
7.1.1. Ten-month-old infants' encoding of object pattern. Part 1 and 2 ... 223
7.1.2. Four- and six-month-old infants' encoding of object pattern ... 225
7.1.3. Six-month-old infants' encoding of object pattern and trajectory ... 227
7.1.4. Six-month-old infants' detection of speech synchrony ... 228
7.2. Theoretical implications ... 230
7.2.1. Implications for the IRH ... 230
7.2.2. Implications for multisensory development ... 237
7.3. Limitations ... 244
7.4. Future research ... 246
7.5. Conclusions ... 249
References ... 250

LIST OF TABLES AND FIGURES

Figure 1.1. The number of articles retrieved by PubMed on four areas of infant development. 34

Table 2.1. Participant characteristics by habituation condition. 76

Figure 2.1. Displays shown to infants and stimuli dimensions 78

Figure 2.2. Timeline detailing the study procedure. 82

Figure 2.3. Individual looking times (in seconds) at the test events. 85

Figure 2.4. (A) Individual looking times (in seconds) in the Last Habituation trial and the No Change Post-test trial, and (B) mean looking time in the first four and the last four habituation trials. 86

Table 2.2. The number of infants who preferred either the Change or the No Change event at Pre-test and who either maintained the same looking preference at Post-test or displayed the opposite preference. 89

Table 2.3. Pearson correlation coefficients between the infants' visual preference for the Change event ($\text{PTLT}_{\text{change}}$), at Pre-test and Post-test, and various participant characteristics. 90

Table 3.1. Participant characteristics by habituation condition. 110

Figure 3.1. Displays shown to the infants and stimuli dimensions. 112

Figure 3.2. Habituation conditions employed in the study. 113

Figure 3.3. Timeline detailing the study procedure. 115

Figure 3.4. (A) Individual looking times (in seconds) at the test events, and (B) looking preference for the Change event across the three test blocks. 118

Figure 3.5. (A) Mean looking time (in seconds) in the first four and the last four habituation trials, and (B) individual looking times (in seconds) in the Last Habituation trial and the first No Change test trial. 119

Table 3.2. Average looking time (log10-transformed data) at the two test events (Change vs. No Change) across habituation conditions. 121

Table 3.3. Pearson correlation coefficients between the infants' visual preference for the Change event ($\text{PTLT}_{\text{change}}$) and various participant characteristics. 124

Table 4.1. Participant characteristics by habituation condition for the 4-month-old age group. 140

Table 4.2. Participant characteristics by habituation condition for the 6-month-old age group. 141
Figure 4.1. Habituation conditions employed in the study. 143
Figure 4.2. Timeline detailing the study procedure. 144
Table 4.3. The number of participants who looked longer at the Change event in Test Block 1 and Wilcoxon signed ranks test results. 146
Figure 4.3. Individual looking times (in seconds) at the test events presented separately for each age group. 147
Figure 4.4. (A) Mean looking time (in seconds) in the first four and the last four habituation trials, and (B) individual looking times (in seconds) in the Last Habituation trial and the first No Change test trial. 148
Table 4.4. Average looking time (log10-transformed data) at the two test events (Change vs. No Change) in Test Block 1 presented separately for each age group of infants, and for the 6-month-old infants in each habituation condition. 152
Table 4.5. Pearson correlation coefficients between infants' visual preference for the Change event (PTLT$_{change}$) and various participant characteristics. 154
Table 5.1. Participant characteristics by habituation condition. 170
Figure 5.1. Habituation conditions employed in the study. 172
Figure 5.2. Test events used in the study. 172
Figure 5.3. Timeline detailing the study procedure. 174
Table 5.2. The number of participants who looked longer at the Change vs. No Change event, and at the Trajectory Change vs. No Change event, in each habituation condition and Wilcoxon signed ranks test results. 177
Figure 5.4. Individual looking times (in seconds) at the test events. 178
Figure 5.5. (A) Mean looking time (in seconds) in the first four and the last four habituation trials, and (B) individual looking times (in seconds) in the Last Habituation trial and the No Change test trial. 179
Table 5.3. Average looking time (log10-transformed data) at the three test events (Trajectory Change vs. Change vs. No Change) presented separately for in each habituation condition. 180
Figure 5.6. Mean looking times (in seconds) at the test events in the previous and the current studies. 183

Table 5.4. Average looking time (log10-transformed data) at the two test events (Change vs. No Change), pooled across the 6-month-old infants in the current study and those in the previous study (Test Block 1; see Chapter 4). 184
Table 6.1. Participant characteristics by voice condition. 202
Figure 6.1. Example of videos shown to infants and stimuli dimensions. 204
Table 6.2. Fundamental frequency characteristics (in Hertz) of the voice recordings and average voice-ratings. 206
Figure 6.2. Schematic overview of the Synchronous and Asynchronous videos and waveforms of the female and male voice recordings. 207
Figure 6.3. Timeline detailing the study procedure. 209
Figure 6.4. Infants' scores on three different measures split by voice condition. 211

LIST OF ABBREVIATIONS

EQM PA	Early Motor Questionnaire Perception-Action section
IRH	Intersensory Redundancy Hypothesis
LT	Total looking time
Mullen VR	Mullen Visual Reception Scale
$\mathbf{PTLT}_{\mathbf{change}}$	Proportional total looking time to the *Change* test event
$\mathbf{PTLT}_{\mathbf{sync}}$	Proportional total looking time to the *Synchronous* test event

CHAPTER 1

Introduction

Humans function in a complex and dynamic environment where they receive information about objects, people, and scenes through various sensory modalities, including all of the classic Aristotelian modalities of vision, audition, touch, smell, and taste, plus more (Fulkerson, 2014; Hellier, 2016).[1] These sensory modalities do not only offer different information about the external world, but they also act as complementary sources of information which improve perception in ambiguous situations. For example, both seeing and hearing a person speaking enhances speech intelligibility in noisy conditions (MacLeod & Summerfield, 1987; Sumby & Pollack, 1954). Research shows that adults integrate information from different sensory modalities in an optimal way by keeping track of how reliable the information provided by each modality is in a given task (Alais & Burr, 2004; Ernst & Banks, 2002; Ernst & Bülthoff, 2004). However, before integrating multisensory information, the brain has to decide which cues originate from the same object/event and should be combined, and which come from different stimuli and should be segregated (Körding et al., 2007; Rohe & Noppeney, 2015b).

To solve this cross-modal binding problem, adults rely on the spatiotemporal and semantic relations between the cues (see Calvert, Spence, & Stein, 2004; Naumer & Kaiser, 2010), as well as on their prior perceptual experience and beliefs about the stimuli (Ernst, 2007; Rohe et al., 2019).[2] However, infants lack both the perceptual experience and precision

[1] Other sensory modalities are proprioception, interoception, perception of temperature, and perception of vestibular information.

[2] Different authors use different terms for describing the problem of how the brain combines and segregates multisensory cues, e.g., "crossmodal binding problem" (Spence, 2011), "correspondence problem" (Ernst, 2007), "causal inference" (Körding et al., 2007; Rohe & Noppeney, 2015a,b).

(Fenwick & Morrongiello, 1998; Lewkowicz, 1996, 2010) that adults have. Therefore, deciding which sensory cues to combine and which to keep separate may be challenging for them. In light of these perceptual problems, the question arises as to whether infants benefit from multisensory stimulation (see A. J. Bremner, Lewkowicz, & Spence, 2012; Lewkowicz & Kraebel, 2004). This thesis will investigate this research question further. It will first review some of the existing empirical findings, and then it will describe the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012). The IRH is a theoretical account of how multisensory information affects cognitive development. Finally, the thesis will present a series of studies on how different kinds of multisensory stimulation affect infants' visual processing and learning of social and physical objects.

The research that I will report in this thesis will focus on audiovisual stimulation because this area of multisensory processing has been studied the most in adults and infants, and arguably it is better understood. Furthermore, the studies will look at the effects of multisensory stimulation on visual processing because vision matures after birth (see Colombo, 2001), whereas auditory processing starts before birth (Lecanuet & Schaal, 1996; Rand & Lahav, 2014). This difference in prenatal perceptual experience between the visual and the auditory systems may explain why, in different multisensory tasks, children rely more on auditory cues than adults do (Innes-Brown et al., 2011; Massaro et al., 1986; Napolitano & Sloutsky, 2004; Nava & Pavani, 2013; Thomas et al., 2017). It is possible that, like children, infants attend more the auditory information which may interfere with their visual processing and learning (Robinson & Sloutsky, 2007a, 2007b, 2008). The studies reported here will investigate this possibility by systematically varying the spatiotemporal and semantic relations between the auditory and the visual cues.

In this chapter, I lay out the justification for the empirical work reported in this thesis. In the first section, I will review the literature on multisensory processing in adults. I will describe the conditions that favour multisensory binding. Then I will explain some of the benefits of congruent multisensory stimulation on adults' behaviour.

In the second section, I will review the empirical evidence on multisensory processing in infants. I will give examples of studies which demonstrate multisensory abilities in early life, and how those abilities develop with age. I will conclude the section by outlining some of the benefits of multisensory perception in infants.

In the third section of this chapter, I will discuss the IRH. The IRH is a theoretical framework which attempts to explain multisensory development in early life. This theory proposes that the sensory stimulation infants receive guides their attention and learning. After describing this theory, I will discuss some of the empirical evidence that is either consistent or inconsistent with the IRH.

In the final section, I will describe the approach that I planned to adopt to test the IRH, and I will provide methodological details about the testing paradigms used in the empirical chapters that follow: (1) habituation of looking behaviour and (2) preferential looking. Finally, I will end the chapter outlining the research questions that the studies reported in this thesis tried to answer.

1.1. Multisensory processing in adults

Most of the research on multisensory processing in humans has focused on adults (see Calvert et al., 2004). This research has revealed that, in certain situations, the input from one sensory modality can affect how adults perceive information in another modality. The Bouncing Discs Illusion reported by Sekuler, Sekular, & Lau (1997) exemplifies such a situation. In their study, Sekuler et al. showed participants visual displays in which two identical discs moved towards each other, met, and then moved apart. Since the discs had very similar trajectories whether they bounced off or streamed through each other, the display was visually ambiguous for participants, who had to report how they perceived the event. When the discs moved in silence, most of the participants reported seeing a streaming event. However, when the participants heard a brief sound (with a sharp onset) that coincided with the meeting of the discs, they reported seeing a bouncing event. The fact that the sound altered how the participants interpreted the visual event, suggests that vision and audition are interconnected and that they influence each other. It is unclear how this interconnection may have come about. However, participants' experience with collision events in the natural world, may have contributed to the formation of this close association (Shams et al., 2004).

The influence of one sensory modality over the other is not unidirectional. Visual perception, too, can alter auditory perception. For example, the syllable /ba/ is heard as /da/ when paired with the lip movement for /ga/, and the syllable /ga/ is perceived as /bga/ when paired with the lip movement for /ba/ (McGurk & MacDonald, 1976). This auditory phenomenon is known as the McGurk Effect, and it represents a particular form of multisensory integration whereby the resulting percept is independent of the visual or the auditory signal (see Partan & Marler, 1999). Empirical evidence suggests that this phenomenon occurs at a neurophysiologic level, in the brain's response to auditory stimuli (Colin et al., 2002; Sams et al., 1991).[3] Therefore, cross-modal interactions can take place both at the lower, perceptual level and at the higher, decision-making level of cognitive processing. Studying multisensory processing in infants might tell us more about how these interactions develop, and what role perceptual experience plays in them.

1.1.1. Rules of binding

While the studies mentioned above show that the brain integrates information across the senses, they do not address how adults decide which multisensory cues to combine. The empirical research conducted on this topic has revealed that the temporal, spatial, and semantic relationships between the cues play a critical role (Calvert et al., 2004; Naumer & Kaiser, 2010). To exemplify this, let us consider speech perception. Speech is a multisensory event in which the auditory and the visual cues are bound together into a single representation of a person speaking. For adults to perceive unified audiovisual speech, the voice modulations heard, and the articulatory lip-movements seen must be synchronous or to occur close in time. Some evidence in this regard comes from Dixon & Spitz (1980). Dixon

[3] Colin et al. (2002) found that, in a sequence of audiovisual speech stimuli (i.e., visual /ba/ - auditory /ba/), an incongruent articulatory lip-movement (i.e, visual /ga/ - auditory /ba/) gives rise to an auditory Mismatch Negativity (MMN), even if the auditory input remains unchanged. MMN is a brain response, recorded via Electroencephalography (EEG), in response to rarely occurring (deviant) stimulus in a sequence of frequent (standard) stimuli.

& Spitz showed participants a video recording of a man reading prose and asked participants to press a button repeatedly to introduce an increasing delay between the video and the sound. The participants had to stop pressing the button when they perceived that the two were out of sync. When the video led the sound, the participants needed an average time lag of 258 ms between the visual and the auditory signals to detect the asynchrony. However, when the sound led the video, the average time lag needed to notice the asynchrony was 131 ms.[4] Similar results were reported by Munhall, Gribble, Sacco, & Ward (1996), who studied the effects of temporal asynchrony on the McGurk Effect. Munhall et al. found that the strength of the McGurk Effect decreased significantly when the video led the sound by 240 ms, and the sound led the video by 60 ms. These findings show that (1) the visual and the auditory speech cues have to coexist for adults to perceive them as united, and (2) the brain tolerates small temporal asynchronies between the speech cues. This leniency towards temporal asynchronies has been reported with non-speech stimuli as well, such as musical instruments and objects (see Vatakis & Spence, 2010, for a review), and it may represent an adaptive response to the inherent difference between the speed of light and that of the sound.

Aside from being concurrent, the audiovisual speech cues have to be colocated or to occur close together in space to be fused into a single representation of an object/event. Evidence that this is the case comes from a frequently encountered phenomenon, the Ventriloquism Effect. The Ventriloquism Effect describes the mislocalization of a sound toward a

[4] When Dixon & Spitz (1980) used an audiovisual recording of a hammer striking a nail, they found the same pattern of responses but the threshold for detecting the asynchrony was shorter (i.e., 188 ms when the video led, and 75 ms when the sound led).

concurrent visual event which occurs in a different spatial location (Bertelson & Radeau, 1981; Howard & Templeton, 1966; Jackson, 1953). For example, when we watch a film, we perceive the voices as coming from the actors' lips and not from the loudspeakers positioned below the TV. However, the research conducted into the effects of spatial separation between the auditory and the visual speech cues has shown that participants no longer perceive the cues as coming from the same source when the degree of separation is too large. To study this phenomenon, Colin, Radeau, Deltenre, & Morais (2001) systematically increased the angle of separation between the audiovisual speech cues to 20º, 40º, 60º and 80º visual angle, such that the participants saw a video of a person speaking in front of them and heard the speaker's voice coming from the side. During the study, the participants had to judge whether the voice they heard came from the person that they saw speaking. In these conditions, the participants reported experiencing the Ventriloquism Effect in only 45%, 14%, 4% and 0% of the trials (see also Radeau & Colin, 1999). Although the participants no longer perceived the face and the voice as belonging to the same person when the angle of separation increased, they still integrated the speech cues into a meaningful percept, and they reported experiencing the McGurk Effect (see also Jones & Munhall, 1997).[5] Therefore, adults need the spatial proximity between the auditory and the visual cues to perceive them as belonging to the same object/event, but not necessarily to experience an interaction between the senses (see Spence, 2013, for a review).

[5] Jones & Jarick (2006) found that when the separation between the cues is larger than 90º visual angle (as it is the case when the sound source is behind the participant), the McGurk Effect diminishes significantly.

A third aspect that adults consider when attributing the auditory and the visual speech cues to the same person is the semantic correlation between the cues. A semantic correlation describes a relationship between specific auditory and visual features, that frequently co-occur within the same object or category (e.g., dogs bark, birds chirp, women have a higher-pitched voice). Evidence that adults are more likely to bind together the semantically congruent audiovisual speech cues than the incongruent cues comes from Vatakis & Spence (2007). In their experiments, Vatakis & Spence showed participants audiovisual speech stimuli that originated from the same location but had varying degrees of temporal asynchrony (i.e., either the video led the sound by 0 to 300 ms or the reverse). Furthermore, the stimuli were gender-matched (i.e., a female face uttered some words concurrently with a female voice) or they were gender-mismatched. On each trial, the participants had to judge if either the auditory or the visual stimulus came first. Vatakis & Spence reported that the participants found it harder to indicate the temporal order of the gender-matched stimuli, and they needed more time between the auditory and the visual stimuli to respond correctly (see also Vatakis, Ghazanfar, & Spence, 2008).[6] The fact that the participants found it difficult to separate the gender-matched stimuli suggests that, at least in the case of speech perception, adults use semantic correspondences to decide whether the incoming sensory signals originate from the same multisensory object/event.

Although the temporal, spatial, and semantic relations between the cross-modal cues help the adults decide what information to combine, it is

[6] Vatakis & Spence (2008) and Vatakis et al. (2008) failed to find similar effects with musical stimuli (e.g., someone playing a note on the piano), object actions (e.g., a ball falling), or monkey calls, which made the authors argue that speech is processed differently.

unclear how the brain binds those signals. The research conducted on multisensory integration in adults has revealed that one sensory modality tends to dominate the other (see Calvert et al., 2004). For example, adults' judgements about the location of a sound source are affected by the location of a concurrent visual stimulus (Bertelson & Radeau, 1981; Howard & Templeton, 1966; Jackson, 1953). Meanwhile, judging the temporal order of visual stimuli is improved or degraded by the auditory input (Shimojo et al., 2001).

One account that has attempted to explain these effects is the Modality Appropriateness Hypothesis (Welch & Warren, 1980). This account argues that the sensory modality that has better resolution dominates the multisensory percept. This sensory dominance occurs more often in spatial localization tasks, where the vision has a higher resolution than the audition. However, more recent studies have shown that the brain monitors the sensory cues and adjusts their influence in the integrated percept based on how reliable each one is in each task. Empirical evidence in this regard comes from Alais & Burr (2004). Alais & Burr presented participants with two compound audiovisual stimuli (e.g., a Gaussian blob accompanied by a click sound), and ask them to judge whether the second compound stimulus was more to the left than the first one. The visual and the auditory cues were spatially aligned in some compound stimuli and misaligned in others. Besides, the authors manipulated the size of the Gaussian blob. This manipulation allowed them to change the resolution of the visual cue. Alais & Burr found that when the cues were misaligned, and the Gaussian blob was big (i.e., the visual cue had low spatial resolution and was less reliable), the participants relied more on the location of the auditory cue for their judgements.

Similar findings have been reported in a visual-haptic integration task (see Ernst & Banks, 2002), which has led Ernst & Bülthoff (2004) to argue that the sensory dominance effect is dependent on the precision and reliability of each sensory estimate. As a result, the sensory modality that provides more precise and reliable estimates weighs more in the integrated percept than the less reliable ones. However, both Ernst & Banks and Alais & Burr (2004) employed tasks which required the participants to fuse two sensory cues because they occurred close in space and time. Outside the laboratory, the brain must choose what sensory signals to integrate. Researching this aspect, Rohe & Noppeney (2015a) found that the brain keeps track of both the cues' reliability and the (causal) structure of multisensory stimulation. In brief, Rohe & Noppeney proposed that adults hold two assumptions about the relations between multisensory cues. One assumption is that the signals originate from multiple independent sources (i.e., the segregation assumption). The other assumption is that the signals come from a common source (i.e., the integration assumption). These two assumptions have different prior probabilities that the organism has acquired through prior knowledge and perceptual experience (see Ernst, 2007; Rohe et al., 2019). Furthermore, their relative likelihood changes online depending on the spatiotemporal and semantic relations between the cues (see also Körding et al., 2007).

When considering the various models of multisensory integration, a common theme becomes apparent. Multisensory integration is a complex cognitive process that relies on perceptual precision, prior experience, working memory, and attention. Since these abilities have not developed fully in infants, it is reasonable to wonder whether infants can differentiate correctly between congruent and incongruent cross-modal cues. A related issue is whether infants benefit from multisensory stimulation as adults do.

1.1.2. Benefits of congruent stimulation

Despite the complexity of the multisensory integration process, when the cross-modal cues are optimally integrated, they prove beneficial for the individual. For example, when adults have to respond to a flashing LED, a burst of white noise or a compound stimulus made up of the flashing LED and the white noise, which are spatiotemporally congruent, adults consistently respond faster to the compound audiovisual stimulus (J. Miller, 1982; Mordkoff & Yantis, 1991; Schröger & Widmann, 1998). The scientists investigating this phenomenon cannot agree on its underlying cognitive mechanism (Hughes et al., 1994; J. Miller, 1982; Raab, 1962; Schröter et al., 2009; Townsend & Nozawa, 1997). However, the various studies conducted on the topic have revealed that providing spatiotemporally congruent audiovisual stimulation speeds up both participants' motor responses and the latency of their eye movements to the targets (Harrington & Peck, 1998; Hughes et al., 1994). A possible explanation for why participants detect faster audiovisual stimuli is that both the manual responses and the saccadic eye movements involve attention switching and target selection. Some research suggests that these cognitive processes occur in the superior colliculus (Song et al., 2011), which is also a site for multisensory integration (Meredith & Stein, 1986; Jay & Sparks, 1987a, 1987b).

Animal studies have found that different sensory modalities send afferent signals to the superior colliculus (King, 1993; Stein & Meredith, 1993; Wallace & Stein, 1996). Most of this input converges onto single neurons located in the deep layer of this brain area, which renders them multisensory. Relative to other types of neurons, the multisensory neurons fire in response to stimuli that occur within the same area of the sensory space, irrespective of their modality (e.g., visual, auditory, etc.).[7] This phenomenon is possible because the neurons in the superior colliculus have spatially registered receptive fields that are anchored (predominantly) to the gaze direction of the animal (Stein et al., 2004). By using one frame of reference, the brain can keep the cross-modal sensory maps aligned. Besides responding to colocated stimuli, the multisensory neurons respond differently to collocated than dislocated stimuli. More specifically, when the brain receives spatially overlapping cross-modal input from the same object/event, these neurons have a higher discharge rate than when the sensory input is not overlapping. In return, this altered response affects the orienting behaviour of the animal and determines what object/event draws its attention. While these brain processes could explain why adults respond faster and more accurately to multisensory stimuli, they cannot explain the other cognitive and behavioural benefits.

Congruent multisensory stimulation does not only reduce reaction times, but it also facilitates perceptual discrimination in adults. Visual motion

[7] The multisensory neurons of kittens and infant monkeys lack the integrative properties of adult animals (Wallace & Stein, 1997, 2001). For example, in kittens, the researchers could detect enhanced discharge rate in response to spatiotemporally congruent stimuli only one month after the animal's birth. Besides developing better integrative properties, the number of multisensory neurons increases with age (see also Stein, 2012a).

perception is one area that benefits from spatiotemporally congruent audiovisual stimulation. To study this, Kim, Peters, & Shams (2012) manipulated the type of sound that the participants heard while they completed a perceptual task known as the random-dot kinematograms (see Hadad, Schwartz, Maurer, & Lewis, 2015, for details). During the study, the participants watched different visual displays in which many small dots moved across the screen. In some of the displays, a subset of the dots moved coherently in one direction while the remaining dots moved randomly. The participants' task was to indicate, after every other display, which one of the two displays had dots that moved together. While the participants completed the task, they heard bursts of white noise that were panned and gave the impression that they moved leftwards or rightwards between the loudspeakers. Kim et al. found that the participants were more accurate at detecting which visual display contained coherent motion when the direction of the sound mirrored that of the coherently moving dots. Interestingly, this was the case both when the congruent sound was informative (i.e., it accompanied only the coherent motion display) and when it was non-informative (i.e., both displays appeared alongside the same sound). Therefore, a spatiotemporally congruent sound enhances perceptual discrimination, while an incongruent sound does not.

Semantically congruent sounds too can help visual perception. In their experiments, Chen & Spence (2010) showed participants line-drawings of different animals and asked them to name the animals. The pictures appeared for only 27 ms, and then they were masked by a patterned rectangle. While the participants watched the stimuli, they heard brief sounds. The sounds were either semantically congruent with the pictures (e.g., a dog barking matched with a picture of a dog), semantically incongruent (e.g., a door creaking paired with an image of a bird), or a burst of white noise. Chen & Spence found that the participants were more accurate at naming the briefly presented pictures when the accompanying sound was semantically congruent than when it was incongruent, or when it was just white noise. Since the sounds and the target pictures coincided (temporally and spatially) throughout the study, these results are somewhat surprising. They suggest that the semantic congruency between the audiovisual cues specifying a familiar stimulus is more important for its identification than the spatiotemporal congruency. Supporting this interpretation is the fact that the participants performed similarly when the sound onset occurred approximately 300 ms after the picture onset (Chen & Spence, Exp. 3).

Although the cognitive mechanism underlying the semantic congruency effect is not fully understood, Murray et al.'s (2004) findings suggest that it is not limited to visual perception and that it occurs in memory as well. In their study, Murray et al. showed participants a sequence of images and asked them to decide whether the stimuli were old (i.e., had appeared previously) or new. The pictures that appeared twice were in the first instance presented on their own (*Visual* condition) or together with a semantically congruent sound (*Audiovisual* condition). The authors found that the participants were more accurate at identifying repeated audiovisual stimuli than visual stimuli, even if the delay between the first and the second presentation of the stimuli averaged 25 s (interval during which participants saw other pictures). Similar results were obtained with longer delays of approximately 50 s (Murray et al., 2005), but not when the sound and the picture were semantically incongruent (Lehmann & Murray, 2005). In other words, adults do not only detect faster semantically congruent audiovisual stimuli, but they also remembered them better.

Arguably, one of the most noticeable benefits of congruent audiovisual stimulation is in speech perception. Adults understand speech better when they both watch and listen to someone speaking than when they only hear them. Sumby & Pollack (1954) made one of the first attempts to demonstrate the effects of bimodal input on speech intelligibility. They asked participants to listen to a list of spoken words in different noisy conditions and to identify the words. The participants could either hear the speaker or they could both see and hear the speaker. Sumby & Pollack found that, in the *Auditory* condition, the number of correctly identified words decreased progressively as the speech signal became noisier. However, in the *Audiovisual* condition, word intelligibility reached 95% or more both when the speech-to-noise ratio was high (i.e., the participants heard the speech signal very well) and low (i.e., the background noise covered the speech signal). Therefore, the audiovisual presentation changed the participants' speech-detection thresholds. Following up on this observation, MacLeod & Summerfield (1987) reported that the speech signal could be 11 dB quieter in the *Audiovisual* condition, and the participants would report the same level of intelligibility.[8] A potential explanation for this enhancement is that the visual lip-reading allows the brain to form a coarse representation of the

[8] In MacLeod & Summerfield (1987), the gain in speech-reception thresholds varied between 6-15 dB across participants.

speech signal in the auditory cortex, which compensates for the degraded auditory input (Bourguignon et al., 2020). While the exact mechanism remains unclear, these findings suggest that adults do benefit from congruent audiovisual stimulation during speech perception.

Without a doubt, congruent cross-modal stimulation is helping adults to perform different cognitive tasks. One explanation for this phenomenon is that combining related cues is adaptive because it improves perception (see Calvert et al., 2004). A more mechanistic explanation for the phenomenon comes from Ernst & Banks (2002). Ernst & Banks asked their participants to estimate the height of a wooden block via vision and touch. In such cases, both sensory modalities provide an estimate of the block's features. These estimates can be more or less precise depending on the spatial resolution of each modality. For example, the distribution of their vision-based responses is less spread as adults have higher visual resolution and make similar estimates (i.e., the visual estimate is more precise and reliable). By contrast, the touch-based responses have a broader distribution, as adults provide more varied estimates. Ernst & Banks found that, when adults combine the visual and the haptic cues, they become more accurate and precise in their responses. Therefore, combining related multisensory information diminishes the noise (or variability) inherent in perceptual judgements, and increases the reliability of the calculated estimates (see also Nardini et al., 2008).

The literature reviewed so far indicates that perception often involves merging cross-modal information. The brain binds together auditory and visual cues when they occur close in time and space and are semantically related. This binding benefits adults' motor responses, visual perception, visual memory, and speech perception. In this review, I touched on some of the cognitive mechanisms that underlie multisensory integration in adults (see also Calvert et al., 2004; Naumer & Kaiser, 2010; Spence & Driver, 2004; Stein, 2012b). As discussed, the more recent accounts of the cross-modal binding problem offer a probabilistic explanation for this perceptual problem. Since these accounts emphasize the role of prior knowledge and experience, they raise the question of whether the infants process multisensory stimulation in the same way as the adults do. In the following section, I will discuss some of the research on multisensory processing in infants.

1.2. Multisensory processing in infants

Infants grow up in a multisensory environment, and they respond to stimuli presented in different sensory modalities, such as vision, audition, touch, from birth (Fantz, 1963; Ockleford et al., 1988; Sann & Streri, 2007). Although infants receive multisensory input, many researchers have formulated theories of perceptual and cognitive development by focusing on a single sensory modality at a time (Baillargeon, 2004; M. H. Johnson & de Haan, 2010; Mareschal et al., 2007; Spelke & Kinzler, 2007). Furthermore, many of the studies that have investigated cognitive development have employed a unimodal research approach (A. J. Bremner et al., 2012; see **Figure 1.1**).

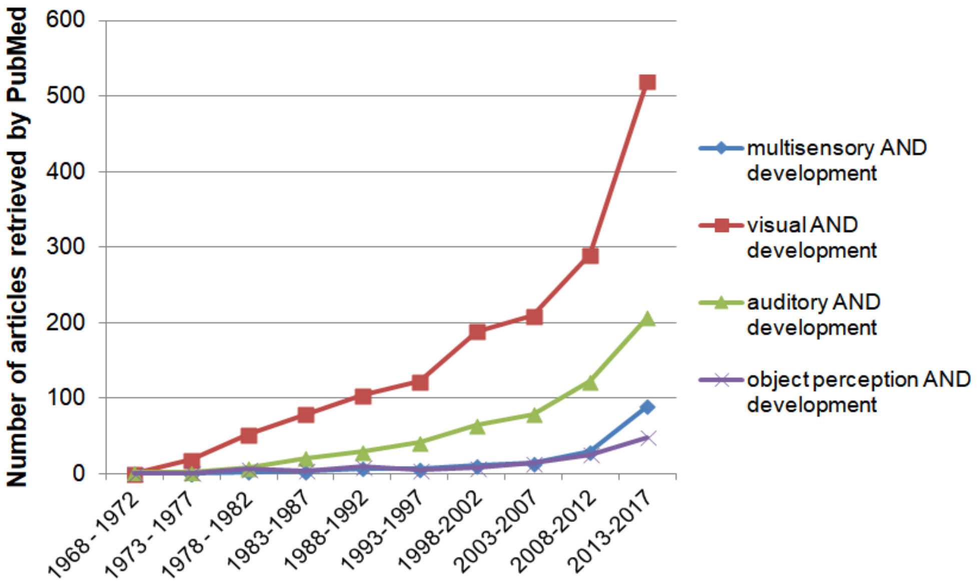

Figure 1.1. The number of articles retrieved by PubMed on four areas of infant development.

The articles focused on: multisensory perception, visual perception, auditory perception, and object perception in infants. I restricted the search to the time window between 1968 - 2017. For the topic of multisensory perception, the search words were: "development", "infancy" (or "infant", "early life"), "multisensory" (or "multimodal", or "crossmodal", or "intermodal", or "intersensory", or "multi-sensory", or "cross-modal", or "inter-modal", or "inter-sensory"). For visual perception, I replaced the search term "multisensory" with "vision" (or "visual"). Same for auditory perception, where the third search term was "audition" (or "auditory"). Finally, for object perception, the search terms were: "development", "infancy" (or "infant", or "early life"), "object", and "perception" (or "exploration", or "representation"). I limited the search to the title and the abstract of the articles.

This research approach has shed light on infants' cognitive abilities (see J. G. Bremner & Wachs, 2010). However, it may have also resulted in an incomplete picture of these abilities. For example, Kelly et al. (2007) reported that, toward the end of the first year of life, infants get better at discriminating between own-race faces and lose the ability to differentiate between other-race faces. This phenomenon is known as the Other-Race Effect, and it reflects infants' perceptual tuning to a specific category of faces with which they have more experience (Kelly et al., 2007; Liu et al., 2015; Xiao et al., 2013). However, most researchers have investigated the Other-

Race Effect by using either static images of faces (Kelly et al., 2007, 2009) or silent videos of moving faces (Wheeler et al., 2011; Xiao et al., 2013). When Minar & Lewkowicz (2018) used faces that articulated audible speech syllables, they found that 10- to 12-month-old infants discriminated between both own-race and other-race faces. Therefore, receiving congruent multisensory stimulation allowed the infants to make perceptual discriminations that otherwise they would not have been able to make based on visual stimulation only. However, when the incoming multisensory stimulation is incongruent, infants fail to learn or differentiate between stimuli despite succeeding on the exact same tasks when the stimulation is congruent (Barr et al., 2010; Begum Ali et al., 2015).

Despite the relatively limited number of studies on infants' multisensory processing, various studies have shown that, from very early on, babies are sensitive to the relations between stimuli presented in different sensory modalities (Bahrick, 1992; Lewkowicz & Ghazanfar, 2006; Nava et al., 2017; Orioli et al., 2018; Patterson & Werker, 1999; N. A. Smith et al., 2017; Walker-Andrews et al., 1991; Walker et al., 2010). Furthermore, a handful of studies have pointed to some rudimentary form of multisensory integration in infants rather than just sensitivity to cross-modal associations (Burnham & Dodd, 2004; Kushnerenko et al., 2008; Neil et al., 2006; Scheier et al., 2003).[9] One such study is that of Scheier et al., who found that infants,

[9] Multisensory integration refers to a cognitive (and neural) representation of an object/event in which information from different sensory modalities is merged and encapsulated. In effect, it is a multisensory representation of an object/event that includes its visual, auditory, tactile, etc. features. Cross-modal associations describe a connection or a link between separate stimuli presented in different sensory modalities. This link implies that the representation of each stimulus is kept separate. Researchers are still debating how best to differentiate between integration and association when investigating multisensory processing in infants who cannot report how they perceive the stimuli (see Stein et al., 2010).

like adults, experience the Bouncing Disks Illusion (see Section 1.1, for details). Scheier et al. (Exp. 1) habituated 4-, 6-, and 8-month-old infants with a video of two discs that moved toward one another and produced a sound when they coincided. After the habituation, the infants watched the familiar video as well as two novel videos. In the new films, the sound occurred either before or after the discs met. The 6- and the 8-month-old babies looked longer at the novel videos, behaviour that Scheier et al. interpreted as evidence that the infants perceived the habituation display as a bouncing event. However, this behavioural change could have also been due to the sound being time-shifted in the novel videos.

To investigate this alternative explanation, Scheier et al. (2003, Exp. 2) conducted a follow-up experiment. In this second experiment, after the habituation, Scheier et al. showed infants the familiar video and a novel one, which omitted the frame where the discs coincided. In effect, the researchers paused the video when the two discs were side-by-side, and this gave a group of adults the overwhelming impression that the discs had bounced off each other. To maintain the audiovisual temporal relationship established during the habituation, in the novel video, Scheier et al. played the sound during the video pause. The authors reasoned that, if the infants had interpreted the habituation display as a bouncing event, then they would look equally long at the novel and the familiar test videos. As expected, neither the 6- nor the 8-month-old infants differentiated between the test videos, which provided corroborating evidence that infants can perceive the Bouncing Discs Illusion from 6 months of age (but see Slater, 2003).

Around the same age, infants also experience the McGurk Effect (McGurk & MacDonald, 1976). As mentioned, the McGurk Effect describes a misperception of a speech sound when a mismatching lip movement accompanies the sound (e.g., adults report hearing /da/ when presented with an auditory /ba/ - visual /ga/ pair). In an attempt to study how infants process conflicting audiovisual speech stimuli, Burnham & Dodd, (2004) habituated two groups of 4.5-month-old infants with a live actor that quietly articulated either /ba/ or /ga/. While the infants watched the actor, they heard a gender-matched voice recording of someone uttering /ba/. After the habituation, the infants listened to some audio recordings of the syllables /ba/ and /da/ while the actor remained silent.[10] The researchers argued that, if the infants in the *Mismatch* habituation condition had experienced the McGurk Effect (i.e., they had perceived /da/ instead of /ba/), they would show a differential response to the test trials because the /ba/ sound was presumably more novel. Same with the infants in the *Match* condition, for whom the /da/ sound was relatively new. The results partially confirmed these predictions. Only the infants in the *Mismatch* condition differentiated between the test trials and looked longer at the silent actor when they heard /da/. Although it is unclear why the infants in the *Mismatch* condition exhibited this looking behaviour given that the /da/ sound was presumably more familiar, Burnham & Dodd interpreted their results as evidence that 4.5-month-old infants perceive the McGurk Effect.

More convincing evidence that young infants experience the McGurk Effect comes from Kushnerenko et al. (2008). Kushnerenko et al. showed a

[10] Burnham & Dodd (2004) used two different pronunciations for the syllable /da/. Specifically, infants heard either /da/ or /tha/ (as in 'that').

group of 5-month-old infants a woman who repeatedly articulated /ba/ or /ga/. In the *Match* trials, the infants saw and heard either the syllable /ba/ or /ga/. In the *Mismatch* trials, the researchers paired the visual /ba/ with the auditory /ga/ and the reverse. While the infants watched the videos, Kushnerenko et al. recorded their brain activity. The authors predicted that since the auditory /ba/ - visual /ga/ pair gives rise to the phonetically legal percept /da/, then the infants' brain response to these stimuli should be like that recorded in response to the *Match* trials. In the auditory /ga/ - visual /ba/ trials (an audiovisual combination that adults perceive as /bga/), the researchers expected the infants' brain to respond differently since this consonant cluster is phonetically illegal at the beginning of English words. The results confirmed the authors' predictions. The auditory /ga/ - visual /ba/ combination triggered a different brain response than the other audiovisual pairs. Specifically, higher voltage readings were recorded at sites over the frontal cortex in the time window from 290 ms to 590 ms after the sound onset. The fact that the brain did not differentiate between the mismatched auditory /ba/ - visual /ga/ pair and the matched /ba/ or /ga/ pairs provides robust evidence that 5-month-old infants experience a phonetically legal percept when presented with an auditory /ba/ - visual /ga/ pair. Measuring the infants' brain response rather than their looking behaviour during the study takes away any doubt as to whether the infants detected the audiovisual mismatch but failed to respond to it. That said, it is unclear whether the phonetically legal percept that infants experience when they see someone articulating /ba/ and they hear /ga/ is identical to the /da/ percept that adults frequently report hearing in these conditions.

1.2.1. Rules of binding

Although it is unclear whether the infants bind multisensory cues, just like the adults, the infants are sensitive to the temporal, spatial, and semantic relationships between incoming sensory stimuli.[11] For example, if an impact sound occurs close in time to a collision event, infants are likely to link the sound to the observed event. Lewkowicz (1996) conducted a systematic investigation into how close in time the auditory and the visual cues have to occur for infants to associate them. The researcher habituated 2- to 8-month-old infants with a video of a bouncing disc that generated a sound whenever it hit a surface and changed its direction of motion. After the habituation, the infants watched a few novel videos in which the impact sound was no longer synchronous with the visual collision (i.e., it occurred before or after the strike). Lewkowicz found that when the sound preceded the visual event by about 350 ms, the infants regained interest in the stimuli. And the same thing happened when the visual event occurred approximately 450 ms before the sound. The fact that the infants did not distinguish between the trials with synchronous audiovisual stimuli and those with offsets shorter than 350 ms suggests that, like the adults, the infants link together different sensory cues that occur within a particular time window from each other. But compared to the time window that the adults use to combine audiovisual stimuli (see Vatakis & Spence, 2010, for a review), the infants' time window seems to be broader (see also Lewkowicz, 2010).

Aside from the temporal proximity rule, infants consider how close in space the stimuli are when they associate them. Evidence regarding this

[11] I labelled this section "Rules of binding" to emphasize the parallelism with Section 1.1.1 (on multisensory processing in adults).

comes from studies on object-sound associations in infants. In one such study, Lawson (1980) familiarized 6-month-old infants with a toy that moved in synchrony with a spatially colocated or a spatially dislocated sound (to achieve this, the researcher either attached a loudspeaker to the toy or positioned it to the side of the display). After the familiarization, the infants saw two stationary toys (one familiar and one novel) and heard either the old tune or a new one playing in the background. Lawson reported that, when the novel song played, the infants looked equally long at both toys, but when the old tune played, the infants looked more at the familiar toy. However, this was the case only after the researcher familiarized the infants with a colocated sound, which made Lawson conclude that the spatial and temporal proximity between the audiovisual cues helped the infants attribute the sound to the object. In a more systematic investigation into the role of colocation in infants' object-sound associations, Fenwick & Morrongiello (1998) reproduced Lawson's findings and showed that 6-month-old infants needed the sound and the object to be within approximately 5^o visual angle of each other for infants to join them together. By comparison, Fenwick & Morrongiello found that 4-month-old infants made object-sound associations even when the audiovisual cues were approximately 10^o apart (i.e., the sound came from somewhere to the side of the object). These findings suggest that very young infants do not need precise spatial colocation between the auditory and the visual cues to link them together. However, as infants grow up, they learn to use spatial proximity to differentiate between related and unrelated auditory and visual stimuli (see also Morrongiello, Fenwick, & Nutley, 1998).

Lastly, when the infants cannot differentiate between related and unrelated audiovisual cues based solely on the spatiotemporal relations between them (e.g., when two people speak in synchrony, or two objects fall on the ground at the same time), they use the semantic correlations between the cues. Essentially, infants employ their previous experience with similar stimuli to judge how to pair the stimuli. Two sets of findings speak to this effect. The first one is that infants can make audiovisual gender matches when they see two people simultaneously uttering something. When shown two side-by-side videos of a man and a woman speaking in synchrony, infants look longer at the person whose voice they hear. If the speakers are the infants' parents, infants can match the faces and voices from 3.5 months of age (Spelke & Owsley, 1979). However, if the speakers are unfamiliar, infants can make gender matches from 6 months of age (Richoz et al., 2017; Walker-Andrews et al., 1991). The second set of findings is that infants make audiovisual number matches. Specifically, 4-month-old infants increase their looking at a video when it displays two balls bouncing in synchrony with a single tone. On the other hand, infants look longer at a video of a ball when its bouncing movement is synchronous with two differently tones (N. A. Smith et al., 2017). N. A. Smith et al. took the increased interest in the mismatching videos as evidence that infants segmented the auditory and the visual scenes and compared the number of elements between them (see also Bahrick, 1987, 1988, 1992; Jordan & Brannon, 2006; Ujiie, Kanazawa, & Yamaguchi, 2020). Since in everyday life, the number of falling objects matches that of the impact-sounds heard, infants may have found the mismatching videos more novel and captivating. Altogether, these results show that infants employ various rules to combine and segregate the multisensory stimulation they receive and that they learn and refine these rules with increasing perceptual experience.

1.2.2. Benefits of congruent stimulation

The findings reviewed above reveal similarities between infants' and adults' multisensory processing. However, they do not show whether spatiotemporally and semantically congruent audiovisual stimulation benefits infants. By definition, infants are immature organisms that are still learning how to integrate multisensory information and, as a result, they may not take advantage of it as adults do (see also Lewkowicz & Kraebel, 2004). That said, a few empirical studies have found that, in some perceptual tasks, infants perform better when they receive correlated audiovisual stimulation than unisensory stimulation. One such perceptual task is the optical superimposition task. When two videos are superimposed, the resulting visual stimulus is a complex one. Visual elements are partially occluded, and their boundaries are hard to detect. When faced with such displays, observers have to select what to look at and decide what belongs together. Bahrick, Walker, & Neisser (1981) found that playing a background sound that was congruent with one of the overlapping videos, helped 4-month-old infants focus their attention on the audible video. This ability became apparent when the researchers separated the superimposed videos, and the infants viewed them side-by-side. In these conditions, the infants looked longer at the video that had been silent during the superimposition. This preference for the previously silent video made Bahrick et al. conclude that, during the superimposition, the background sound made the audible video stand out and be more visible to the infants.

Bahrick et al.'s (1981) results show that spatiotemporally and semantically correlated stimulation enhances infants' visual perception. However, there is evidence that an arbitrary sound can also benefit perception if it is spatiotemporally congruent with the visual target. To study the effects of auditory stimulation on infants' visual discrimination, Wada et al. (2009) showed infants a few sequences of four briefly presented images. In each image-sequence, one picture contained an illusory contour figure, a stimulus that attracts infants' attention in a visual display (Otsuka et al., 2004; Otsuka & Yamaguchi, 2003). During the presentation, the infants heard brief sounds that coincided with the onset of each image. The sounds were either rare (they occurred only once during each sequence) or frequent. The 7-month-old infants looked longer at the side of the display containing the illusory contour figure when the accompanying sound was rare and failed to display any looking preference than when the sound was frequent. These results provide evidence that a concomitant salient sound enhances the discrimination of a visual target even if the two are semantically unrelated. While the cognitive mechanism behind this visual enhancement is unclear, one possibility is that the rare sound acts as an alerting signal for infants, who may become more attentive as a result.

The effects of congruent audiovisual stimulation are not specific to visual discrimination. Auditory perception, too, is enhanced when infants see and hear a target versus when they only listen for it. This perceptual facilitation was demonstrated by Morrongiello & Rocca (1987), who studied the head orienting behaviour of infants aged between 6- and 18-months-old in response to auditory and audiovisual stimuli. Morrongiello & Rocca placed each infant in front of a semicircular array of loudspeakers which delivered, in random order, short sequences of auditory clicks. During the *Auditory* trials, the researchers presented the clicks on their own, and during the *Audiovisual* trials, a spatiotemporally congruent light accompanied the clicks. Morrongiello & Rocca found that the infants' head orienting responses were more precise during the audiovisual trials (see also Neil et al., 2006).[12] On average, the infants turned their heads within 4º to 6º of the *Audiovisual* targets, and within 6º to 16º of the *Auditory* targets. Besides, the accuracy of the target localization remained stable across the ages in the case of bimodal targets, and it improved progressively in the case of unimodal targets. Thus, by 18 months of age, the infants were equally accurate at localizing targets based on auditory cues alone as well as audiovisual cues. These results are evocative of those reported by studies on sound localization in ferrets and barn owls (Hammond-Kenny et al., 2017; Whitchurch & Takahashi, 2006). The studies conducted on the development of sensory maps in the superior colliculus of these animals has revealed that correlated audiovisual stimulation helps refine the animals' auditory space maps (King, 2004; King et al., 2004). Therefore, the progressive improvements seen in sound localization in human infants may be due to the

[12] Neil et al. (2006) looked at infants' response latency to visual, auditory, and congruent audiovisual stimuli positioned at various degrees of eccentricity from the midline. They found that, when the stimuli appeared at 25º eccentricity, 2 to 10-month-old infants oriented faster to audiovisual stimuli than auditory stimuli. Similarly, when the stimuli appeared at 45º eccentricity, almost all the infants responded quicker to the audiovisual stimuli. Interestingly, at both eccentricities, 8 to 10-month-old infants responded faster to audiovisual stimuli than to the visual-only stimuli. Based on these findings, Neil et al. concluded that a rudimentary form of multisensory integration occurs in infants aged between 8 to 10 months old but not younger.

increasing experience with audiovisual objects/events that the infants acquire as they grow up.

Simultaneously seeing and hearing objects supports not only infants' perception but their memory as well. A common task that requires infants' memory is the interpretation of occlusion events. These events describe situations in which an observer no longer has visual access to a tracked object because another object blocks the view. When this happens, the observer must fill in the perceptual gap by representing the occluded object for a brief interval of time. Researching this interval, S. P. Johnson, Bremner, et al. (2003) found that the 4-month-old infants could represent an object for about 400 ms, while the 6-month-old infants could do that for roughly 600 ms. Since S. P. Johnson, Bremner, et al. had used visual-only stimuli, J. G. Bremner, Slater, Johnson, Mason, & Spring (2012) decided to extend the research to audiovisual stimuli. To this end, J. G. Bremner et al. (Exp. 1) habituated a group of 4-month-old infants with a video of a ball that oscillated behind a screen. While it moved, the ball appeared to generate a tune (i.e., the audiovisual cues were spatiotemporally congruent). After the habituation, the scientists removed the occluding screen and recorded how long the infants looked at two test videos presented in silence. In one video, the ball oscillated uninterrupted across the display, while in the other, the ball disappeared midway through the translation, just as it had done when the screen was present. The results showed that the infants looked longer at the test video that displayed the discontinuous trajectory, which the authors interpreted as evidence that the infants had represented the occluded trajectory of the ball as continuous during the habituation. In J. G. Bremner et al., the ball disappeared for approximately 600 ms at a time, an interval that proved too long for the 4-month-old infants in S. P. Johnson, Bremner, et al.'s study to represent the ball throughout the occlusion. Therefore, J. G. Bremner et al.'s results suggest that congruent audiovisual stimulation benefits 4-month-old infants' perception and memory of occlusion events (see also Kirkham, Wagner, Swan, & Johnson, 2012).

I started the literature review by presenting some studies which suggest that infants, like adults, are susceptible to the Bouncing Discs Illusion and the McGurk Effect (McGurk & MacDonald, 1976; Sekuler et al., 1997). I then presented empirical evidence that infants link together auditory and visual stimuli if they correlate (sufficiently) in time, space, and semantically. In this regard, I described studies which found that infants associate stimuli over a wider spatiotemporal gap than adults and that this gap reduces with age. Lastly, I discussed some of the benefits of congruent audiovisual stimulation on infants' perception and memory. The studies reviewed reveal similarities in the way infants and adults process multisensory information. However, some researchers (Bahrick & Lickliter, 2000, 2002, 2012) argue that infants benefit from congruent multisensory stimulation in some cognitive tasks, but not others. I will cover this theoretical perspective in the following section.

1.3. Intersensory Redundancy Hypothesis (IRH)

According to its proponents, the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012) is a developmental theory of selective attention, which argues that the sensory stimulation infants receive guides their perception and learning. More specifically, the IRH claims that when infants receive congruent or redundant multisensory stimulation, they attend to the object/event properties that are specified by multiple sensory modalities versus the properties that infants perceive through only one modality. In this section, I will describe the theory, and I will discuss some of the empirical evidence that is either consistent or inconsistent with it. By doing so, I hope to highlight the fact that testing the IRH could offer an insight into whether infants benefit from multisensory stimulation.

Before detailing the theory, I consider it is necessary to clarify what the phrase *intersensory redundancy* means. According to Bahrick & Lickliter (2000, p. 190) "*Intersensory redundancy* refers to [the] spatially coordinated and concurrent presentation of the same information (e.g., rate, rhythm, intensity) across two or more sense modalities". In essence, the spatial and the temporal information in auditory and visual stimuli is redundant when the stimuli are spatiotemporally congruent (such as when the impact between a falling object and a surface is both seen and heard). According to the IRH, the fact that the two stimuli coincide and originate from the same location in space makes the observer perceive them as a singular audiovisual object/event, which is defined by redundant or superfluous multisensory cues. Bahrick & Lickliter's definition of intersensory redundancy is consistent with J. J. Gibson's (1966, 1979) and E. J. Gibson's (1969) views of perception and perceptual development, but it overlooks the cross-modal binding problem. J. J. Gibson defines perception as a direct, unmediated process, which works in a bottom-up fashion. Meanwhile, E. J. Gibson argues that such a perceptual process is also naturally available crossmodally, whereby “amodal” features of perceptual stimulation are accessible independently of the infants' experience, and the senses which deliver them.

This assumption that amodal features of objects/events exist is the basis of Bahrick & Lickliter (2000; 2002; 2012) claim that *intersensory redundancy* is available in early childhood, and that infants can perceive automatically when multisensory stimulation is redundant or not. The empirical evidence reviewed in Section 1.2 offers a more nuanced view of multisensory perception in infants. Infants can pick up rudimentary spatiotemporal relations between auditory and visual cues, but these relations do not have to be as precise as adults need them to be to bind them (see Fenwick & Morrongiello, 1998; Lewkowicz, 1996, 2010).[13] In a complex environment, where auditory and visual stimulation coexists and is uninterrupted, the precision of the spatiotemporal associations between audiovisual cues is essential and, therefore, picking-up audiovisual redundancy may not be that easy for infants as the IRH argues. Having made this conceptual clarification, I will proceed to describe the IRH.

Audiovisual objects/events have numerous properties or features. They have visual (e.g., shape, colour, pattern), auditory (e.g., volume, pitch, timbre, intonation) and dynamic features (e.g., rhythm, tempo, trajectory, duration). The IRH divides these features into modality-specific properties and amodal properties. The modality-specific properties are those that are

[13] There is also evidence that infants process the auditory and the visual stimuli separately before they combine them into a multisensory representation. Chen & Westermann (2018) report that 10- and 15-month-old infants exhibit pupil dilation in response to both perceptual novelty (when either the visual or the auditory stimulus in a multisensory pair changes) and association novelty (when old visual and auditory stimuli are paired up in a new way). Interestingly, pupil dilation in response to perceptual novelty occurs earlier than pupil dilation in response to association novelty. This time difference in pupil dilation suggests that multisensory perception is not automatic and that infants process the auditory and visual stimuli separately before binding them together into a unitary representation.

specified by only one sensory modality (e.g., shape, pattern, colour, pitch, timbre). Meanwhile, the amodal properties are those that are defined concurrently by multiple sensory modalities (e.g., rhythm, tempo, trajectory, duration, location). Given this wide range of features that infants can attend to while exploring an audiovisual object/event, the IRH argues that there must be a serial processing of these properties (Bahrick & Lickliter, 2012). The reason being that infants have limited attentional resources which impedes them from processing multiple object/event properties at the same time. Furthermore, Bahrick & Lickliter make the case that young infants do not have endogenous control over their attention and, as a result, the saliency of the stimuli encountered guides their attention. Based on these two assumptions, the IRH claims that the type of stimulation infants receive determines the order in which infants process different object/event properties.

1.3.1. Intersensory facilitation

The IRH consists of four predictions. The first two predictions address the effects of unimodal and bimodal stimulation on young infants, and the remaining two predictions deal with the implications of the IRH across the life span. The first prediction is the *intersensory facilitation* prediction. It holds that infants learn better the amodal properties of an object/event when they receive congruent bimodal stimulation than unimodal stimulation. Furthermore, since the bimodal stimulation enhances the saliency of the amodal properties, infants detect and learn these properties before the modality-specific properties.[14] Some support from this prediction comes from Bahrick & Lickliter (2000), who studied how 5-month-old infants learn the rhythm of a dynamic event. During the study, the infants watched a video of a hammer repeatedly striking a surface. Some infants watched the video in silence (*Unimodal* condition), while other infants watched the video accompanied by a tapping sound. The sound was either synchronous with the video (*Congruent-Bimodal* condition) or asynchronous (*Incongruent-Bimodal* condition). When the infants habituated to the movement of the hammer, they watched in silence the same hammer striking the surface with a novel rhythm. Since the test and the habituation stimuli were identical in all but rhythm, longer looking at the former indicated that the infants noticed the rhythm change. Bahrick & Lickliter found that the infants in the *Congruent-Bimodal* condition looked longer at the test stimuli, while those in the *Unimodal* and *Incongruent* conditions did not. These results suggest that the infants learned the amodal properties of a dynamic event when they received congruent bimodal stimulation but not unimodal or incongruent stimulation (see also Bahrick, Flom, & Lickliter, 2002; Bahrick, McNew, Pruden, & Castellanos, 2019; J. G. Bremner et al., 2012; Flom & Bahrick, 2007; Kirkham, Wagner, et al., 2012). However, it is unclear whether, in this study, the infants learned the amodal properties of the tapping event better than its modality-specific properties because the authors did not test this aspect.

[14] In Bahrick & Lickliter (2012, p. 193) own words: "Rather, intersensory redundancy promotes attention to certain properties of stimulation (amodal) at the expense of other properties (modality-specific)."

1.3.2. Unimodal facilitation

The second prediction of the IRH is the *unimodal facilitation* prediction. This prediction holds that infants detect and learn the modality-specific properties of an object/event in conditions of unimodal stimulation but not congruent bimodal stimulation. According to the IRH, this facilitation occurs because, when there is no bimodal stimulation, there is little competition for attention from the amodal properties, which allows the infants to focus on the modality-specific properties. Consistent with this prediction, Bahrick, Lickliter, & Flom (2006) found that 3- and 5-month-old infants can discriminate the orientation of an object after they receive unimodal or asynchronous audiovisual stimulation but not after synchronous stimulation. As with Bahrick & Lickliter's (2000) study, the scientists habituated the infants to a video of a hammer repeatedly striking a surface. After the habituation, the infants watched a novel video in which the orientation of the hammer changed. Specifically, the hammer tapped downwards during the habituation and upwards at test. Bahrick et al. reported that only the infants in the unimodal and the incongruent bimodal conditions looked longer at the novel test stimuli. In other words, the visual-only and the incongruent stimulation promoted infants' processing of the visual properties of the tapping event, while the congruent stimulation hampered it (see also Bahrick, Hernandez-Reif, & Flom, 2005; Bahrick, Krogh-Jespersen, Argumosa, & Lopez, 2014; but see Minar & Lewkowicz, 2018).

1.3.3. Developmental and episodic changes

The third prediction addresses developmental changes in infants' selective attention. According to the IRH, the effects of stimulation described above are more robust in younger than older infants because, with age, infants gain more perceptual experience, their cognitive processing becomes more efficient, and their attention more flexible. These changes allow the infants to detect both the amodal and the modality-specific properties of objects/events irrespective of the stimulation that they receive. Various studies have found support for this prediction. For example, when 8-month-old infants are viewing the hammer tapping events described above, they detect both the modality-specific and the amodal properties of the events, irrespective whether they only watch the events or they synchronously watch and hear them (Bahrick et al., 2006; Bahrick & Lickliter, 2004). A similar improvement occurs in infants' processing of audiovisual speech stimuli (Flom & Bahrick, 2007), which suggests a more generalized improvement in infants' cognitive abilities with increasing age.

An extension of this third prediction is that, during an episode of exploration and throughout the infants' development, when congruent multisensory stimulation is available, infants first detect the amodal audiovisual relations, and then the arbitrary object-sound associations (Bahrick & Lickliter, 2002, 2012). Essentially, the amodal information constrains infants' exploration of the multisensory object/event and promotes the learning of the arbitrary/semantic relations. For example, 7-month-old infants learn the incidental association between a speech sound and an object when the latter moves in synchrony with the sound, but not when it is still or moves asynchronously (Gogate & Bahrick, 1998; see also Slater, Quinn, Brown, & Hayes, 1999). Furthermore, 2-month-old infants do not detect the arbitrary associations between speaking faces and their voice, but the 4- and 6-month-old infants do (Bahrick et al., 2005; see also Bahrick, 1994; but see Bahrick, 1992, Exp. 2 & 3; cf. Morrongiello, Fenwick, & Chance, 1998).[15] Since 3-month-old infants can detect amodal relations (see Bahrick & Lickliter, 2004) but not arbitrary ones, Bahrick et al. argue that this may reflect a developmental shift in cognitive processing from (global) amodal to (specific) arbitrary/semantic associations.

1.3.4. Task difficulty and processing expertise

The fourth prediction of the IRH holds that a mature observer experiences either *intersensory* or *unimodal facilitation* if the task demands are high and the stimuli are new. In other words, both the adults and the children (who have more perceptual experience than the infants) should differentiate and learn better the amodal properties of an object/event presented during a difficult task if they receive congruent audiovisual stimulation but not unimodal stimulation. Conversely, mature perceivers should discriminate and encode the modality-specific properties of the same object/event when the sensory input is unimodal. Support for this prediction comes from Bahrick et al. (2014), who found that preschool children remembered better the faces of people they saw in silent videos or in videos where the background voice was asynchronous. But they failed to memorize the faces when the voice heard was synchronous with the video (but see

[15] In Bahrick et al. (2005), two face-voice pairs appeared alternatively until the infants habituated. After the habituation, infants watched two novel events in which the duos crossed over (e.g., face A-voice B, face B-voice A). Only the older infants displayed visual recovery to the novel events, which the authors interpreted as evidence that only they had detected the arbitrary/semantic audiovisual relations during the habituation (see also Bahrick, 1994). In contrast, Morrongiello, Fenwick, & Chance (1998) habituated newborn infants with only one audiovisual pair, and they found that infants exhibited visual recovery when the object-sound association changed (see also Bahrick, 1992, Exp. 2 & 3). Therefore, it is debatable whether the 2-month-old infants in Bahrick et al. failed to make the face-voice associations because they attend to the amodal properties of the event or due to memory limitations.

Murray & Sperdin, 2010). This pattern of results reflects a *unimodal facilitation effect* which Bahrick et al. attributed to the high task demands. During the study, the children first watched six videos of different people speaking, and then they saw two side-by-side static faces and judged which one of them had appeared in the videos. Since each video lasted only 4 s, the children had only a limited interval of time to memorize the faces, which increased the difficulty of the task and led to *unimodal facilitation* (see also Barakat, Seitz, & Shams, 2015).

1.3.5. Empirical findings inconsistent with the IRH

Having reviewed the predictions of the IRH, I would like to present some findings that are inconsistent with the theory. The studies that I will describe below did not set-out to test the IRH. However, the auditory and the visual stimuli used were spatiotemporally congruent in some of the experimental conditions. Therefore, the findings of these studies are relevant for any debate about the IRH. As I mentioned above, the IRH defines *intersensory redundancy* as spatiotemporally congruent bimodal stimulation. Therefore, disrupting either the spatial or the temporal relations between auditory and visual stimuli results in non-redundant stimulation. One study that investigated the effect of spatial colocation on infants' perceptual learning is that of J. G. Bremner et al. (2011). J. G. Bremner et al. habituated 2-, 5-, and 8-month-old infants with a video of a ball that moved left-right across the display. While the infants watched the video, they heard a synchronous tune. The sound was either colocated (i.e., it moved together with the ball) or dislocated (i.e., it moved in the opposite direction from the ball). After the habituation, the infants watched two videos, one familiar and one novel. The former maintained the spatiotemporal relations established between stimuli, but the latter disrupted them. J. G. Bremner et al. found that, in the colocated sound condition, all the infants looked longer at the test video in which the ball and the sound moved in opposite directions. However, in the dislocated sound condition, only the 2-month-old infants preferred the test video in which the ball and sound moved together. These results suggest that the 2-month-old infants learned the rhythm of the auditory and the visual stimuli both when they were redundant and non-redundant (during the habituation phase). Since the rhythm is an amodal property of an object/event, J. G. Bremner et al.'s findings do not support the *intersensory facilitation* prediction of the IRH.

Other findings which are inconsistent with the IRH are those reported by Kirkham, Richardson, Wu, & Johnson (2012). In a series of experiments looking at infants' encoding of the location of audiovisual events, Kirkham, Richardson, et al. familiarized different groups of infants with two cartoon characters that appeared alternatively on the screen and moved in synchrony with a background sound. Each character had a unique tune, danced in a particular way, and occupied a specific rectangle on the screen. In essence, the infants watched two multisensory events. After the familiarisation, the rectangles were left empty, and the tunes associated with each character played sequentially in the background. The results revealed that, when the infants saw two cartoon characters during the familiarization, they looked longer at the empty rectangle associated with the tune heard. However, when the infants saw only one cartoon character, they did not exhibit any preference for either rectangle (see also Richardson & Kirkham, 2004). The IRH proposes that infants encode the location (an amodal property) of a redundantly-specified audiovisual object/event irrespective of its visual properties. In Kirkham, Richardson, et al.'s study, the characters moved in synchrony with the background tune both when one and two characters appeared thus *intersensory redundancy* was present in both conditions. However, the infants learned the location of the multisensory events only when they saw two visually distinct characters. Therefore, the visual properties of audiovisual objects/events play a role in the spatial indexing of these objects/events, which conflicts the IRH.

As illustrated, the IRH is a theory of cognitive development that emphasizes the role of multisensory stimulation in infants' development. It argues that young infants attend to different properties of an object/event depending on whether multiple sensory modalities specify the object/event concurrently or not. The theory makes specific predictions about what infants learn in different sensory contexts, which has enabled scientists to test the theory. So far, various studies have lent their support to the IRH, but some are inconsistent with it. Given the inconsistent empirical findings and the fact that the IRH is backed primarily by research on the effects of multisensory stimulation on infants' processing of impact events, I decided to test this theory by looking at other types of events – object occlusion and audiovisual speech. In the next section, I will detail the approach that I planned to adopt in this regard.

1.4. How I planned to test the IRH

As mentioned above, one of the tenets of the IRH is that concurrent multisensory stimulation constrains infants' perception and learning. Considering the emphasis that the IRH is placing on multisensory stimulation, I set out to test the theory. I planned to test aspects of the IRH across the physical and social domains of perception, and across different age groups of infants. By doing so, I hoped to gain a better understanding of whether multisensory stimulation benefits infants' cognitive processing, and if so, how? In this section, I will describe the topics I sought to investigate and the methods I planned to use.

1.4.1. Testing the unisensory facilitation prediction

In Section 1.3.5, I presented some empirical findings that are inconsistent with the IRH. One point of contention is the extent to which infants process modality-specific object/event properties when exposed to redundant multisensory stimulation. The IRH holds that young infants learn better the modality-specific properties of objects/events when they perceive them through one versus multiple sensory modalities (i.e., the *unimodal facilitation* prediction). To test this prediction, I decided to look at how different types of sensory stimulation affect infants' learning of object pattern.

I chose to investigate visual pattern because it is a modality-specific object property that infants attend to from the first few days of life. For example, newborn infants look longer at images depicting schematic faces and concentric circles than plain-coloured patches (Fantz, 1963). Furthermore, infants start to use object pattern information to disambiguate occlusion events during their first year of life. Wilcox (1999) documented this development in a series of experiments conducted with infants aged between 4.5- and 7.5-months-old. The infants watched an occlusion event in which one or two objects moved horizontally, back and forth, behind a screen. Some infants saw one ball that swayed behind the screen, while others saw two balls - a dotted ball that disappeared behind the screen and a striped ball that reappeared (and the reverse). Wilcox found that, when the occluding screen was narrow, the infants who saw two differently patterned balls looked longer at the experimental display than the infants who saw only one ball. When the occluding screen was broad, the infants looked equally long at the one-ball and the two-ball occlusion events. These results suggest that the infants detected the change in the ball's pattern only when the spatiotemporal gap between the successive reappearances of the ball(s) was small. Interestingly, only the 7.5-month-old infants displayed this looking pattern. The 4.5-month-old infants did not differentiate between the one-ball and the two-ball occlusion events in either screen condition (see also Wilcox & Chapa, 2004; Wilcox, Smith, & Woods, 2011).

Considering this development in infants' use of object pattern information, I reasoned that manipulating the pattern would be ideal for capturing the effects of multisensory stimulation on different age groups of infants. To this end, I planned to employ an infant-controlled habituation paradigm. The infant habituation is a testing procedure during which a stimulus is (repeatedly) presented until the infant loses interest in it. With each presentation, the researcher records the infant's looking time at the picture/video/object to gauge their interest. When the infant reaches a specific looking criterion (e.g., in the last two trials, the infant looked at the picture/video/object for less than half of their looking time in the first two trials), the presentation is interrupted, and the researcher shows a new stimulus. The novel stimulus can vary in shape, pattern, colour, rhythm, tempo, sound, etc. and, if the infant regains interest in it, it suggests that they detected the change and discriminated between the current and the previous stimulus (see Colombo & Mitchell, 2009; Oakes, 2010, for reviews). This testing method allows the scientist to control the stimulation that infants receive during the habituation phase, and to probe for the stimulus properties that the infants discriminated and learned.

In my study, I intended to use a similar testing set-up as Wilcox (1999). As detailed above, Wilcox showed an object that moved silently across the display (i.e., unimodal stimulation). In contrast, I wanted to present it either unimodally or bimodally. Because I found it too complex to manipulate the sensory information that the infants received about an actual object while I maintained constant its speed, trajectory, and interval of occlusion, I planned to use a 3D computer-generated animation. Other studies have shown that 4-month-old infants attend to such computer-generated animations (J. G. Bremner et al., 2012; Kirkham, Wagner, et al., 2012), and they even detect the spatiotemporal relations between the movement of an object across the display and a musical sound that accompanies it (see also J. G. Bremner et al., 2011). More specifically, in such experimental conditions, the infants can differentiate between a panned sound that appears to move together with the object across the display and an equally balanced sound. Based on the IRH, I predicted that the infants would encode an object's pattern only when they see the object (i.e., unimodal stimulation) or the movement of the object is incongruent with the sound (i.e., incongruent stimulation), but not when the sound and the ball are spatiotemporally congruent.

1.4.2. Testing the developmental prediction

Because I wanted to capture the developmental change in infants' attention to object pattern and the effects of multisensory stimulation on different ages of infants, I planned to conduct the studies with 4-, 6-, and 10-month-old infants. In the process, I intended to test the developmental prediction of the IRH, which argues that the effect of multisensory stimulation is more noticeable in younger than older infants. These age groups seemed the most appropriate given Wilcox's (1999) findings, which suggested that infants' ability to use pattern information to individuate briefly occluded objects develops between 4 and 7 months of age (see also Needham, 1999). Based on these findings, I hypothesized that only the 6- and the 10-month-old infants would learn the object's pattern. Furthermore, I expected that the 6-month-old infants would do so only when the object is presented unimodally, as per IRH's predictions.

1.4.3. Testing the intersensory facilitation prediction

Another aspect that I wanted to test was whether spatiotemporally congruent stimulation affects only infants' learning of the modality-specific object properties, or it has a broader impact on their cognitive processing. The IRH argues that, when infants receive congruent multisensory stimulation, they pay attention to the amodal properties of an object instead of its modality-specific properties. The amodal property I intended to study was object trajectory. I wanted to look at this object property because

Kirkham, Wagner, et al. (2012) found that 4-month-old infants learn better how an object is moving when they receive spatiotemporally congruent audiovisual information about its motion path. Based on Kirkham, Wagner, et al.'s findings and the IRH, I expected 6-month-old infants to learn the trajectory of an object that is specified concurrently by both vision and audition. At the same time, I predicted that the infants would encode the pattern on an object only when they see the object (i.e., unimodal stimulation). I set out to compare how infants learned these two object properties when they received either unimodal or bimodal information because I wanted to understand whether infants prioritise the learning of a particular object property in some sensory contexts, as the IRH argues.

1.4.4. Testing the interaction between amodal and arbitrary relations

The last aspect that I planned to research was the effect of face-voice gender correspondences on infants' speech processing. As I mentioned above, the IRH argues that the detection of amodal (i.e., spatiotemporal) audiovisual relations enables infants to focus on unitary audiovisual events and constrains their learning of the arbitrary (i.e., semantic) relations. In other words, infants learn to associate a face with a voice through their repeated encounter with synchronous audiovisual speech (Bahrick et al., 2005). I reasoned that, if this is indeed the case, then infants' perception of audiovisual speech synchrony (an amodal relation) should remain unaltered irrespective whether the voice characteristics and the facial features of a speaker are gender-matched (an arbitrary relation) or not.[16] Both audiovisual

[16] There is inconsistency in how researchers classify face-voice gender correspondences. For example, Bahrick et al. (2005, p. 543) classify them as amodal relations. By contrast, Walker-Andrews (1994, p. 48) and Bahrick, Netto, & Hernandez-Reif

speech synchrony and gender associations occur naturally in the environment, and infants are sensitive to them from a very young age. For example, when 4.5-month-old infants see two speakers, one that says /a/ and the other one /i/, they look longer at the person who articulates the audible vowel (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999). Similarly, when 6-month-old infants see a man and a woman uttering something, they prefer to look at the gender-matched speaker irrespective whether their facial movements are synchronous or not (Richoz et al., 2017; Walker-Andrews et al., 1991; but see Patterson & Werker, 2002; Exp. 5).[17]

Interestingly, when audiovisual gender information and speech synchrony are conflicting, young infants do not show any looking preference. For example, when 4.5-month-old infants simultaneously watch a video of a man saying /a/ and another one of a woman saying /i/, while they hear a voice recording of a man articulating /i/, infants look at both videos for a similar interval of time (Patterson & Werker, 2002, Exp. 4). The fact that infants do not prefer the synchronously speaking woman suggests that, contrary to what the IRH argues, infants do not prioritise the processing of audiovisual speech synchrony over that of gender correspondences. Adults too are susceptible to audiovisual gender information when processing speech synchrony. As mentioned earlier in the chapter, Vatakis & Spence (2007) found that adults need more time between the gender-matched than between the gender-mismatched auditory and visual speech cues to judge

(1998, p.1263) state that gender relations are neither exclusively amodal nor arbitrary. However, Bahrick & Lickliter (2002, 2012) argue that face-voice associations are arbitrary.

[17] The 6-month-old infants in Patterson & Werker's (2002) study may have failed to match the face and the voice by gender because the auditory stimuli were too short. Patterson & Werker used isolated vowel sounds, whereas Walker-Andrews et al. (1991) and Richoz et al. (2017) used nursery rhymes.

their order correctly. These findings, together with those of Patterson & Werker, put into question whether the perception of audiovisual speech synchrony is direct or unmediated as the IRH argues and suggest that arbitrary audiovisual gender correspondences play a significant role in the perception of speech synchrony in both adults and infants.

To study the effect of audiovisual gender information on infants' speech perception, I planned to use a variant of the preferential looking paradigm (Fantz, 1958). In a preferential looking study, the infant sees two stimuli, placed side-by-side in front of them, and the researcher records their looking behaviour. The two pictures/videos appear for a set duration, and swap position across trials (this is to control for any side-bias). If the infant looks significantly longer at one picture/video, it suggests that they differentiated between the stimuli and preferred one over the other. In the variant that I intended to use (i.e., the intermodal preferential looking paradigm), the infant hears a sound while they watch the two pictures/videos. If the sound matches one of the visual stimuli and the infant detects this correspondence, they look longer at the matching picture/video than at the non-matching one (Golinkoff, Ma, Song, & Hirsh-Pasek, 2013).

I planned to investigate the effect of face-voice gender correspondences on infants' speech processing in 6-month-old infants because, at this age, infants match faces and voices by gender (Richoz et al., 2017; Walker-Andrews et al., 1991). Furthermore, 6-month-old infants can detect audiovisual synchrony when the stimuli are short, repetitive utterances (Baart et al., 2014; Kuhl & Meltzoff, 1982; Patterson & Werker, 1999). The study aimed to find out whether infants can detect speech synchrony irrespective whether the face and voice characteristics indicate speakers of the same gender or opposite genders. Given that the empirical evidence is inconsistent with the IRH, I predicted that the infants would be able to detect the speech synchrony only when the audiovisual speech cues are gender-matched.

1.5. Thesis overview

In this thesis, I set out to test the IRH. In doing so, I aimed to answer the following questions: (1) Does multisensory stimulation affect infants' visual processing and learning? (2) Do the effects change with age? (3) Do infants use their prior knowledge about face-voice gender correspondences to inform their speech processing? In the empirical chapters that follow, I will report several studies that I conducted on these topics.

Chapter 2 and Chapter 3 focus on whether multisensory stimulation affects 10-month-old infants' encoding of (visual) object pattern. According to the IRH, infants' processing of the modality-specific object properties, like the pattern, is hindered when the object that the infants are exploring is specified concurrently by multiple sensory modalities. The studies I report in these two chapters tried to investigate the *unimodal facilitation* prediction of the IRH. Based on the theory, I predicted that the infants would learn the pattern on an object and detect changes in this object property only if they receive unimodal or incongruent audiovisual stimulation.

Chapter 4 looks at the same issue, but from a developmental perspective. In this study, I wanted to find out whether 4- and 6-month-old infants learn the pattern on an object when they receive unimodal or bimodal stimulation. The IRH argues that the effects of congruent multisensory stimulation are more pronounced in younger than in older infants. Furthermore, some empirical studies suggest that 4-month-old infants do not spontaneously attend to the pattern on an object (see Wilcox, 1999; Needham 1999). Therefore, I predicted that only the 6-month-old infants who receive unimodal stimulation would encode the object's pattern.

Chapter 5 investigates another prediction of the IRH, namely the *intersensory facilitation* prediction. According to the IRH, spatiotemporally congruent audiovisual stimulation increases the saliency of those object properties that are specified by multiple sensory modalities (i.e., amodal properties such as onset, duration, tempo, rhythm, trajectory). As a result, infants attend to and process better these amodal properties instead of the modality-specific properties (e.g., colour, pattern, shape). In this study, I assessed 6-month-old infants' learning of both object pattern and object trajectory. I predicted that, if infants receive spatiotemporally congruent audiovisual information about an object, they learn only its trajectory. Instead, if the infants receive unimodal information, they encode only the object's pattern.

The last empirical chapter, Chapter 6, tries to answer a slightly different question about multisensory processing in infants. Namely, how does prior knowledge about face-voice gender correspondences affect infants' speech perception? The IRH differentiates between amodal (spatiotemporal) and arbitrary (semantic) relations between cross-modal cues. Furthermore, it argues that infants detect the former associations before the latter. In the case of speech perception, the synchrony between the lip movements and the voice modulations represents an amodal relation. Instead, the association between the face and the voice is an arbitrary relation. This last study investigated whether 6-month-old infants detect speech synchrony irrespective whether the audiovisual speech cues are gender-matched or gender-mismatched. Since there is some empirical evidence which suggests that face-voice gender correspondences affect infants' speech perception (Patterson & Werker, 2002), I predicted that only the infants who receive gender-congruent information would detect the audiovisual speech synchrony.

CHAPTER 2

Effects of multisensory stimulation on 10-month-old infants' encoding of object pattern. Part 1

2.1. Introduction

As discussed in Chapter 1, an area of infant cognition that seems to benefit from concurrent multisensory stimulation is the perception of occlusion events. When objects/people move across space and time, they often disappear behind other items. Understanding that a briefly occluded object does not just disappear and then reappear a few moments later in a different place but that it continues its trajectory behind the occluder is a representational skill that emerges over the first few months of life (J. G. Bremner et al., 2012, 2013; S. P. Johnson, Bremner, et al., 2003; Tham et al., 2019).

A significant factor in the development of this ability is the size of the spatiotemporal gap between the successive reappearances of the visually tracked object. For example, 6-month-old infants can perceive the trajectory of the briefly hidden object when the spatiotemporal gap is ~667 ms long. By contrast, 4-month-old infants can do that only when the gap is less than ~400 ms long (S. P. Johnson, Bremner, et al., 2003). Interestingly, when a temporally synchronous and spatially collocated sound accompanies the movement of the object, the 4-month-old infants can represent the occluded object for as long as the 6-month-old infants do (J. G. Bremner et al., 2012).

To investigate the role of multisensory information in infants' processing of occlusion events, J. G. Bremner et al. (2012) showed a group of 4-month-old infants a video of a ball that moved horizontally behind a screen. While the infants watched this occlusion event, they either heard a musical sound that was spatiotemporally congruent with the ball, and it appeared to originate from the ball, or the sound was incongruent and seemed to come from another object in the display. Furthermore, each time the ball disappeared behind the screen, it remained hidden for more than 600 ms at a time. To study whether the infants represented the trajectory of the occluded ball, after the habituation, the researchers removed the screen, and they measured how long the infants looked at two test videos. In one video, the familiar object moved across the display along a continuous trajectory. In the other video, the object moved along a discontinuous trajectory (i.e., the ball disappeared briefly half-way along the path). Given that the latter video was perceptually more similar to the occlusion event, the authors reasoned that longer looking at this test event indicated that the infants had represented the continuous trajectory of the ball behind the screen. J. G. Bremner et al. found that only those infants who watched the habituation display alongside a congruent musical sound looked longer at the test video which displayed the discontinuous trajectory. Since the occlusion interval used in this study was longer than what was previously thought to be the interval of time for which 4-month-old infants could represent a hidden object (S. P. Johnson, Bremner, et al., 2003), J. G. Bremner et al.'s results show that the congruent audiovisual cues helped the infants represented the ball over an extended period of occlusion (see also Tham et al., 2019).

Further evidence that multisensory congruent information about an object boosts infants ability to process occlusion events comes from an eye-tracking study by Kirkham, Wagner, Swan, & Johnson (2012). Kirkham, Wagner, et al. investigated the visual scanning pattern of two groups of 4-month-old infants who watched various objects swaying behind a screen.

One group of infants watched the objects move together with a spatiotemporally congruent sound, and the other group watched them move together with an incongruent sound. To measure the effect of sound type on infants' visual scanning, the authors recorded the number of saccades that the infants made to the area where the objects re-emerged from behind the screen and the latency of those saccades. Based on these two measures, the researchers split the infants' saccades between anticipatory saccades (initiated in less than 150 ms from the object's reappearance) and reactive saccades (initiated after 150 ms). Kirkham, Wagner, et al. found that, out of all the saccades that the infants made to the target area in the congruent sound condition, 50.3% were anticipatory saccades compared to 29.1% in the incongruent sound condition. The improved visual tracking in the congruent sound condition suggests that combining the visual and auditory motion cues of an object allows young infants to: (1) represent the occluded object for longer, and (2) make more accurate judgements about when and where the object will reappear.

One theory that might explain how congruent cross-modal cues support the processing of occlusion events in infants is the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012). According to IRH, concurrent multisensory stimulation that specifies the same object/event attracts infants' attention toward the object/event and guides how infants process its properties. More specifically, the IRH proposes that infants process first those object/event properties that are specified by different sensory modalities (i.e., amodal properties), such as the duration, spatial location, tempo, or rhythm of the object/event. After processing these properties, the infants attend to the object/event properties that are specified by only one sensory modality (modality-specific properties), such as the colour, pattern, pitch, or timbre of the object/event. Furthermore, the IRH argues that the processing precedence given to the amodal properties constrains infants' learning and memory about the multisensory object/event and guides their future exploration of that item. In other words, when infants explore a multisensory object/event, they learn its amodal properties at the expense of its modality-specific properties.

The findings described above (J. G. Bremner et al., 2012; Kirkham, Wagner, et al., 2012) lend their support to the IRH. The infants learned the trajectory of an object (i.e., an amodal property) when both vision and audition specified it. However, it is unclear whether this was the only object property that the infants learned during the familiarization, or they also encoded the object's colour and pattern. The colour and the pattern are two modality-specific properties that define the identity of an object. The IRH predicts that infants process these modality-specific properties when they perceive the object through one sensory modality. Before testing this hypothesis, it is necessary to understand what properties infants encode when they watch occlusion events in silence (i.e., unimodally).

Research shows that when infants watch a silent occlusion event, they learn both the trajectory of the object (S. P. Johnson, Bremner, et al., 2003) and its shape, pattern and colour (Wilcox, 1999; Wilcox & Baillargeon, 1998b) if the spatiotemporal gap between the successive sights of the occluded object is short. In a study similar to those described above, Wilcox & Baillargeon (1998b) showed 4-month-old infants an occlusion event in which one or two objects oscillated behind a screen. In the one-object condition, the infants saw a ball swaying behind the screen. In the two-object condition, the infants saw a ball disappearing behind the screen, and a box reappearing from behind it. Wilcox & Baillargeon (1998b) found that when the occluding screen was narrow, and the spatiotemporal gap between the successive reappearances of the object(s) was small, the infants in the two-object condition looked longer at the visual display than those in the one-object condition. However, when the occluding screen was wide, and the occlusion lasted more than 1 s, the infants looked equally long at the display irrespective whether they were in the one-object or two-object condition.[18]

Wilcox & Baillargeon (1998b) took the results as evidence that infants computed online the combined width of the ball and the box and compared it to the width of the occluding screen. They argued that, since the combined width of the objects was broader than that of the narrow-screen, the infants must have inferred that only one item could have hidden behind the screen. As a result, the infants were surprised to see reappearing from behind the narrow screen the box and not the ball (see also Wilcox, 1999). Although possible, it is unlikely that 4-month-old infants performed such complex computations during the presentation of the stimuli. More likely it is that, in the wide-screen condition, infants represented the ball for a shorter period than the occlusion interval. Hence, when the box re-emerged from behind the screen, the infants did not notice that the object had changed because

[18] Wilcox & Baillargeon (1998b) do not report how long the occlusion of the ball lasted. However, based on the width of the occluding screens (narrow-screen: 15.5 cm; wide-screen: 30 cm), the diameter of the ball (10.25 cm), and the speed of the ball (~12 cm/s), I estimated that in the narrow-screen test event the ball was occluded for ~438 ms, while in the wide-screen test event the occlusion lasted ~1646 ms. To reach these estimates, I subtracted the diameter of the ball from the width of the occluding screen. Then I divided the difference by the speed of the ball.

they could not remember what they had seen disappearing behind the screen. Putting aside the conflicting interpretations of the results, Wilcox & Baillargeon's findings and the other findings reviewed above provide evidence that 4-month-old infants learn various object properties (e.g., object trajectory and shape) when they watch a silent occlusion event.

This study aimed to find out whether receiving multisensory information about an occlusion event affects infants' learning of the occluded object's pattern. The IRH predicts that infants learn the modality-specific object properties (e.g., object's shape, pattern, colour) when they perceive an object though only one sensory modality (i.e., unimodal or bimodal-incongruent presentation). To test the IRH, I habituated two groups of 10-month-old infants with a ball that oscillated behind a box. In the *Congruent (Dynamic Sound)* condition, a dynamically panned musical sound accompanied the movement of the ball. Whereas, in the *Incongruent (Static Sound)* condition, the infants listened to an evenly balanced sound. After the habituation, the infants watched two silent test events. In one test event, the ball changed its pattern during the occlusion (i.e., it was half red and half green before it disappeared, and it reappeared chequered red and green; *Change* event). In the other event, the ball remained unchanged during the occlusion (*No Change* event). In the former test event, the pattern on the ball changed during the occlusion to reduce the working memory load (the occlusion interval was shorter than the inter-trial interval) and to consolidate infants' representation of the familiar object. To test for any intrinsic preferences, the infants saw the test events before the habituation (*Pre-test*) and after (*Post-test*).

I predicted that, at *Pre-test*, the infants would look equally long at the two test events. At *Post-test*, the habituation condition would have differential effects on the infants' looking behaviour. As per the IRH, the infants in the *Incongruent (Static Sound)* condition would look longer at the *Change* event. Whereas the infants in the *Congruent (Dynamic Sound)* condition would show no difference between test events.

2.2. Methods

2.2.1. Design

Infants were randomly assigned to one of two habituation conditions: *Congruent (Dynamic Sound)* or *Incongruent (Static Sound)*. To study the effects of multisensory stimulation on infants' encoding of object pattern, I employed a test re-test paradigm, in which the infants' looking times at two test events (*Change* vs. *No Change*) was measured both before (*Pre-test*) and after the habituation (*Post-test*). The looking time was defined as the total interval of time that the infants looked at the screen between the beginning and the end of the trial. In a nutshell, the study employed a 2 x 2 x 2 mixed design with Habituation Condition (*Congruent* vs. *Incongruent*) as a between-subjects factor, and Test Time (*Pre-test* vs. *Post-test*) and Test Event (*Change* vs. *No Change*) as within-subjects factors.

2.2.2. Participants

Fourteen 10-month-old infants (*M* = 305.36 days, range = 287-321 days, 5 females) participated in the study. Seven additional babies were tested (i.e., 33% of the total *N* = 21), but were excluded because of fussy behaviour which led to the non-completion of the study (*n* = 1), failure to reach habituation criteria (*n* = 5), and experimenter error (*n* = 1).[19] Participants were recruited via an invitation letter sent to families with young babies, living in South-East London. Those families that volunteered to take part in the research were contacted again by telephone and invited for a visit to our lab when the babies reached the age range of the study. The Ethics Committee board of Goldsmiths, University of London granted full approval for this research and the caregivers involved provided consent for their children to participate in the study.

Based on parental report, none of the infants had sight or hearing problems, and all the infants had been full-term pregnancies (i.e., 37 weeks or more). Most of the infants were White-Caucasian and came from middle-class families. The infants were randomly assigned to one of the two habituation conditions (*n* = 7 per condition). There were no significant differences between groups in infants' score on the Mullen Visual Reception Scale (Mullen VR; Mullen, 1995), score on the Early Motor Questionnaire – Perception-Action section (EMQ PA; Libertus & Landa, 2013), age, and accumulated looking time during the habituation (see **Table 2.1**). For the Mullen VR scale, I used the standardized T-score, whereas, for the EMQ PA, I used the raw-score because this measure has not been standardized. The EMQ PA raw-score is calculated by summing up the ratings that the caregivers provide for all the statements in one section. Since the ratings vary between -2 (the child does not show the behaviour described) and 2 (the child exhibits the behaviour), the calculated raw-score can be a negative number.

[19] The rate of attrition for visual habituation studies is 22% (Slaughter & Suddendorf, 2007; Wachs & Smitherman, 1985).

Table 2.1. Participant characteristics by habituation condition.

	Habituation Condition		T-Test		
	Congruent (Dynamic Sound) (*n* = 7)	**Incongruent (Static Sound)** (*n* = 7)	***t*-value**	***df***	***p*-value** (2-tailed)
Age (days)	299.71 (13.61)	311.00 (10.95)	1.71	12	.11
Mullen VR T-score	65.29 (6.65)	68.86 (7.31)	.96	12	.36
EMQ PA raw-score	14.43 (7.46)	13.00 (5.66)	.40	12	.69
Habituation LT (s)	131.75 (86.40)	87.86 (20.96)	1.31	6.70#	.24

Note. Values represent mean (SD). Age, infants' age (in days) at test; Mullen VR T-score, standardized score on the Mullen Visual Reception Scale; EQM PA raw-score, raw score on the Early Motor Questionnaire Perception-Action section; Habituation LT (s), accumulated looking time during the habituation (seconds). #For the Habituation LT (s) comparison, the degrees of freedom were adjusted because Levene's test for equality of variance was significant, F = 5.82, p = .03.

2.2.3. Apparatus and Stimuli

A PC, a 24” BenQ screen (resolution 1920 x 1080), and two loudspeakers placed right below the screen, at 50 cm from each other (as measured from the centre of one loudspeaker to the other), were used to present the stimuli. Infants’ looking to the stimuli was recorded using a surveillance video camera positioned under the screen and hidden from the infants’ view. The video recording was presented live on a second screen, located outside the testing booth, and was used by the researcher to judge the infants’ looking behaviour during the study. The presentation of the stimuli was controlled via an in-house computer script, using MATLAB 2017b and Psychtoolbox 3.0.13 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). The script also recorded the infants’ looking time at the screen and calculated habituation criteria for each infant.

To code the infant's looking behaviour during the study, the researcher pressed a key on the keyboard whenever the infant looked at the screen. To minimize coding bias, the researcher remained blind both to the experimental condition (*Congruent* vs. *Incongruent*) and the test event (*Change* vs. *No Change*) being watched by the infant. For n = 8 randomly selected video recordings, the infants' looking behaviour during the test trials was coded offline by a second coder who was blind to the study hypothesis, experimental condition, and test event. A two-way mixed intra-class correlation analysis with absolute-agreement (Trevethan, 2017) revealed an excellent inter-rater agreement on infants' total looking duration during the test trials, $ICC_{2,1}$ = .99.

During the habituation and the test trials, the infants watched an animation which depicted a room with a centrally located blue box and a colourful ball that oscillated behind the box (see **Figure 2.1,** for examples of stimuli and stimuli dimensions). The animation was designed in Blender (www.blender.org), and it was purposefully created to give the impression that these were actual 3D objects. In the animation, the ball moved with a constant speed of ~12.15^{o}/s and covered 30.37^{o} degrees of visual angle (as seen from 70 cm distance), from one side of the box to the other side, in 2.5 s. During this interval, the ball was fully visible for 533 ms on each side of the box and was fully occluded for 834 ms. The transition from full visibility to full occlusion or the reverse lasted 300 ms.

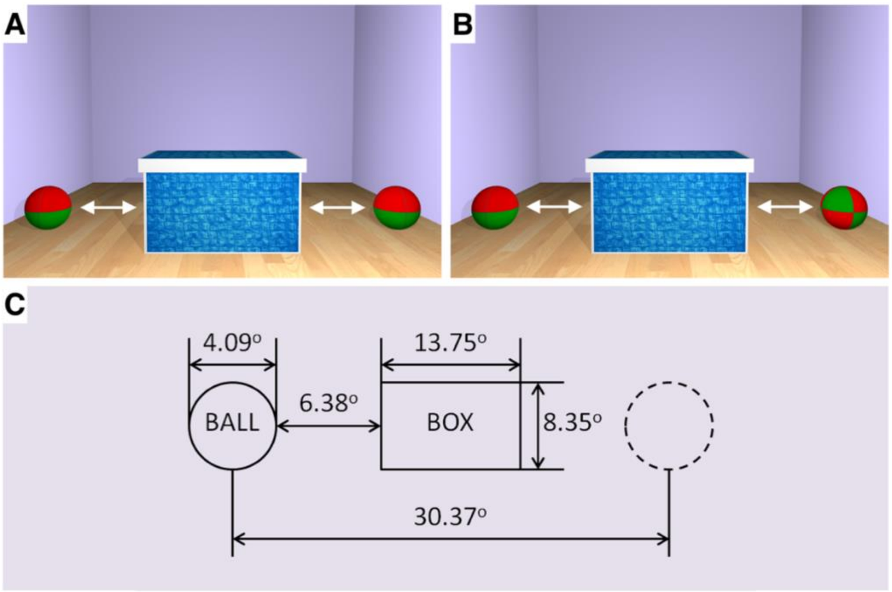

Figure 2.1. Displays shown to infants and stimuli dimensions.

(A) Habituation display and No Change test display. The ball kept its pattern during the occlusion (B) Change test display. The ball changed its pattern during the occlusion. (C) Schematic drawing of the display. Numbers represent the stimuli dimensions in degrees of visual angle, as seen from 70 cm distance (i.e., the infants' viewing distance).

The test trials were presented in silence, while the habituation trials were accompanied by a musical sound created based on a 1 s excerpt from Reich (1978, track 2). This excerpt was repeated 60 times to make a tune that lasted 60 s, had a tempo of 144 bpm, and a pitch of 440 Hz, as measured using the MIRtoolbox 1.7.2 (Lartillot & Toiviainen, 2007). In the *Congruent (Dynamic Sound)* habituation condition, the tune started and stopped when the animation started and stopped. Furthermore, the sound signal changed dynamically between the Left and the Right loudspeaker to mimic the translational motion of the ball from one side of the screen to the other side. Thus, when the ball was visible on the left-hand-side of the screen, the sound originated mainly from the Left loudspeaker, and as it progressed to the right-hand-side of the screen, the sound increasingly originated from the Right loudspeaker. To achieve this effect, the tune was edited in Audacity (www.audacityteam.org). In the *Incongruent (Static Sound)* habituation condition, the music onset and offset coincided with the trial onset and offset. However, the sound signal was evenly balanced across loudspeakers, and it did not change location as the ball moved across the screen. To ensure that the sound was equally loud across conditions, the amplitude of the two musical pieces was normalized using root-mean-square amplitude. When measured at 70 cm distance from the screen and the loudspeakers (i.e., the infants' viewing distance), the amplitude of the sound varied between 54 and 59 dB, reflecting inherent melodic variations in the tune.

Irrespective of the habituation condition, the infants watched (in silence) two test events both before and after the habituation. In the *No Change* test event, the infants saw a half red and half green ball that disappeared behind the box and reappeared on the other side of the box after a brief period (see **Figure 2.1A**). In the *Change* test event, the infants watched a half red and half green ball that disappeared behind the box, and a chequered red and green ball that reappeared from behind the box (see **Figure 2.1B**). Note that the *No Change* test display was identical to the habituation display.

To measure the infants' visual-perceptual skills, I conducted the Mullen VR scale (Mullen, 1995) and asked the caregivers to fill in the EMQ (Libertus & Landa, 2013). The Mullen VR scale is a set of standardized tasks designed to assess visual discrimination and visual memory in children aged between 0 and 45 months old. During the tasks, the experimenter shows different objects and images to the children and observes how the children respond to them. The tasks are organized by difficulty, such that basic visual skills, like focusing attention on a particular object or tracking a moving object through space, are assessed first. Success on these tasks is measured in terms of whether the child displays the desired skill or not, and the assessment is terminated when the child fails to score points on three consecutive tasks.

The EMQ, on the other hand, is a parental-report measure of the children's motor skills. It covers the ages of 2 to 24 months, and it consists of statements about the child's skills in everyday contexts. Parents have to rate each statement on a 5-point scale from -2 (the child does not show the skill) to +2 (the child shows the skill). The statements are organized in three sections: gross motor skills, fine motor skills, and perception-action skills, and the raw scores for each section are calculated by summing up the responses. When it was validated, Libertus & Landa (2013) found that parents' ratings on the EMQ Perception-Action section were positively correlated with the children's raw scores on the Mullen VR scale (based on *N* = 94 children aged 3 to 24 months old), thus suggesting that the two instruments may be measuring similar skills.

2.2.4. Procedure

The study was conducted in a dimly lit room with few visual distracters. During the study, the infants sat on their parents' lap, at ~70 cm from the stimuli presentation screen, and their eyes were aligned with the centre of the screen. Before the study started, the parents were instructed to look at their child's head and to refrain from redirecting their child's attention to the screen during the study. The presentation of the stimuli was controlled by the researcher, who remained outside the testing booth during the study.

The study started with 2 pre-test trials, followed by up to 12 habituation trials, and 2 post-test trials (see **Figure 2.2**). The habituation trials were presented until the infant's accumulated looking time in the last 4 trials (starting from the second trial) was less than half of the infant's accumulated looking time in the first 4 trials. All the infants completed at least 5 habituation trials. I used this infant sliding window of 4 trials, because I wanted to replicate J. G. Bremner et al.'s (2012) study.[20] Each trial lasted for 60 s or until the infants looked away from the screen for more than 2 s. If the trial duration was shorter than 2 s, the trial was repeated. This minimum trial duration interval was chosen because it took the ball ~2 s to travel across the display and re-emerge on the other side of the box. At the beginning of the trial, a 4^o x 4^o audiovisual looming animation attracted the infant's attention to the centre of the screen. The attention-getter had a minimum duration of 1.5 s and lasted until the infant looked at the screen.

Upon completing the study, which lasted ~7 minutes, the infants took a 5-minutes break, and then completed the Mullen VR Scale (Mullen, 1995), which took ~10 minutes. Finally, the parents were asked to fill in the EMQ (Libertus & Landa, 2013) while the researcher entertained the infants. At the end of their visit to our lab, the families were debriefed. As a thank you for

[20] Many infant studies use a sliding window of 3 trials, instead of 4 trials (J. G. Bremner et al., 2011; S. P. Johnson & Aslin, 1995; Spelke, Kestenbaum, Simons, & Wein, 1995; see also review by Oakes, 2010).

their participation in the study, the families received a certificate and a baby t-shirt.

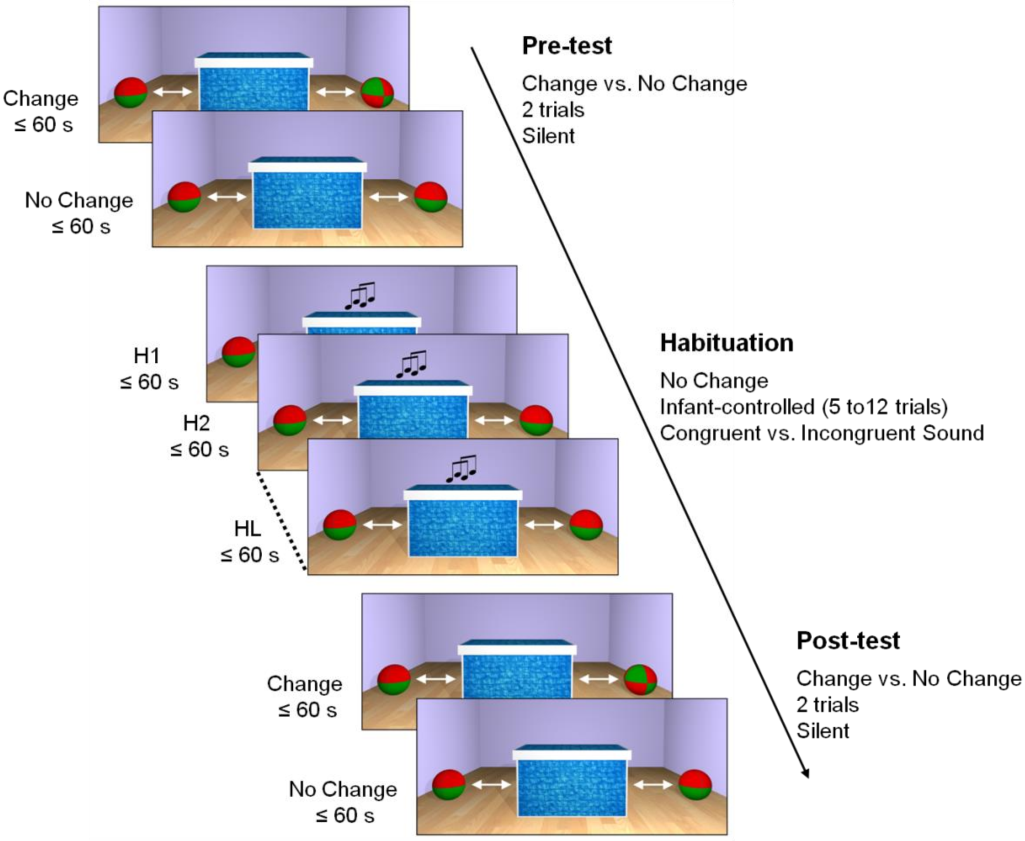

Figure 2.2. Timeline detailing the study procedure.

During Pre-test and Post-test, the infants watched two test events. One event depicting a ball that changed its pattern during the occlusion (Change test event), and the other one depicting a ball that kept its pattern during the occlusion (No Change test event). All the test events were presented in silence, and their presentation order was counterbalanced across participants. The habituation display was the same as the No Change test display. The habituation trials were accompanied by a musical sound that was either spatiotemporally congruent with the ball and "moved" left-right together with the ball (Congruent condition) or it was independent of the ball and remained located in the centre of the display throughout the trial (Incongruent condition). Habituation criterion was reached when the infants' accumulated looking time in the last 4 habituation trials was less than half of their accumulated looking time in the first 4 habituation trials (starting with the second trial). The infants completed between 5 and 12 habituation trials. Each trial lasted 60 s or until the infants looked away from the screen for more than 2 s. In between the trials, an audiovisual animation was presented at the centre of the screen.

2.3. Results

Figure 2.3 depicts the infants' individual looking times (in seconds) at the test events, across the *Congruent (Dynamic Sound)* and *Incongruent (Static Sound)* habituation conditions. In both conditions, the infants watched the *Change,* and the *No Change* test events for equally long, both before the habituation (*Pre-test*: 6 out of 14 infants preferred the *Change* event; Wilcoxon signed ranks test z = .09, p = .93, *ns*) and after the habituation (*Post-test*: 8 out of 14 infants; z = .53, p = .59, *ns*).

This was confirmed by a 2 (Test Event: *Change* vs. *No Change*) x 2 (Test Time: *Pre-test* vs. *Post-test*) x 2 (Habituation Condition: *Congruent* vs. *Incongruent*) mixed ANOVA conducted on $\log_{10}$-transformed data. The Test Event and the Test Time were manipulated within-subjects, the Habituation Condition was manipulated between-subjects. Log_{10}-transformed data rather than raw looking time data was used because the raw data was positively skewed (see also Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016). The analysis showed a main effect of Test Time, as the infants looked longer at the *Pre-test* (M = 1.28, SD = .22) than the *Post-test* trials (M = 1.02, SD = .27), F(1, 12) = 7.57, p = .02, η_p^2 = .39. No other main effects or interactions reached significance - Test Event: F(1, 12) = .26, p = .62, *ns*; Habituation Condition: F(1, 12) = .004, p = .951, *ns*; Test Time x Habituation Condition: F(1, 12) = 2.25, p = .16, *ns*; Test Event x Habituation Condition: F(1, 12) = .28, p = .61, *ns*; Test Time x Test Event: F(1, 12) < .001, p > .99, *ns*; Test Time x Habituation Condition x Test Event: F(1, 12) = .93, p = .35, *ns*.[21]

[21] I found similar results when Test Event Order at *Post-test* (*Change* or *No Change* was presented first) was included in the analysis as a between-subjects factor. The main

Given that the habituation and the No Change test displays were identical except for the sound, I compared the infants' looking times to the *Last Habituation* trial and the *No Change Post-test* trial (see **Figure 2.4A**).[22] This analysis was a way to check that the infants had associated the sound to the visual display and that the habituation interval influenced the infants' looking behaviour during the test events. For this analysis, I conducted a 2 (Trial Type: *Last Habituation* vs. *No Change Post-test*) x 2 (Habituation Condition: *Congruent* vs. *Incongruent*) mixed ANOVA on $\log_{10}$-transformed looking time data. Trial Type was manipulated within-subjects, Habituation Condition was manipulated between-subjects. The infants' looking time increased from the *Last Habituation* trial (M = .68, SD = .13) to the *No Change Post-test* trial (M = 1.01, SD = .31), $F(1, 12) = 20.56$, $p = .001$, η_p^2 = .63. No other main effects or interactions were found - Habituation Condition: $F(1, 12) < .001$, $p > .98$, *ns*; Trial Type x Habituation Condition: $F(1, 12) = 1.96$, $p = .19$, *ns*. The recovery in infants' attention from the *Last Habituation* trial to the *No Change Post-test* trial suggests that the infants may have associated the musical sound with the visual display, and they detected when the sound stopped at test.

effect of Test Event Order was non-significant, $F(1, 10) = 2.96$, $p = .12$, *ns*, and neither was any interaction between Test Event Order and other factors, all $F < .84$, all $p > .38$, *ns*.

[22] I chose the Last Habituation trial for this comparison because, presumably by this trial the infants have learned the sound-display association.

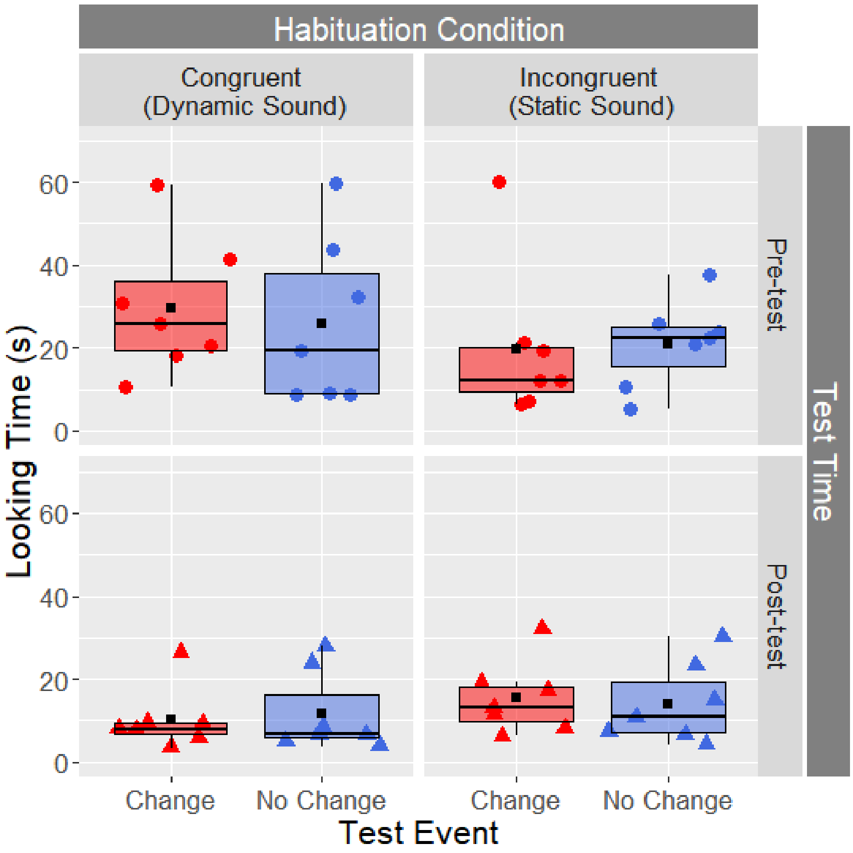

Figure 2.3. Individual looking times (in seconds) at the test events.

Change event (in red), the ball changed its pattern during the occlusion. No Change event (in blue), the ball kept its pattern during the occlusion. Looking times was measured before (Pre-test) and after the habituation (Post-test). The left panels depict looking times for infants in the Congruent (Dynamic Sound) habituation condition, where the musical sound was spatiotemporally congruent with the movement of the ball. The right panels depict looking times for infants in the Incongruent (Static Sound) habituation condition, where the musical sound was independent of the ball's movement and was located in the centre of the display. *Note*: Black dots represent mean values.

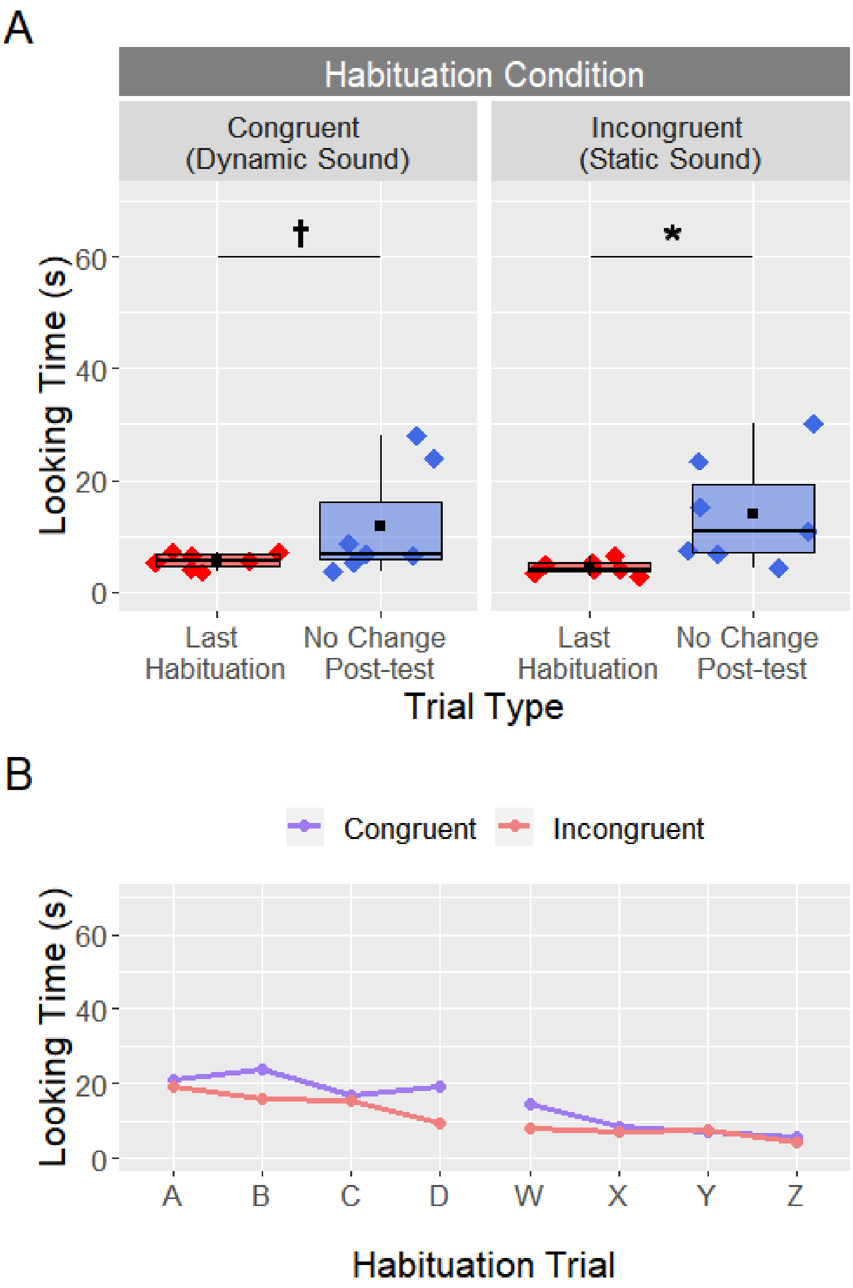

Figure 2.4. (A) Individual looking times (in seconds) in the Last Habituation trial and the No Change Post-test trial, and (B) mean looking time in the first four and the last four habituation trials.

The habituation display and the No Change test display were identical, and they depicted a ball that kept its pattern during the occlusion. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. The Post-test trial was presented in silence. *Note*: Black dots represent mean values. $^*p$ < .05, $^{\dagger}p$ < .10, 2-tailed.

2.3.1. Exploratory Analysis

Although the current study was not designed to measure how infants respond to possible vs. impossible events, there are some similarities between the current study and studies that have employed the "violation of expectation" testing paradigm (Baillargeon, Spelke, & Wasserman, 1985; Spelke, Breinlinger, Macomber, & Jacobson, 1992). For example, in the current study, the possible event was the *No Change* event, while the impossible event was the *Change* event. Furthermore, at *Post-test*, the *No Change* test event was also the more familiar event, while the *Change* event was the more novel event. One of the major criticisms that violation of expectation studies have faced is that it is difficult to disentangle, at *Post-test*, whether the infants' looking preference is due to their reasoning about the impossibility of an event occurring, or due to the perceptual similarity between the possible event and the habituation event (Munakata, 2000; L. B. Cohen, 2004). In other words, it is unclear whether the infants' *a priori* looking preference is maintained or it changes because the infants habituate to a particular visual stimulus.

Our main analysis did not find an interaction between Test Event (*Change* vs. *No Change*) and Test Time (*Pre-test* vs. *Post-test*), suggesting that the habituation did not have differential effects on the infants' looking to the two test events. However, this analysis cannot reveal whether there was a relationship between the infants' looking preference at *Pre-test* (*Change* or *No Change*) and the maintenance of this preference at *Post-test* (*Same* or *Opposite*). I reasoned that, if perceptual familiarity with the *No Change* event does affect the infants' looking preference, then out of those infants who preferred the *No Change* event at *Pre-test* there would be more infants that display the opposite preference at *Post-test* than infants that display the same preference. This is because the infants are familiarized with the *No Change* event during the habituation. On the other hand, if the infants preferred the *Change* event at *Pre-test*, their preference should remain unchanged at *Post-test* (i.e., more infants should display the same preference at *Post-test* than the opposite preference) because they do not watch the *Change* event during the habituation. To answer this question, I calculated the infants' proportion of total looking time at the *Change* event ($\text{PTLT}_{\text{change}}$) at *Pre-test* and *Post-test* by dividing the infants' looking time (in seconds) at the *Change* event by their accumulated looking time at both events, $\text{PTLT}_{\text{change}} = \text{LT}_{\text{change}}/(\text{LT}_{\text{change}} + \text{LT}_{\text{no change}})$. If $\text{PTLT}_{\text{change}}$ was above .50, I reasoned that the infant preferred the *Change* event (see **Table 2.2**, for observed frequencies). A Fisher's exact test yielded a non-significant result (p > .99, *ns,* 2-tailed), suggesting that irrespective of the infants' looking preference at *Pre-test* (i.e., *Change* or *No Change*), the same proportion of infants displayed a shift in their looking preference from *Pre-test* to *Post-test*.[23]

To better understand whether more infants displayed a shift in looking preference from *Pre-test* to *Post-test* than what would have been expected by chance, I conducted a binomial test. The binomial test indicated that the proportion of infants who shifted their preference was 71% and that this was not significantly different than the expected 50% chance (p = .18, *ns*, 2-tailed). Furthermore, a point-biserial correlation showed that there was no

[23] Fisher's exact test was used instead of a Chi-Square test of independence because the expected frequencies in three cells were less than 5.

relationship between the infants' accumulated looking time during the habituation (in seconds), and whether the infants displayed the *Same* or the *Opposite* looking preference at *Post-test*, $r_{pb}(12) = .11$, $p = .70$, *ns,* 2-tailed.

Table 2.2. The number of infants who preferred either the Change or the No Change event at Pre-test and who either maintained the same looking preference at Post-test or displayed the opposite preference.

Pre-test Preference	Post-test Preference		Total N
	Same	Opposite	
No Change	2	6	8
Change	2	4	6
Total N	4	10	14

Note. No Change event, the ball kept its pattern unchanged during the occlusion; Change event, the ball changed its pattern during the occlusion. Infants' looking preference was calculated by dividing infants' looking time at the Change event by their accumulated looking time at both the Change and No Change events, $PTLT_{change} = LT_{change}/(LT_{change} + LT_{no\ change})$. Pre-test Change Preference, if $PTLT_{change}$ exceeded .50 at Pre-test (before habituation); Pre-test No Change Preference, if $PTLT_{change}$ was less than .50 at Pre-test; Post-test Same Preference, if $PTLT_{change}$ exceeded (or was less than) .50 at both Pre-test and Post-test; Post-test Opposite Preference, if $PTLT_{change}$ was less than .50 at Pre-test and more than .50 at Post-test, or the reverse*.* Fisher's exact test yielded a non-significant relationship between infants' Pre-test Preference and their Post-test Preference, $p = 1.00$, *ns*, 2-tailed.

This exhaustive analysis of infants' looking preferences revealed that although the habituation led to a shift in the looking preference of some infants, it did not do so in more infants than would be expected by chance. Furthermore, if the habituation had resulted in an increased familiarity with the *No Change* event, then only those infants that showed a *No Change* preference at *Pre-test* should have shifted their preference for the more novel, *Change* event, at *Post-test*. However, this was not the case in the current study. Finally, the amount of perceptual experience that the infants accumulated with the *No Change* event during the habituation did not account for whether the infants displayed a shift in their looking preference between the *Pre-test* and the *Post-test*. This additional analysis complements the mixed ANOVA reported above and suggests that, contrary to our expectations, the habituation did not affect significantly how the infants responded at *Post-test* compared to *Pre-test.*

Table 2.3. Pearson correlation coefficients between the infants' visual preference for the Change event (PTLT$_{change}$), at Pre-test and Post-test, and various participant characteristics.

	Correlation with Pre-test PTLT$_{change}$	**Correlation with Post-test PTLT$_{change}$**
Age (days)	-.22	.21
Mullen VR T-score	.28	-.04
EMQ PA raw-score	.49†	-.45

Note. No Change event, the ball kept its pattern unchanged during the occlusion; Change event, the ball changed its pattern during the occlusion. Infants' looking preference was calculated by dividing infants' looking time at the Change event by their accumulated looking time at both test events, PTLT$_{change}$ = LT$_{change}$/(LT$_{change}$ + LT$_{no\ change}$). Pre-test PTLT$_{change}$, PTLT$_{change}$ score at Pre-test (before habituation); Post-test PTLT$_{change}$, PTLT$_{change}$ score at Post-test (after habituation); Age, infants' age (in days) at test; Mullen VR T-score, standardized score on the Mullen Visual Reception Scale; EQM PA raw-score, raw score on the Early Motor Questionnaire Perception-Action section. Positive correlation coefficients indicate that the infants with a stronger visual preference for the Change event (i.e., higher PTLT$_{change}$ score) were older or scored higher on a particular measure. $^{\dagger}p$ < .10, 2-tailed.

A second aspect that this exploratory analysis sought to understand was whether there was a relationship between the infants' ability to discriminate between the event and the infants' age, respectively their general visual-perceptual skills. The infants' visual-perceptual skills were measured via the Mullen VR scale (Mullen, 1995) and the EMQ PA section (Libertus & Landa, 2013). Both measures include items that tap into the infants' ability to process occlusion events and to discriminate between visual patterns. Therefore, I reasoned that the infants' score on these measures may be related to their visual preference for the *Change* event

($\text{PTLT}_{\text{change}}$). As can be seen in **Table 2.3**, none of the correlations reached significance, all $r(12) \leq .49$, all $p \geq .08$, *ns,* 2-tailed. This suggests that even if some of the infants had better visual-perceptual skills, they did not display a stronger preference for the *Change* event. That being said, these results should be interpreted with caution given the lack of power for this kind of analysis with this sample size.

Finally, to understand why the infants failed to show a differential response between test events at *Post-test*, I split the group of infants between those who preferred the novel event at *Post-test* (*Novelty* group, $n = 8$), and those who preferred the familiar event (*Familiarity* group, $n = 6$). The infants in the *Novelty* group had a $\text{PTLT}_{\text{change}}$ score above .50, while infants in the *Familiarity* group had a $\text{PTLT}_{\text{change}}$ score below .50. I then performed a 2 (Group: *Novelty* vs. *Familiarity*) x 2 (Test Event: *Change* vs. *No Change*) mixed ANOVA on $\log_{10}$-transformed looking time data (for the *Post-test* trials only). The Group was manipulated between-subjects, and the Test Event was manipulated within-subjects. The analysis yielded a significant interaction between Group x Test Event, $F(1, 12) = 15.98$, $p = .002$, $\eta_p^2 = .57$. No other effects were significant - Test Event: $F(1, 12) = .001$, $p = .98$, *ns*; Group: $F(1, 12) = .18$, $p = .68$, *ns*. I followed-up this interaction with 2 paired *t*-tests, one for each group. The *Novelty* group looked significantly longer at the *Change* event ($M = 1.07$, $SD = .22$) than at the *No Change* event ($M = .91$, $SD = .30$), $t(7) = 3.37$, $p = .01$, Cohen $d_z = 1.19$. Meanwhile the *Familiarity* group looked equally long at the *Change* event ($M = .98$, $SD = .33$) and the *No Change* event ($M = 1.13$, $SD = .30$), $t(5) = 2.36$, $p = .07$, *ns*. Thus, even if some infants preferred the more familiar test event, their looking preference was not statistically significant. Furthermore, the difference in the infants' looking time at the *Change* event between the *Novelty* group and the *Familiarity* group was not statistically significant, $t(12) = .60$, $p = .56$, *ns*. This suggests that it was the infants' response to the *No Change* event that differentiated between the *Novelty* group and the *Familiarity* group.

2.4. Discussion

Several studies have shown that young infants benefit from having spatiotemporally congruent audiovisual information when processing occlusion events. More specifically, when both the auditory and the visual cues define the same object that is moving behind a screen, the infants represent the occluded object for longer (J. G. Bremner et al., 2012). Furthermore, they anticipate better when and where the object will reappear (Kirkham, Wagner, et al., 2012). The study reported in this chapter investigated whether the infants learn the pattern on a briefly occluded object when the audiovisual cues provided are congruent and specify the same object compared to when they are incongruent and indicate different objects. The results showed that the infants watched the *Change* and *No Change* events for a similar interval of time. Despite the lack of a differential response to the two test events, the infants' looking times decreased significantly from *Pre-test* to *Post-test*, suggesting that the infants learned something about the occlusion event during the habituation. It is unclear what the infants learned, but I can say that they may have associated the musical sound with the visual display because their looking time increased from the *Last Habituation* trial to the *No Change Post-test* trial (when the background sound stopped). Besides, the habituation condition did not have differential effects on the infants' responses to the *Post-test* occlusion events, which contradicts my prediction that the infants in the *Incongruent (Static Sound)* condition would look longer at the *Post-test Change* event. Finally, the exploratory analysis conducted on the infants' proportional looking time at the *Change* event showed that the infants' looking preference was independent of their age or general visual-perceptual skills. Moreover, the looking time that the infants accumulated during the habituation did not account for a switch in their looking preference between *Pre-test* and *Post-test*, and nor did the infants' looking preference at *Pre-test*.

This pattern of results is inconsistent with the IRH (Bahrick & Lickliter, 2000, 2002, 2012) which argues that, in the absence of redundant multisensory stimulation that defines the same object/event, the infants attend to the modality-specific object/event properties (e.g., object shape, pattern, colour, and pitch). In the *Incongruent (Static Sound)* habituation condition, the ball was specified only by the vision, while the musical sound appeared to originate from the box. Therefore, in this habitation condition, the infants should have attended to and learned the pattern on the ball. Contrary to what I expected, the infants looked equally long at the *Change* event (the novel event) and the *No Change* event (the familiar event). The IRH also states that, in the presence of multisensory information that specifies the same object/event, the infants pay attention to the object/event properties that are highlighted concurrently by multiple senses. In the *Congruent (Dynamic Sound)* condition, both the visual and auditory cues specified the trajectory of the ball during the habituation. Hence, I expected the infants to attend to the ball's trajectory and not to its pattern. The results showed that, at *Post-test*, the infants in this habituation condition did not differentiate between the *Change* and the *No Change* test events. A possible explanation is that the infants failed to notice the pattern change because they paid attention to the trajectory of the ball. However, this interpretation of this null finding can only be speculative since the looking data acquired does not provide any information about the infants' visual scanning pattern (e.g., the number of anticipatory saccades) during the study.

Infants' failure to respond to the ball pattern change at Post-test was somewhat surprising. Wilcox & Baillargeon (1998b) have found that infants as young as four months old detect and respond to changes in the surface features of a briefly occluded object. Based on this finding, I had expected that the 10-month-old infants in my study would be able to differentiate between the Change and No Change test events. After testing $N = 14$ infants, I conducted a power analysis. I found that, at *Post-test*, the difference in looking times between the two test events was small – equivalent to Cohen $d_z = .11$. To demonstrate a statistically significant difference between test events with a 2-tailed *t*-test (alpha = .05) with Cohen $d_z = .11$ and power =.75, I would have had to test $N = 559$ infants. Considering that other infant-habituation studies have obtained statistically significant results with $N < 30$ infants (see Csibra et al., 2016), I concluded that the difference between test events was too small and reflected infants' inability to differentiate between the *Change* and *No Change* test events. As a result, I stopped testing.

I have identified three alternative explanations for the results: (1) the infants detected the change in pattern, but there were individual differences in how they responded to the change - some infants showing a novelty preference (i.e., they looked longer at the *Change* event) and other infants showing a familiarity preference (i.e., they looked longer at the *No Change* event); (2) the infants detected the change in pattern, but did not respond to it because they did not consider it an important outcome of the event; (3) the infants did not detect the change in the ball's pattern.[24] Based on the data collected, I believe that the first explanation is unlikely. While it is the case that, at *Post-test*, 8 out of 14 infants looked longer at the *Change* event and the rest looked longer at the *No Change* event. The increase in looking time from one event to the other was statistically significant only in the group of infants that preferred the *Change* event (i.e., that displayed a novelty preference). The infants who preferred the *No* Change event did not look significantly more at the *No Change* event than the *Change* event. Therefore, I argue that the split in infants' looking preferences cannot fully explain the non-significant difference in looking times between the two test events at *Post-test*.

The second explanation, that the infants may have detected the change in the ball's pattern at *Post-test* but that they did not respond to it because they did not consider it important is also unlikely. In a series of studies conducted with infants aged between 4.5 and 11.5 months old,

[24] There is also a fourth explanation, namely that the infants' looking behaviour during the test trials was off-task. The analysis I conducted between the *Last Habituation* trial and the *No Change Post-Test* trial suggests that the infants did pay attention to the stimuli during the *Post-test* trials. Furthermore, at the beginning of the study, I have set-up a minimum looking duration of 2 s per trial to ensure that the infants were maintaining their attention to the stimuli.

Wilcox (1999) found that 7.5-month-old infants responded to a change in either the pattern or the shape of a briefly occluded object, while 4.5-month-old infants responded only to a change in the object's shape. More specifically, the 7.5-month-old infants looked longer at an occlusion event when it depicted a ball that changed its pattern (from dots to stripes) during the occlusion than a ball that remained unchanged. Interestingly, this differential response was observed only when the spatiotemporal gap between the successive reappearances of the ball was short - i.e., less than 1 s long. In other words, infants that are older than 7.5 months old find changes in an occluded object's pattern significant and they respond to these changes, but whether they detect the pattern changes or not depends on how long the occlusion event lasts.

Based on Wilcox's (1999) findings, I argue that the most likely explanation for the null finding is that, at *Post-test*, the infants did not detect the change in the ball's pattern - from half red and half green to chequered red and green and the reverse. Something which speaks to this interpretation of the results is the fact that, even if some infants preferred the *Change* event over the *No Change* event, their looking time at the *Change* event was comparable to the looking time of those infants who preferred the *No Change* event. If some infants had indeed detected the change in the ball's pattern, these infants should have looked longer at the *Change* event than those infants who did not notice the pattern change. However, this was not the case. What seemed to differentiate between the group that looked longer at the *Change* event and the group that looked longer at the *No Change* event was the infants' looking time at the *No Change* event. In other words, whether the infants recognized the silent *No Change* event as the familiar event.

What factors may have contributed to the infants' failure to detect the pattern change? One factor may have been the relatively long interval of occlusion used in the current study - the occlusion lasted 834 ms. Previously, S. P. Johnson, Bremner, et al. (2003) found that infants represent the trajectory of an occluded object if the object is out of sight for ~400 ms when the infants are four months old, respectively ~667 ms when the infants are six months old. While these findings suggest that infants of different ages can represent occluded objects for different intervals of time, they also indicate how long that interval might be. Consequently, I planned to conduct a follow-up study (reported in Chapter 3) and use a shorter occlusion interval.

Another contributing factor may have been the pattern manipulation employed - the ball changed from half red and half green to chequered red and greed, and the reverse. That infants habituate faster and exhibit more visual recovery to some patterns than others has been reported in various studies (Bornstein et al., 1981; Bornstein & Krinsky, 1985; Humphrey et al., 1986). More specifically, the infants seem to encode faster and better symmetrical patterns with a vertical axis of symmetry (e.g., left-half red, right-half green) than symmetrical patterns with a horizontal axis of symmetry (e.g., top-half red, bottom-half green) or asymmetrical patterns (e.g., chequered red and green). In our study, the infants were habituated with a ball whose pattern had a horizontal axis of symmetry which may have posed additional demands on the infants' visual processing and learning abilities.

To address this potential limitation, in the follow-up study that I planned to conduct, I intended to habituate the infants with a dotted ball. I reasoned that a dotted pattern would control for the presence of a vertical axis of symmetry. Furthermore, Wilcox (1999) found that 7.5-month-old infants look longer at a ball that changes pattern from dots to stripes and the reverse, which suggests that such a pattern change is easy to detect for infants.

A third factor might have been the presence of a background sound. (Robinson & Sloutsky, 2007b) found that 14-month-old infants have difficulties detecting which one of two side-by-side visual streams changes repeatedly and which remains unchanged when they hear a computer-generated sound alongside the visual stimuli. Hearing an unrelated sound affects infants' performance on other tasks as well (Barr et al., 2010; Lejeune et al., 2016; Robinson & Sloutsky, 2007a, 2008). Therefore, in the follow-up study, I planned to include a unisensory habituation condition.

To conclude, this study gives an insight into the representational skills of 10-month-old infants. The results suggest that, if infants receive sound support that specifies the location of an object, this does not boost nor hinders their memory for the surface features of the object. As outlined above, multiple factors could have contributed to the lack of a differential response to the two test events. To address the various limitations of the current study, and to gain a better insight into the role of spatiotemporal congruency in the infants' processing of occlusion events, I planned to conduct a follow-up study (reported in Chapter 3) with another group of 10-month-old infants. In the follow-up study, I intended to use a shorter occlusion interval. Furthermore, I planned to employ a more salient pattern change and include a unimodal habituation condition.

CHAPTER 3

Effects of multisensory stimulation on 10-month-old infants' encoding of object pattern. Part 2

3.1. Introduction

In my previous study, I was unable to conclude whether congruent multisensory stimulation affects 10-month-old infants' processing of the modality-specific object properties. Irrespective whether the infants heard a spatiotemporally congruent or incongruent sound while they watched a ball moving across the display, they failed to notice that the ball changed its pattern during the test trials. While these results are inconsistent with the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012), they are somewhat surprising. Infants younger than ten months old can detect pattern changes in similar experimental tasks (Wilcox, 1999; Wilcox et al., 2011; Wilcox & Chapa, 2004). In Chapter 2, I laid out three possible factors that may have hindered the infants' ability to detect the change in object pattern: (1) a long interval of occlusion, (2) the use of a pattern with a horizontal axis of symmetry, and (3) possible auditory distraction. To address these limitations, I shortened the occlusion interval to 634 ms (instead of 834 ms, as in Chapter 2). Furthermore, I decided to present a more salient pattern change in which the pattern on the ball changed from dots to stripes (rather than from half red and half green to chequered red and green, as in Chapter 2). Finally, I added a visual-only habituation condition to control for possible auditory distraction effects (see Chapter 2).

The IRH argues that the sensory stimulation that infants receive guides their attention toward some object/event properties and away from other properties. It proposes that, when a sound is spatiotemporally congruent with an object, the infants learn those object/event properties that are specified by both vision and audition (i.e., amodal properties), such as the trajectory, location, duration, rhythm, and tempo of the object/event. Furthermore, the infants fail to encode those object/event properties that are specified only by vision or only by audition (i.e., modality-specific properties), such as the shape, pattern, colour, orientation, and pitch of the object/event. According to the IRH, this prioritization is happening because the infants have limited attentional resources. As a result, they have to process the various properties in a serial order rather than in parallel. Since the amodal properties are more salient, they are processed first.

The visual shape, pattern, and colour of an object are modality-specific properties that define an object's identity. Furthermore, infants used them to individuate objects in a visual display (Wilcox, 1999; Wilcox & Baillargeon, 1998a, 1998b; Xu & Carey, 1996). What the IRH seems to be proposing is that infants fail to encode the identity of an object when a congruent sound accompanies it. Such a prediction is at odds with the mounting evidence that infants associate a sound to an object when the sound is synchronous and colocated (Fenwick & Morrongiello, 1998; Lawson, 1980; Morrongiello, Fenwick, & Nutley, 1998; Richardson & Kirkham, 2004). To make such object-sound associations, infants must process and learn the visual features of the multisensory objects. That infants form such associations from the first few hours of life is apparent from a study conducted by Morrongiello, Fenwick, & Chance (1998). Morrongiello, Fenwick, & Chance first habituated a group of newborn infants with two toys, each fitted with a miniature loudspeaker. Whenever the toys appeared, one of them generated a rhythmical rattle sound that lasted until the infants looked away from the display. After the habituation, the sound was swapped between the toys - the toy that had been silent during the habituation produced the rattling sound while the other toy remained silent. The researchers found that the infants looked significantly longer at the toys during the test trials than during the last habituation trial. This regained interest in the stimuli suggests that the newborn infants formed object-sound associations during the habituation, even though the object's location was presumably more salient because it was specified concurrently by both vision and audition.

Further evidence that the spatiotemporal congruency between the auditory and the visual cues facilitates infants' visual processing comes from a series of experiments conducted by Lawson (1980; see also Fenwick & Morrongiello, 1998; Morrongiello, Fenwick, & Nutley, 1998). In her experiments, Lawson manipulated either the spatial relationship between a sound and an object or the synchrony between them to see how these factors affected 6-month-old infants' ability to make object-sound associations. The experiments had a familiarisation phase and a test phase. During the familiarisation, the infants saw an object that oscillated left-right or up-down while a sound played in the background. The sound was either colocated (i.e., the object emitted the sound) or dislocated (i.e., the sound originates from somewhere else in the display), in addition to being synchronous (i.e., it played when the object changed its direction) or asynchronous. During the test phase, the infants saw two stationary objects (one was familiar and the other one was novel) and heard either the old or a new sound playing in the background. The results showed that, when the familiar sound played, the infants preferred to look at the familiar object. However, only the infants in the synchronous and colocated sound conditions displayed this response. The infants in the asynchronous and dislocated sound conditions looked equally long at the stationary objects irrespective whether they listened to the familiar or the novel sound. This pattern of results suggests that: (1) correlated auditory and visual cues help 6-month-old infants process the visual features of the objects and form object-sound associations, and (2) incongruent multisensory stimulation hinders infants' cognitive processing.

Other findings which contradict the IRH are those reported by Kirkham, Richardson, Wu, & Johnson (2012). In a series of experiments conducted with infants aged 3, 6, and 10 months old, Kirkham, Richardson, et al. found that the infants learned the spatial location of different multisensory objects/events when the visual features of the objects/events were distinct but not when they were similar. The authors first familiarized the infants with two cartoon characters that moved in synchrony with two musical sounds. Each character moved in a specific way, made a distinct musical sound, and occupied a particular rectangle on the screen, thus forming a unique multisensory event with a precise spatial location. After the familiarization, the rectangles where the characters had appeared remained empty while the familiar tunes played in the background. The results showed that, if the infants saw two visually distinct characters during the familiarization, both the 3- and the 6-month-old infants looked longer at the rectangle that was associated with the tune heard. However, if they saw only one character that was associated with both sounds and both locations, neither the 3- nor the 6-month-old infants showed evidence of having formed sound-location pairs. While the younger infants failed to spatially index the multisensory events when only one character appeared in the study, the 10-month-old infants learned the location of the events both when one and multiple characters were used. These results show that infants do not only process the visual properties of multisensory objects/events, but they also use these properties to track the spatial location of the objects/events.

The studies mentioned above suggest that infants benefit from having congruent audiovisual information when processing objects/events. However, it is unclear whether the congruent stimulation facilitates infants' processing of some object/event properties, while it hinders the processing of other features, as the IRH argues. To test this prediction, in my previous study (see Chapter 2), I manipulated the conditions in which a group of 10-month-old infants learned the pattern on a briefly occluded object. The results of that study did not allow me to conclude anything about the role of multisensory stimulation in infants' processing of object pattern because the infants did not detect the pattern change in the test events. Therefore, I decided to conduct a follow-up study in which I tried to address the limitations of the former study. In the present study, I shortened the occlusion interval from 834 ms to 634 ms, and I used a dotted ball that changed into a striped ball and the reverse.[25] Furthermore, I added a visual-only habituation condition which allowed me to assess whether auditory

[25] I used an occlusion interval of 634 ms because S. P. Johnson et al. (2003) found that 6-month-old infants can represent an occluded object when the occlusion interval lasts 667 ms. Therefore, I assumed that 10-month-old infants can do that as well, given that they are older and have more experience. Similar with the dots to stripes pattern change. Wilcox (1999) found that 7.5-month-old infants can detect this change. Consequently, I assumed that 10-month-old infants would be able to do that too.

input distracted the infants from the visual display (see Chapter 2, for the rationale behind the changes).

This study aimed to find out whether 10-month-old infants learn the pattern on an occluded object when they receive unimodal or multimodal stimulation. For this purpose, I habituated three groups of infants with a dotted ball that moved horizontally behind a box. One group of infants watched the occlusion event in silence (*Visual-Only* condition), while the other two groups watched the occlusion event alongside an auditory cue (a musical sound). In the *Congruent (Dynamic Sound)* condition, the sound was spatiotemporally congruent with the movement of the ball. In the *Incongruent (Static Sound)* condition, the sound was incongruent with the ball. After the habituation, the infants watched in silence two occlusion events: an event in which the ball kept its pattern during the occlusion (*No Change* event) and an event in which the ball changed from dots to stripes and the reverse during the occlusion (*Change* event).

Based on the IRH and the previous empirical research (Wilcox, 1999; Wilcox et al., 2011; Wilcox & Chapa, 2004), I predicted that the infants in the *Visual-Only* condition would look longer at the *Change* event. I also hypothesized that the infants in the *Incongruent (Static Sound)* condition would prefer looking at the *Change* event. Lastly, the infants in the *Congruent (Dynamic Sound)* condition would show no difference between the test events (but see Lawson, 1980).

3.2. Methods

3.2.1. Design

To study how multisensory stimulation affects infants' learning, I employed an infant-controlled habituation paradigm. The infants were randomly assigned to one of the 3 habituation conditions: *Visual-Only*, *Congruent (Dynamic Sound)* and *Incongruent (Static Sound)*. During the test phase, all the infants watched in silence 2 occlusion events: *Change* and *No Change*. The test events were presented in alternation, three times each (i.e., 3 test blocks), for a total of 6 test trials. I decided to show each test event three times instead of only once, like in the previous study (see Chapter 2), because I wanted to reduce the noise in the looking time data. Young infants can be easily distracted by contextual factors during psychological experiments (e.g., a sound coming from outside the testing booth). If they are distracted during a specific test trial, their looking time to the stimuli can be shorter and may not reflect their actual level of interest. Increasing the number of test trial per test event and averaging across trials allowed me to increase the signal-to-noise ratio. Therefore, this study a 3 x 2 x 3 mixed design with *Habituation Condition* (Visual-Only vs. Congruent vs. Incongruent) as a between-subjects factor, and *Test Event* (Change vs. No Change) and *Test Block* (Block 1 vs. Block 2 vs. Block 3) as within-subjects factors. Test Block was included as a factor to account for practice effects and fatigue. The dependent variable was the infants' looking time at each test event, defined as the total interval of time that the infants looked at the screen between the beginning and the end of the trial.

3.2.2. Participants

Thirty-nine 10-month-old infants (*M* = 304.38 days, range = 258 - 321 days, 22 females) participated in the study. Twelve additional infants were tested (i.e., 23.53% of the total *N* = 51), but their data was not included in the analysis due to fussiness, which resulted in the infants not completing the study (*n* = 4), failure to habituate (*n* = 5), and experimenter error (*n* = 3). The infants were recruited in the same way as described in Chapter 2. All the infants had been full-term (i.e., gestation age: 37 weeks or more) and none of them had experienced sight or hearing problems, as reported by the caregivers. Most of the infants were White-Caucasian and came from middle-class families. Before the study, the infants were randomly assigned to one of the three habituation conditions (*n* = 13 infants in each habituation condition). Analysis of the infants' scores on the Mullen Visual Reception Scale (Mullen VR; Mullen, 1995) and the Early Motor Questionnaire – Perception-Action section (EMQ PA; Libertus & Landa, 2013), as well as the infants' age, accumulated looking time during the habituation, and initial and terminal level of attention showed no significant differences between the groups (see **Table 3.1**).

To estimate the number of participants needed in each habituation condition I conducted a power analysis. The analysis was carried out in G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) and it was based on the infant looking time data reported by Wilcox (1999, Exp. 3B). Specifically, I compared the looking times of the Same-Pattern Narrow-Screen group (M = 28.1; SD = 8.0) with those of the Different-Pattern Narrow-Screen group

(M = 41.4; SD = 8.9).[26] The effect size in this study was Cohen d_s = 1.57, which is considered to be a large effect using J. Cohen's (1988) criteria. Given that at least one another study (Wilcox & Baillargeon, 1998b) found similar differences between groups when other features of the object (e.g., shape and colour) were manipulated, I expected to find a large difference between the infants' looking times at the two test events. The projected sample size for a paired samples t-test (2-tailed), with effect size = 1.5, alpha = .05, and power = .75, was approximately *N* = 6 infants. Although the power analysis revealed that I needed at least n = 6 infants for each habituation condition, I decided to test n = 13 infants in each condition. This was because Wilcox's (1999) study had a small sample size (n = 6 per group), and the magnitude of the effect may have been overestimated as a result (Cumming, 2012; Lakens, 2013).

[26] Wilcox (1999, Exp. 3B) familiarized 7.5-month-old infants with an occlusion event in which a ball that oscillated behind an occluding screen either changed its pattern during the occlusion (*Different-Pattern*) or kept its pattern (*Same-Patten*). During the test phase, Wilcox either increased the width of the occluding screen (*Wide-Screen*) or decreased it (*Narrow-Screen*). Given the similarity in pattern manipulation between Wilcox's study and the Visual-Only condition in the current study, as well as the length of the occlusion interval in the *Narrow-Screen* condition (~438 ms), I calculate the difference in infants' looking times (to the test trials) between the *Same-Pattern Narrow-Screen* group (*n* = 6) and the *Different-Pattern Narrow-Screen* group (*n* = 6) and used that to estimate for the expected effect size in my study.

Table 3.1. Participant characteristics by habituation condition.

	Habituation Condition			**One-Way ANOVA**		
10-month-olds	**Visual-Only** (*n* = 13)	**Congruent (Dynamic Sound)** (*n* = 13)	**Incongruent (Static Sound)** (*n* = 13)	***F*-value**	***df1, df2***	***p*-value** (2-tailed)
Age (days)	307.92 (9.23)	301.69 (7.83)	303.54 (11.28)	1.46	2, 36	.25
Mullen VR T-score	60 (8.16)	61.54 (5.41)	63.08 (3.86)	.83	2, 36	.44
EMQ PA raw-score	9.25 (10.58)	12.27 (14.15)	9.69 (12.65)	.20	2, 33[#]	.82
Habituation LT (s)	154.82 (66.02)	145.60 (64.99)	127.34 (31.74)	.80	2, 36	.46
Pre-test LT (s)	49.75 (15.38)	43.60 (20.51)	53.40 (11.60)	1.21	2, 36	.31
Post-test LT (s)	59.26 (1.41)	50.26 (13.52)	55.82 (11.54)	2.53	2, 36	.09

Note. Values represent mean (SD). Age, infants' age (in days) at test. Mullen VR T-score, standardized score on the Mullen Visual Reception Scale. EQM PA raw-score, raw score on the Early Motor Questionnaire Perception-Action section. Habituation LT (s), accumulated looking time during the habituation (seconds). Pre-test LT (s), infants' looking time at a 60 s control video that the indicated infants' level of attention before the study (seconds). Post-test LT (s), infants' looking time at the same 60 s control video that indicated the infants' level of attention at the end of the study (seconds). There were no significant differences between the three groups of infants. [#]Caregivers forgot to fill in the EMQ for three infants: *Visual-Only* (*n* = 1) and *Congruent* (*n* = 2).

3.2.3. Apparatus and Stimuli

I used the same testing set-up as the one described in Chapter 2: a PC, a 24” BenQ video screen (resolution 1920 x 1080), and two loud-speakers placed under the screen, at ~2 cm below the screen and ~50 cm from each other. The infants' looking during the study was video recorded using a surveillance video-camera system which was hidden from the infants' view. The video recording was presented live on a second screen, located outside the testing booth, and was used by the researcher to judge the infants’ looking behaviour during the study. The presentation of the stimuli and the researcher's coding were controlled via an in-house computer script.

During the study, the researcher was blind to the habituation condition (*Visual-Only* vs. *Congruent* vs. *Incongruent*) and the test event (*Change* vs. *No Change*) that the infants watched. A third of the videos (n = 13) were randomly selected and coded offline by a second coder.[27] The second coder was blind to the habituation condition, the test events, and the experimental hypotheses. A two-way mixed intra-class correlation analysis with absolute-agreement (Trevethan, 2017) conducted on the infants' total looking time during the test trials showed an excellent inter-rater agreement, $\text{ICC}_{2,1}$ = .99.

The stimuli were like those described in Chapter 2. The animation that the infants watched during the habituation depicted a room with a black and white dotted ball that moved behind a blue box. The ball appeared on each side of the box, and it was occluded when behind the box (see **Figure 3.1**, for stimuli examples and dimensions). Each journey made by the ball from

[27] The n = 13 videos coded offline were: n = 4 from the *Visual-Only* condition, n = 4 from the *Congruent* condition, and n = 5 from the *Incongruent* condition.

one side of the display to the other side lasted 2.5 s and covered 32.09^{o} degrees of visual angle (as seen from 70 cm distance). The speed of the ball was ~12.84^{o}/s, and it remained constant during the presentation. In the 2.5 s that the ball took to translate across the display, it was visible on either side of the box for 533 ms, it transitioned from full visibility to full occlusion (and the reverse) in 400 ms, and it remained fully occluded for 634 ms. The infants in the *Visual-Only* habituation condition watched the animation in silence, while the infants in the *Congruent (Dynamic Sound)* condition, and those in the *Incongruent (Static Sound)* condition, watched the animation accompanied by a musical sound (see **Figure 3.2**).

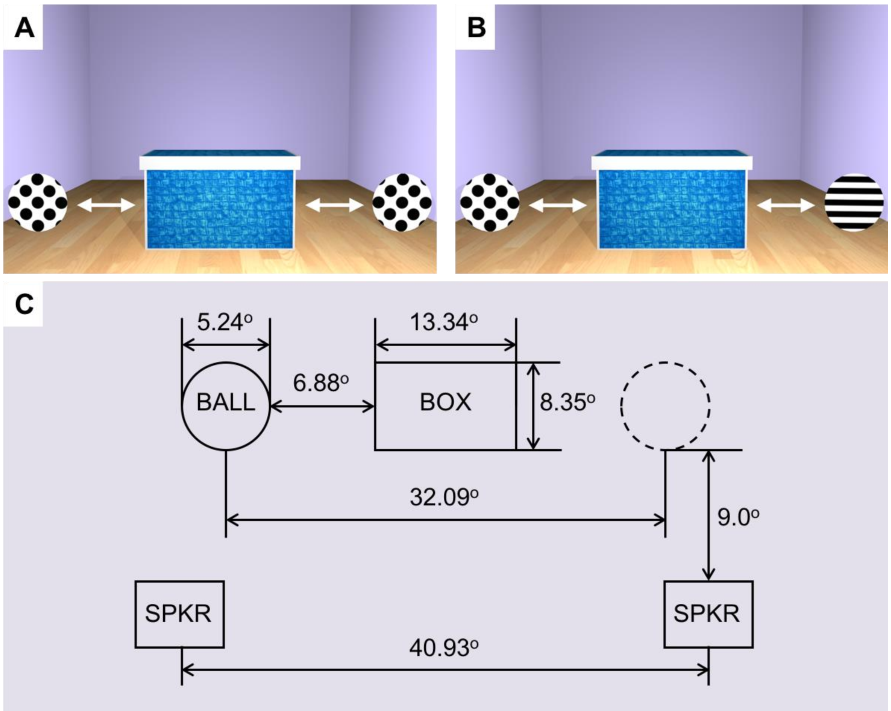

Figure 3.1. Displays shown to the infants and stimuli dimensions.

(A) Habituation display and No Change test display. The ball kept its pattern during the occlusion. (B) Change test display. The ball changed its pattern during the occlusion. (C) Schematic drawing of the display. Numbers represent the stimuli dimensions in degrees of visual angle, as seen from 70 cm distance (i.e., infants' viewing distance). *Note*: The acronym SPKR stands for "loudspeaker".

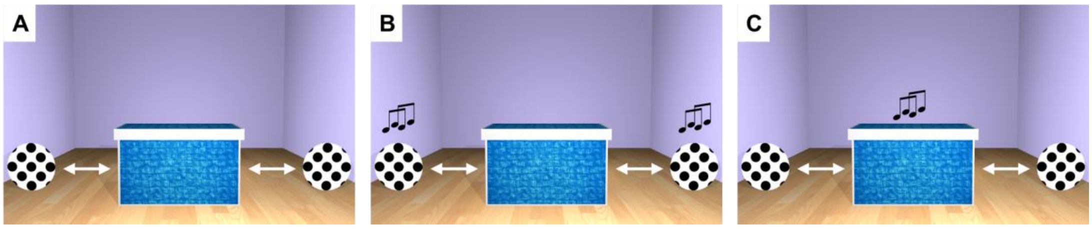

Figure 3.2. Habituation conditions employed in the study.

(A) Visual-Only condition. The habituation display was presented in silence. (B) Congruent (Dynamic Sound) condition. The habituation display was accompanied by a musical sound. Twelve adults reported that the musical sound appeared to originate from the ball. (C) Incongruent (Static Sound) condition. The same group of adults reported that the musical sound appeared to originate from the box.

The sound manipulation employed was the same as in Chapter 2. To assess whether the changes in sound panning were noticeable, twelve adults were asked to watch the animation while listening either to the dynamic or to the static sound. After 60 s, the adults were asked to indicate whether the sound was dynamic or static and whether it originated from the ball or the box. To minimize the demand characteristics, the adults were not informed about the purpose of the study, and the order of the stimuli presentation was counterbalanced. All 12 adults reported that the dynamic sound appeared to originate from the ball and that it moved together with the ball from one side of the display to the other. Meanwhile, the static sound gave the impression that it originated from the box and that it remained located in the centre of the display during the presentation.

After the habituation, all the infants watched in silence two occlusion events: (1) the familiar occlusion event (*No Change;* see **Figure 3.1A**), in which the dotted ball kept its pattern during the occlusion, and (2) a novel occlusion event (*Change*; see **Figure 3.1B**), in which the ball changed from dots to stripes (and the reverse) during the occlusion. No other visual parameters changed between habituation and test. Finally, as with the previous study, the infants' visual perceptual skills were measured via the Mullen VR scale (Mullen, 1995) and the EMQ (Libertus & Landa, 2013). For a description of these instruments, please see Chapter 2.

3.2.4. Procedure

The infants were brought to the lab by their caregivers, who also provided informed consent. The study was conducted in a dimly lit room with few distracters. During the study, the infants sat on their parents' lap, at ~70 cm from the presentation screen and the loudspeakers. The infants' eyes were aligned with the middle of the screen, and the parents were asked to refrain from looking at the screen or from interacting with the infants during the study. The researcher remained outside the testing booth and controlled the presentation of the stimuli via a computer script. A surveillance video system allowed the researcher to monitor the infants' looking behaviour during the study and to video record it for subsequent secondary coding.

The study began with a single pre-test trial in which a video from the TV series "In the Night Garden" was played for up to 60 s. After the pre-test trial, the infants watched up to 12 habituation trials, followed by 6 test trials (see **Figure 3.3**). The study ended with a single post-test trial during which the video that the infants watched at pre-test was played again. The video presented at pre-test and post-test was purposefully different from the animation used during the study and it allowed me to measure the infants' initial and terminal level of attention. If an infant watched less than 5 s of the post-test video (i.e., less than 8% of its entire duration), I considered that they were tired and I excluded their data from the analysis.

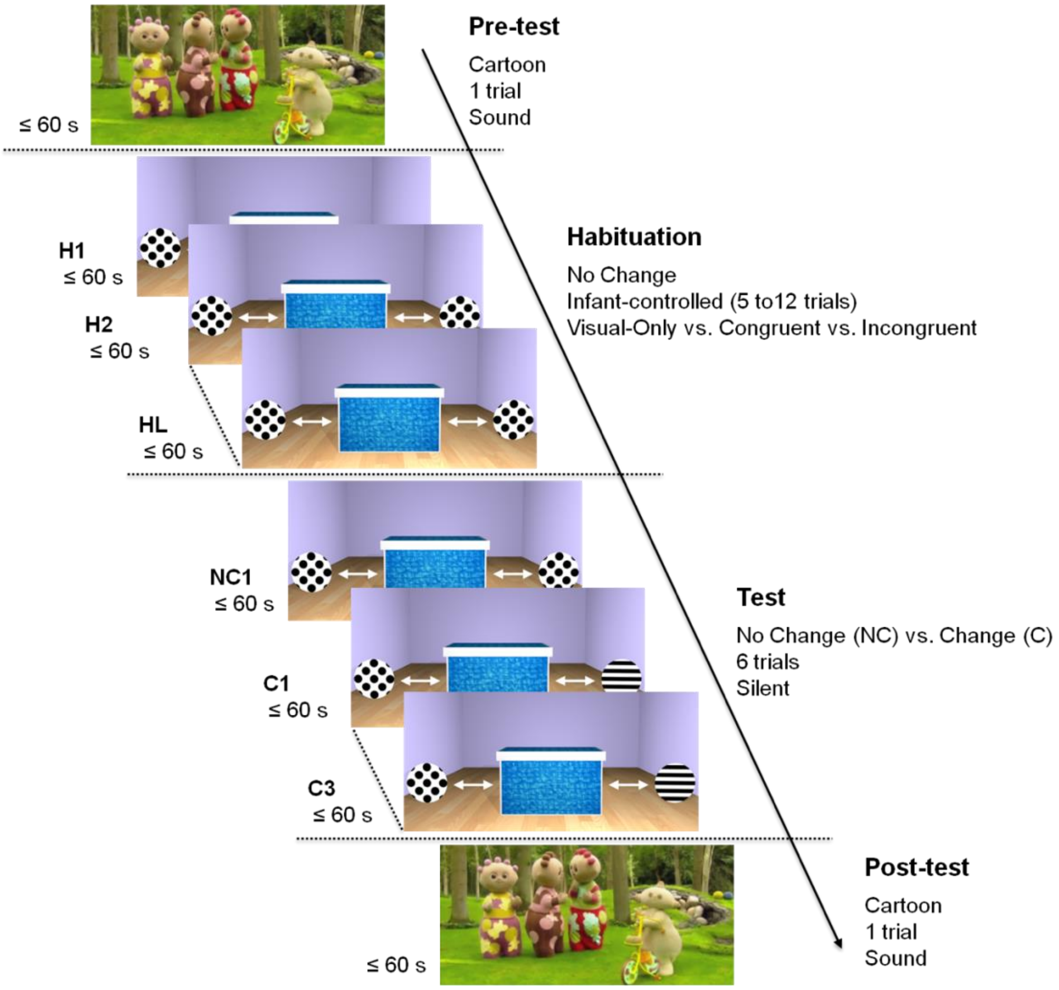

Figure 3.3. Timeline detailing the study procedure.

During Pre-test and Post-test, the infants watched a video from the TV series "In the Night Garden". During the habituation, the infants watched an animation of a dotted ball that moved horizontally behind a box. The habituation display was either presented in silence (Visual-Only), or it was accompanied by a musical sound. The sound was either spatiotemporally congruent with the ball, and "moved" left-right together with the ball (Congruent condition), or it was independent of the ball, and it remained located in the centre of the display throughout the trial (Incongruent condition). The habituation criterion was reached when the infants' accumulated looking time in the last 4 habituation trials was less than half of their accumulated looking time in the first 4 habituation trials. The infants completed between 5 and 12 habituation trials. At test, the infants watched two occlusion events in silence: the familiar event, in which the ball remained unchanged during the occlusion (No Change test event), and a novel event in which the ball changed its pattern during the occlusion (Change test event). The two test events were presented in alternating order, three times each, and about half of the infants watched the Change event first, while the other half watched the No Change event first. Each trial lasted 60 s or until the infants looked away from the screen for more than 2 s. In between the trials, an audiovisual animation was presented at the centre of the screen.

Each infant completed between 5 and 12 habituation trials. The habituation trials were presented until the infant's accumulated looking time in the last 4 habituation trials was less than half of their accumulated looking time in the first 4 habituation trials (starting from the second trial). This habituation criterion was calculated online, and it was based on the researcher's online judgements about the infant's looking behaviour. If an infant did not reach the habituation criterion within 12 trials, it was considered that they had failed to habituate, and their data was excluded from the analysis.

When the habituation terminated, the test trials started (6 trials in total). During each test trial, the infants watched the 2 test events: *Change* or *No Change.* The events were presented in silence, three times each, and in alternating order. About half of the infants viewed the *Change* event first (*n* = 21), and the other half viewed the *No Change* event first.[28] Note that the *No Change* test display was identical to the habituation display.

All the trials lasted 60 s or until the infant looked away from the screen for more than 2 s. If the infant looked at the screen for less than 2 s before looking away, the trial was repeated. This minimum looking time interval of 2 s was chosen because it took the ball ~2 s to translate from one side of the box to the other side. Before each trial, a 4° x 4° audiovisual looming animation attracted the infant's attention to the centre of the screen. The attention-getter was presented for at least 1.5 s and lasted until the infant looked back to the screen.

[28] The number of infants who watched the *Change* test event first was equally distributed across habituation conditions: *Visual-Only* (*n* = 7), *Congruent* (*n* = 7), and *Incongruent* (*n* = 7).

The study took ~7 minutes to complete. Upon finishing the study, the infants took a 5-minutes break and then performed several visual-motor tasks that were part of the Mullen VR Scale (Mullen, 1995). These tasks took ~10 minutes and their difficulty increased as the infant progressed through the tasks. When the infant failed to score points on 3 consecutive tasks, it was considered that ceiling level was reached, and the assessment was terminated. Finally, the caregivers were asked to fill in the EMQ (Libertus & Landa, 2013) while the researcher entertained the infants. At the end of the study, the families were debriefed and received a baby t-shirt and a certificate as a reward for their participation.

3.3. Results

Individual looking times (in seconds) at the two test events, *Change* and *No Change*, are displayed in **Figure 3.4A**. In each habituation conditions, 11 out of 13 infants looked longer at the *Change* event than to the *No Change* event (*Visual-Only*: Wilcoxon signed ranks test z = 2.55, p = .01; *Congruent*: z = 2.97, p = .003; *Incongruent*: z = 2.97, p = .003). Displayed in **Figure 3.5A** are the infants' average looking times in the first 4 and the last 4 habituation trials.

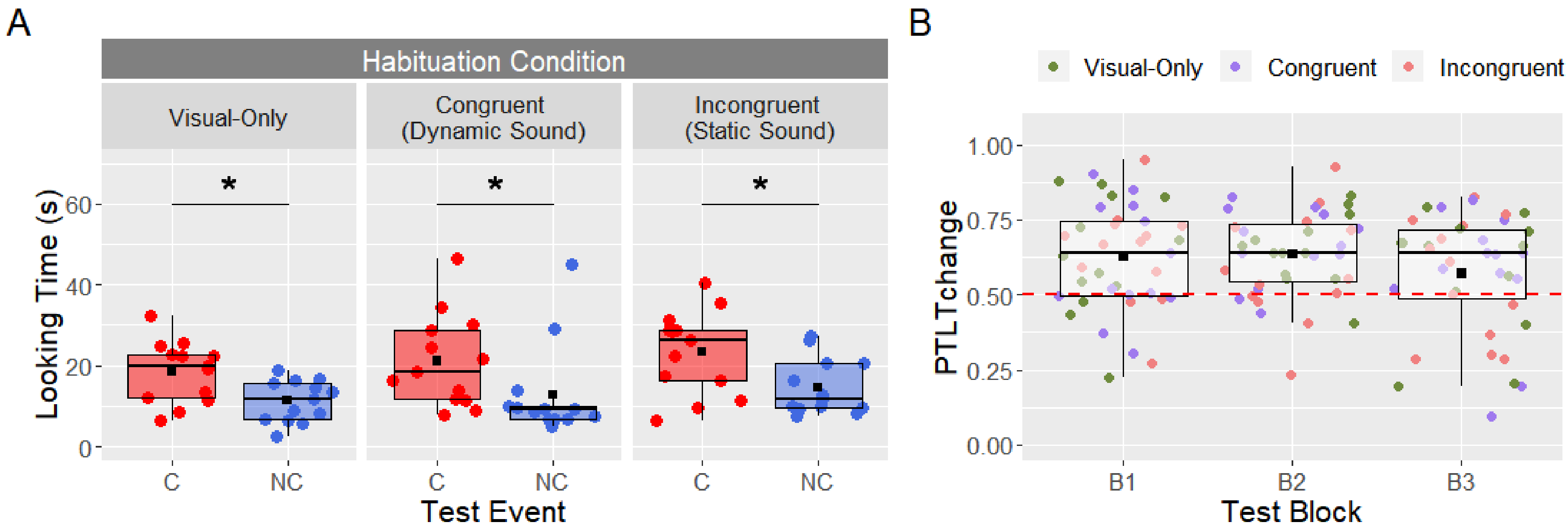

Figure 3.4. (A) Individual looking times (in seconds) at the test events, and (B) looking preference for the Change event across the three test blocks.

(A) Change event (C, in red), the ball changed its pattern during the occlusion. No Change event (NC, in blue), the ball kept its pattern during the occlusion. Infants' looking times were averaged across all 3 test blocks. In the Visual-Only condition, the animation was presented in silence. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. (B) Infants' looking preference ($\text{PTLT}_{\text{change}}$) was calculated by dividing infants' looking time at the Change event by their accumulated looking time at both events, $\text{PTLT}_{\text{change}} = \text{LT}_{\text{change}}/(\text{LT}_{\text{change}} + \text{LT}_{\text{no change}})$. If $\text{PTLT}_{\text{change}}$ exceeded .50, the infants looked more at the Change event. *Note*: Black dots represent mean values. **p* < .05, 2-tailed.

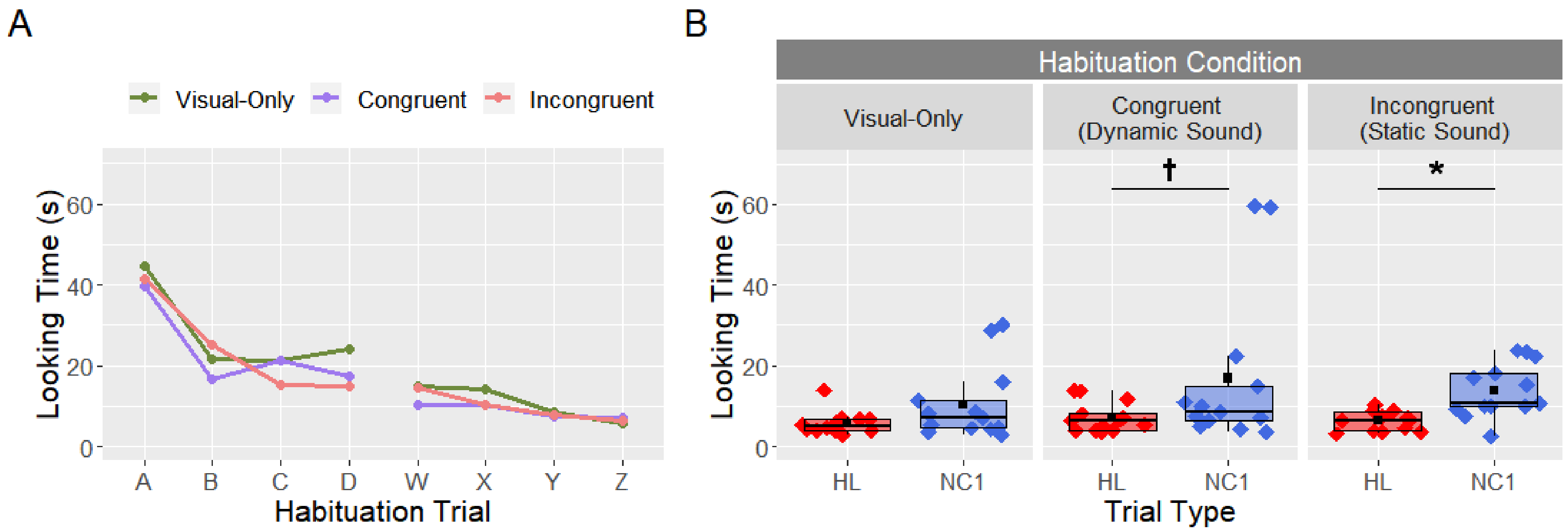

Figure 3.5. (A) Mean looking time (in seconds) in the first four and the last four habituation trials, and (B) individual looking times (in seconds) in the Last Habituation trial and the first No Change test trial.

(A) In the Visual-Only condition, the animation was presented in silence. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. (B) The last habituation trial (HL, in red) and the first No Change test trial (NC1, in blue) were visually identical and depicted a ball that remained unchanged during the occlusion. *Note*: Black dots represent mean values. *p < .05, $^{\dagger}p$ < .10, 2-tailed.

These observations were confirmed by a 3 (Habituation Condition: *Visual-Only* vs. *Congruent* vs. *Incongruent*) x 2 (Test Event: *Change* vs. *No Change*) x 3 (Test Block: *Block 1* vs. *Block 2* vs. *Block 3*) mixed ANOVA. Habituation Condition was manipulated between-subjects, and Test Event and Test Block were manipulated within-subjects. Test Block was included as a factor in the analysis to account for practice and fatigue. The analysis was conducted on $\log_{10}$-transformed data because the raw data was skewed (see also Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016). There was a main effect of Test Event, as the infants looked longer at the *Change* event (M = 1.21, SD = .23) than the *No Change* event (M = .99, SD = .22), F(1, 36) = 40.35, p < .001, η_p^2 = .53. No other main effects or interactions reached significance - Test Block: F(2, 72) = 1.07, p = .35, *ns*; Habituation Condition: F(2, 36) = 1.02, p = .37, *ns*; Test Block x Habituation Condition: F(4, 72) = 1.10, p = .36, *ns*; Test Event x Habituation Condition: F(2, 36) = .08, p = .92, *ns*; Test Block x Test Event: F(2, 72) = 1.83, p = .17, *ns*; Test Block x Habituation Condition x Test Event: F(4, 72) = .18, p = .95, *ns*.[29]

To test the *a priori* hypotheses that the infants would look longer at the *Change* event compared to the *No Change* event in the *Visual-Only* and the *Incongruent (Static Sound)* conditions, I performed 3 sets of paired-samples *t*-tests. As it can be seen in **Table 3.2**, the effect sizes across comparisons were large. This suggests that, in all the three habituation

[29] When I included Test Event Order in the analysis, as a between-subjects factor, I found a significant 4-way interaction between Test Block x Habituation Condition x Test Event x Test Event Order, F(4, 66) = 4.39, p = .003, η_p^2 = .21. However, when the data set was split between infants who watched the *Change* event first and infants who watched the *No Change* event first, the only statistically significant effect was the main effect of Test Event [*No Change First* group: F(1, 18) = 18.14, p < .001, η_p^2 = .50; *Change First* group: F(1, 15) = 23.15, p < 001, η_p^2 = .61], which is similar to what I found in the main analysis.

conditions, the infants encoded the pattern on the ball during the habituation and they responded in a similar way to the *Change* event.

Table 3.2. Average looking time (log10-transformed data) at the two test events (Change vs. No Change) across habituation conditions.

	Test Event		**T-test**			
	Change	**No Change**	***t*-value**	***df***	***p*-value** (2-tailed)	**Cohen d_z**
Visual-Only	1.17 (.21)	.94 (.23)	3.92	12	.002*	1.09
Congruent (Dynamic Sound)	1.18 (.23)	.96 (.24)	4.02	12	.002*	1.12
Incongruent (Static Sound)	1.26 (.25)	1.06 (.20)	3.14	12	.009*	.87

Note. Values represent mean (SD). Change test event, the ball changed its pattern during the occlusion. No Change test event, the ball maintained its pattern during the occlusion. Visual-Only condition, the animation was presented in silence. Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. Looking time was averaged across the 3 test blocks. **p* < .05, 2-tailed.

Since the habituation display and the *No Change* test display were identical, the only difference being that for some infants the habituation trials were accompanied by a musical sound meanwhile the test trials were presented in silence, I checked whether the infants' looking time changed between the *Last Habituation* trial and the first *No Change* test trial (*No Change 1*; see **Figure 3.5B**). I chose only the *No Change 1* trial rather than an average of all three *No Change* trials to maintain the same level of signal to noise ratio within the comparison. A 2 (Trial Type: *Last Habituation* vs. *No Change 1*) x 3 (Habituation Condition: *Visual-Only* vs. *Congruent* vs. *Incongruent*) mixed ANOVA was conducted on $\log_{10}$-transformed looking time data. Habituation Condition was manipulated between-subjects, and Trial Type was manipulated within-subjects. Across all the habituation conditions, the infants' looking time increased from the *Last Habituation* trial (M = .76, SD = .19) to the *No Change 1* trial (M = 1.00, SD = .33), $F(1, 36)$ = 16.21, $p < .001$, $\eta_p^2 = .31$. No other main effects or interactions were found - Habituation Condition: $F(2, 36) = 1.49$, $p = .24$, *ns*; Trial Type x Habituation Condition: $F(2, 26) = .46$, $p = .64$, *ns*. The recovery in the infants' attention from the *Last Habituation* trial to the *No Change 1* trial: (1) was not statistically significant in the *Visual-Only* condition, $t(12) = 1.58$, $p = .14$, *ns*; (2) was marginally significant in the *Congruent* condition, $t(12) = 1.88$, p = .09, *ns*; and (3) was highly significant in the *Incongruent* condition, $t(12)$ = 4.74, $p < .001$, Cohen $d_z = 1.32$. This pattern of results suggests that the infants in the two multisensory conditions may have associated the musical sound with the visual display during the habituation, and the sudden termination of the sound may have surprised them.

3.3.1. Exploratory Analysis

Given that processing occlusion events and discriminating between visual patterns are visual-perceptual skills that are also measured by the Mullen VR scale (Mullen, 1995) and the EMQ PA section (Libertus & Landa, 2013), I reasoned that the infants' looking preference for the *Change* event would be positively correlated with their score on these two scales. The infants' looking preference was calculated by dividing the infants' looking time at the *Change* event by their accumulated looking time at both the *Change* and the *No Change* event, $PTLT_{change} = LT_{change}/(LT_{change} + LT_{no\ change})$. If $PTLT_{change}$ exceeded .50, the infants looked more at the *Change* event. As it can be seen in **Figure 3.4B**, $PTLT_{change}$ remained similar across the test blocks, $F(2, 76) = 1.55$, $p = .22$, *ns.*

$\text{PTLT}_{\text{change}}$ was positively correlated with infants' EMQ PA (Libertus & Landa, 2013) and marginally correlated with Mullen VR scores (Mullen, 1995; see **Table 3.3**).[30] These results suggest that theinfants' ability to orient towards and maintain attention to the *Change* event is related to their general visual-perceptual skills. One possible explanation for such a relationship is that the same visual-perceptual skills that underlie the infants' ability to detect the change in the ball's pattern may allow them to solve some of the EMQ PA and Mullen VR tasks that are performed with 10-month-old infants. Specifically, in the present study, the infants had to remember what the occluded ball looked like, and they had to sustain their attention to the occlusion event for at least 2.5 s to detect that the ball changed its pattern during the occlusion. Similarly, searching for a partially and/or a fully hidden object (tasks that are present in both the EMQ PA and the Mullen VR) requires the infants to remember how the object looks like and to remain focused on the task until they successfully retrieve the object. Therefore, it is not surprising that those infants who detected the change in the ball's pattern in my study were also the ones who scored higher on the EMQ PA and the Mullen VR scales.

[30] The correlation coefficient between EMQ PA and $\text{PTLT}_{\text{change}}$ in the current study, *r*(34) =.44, *p* = .007, was similar to that between EMQ PA and $\text{PTLT}_{\text{change}}$ at *Pre-test* in Chapter 2, *r*(12) = .49, *p* = .08. A Fisher's Z-transformation yielded a non-significant difference between correlations, *Z* = .18, *p* = .86, *ns*. Therefore, in my previous study, I may have failed to find a significant correlation between EMQ PA and $\text{PTLT}_{\text{change}}$ at *Pre-test* because of a lack of statistical power. A possible reason for why in my previous study EMQ PA score may have been positively correlated with *Pre-test Preference*, and negatively correlated with *Post-test Preference, r*(12) = -.45, *p* = .11, could be because the infants with better visual-perceptual skills may have noticed the change in pattern at *Pre-test* and may have maintained their attention for longer at the *Change* event. As a result, they may have encoded the *Change* event better, and they may have recognized it when it was presented again at *Post-test* (as a result, they looked less at the *Change* event).

Table 3.3. Pearson correlation coefficients between the infants' visual preference for the Change event ($PTLT_{change}$) and various participant characteristics.

	N	Correlation with $PTLT_{change}$
Age (days)	39	-.04
Mullen VR T-score	39	.29†
EMQ PA raw-score	36	.44*

Note. No Change event, the ball kept its pattern unchanged during the occlusion. Change event, the ball changed its pattern during the occlusion. $PTLT_{change}$, infants' looking preference for the Change event. The infants' looking preference was calculated by dividing the infants' looking time at the Change event by their accumulated looking time at both test events, $PTLT_{change} = LT_{change}/(LT_{change} + LT_{no\ change})$. Age, infants' age (in days) at test. Mullen VR T-score, standardized score on the Mullen Visual Reception Scale. EQM PA raw-score, raw score on the Early Motor Questionnaire Perception-Action section. Positive correlation coefficients indicate that the infants with a stronger visual preference for the Change event (i.e., higher $PTLT_{change}$ score), were older or scored higher on a particular measure. **p* < .05, †*p* < .10, 2-tailed.

3.4. Discussion

In this study, I sought to find out whether spatiotemporally congruent audiovisual stimulation affects 10-month-old infants' ability to encode the pattern on briefly occluded objects. I found that, across all the three habituation conditions (*Visual-Only*, *Congruent*, and *Incongruent*), the infants looked longer at the *Change* test event than the *No Change* test event. This finding suggests that the infants encoded the pattern on the briefly occluded object during the habituation and detected when this changed during the test trials.[31] I also found that, in the two multisensory conditions, the infants looked marginally (in the *Congruent* condition) or significantly (in the *Incongruent* condition) longer at the first *No Change* test trial than the *Last*

[31] The infants' looking time increased by ~6 s (corresponding to a 65% increase) from the *No Change* event to the *Change* event. I calculated the percentage change by transforming the Test Event means (calculated on log10-transformed data) into seconds via an exponential function. Then I subtracted the difference between the means and divided that by the *No Change* test event mean. Finally, I multiplied the quotient by 100.

Habituation trial. The two were visually identical, which indicates that the infants associated the musical sound to the visual display during the habituation and that they noticed that the sound stopped during the test trials. Finally, the exploratory analysis conducted on the infants' proportional looking time at the *Change* event showed that the infants' preference score was related to their visual-perceptual skills as measured via the EMQ PA section (Libertus & Landa, 2013) and the Mullen VR scale (Mullen, 1995). Both instruments measure infants' ability to: (1) track objects in space, (2) detect similarities in object pattern, and (3) retrieve partially and fully hidden objects. Therefore, the association between the infants' looking behaviour during the study and their scores on these scales provides evidence that the experimental task tapped into the infants' visual-perceptual skills.

These results partially support the hypotheses of this study. They show that 10-month-old infants encode the pattern on an object and use it to interpret occlusion events. To process this object property, the infants must detect the regularities between the elements present on the surface of that object. There is evidence that infants as young as a few days old process these regularities, and they even prefer patterned surfaces over plainly coloured surfaces (Fantz, 1961, 1963). However, it is not until sometime between 5.5 and 9.5 months of age that infants start to use pattern information to segment objects and individuate objects in a visual display (Needham, 1999; Needham & Baillargeon, 1997; Needham & Kaufman, 1997; Wilcox, 1999; Wilcox & Chapa, 2004).[32] It is unclear why the infants fail to use pattern information to disambiguate visual displays at a younger

[32] Object segmentation refers to organizing the surfaces into discrete visual objects. Object individuation means establishing the correspondence between two sequentially presented items.

age. However, the results of the current and those of the previous study (see Chapter 2) suggest that, even if the infants start to use the pattern as a cue to specific objects, they find it easier to process some object patterns than others (Bornstein et al., 1981; Bornstein & Krinsky, 1985; Humphrey et al., 1986).[33]

At the same time, the results contradict the hypothesis that infants would respond differently to the change in the ball's pattern depending on whether they were in the *Congruent (Dynamic Sound)* or the *Incongruent (Static Sound)* habituation condition. Lawson (1980) had reported that spatiotemporal congruency modulates infants' ability to form object-sound associations. These associations require infants to process: (1) the visual features of the item, (2) the acoustic properties of the sound, and (3) to attribute the sound to the object. In Lawson's study, the infants failed to associate an incongruent, unmodulated sound with a periodically moving object. This failure may have been because the unrelated sound distracted the infants. Other studies have also reported that auditory input hinders infants' visual processing and learning (Erickson & Newman, 2017; Robinson & Sloutsky, 2007a, 2007b, 2008). For example, Barr, Shuck, Salerno, Atkinson, & Linebarger (2010) found that infants are less likely to imitate an object-directed action if the demo video depicting it has background music than if the video has background music and video-matched sound effects. Despite the accumulating evidence that a task-irrelevant sound distracts

[33] An alternative explanation for the differences in results across studies could be that, in the current study, the data was less noisy because there were three test trials per test event that I averaged together. Whereas, in the previous study (see Chapter 2), there was only one test trial per test event. Averaging multiple test trials may have decreased the contribution of random, task-irrelevant variability within the data. At the same time, it may have allowed the differences in infants’ looking behaviour between test events to stand out.

infants from visual tasks, the results of this study suggest that the infants learned the pattern on the ball even in the *Incongruent* condition.

Equally possible is that the infants did not differentiate between the spatiotemporally congruent and the incongruent sound and that they attributed both sounds to the ball. Without direct access to the infants' subjective experience, I can only assume that the 10-month-old infants in my study could distinguish between the dynamically panned sound used in the *Congruent (Dynamic Sound)* condition and the equally balanced sound used in the *Incongruent (Static Sound)* condition. The sound manipulation employed in the current study was very similar to that employed by J. G. Bremner, Slater, Johnson, Mason, & Spring (2012). This study found that a spatiotemporally congruent sound helped 4-month-old infants represent an occluded object for longer. It would be surprising if older infants lost the ability to differentiate between the two sounds given that infants get better at localizing sounds in space as they grow older (Morrongiello, 1988). Besides, with increasing age, infants' ability to detect discrepancies between the location of a sound and that of an object improves (Fenwick & Morrongiello, 1998; Morrongiello, Fenwick, & Nutley, 1998). As to whether the infants attributed both types of sound to the ball, that is indeed possible. Adults tend to attribute a static sound to a moving visual object (Soto-Faraco et al., 2002). Meanwhile, 8-month-old infants do not prefer a colocated audiovisual event after they watch a dislocated event (e.g., a ball that moves in the opposite direction from the sound; J. G. Bremner et al., 2011). This finding suggests that even if the sound was dislocated, the 8-month-olds may have perceived it as colocated. Therefore, the infants in the present study may have attributed the incongruent sound to the ball instead of categorizing it as background noise (but see Lawson, 1980, for evidence against this argument).

An alternative explanation for the similarity in infants' response to the *Change* event compared to the *No Change* event across habituation conditions could be that the heightened attention to the former occlusion event reflects a violation of expectation (Baillargeon, 2004). According to this explanation, the infants could have had an *a priori* expectation that briefly occluded objects do not change pattern during the occlusion. Therefore, when an object changed during the occlusion, the infants detected the change and responded to it. Such an explanation implies that the habituation trials may have been unnecessary (Wang et al., 2004). Although this is possible, I think that the results are explained better by the perceptual similarity between the habituation display and the *No Change* test display (Bogartz et al., 2000; A. J. Bremner & Mareschal, 2004; Cashon & Cohen, 2000; L. B. Cohen, 2004; Munakata, 2000).

Evidence that familiarizing the infants with the stimuli affected their looking behaviour during the test trials comes from the comparison between the *Last Habituation* trial and the first *No Change* trial. The habituation and the *No Change* test displays were identical, except for the sound. In the *Visual-Only* habituation condition, there was no difference between the infants' looking time at these trials. However, in the *Congruent* and the *Incongruent* conditions, the infants looked marginally, respectively significantly longer at the first *No Change* test event. These results suggest that the infants in the multisensory conditions associated the sound with the visual display, and they detected when the sound stopped. Therefore, the habituation period did affect the infants' looking behaviour. However, without a baseline condition, in which the infants are exposed only to the test trials, I cannot argue decisively against an explanation based on "violation of expectation".

The fact that all the infants, irrespective of the habituation condition, encoded the surface features of the ball is inconsistent with the IRH (Bahrick & Lickliter, 2000, 2002, 2012). The IRH argues that, given the infants' limited attentional resources, infants pay more attention to those object/event properties that are specified by multiple sensory modalities instead of those that are defined by only one modality. In the *Congruent (Dynamic Sound)* habituation condition, the location and the trajectory of the ball were presumably more salient, because both the audition and the vision pointed to them. In the *Visual-Only* habituation condition, the visual properties of the ball (e.g., the pattern, shape, and colour) had no competition from the amodal properties (e.g., location, trajectory, tempo). Same in the *Incongruent (Static Sound)* condition, where the two sensory modalities specified different objects. Therefore, according to the IRH, only the infants in the *Visual-Only* and *Incongruent* conditions should have learned the ball's pattern. The results showed no differences between the habituation conditions. The size of the effect for the *Change* vs. *No Change* comparison was similar across habituation conditions, which suggests that all the infants encoded equally well the pattern on the ball. While these results are inconsistent with the *unimodal facilitation* prediction of the IRH, the IRH also argues that the effect is more robust in younger than older infants. According to the IRH, this is because the younger infants have fewer attentional resources.

This study investigated the effects of audiovisual spatiotemporal congruency on infants' learning of object pattern. Humans live in a dynamic, multisensory environment in which different sensory cues specify either the same object/event or two separate items. Determining which sensory inputs go together and integrating them into a unified percept is a demanding process that could distract the infants. The results of this study suggest that 10-month-old infants encode the pattern on an object irrespective whether they perceive the object via one or multiple sensory modalities. While older infants do not seem to be affected by multisensory information, it remains to be seen whether this is also the case with younger infants. The younger infants have less experience with cross-modal stimulation and less attentional control than older infants (Colombo, 2001). To test this hypothesis, I planned a follow-up study (reported in Chapter 4) with two younger groups of infants: a group of 4-month-olds and a group of 6-month-olds.

CHAPTER 4

Effects of multisensory stimulation on 4- and 6-month-old infants' encoding of object pattern

4.1. Introduction

The study reported in Chapter 3 showed that spatiotemporally congruent stimulation does not interfere with 10-month-old infants' encoding of visual object pattern. The pattern is a modality-specific object property that is defined only by vision. According to the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012), infants learn better the modality-specific object properties when they perceive the object through only one sensory modality. The study did not find differences between habituation conditions. Therefore, I concluded that the infants encoded equally well the pattern on the ball irrespective of the kind of stimulation they received: unisensory, multisensory - congruent, and multisensory - incongruent. While the results are inconsistent with the IRH, 10-month-old infants may have enough perceptual experience, attentional resources, and processing capacity to attend to and learn multiple object properties in a short episode of exploration. In line with this argument, the IRH predicts that the prioritisation of specific object properties during multimodal processing is more likely in younger infants.

Evidence that younger infants struggle to encode the modality-specific properties (e.g., shape, pattern, colour, pitch) of those objects that they encounter in multisensory contexts comes from Bahrick (1994). In this study, 3-, 5- and 7-month-old infants watched two films in which two different objects struck a surface irregularly. Object A made a high-pitch impact sound, and object B produced a low-pitch sound. After the habituation, Bahrick showed the same films to one group of infants (control group), and two novel films to another group (experimental group). In the novel films, the author switched the object-sound pairs (i.e., object A made a low-pitch sound, and object B generated a high-pitch sound). The 3- and 5-month-old infants responded similarly irrespective whether they watched the familiar or the novel films. The 7-month-old infants who watched the novel films regained interest in the stimuli, while those in the control group did not. This pattern of results suggests that the two younger groups of infants did not form object-sound associations during the habituation, presumably because they did not attend to the modality-specific object properties. Meanwhile, the 7-month-old infants formed object-sound associations.

Similarily, Lewkowicz (2004a, 2004b) found that while younger infants struggle to keep track of the serial order of three falling objects, older infants succeed on this task. In his experiments, Lewkowicz (2004a) showed infants three items that fell on to a ramp, then rolled down to the bottom of the ramp and came to a stand-still side-by-side (with the first item furthest away from the ramp). During the habituation, each object made a specific impact sound when it landed on the ramp, but it moved in the same direction and with the same speed across the display. In the test trials, the order of the stationary objects was either the same as the order in which they were dropped (familiar test trial) or different (novel test trial). The 4-month-old infants looked equally long at all the test trials. The 8-month-old infants looked longer at the novel trials. These results support the IRH and suggest that only the older infants paid attention to the modality-specific properties of the objects and learned the order in which they fell on the ramp. However, equally possible is that the infants may have tried to categorise the collision events rather than encode each falling object.

Research into how infants organise perceptual information and form categories has consistently shown that, when infants perceive a series of stimuli during the same episode of exploration, they try to detect the recurring features of the items and try to form perceptual categories (Mareschal et al., 2002; C. L. Miller et al., 1982; Quinn et al., 2001; Younger, 1985). Furthermore, there is evidence that infants do not categorise only static images or brief auditory stimuli. They also classify dynamic events, such as occlusion, containment, or covering events, which pose different cognitive challenges for infants (Baillargeon & Wang, 2002). In light of this research, the younger infants in Bahrick' s (1994) and Lewkowicz's (2004a, 2004b) studies may have failed to form object-sound associations because they engaged in building a perceptual category of "falling objects" or "collision events". To do this, the infants may have focused on those object properties that were common across the seen objects, which also happened to be the properties that were specified by both vision and audition (i.e., amodal properties, such as the duration of motion, speed, trajectory, tempo, and rhythm).

Support for this argument comes from three of the control experiments conducted by Lewkowicz (2004a, 2004b). In one of the experiments (Exp. 3 in 2004a), Lewkowicz removed the impact sound that the objects produced when they landed on the ramp. Lewkowicz found that the 4-month-old infants still failed to encode the serial order of the items, even though the IRH would have predicted otherwise because the presentation was unimodal. However, when Lewkowicz concealed the collision events by placing a screen in front of the landing area (Exp. 2 in

2004a; Exp. 4 in 2004b), the 4-month-old infants encoded the order of the objects and looked longer at the novel test trials. Therefore, when the recurring properties of the objects/events were no longer visible, the young infants were able to attend to the modality-specific properties of the objects.

Additional support for the object/event categorisation argument comes from Bahrick (1992). Bahrick habituated two groups of 3.5-month-old infants with only one object that struck a surface irregularly. After the habituation, the author presented the infants with either a new impact sound (Exp. 2) or a new item (Exp. 3). The author found that the infants discriminated the novel sound and object, and regained interest in the stimuli during the test trials. These results suggest that, when infants encounter a series of items, they overlook the unique features of those items and engage in a categorisation process.

While the findings reviewed above may have alternative explanations than the one proposed by the IRH, the study reported by Bahrick, Lickliter, & Flom (2006) provides more convincing evidence that multisensory congruent stimulation distracts young infants away from the visual properties of an object/event. More specifically, Bahrick et al. habituated different groups of infants aged 3, 5, and 8 months old with a hammer that repeatedly struck a surface. While some of the infants watched the hammer tapping in silence (*Unisensory* condition), the other infants both watched and heard it tapping (*Multisensory Congruent* condition). After the habituation, the authors changed the orientation of the hammer (i.e., it tapped upwards instead of downwards) and recorded how the infants responded to the change. If the infants regained interest in the new stimuli, the authors assumed that they had attended to the orientation of the hammer (i.e., a modality-specific property) during the habituation. Bahrick et al. found that, in the *Unisensory* condition, all the three age groups of infants displayed visual recovery. However, in the *Multisensory Congruent* condition, only the 8-month-old infants regained interest in the stimuli.[34] The authors took the results as evidence that the younger groups of infants failed to encode the orientation of the hammer because they focused on the amodal properties of the tapping event. However, an alternative explanation could be that the infants were distracted by the tapping sound. With regards to this, Bahrick et al. conducted another experiment with 3-month-old infants in which the sight and the sound of the hammer were no longer synchronous. In this experiment, the infants learned the orientation of the hammer and displayed visual recovery at test. Therefore, the kind of stimulation that infants receive seems to affect whether they encode the modality-specific object/event properties.

Given that the findings reviewed above suggest that multisensory stimulation impacts more younger infants than older infants, I decided to conduct the study reported in Chapter 3 with two younger age groups of infants: a group of 4-month-olds and a group of 6-month-olds. I opted for these two age groups because previous research has found that 4-month-old infants do not abstract colour and pattern information when they explore objects manually and visually (Hernandez-Reif & Bahrick, 2001). Furthermore, 4-month-old infants do not spontaneously use colour and pattern information to segment objects in a display (Needham, 1999) or to

[34] In the *Multisensory Congruent* condition, Bahrick et al. (2006) kept the tapping sound during the test trials.

individuate moving objects (Wilcox, 1999). By contrast, 6-month-old infants can abstract colour and pattern information when they look at an item that they are manually exploring (Hernandez-Reif & Bahrick, 2001). Additionally, 5.5-month-old infants can learn to attend to the pattern and colour of moving objects if they first watch an actor performing various actions with differently patterned objects (Wilcox & Chapa, 2004).[35]

As with the study reported in Chapter 3, there were three habituation conditions: *Visual-Only*, *Congruent (Dynamic Sound)* and *Incongruent (Static Sound)*. During the habituation, the infants watched a dotted ball that moved horizontally behind a box (I used the same stimuli as in Chapter 3). Depending on the habituation condition, the infants heard a musical sound which accompanied the movement of the ball. During the test trials, the infants watched two silent occlusion events: (1) the *Change* event, which depicted a change in the ball's pattern during the occlusion, and (2) the No *Change* event, in which the ball remained unchanged during the occlusion. I predicted that the 6-month-old infants, but not the 4-month-old infants, would detect the change in the ball's pattern and look longer at the *Change* event. Furthermore, based on the IRH, I predicted that the 6-month-old infants in the *Visual-Only* and *Incongruent (Static Sound)* conditions would look longer at the *Change* event. Meanwhile, the infants in the *Congruent (Dynamic Sound)* condition would display no difference between the test events (see also Bahrick et al., 2006).

[35] Wilcox, Smith, and Woods (2011) were able to prime 4.5-month-old infants to attend to the pattern and colour of briefly occluded objects by placing differently patterned items adjacent to each other, and by demonstrating different actions with each one of them.

4.2. Methods

4.2.1. Design

The study employed an infant-controlled habituation paradigm. There were two age groups of infants: *4-month-olds* and *6-month-olds*, who were randomly assigned to one of the following habituation conditions: *Visual-Only*, *Congruent (Dynamic Sound)*, and *Incongruent (Static Sound)*. During the test trials, the infants watched 2 occlusion events in silence: *Change* and *No Change*. The test events were presented twice each (i.e., 2 test blocks), in alternation, for a total of 4 test trials.[36] This resulted in a 2 x 3 x 2 x 2 mixed study design with *Age Group* (4-month-old vs. 6-month-old) and *Habituation Condition* (Visual-Only vs. Congruent vs. Incongruent) as between-subjects factors, and *Test Event* (Change vs. No Change) and *Test Block* (Block 1 vs. Block 2) as within-subjects factors. Test Block was included as a factor to account for practice effects and fatigue. The dependent variable was the infants' looking time at each test trial, defined as the interval of time that the infants looked at the stimuli between the beginning and the end of the trial.

4.2.2. Participants

The final sample consisted of *N* = 72 infants: *N* = 36 four-month-olds (*M* = 125.83 days, range = 108 - 139 days, 15 females), and *N* = 36 six-month-olds (*M* = 188.97 days, range = 170 - 203 days, 11 females). Twenty-five additional infants were tested (i.e., 25.77% of the total *N* = 97; *n* = 11 four-month-olds, and *n* = 14 six-month-olds), but their data was not included

[36] I opted for 4 test trials instead of 6 because a quarter (*n* = 9) of the *4-month-old* infants were too tired/bored and completed only 4 test trials (instead of 6 trials as the 10-month-old in Chapter 3).

in the analysis because they failed to complete at least 4 test trial (*n* = 11), did not habituate (*n* = 8), were too tired during the test trials as confirmed by their short looking time at the post-test control video - i.e., less than 5 s (*n* = 1), were too young/old (*n* = 2) and equipment failure (*n* = 3). The infants were full-term babies (i.e., gestational age: 37 weeks or more), did not have any sight or hearing problems, and the majority came from middle-class, White-Caucasian families. The participants were recruited in the same way as described in Chapter 2, via invitation letter and follow-up phone calls. The infants participated in the study only once, such that if an infant was tested at 4 months, the family was no longer invited for the study when the infant reached 6 months old.

Before the study, the infants were randomly assigned to one of the three habituation conditions. This resulted in *n* = 12 infants per habituation condition, in each age group. There were no significant differences between groups as indicated by the infants' scores on the Mullen Visual Reception Scale (Mullen VR; Mullen, 1995) and the Early Motor Questionnaire – Perception-Action section (EMQ PA; Libertus & Landa, 2013). Furthermore, the infants' age, accumulated looking time during the habituation, and level of attention at the beginning and the end of the study were similar across conditions (see **Table 4.1** and **Table 4.2**).

Table 4.1. Participant characteristics by habituation condition for the 4-month-old age group.

	Habituation Condition			One-Way ANOVA		
4-month-olds	**Visual-Only** (*n* = 12)	**Congruent (Dynamic Sound)** (*n* = 12)	**Incongruent (Static Sound)** (*n* = 12)	***F*-value**	***df1, df2***	***p*-value** (2-tailed)
Age (days)	125.42 (6.74)	128.25 (10.11)	123.83 (9.90)	.73	2, 33	.49
Mullen VR T-score	54.42 (5.73)	54.42 (4.89)	57.08 (4.17)	1.15	2, 33	.33
EMQ PA raw-score	-21.67 (4.48)	-20.45 (3.64)	-19.27 (5.57)	.77	2, 31[#]	.47
Habituation LT (s)	218.86 (117.50)	219.60 (68.22)	198.08 (49.38)	.26	2, 33	.78
Pre-test LT (s)	59.49 (1.39)	57.39 (5.81)	59.69 (.48)	1.62	2, 33	.21
Post-test LT (s)	51.71 (18.05)	55.51 (15.08)	56.63 (10.97)	.36	2, 33	.70

Note: Values represent mean (SD). Age, infants' age (in days) at test. Mullen VR T-score, standardized score on the Mullen Visual Reception Scale. EQM PA raw-score, raw score on the Early Motor Questionnaire Perception-Action section. Habituation LT (s), accumulated looking time during the habituation (seconds). Pre-test LT (s), infants' looking time at a 60 s control video that indicated the infants' level of attention before the study (seconds). Post-test LT (s), infants' looking time at the same control video that indicated the infants' level of attention at the end of the study (seconds). There were no significant differences between conditions. [#]The caregivers forgot to fill in the EMQ for two infants: *Congruent* (*n* = 1) and *Incongruent* (*n* = 1). See Chapter 2 (Section 2.2.2), for details on why EMQ PA raw-score value is negative.

Table 4.2. Participant characteristics by habituation condition for the 6-month-old age group.

6-month-olds	Habituation Condition			One-Way ANOVA		
	Visual-Only	Congruent (Dynamic Sound)	Incongruent (Static Sound)	*F*-value	*df1, df2*	*p*-value (2-tailed)
	(*n* = 12)	(*n* = 12)	(*n* = 12)			
Age (days)	189.33 (9.77)	188.67 (10.09)	188.92 (8.63)	.02	2, 33	.99
Mullen VR T-score	62.42 (7.50)	67.50 (5.35)	66.42 (5.92)	2.15	2, 33	.13
EMQ PA raw-score	-7.91 (10.20)	-10.00 (8.50)	-4.82 (6.78)	1.05	2, 31#	.36
Habituation LT (s)	168.18 (78.33)	179.86 (60.94)	161.66 (95.26)	.16	2, 33	.85
Pre-test LT (s)	43.89 (20.07)	51.27 (12.62)	51.94 (16.94)	.85	2, 33	.44
Post-test LT (s)	46.96 (18.61)	50.19 (14.25)	53.27 (14.64)	.47	2, 33	.63

Note. Values represent mean (SD). Age, infants' age (in days) at test. Mullen VR T-score, standardized score on the Mullen Visual Reception Scale. EQM PA raw-score, raw score on the Early Motor Questionnaire Perception-Action section. Habituation LT (s), accumulated looking time during the habituation (seconds). Pre-test LT (s), infants' looking time at a 60 s control video that indicated infants' level of attention before the study (seconds). Post-test LT (s), infants' looking time at the same control video that indicated the infants' level of attention at the end of the study (seconds). There were no significant differences between conditions. #The caregivers forgot to fill in the EMQ for two infants: *Visual-Only* (*n* = 1) and *Incongruent* (*n* = 1). See Chapter 2 (Section 2.2.2), for details on why EMQ PA raw-score value is negative.

A power analysis was carried out in G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) to estimate the number of participants needed in each habituation condition. The analysis was based on the looking time data reported in Chapter 3, pooled across habituation conditions.[37] The difference between the *Change* and the *No Change* test events in that study had an effect size corresponding to Cohen d_z = 1.04. The projected sample size for a paired samples t-test (2-tailed), with effect size = 1.04, alpha = .05, and power = .75, was minimum *N* = 9 infants. Therefore, I decided to test *n* = 12 infants per habituation condition, in each age group.

4.2.3. Apparatus and Stimuli

I used the same testing set-up and stimuli as described in Chapter 3. I first habituated the infants to a visual display presented either in silence or accompanied by a musical sound (see **Figure 4.1**). Then I showed infants two test events: *Change* and *No Change* (see **Figure 4.2**). The infants' looking behaviour was coded online, and a third of the video-recordings (*n* = 24) were coded offline by a research assistant. Both the experimenter and the research assistant were blind to the habituation condition and the test events that the infants were watching. Furthermore, the research assistant was unaware of the study hypotheses. For the selection of the video-recordings, I employed a stratified random sampling method, such that *n* = 12 videos were selected from each age group (*n* = 4 per habituation condition). A two-way mixed intra-class correlation analysis with absolute-

[37] I decided to base my power analysis on the pooled data from Chapter 3 because I did not find any effect of Habituation Condition in my previous study. However, even if I had been more conservative and had used only the data from the *Incongruent* habituation condition, which was the condition with the smallest difference between test events (Cohen d_z = .87), the expected effect size would still have been large.

agreement (Trevethan, 2017) yielded an excellent inter-rater agreement, $ICC_{2,1} = .99$, on the infants' total looking time during the test trials.

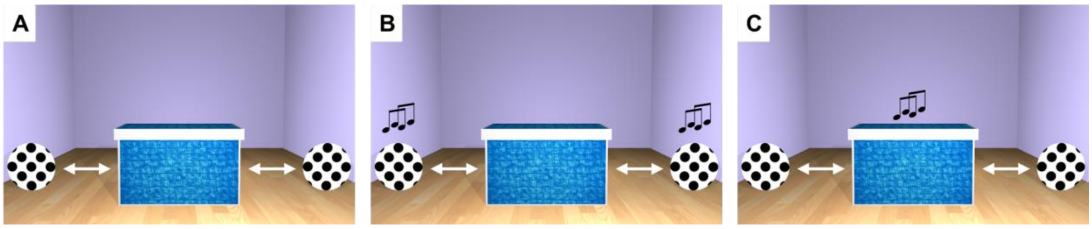

Figure 4.1. Habituation conditions employed in the study.

(A) Visual-Only condition. The habituation display was presented in silence. (B) Congruent (Dynamic Sound) condition. The habituation display was accompanied by a musical sound that appeared to originate from the ball. (C) Incongruent (Static Sound) condition. The habituation display was accompanied by a musical sound that appeared to originate from the box.

4.2.4. Procedure

The study procedure was the same as that described in Chapter 3, (see **Figure 4.2**). In a nutshell, there was a pre-test trial, followed by 5 to 12 habituation trials, then 4 test trials, and finally a post-test trial. About half of the infants viewed the *Change* test event first (n = 35), and the other half viewed the *No Change* test event first.[38] As per my previous study, if an infant watched less than 5 s of the post-test control video (i.e., less than 8% of the entire video), it was assumed that they were tired during the test trials and their data was excluded from the analysis (n = 1).

[38] The number of infants who watched the *Change* test event first was equally distributed across age groups and habituation conditions. Four-month-old: *Visual-Only* (n = 5), *Congruent* (n = 7), and *Incongruent* (n = 5). Six-month-old: *Visual-Only* (n = 6), *Congruent* (n = 7), and *Incongruent* (n = 5).

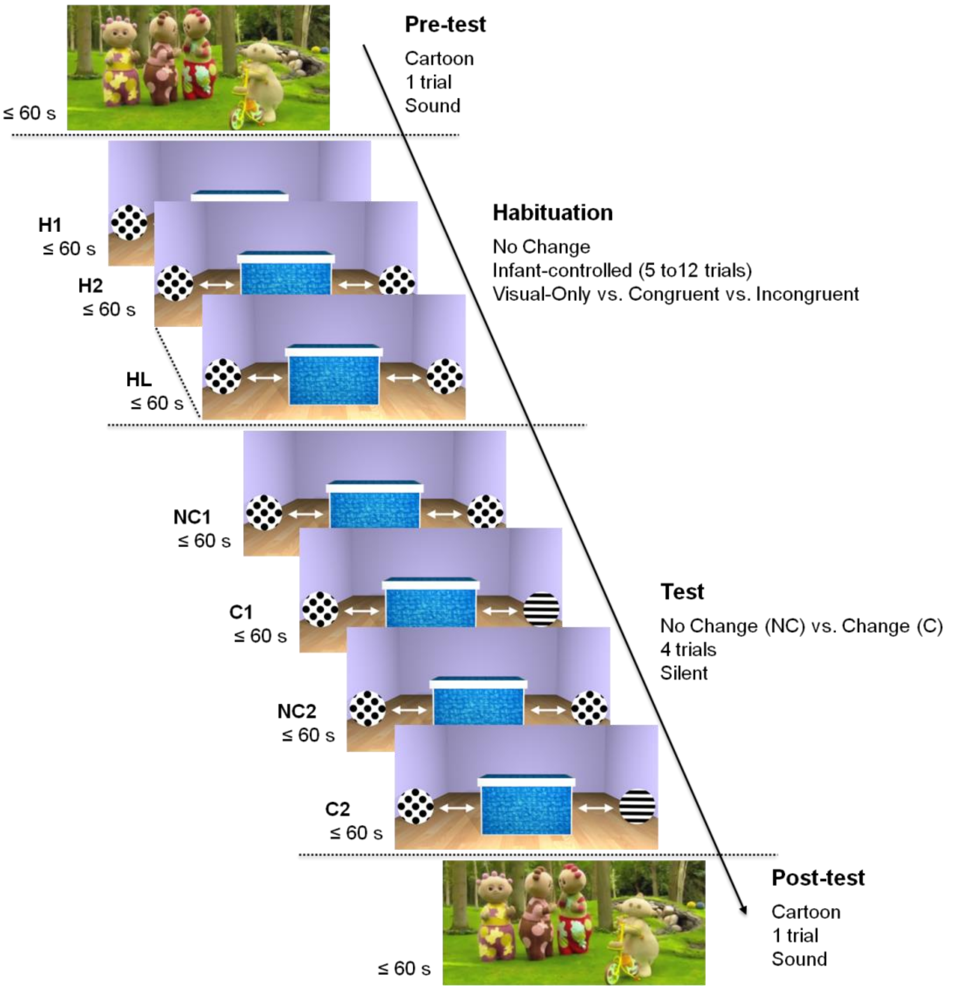

Figure 4.2. Timeline detailing the study procedure.

During Pre-test and Post-test, the infants watched a video from the TV series "In the Night Garden". During the habituation, the infants watched an animation of a dotted ball that moved horizontally behind a box. The habituation display was either presented in silence (Visual-Only), or it was accompanied by a musical sound. The sound was either spatiotemporally congruent with the ball, and "moved" left-right together with the ball (Congruent condition), or it was independent of the ball, and it remained located in the centre of the display throughout the trial (Incongruent condition). The habituation criterion was reached when the infants' accumulated looking time in the last 4 habituation trials was less than half of their accumulated looking time in the first 4 habituation trials. The infants completed between 5 and 12 habituation trials. At test, the infants watched two occlusion events in silence: the familiar event, in which the ball remained unchanged during the occlusion (No Change test event), and a novel event in which the ball changed its pattern during the occlusion (Change test event). The two test events were presented in alternating order, twice each, and about half of the infants watched the Change event first, while the other half watched the No Change event first. Each trial lasted 60 s or until the infants looked away from the screen for more than 2 s. In between the trials, an audiovisual animation was presented at the centre of the screen.

It took the infants ~7 minutes to complete the study. After the study, the infants had a short break (~5-minutes long) and then performed several age-appropriate visual-motor tasks from the Mullen VR scale (Mullen, 1995). The difficulty of the tasks increased as the infant progressed, and the testing was terminated when the infant was unsuccessful on 3 consecutive tasks. Subsequently, the caregivers filled in the EMQ (Libertus & Landa, 2013) while the researcher entertained the infants. Upon completing the questionnaire, the families were debriefed and were rewarded with a baby t-shirt and a certificate for participating in the study.

4.3. Results

Individual looking times (in seconds) at the two test events: *Change* and *No Change*, across habituation conditions and age groups, are displayed in **Figure 4.3**. There was a main effect of Test Block, 47 out of 72 infants looked longer at the stimuli in *Test Block 1* than in *Test Block 2*, Wilcoxon signed ranks test z = 2.39, p = .02. Therefore, the figures present the results separately for each test block. In *Test Block 1*, the infants looked longer at the *Change* event. In *Test Block 2*, the infants looked equally long at the two test events. As it can be seen in **Table 4.3**, the number of infants who looked longer at the *Change* event in *Test Block 1* was higher in the *Visual-Only* condition than in the other two habituation conditions.

These observations were confirmed by a 2 (Age Group: *4-month-old* vs. *6-month-old*) x 3 (Habituation Condition: *Visual-Only* vs. *Congruent* vs. *Incongruent*) x 2 (Test Event: *Change* vs. *No Change*) x 2 (Test Block: *Block 1* vs. *Block 2*) mixed ANOVA. Age Group and Habituation Condition were manipulated between-subjects, and Test Event and Test Block were manipulated within-subjects. The analysis was conducted on $\log_{10}$-transformed data because the raw data was positively skewed (see also Csibra, Hernik, Mascaro, Tatone, and Lengyel, 2016).

Table 4.3. The number of participants who looked longer at the Change event in Test Block 1 and Wilcoxon signed ranks test results.

Age Group	Habituation Condition	Group size	Preferred the Change Event	Wilcoxon Test	
		N	***n***	***z*-value**	***p*-value** (2-tailed)
4-month-old					
	Visual-Only	12	8	.71	.48
	Congruent (Dynamic Sound)	12	6	.08	.94
	Incongruent (Static Sound)	12	7	.55	.58
6-month-old					
	Visual-Only	12	9	2.51	.01*
	Congruent (Dynamic Sound)	12	9	1.18	.24
	Incongruent (Static Sound)	12	6	.16	.88

Note. Change test event, the ball changed pattern during the occlusion. No Change test event, the ball maintained its pattern during the occlusion. Visual-Only condition, the animation was presented in silence. Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. **p* < .05, 2-tailed.

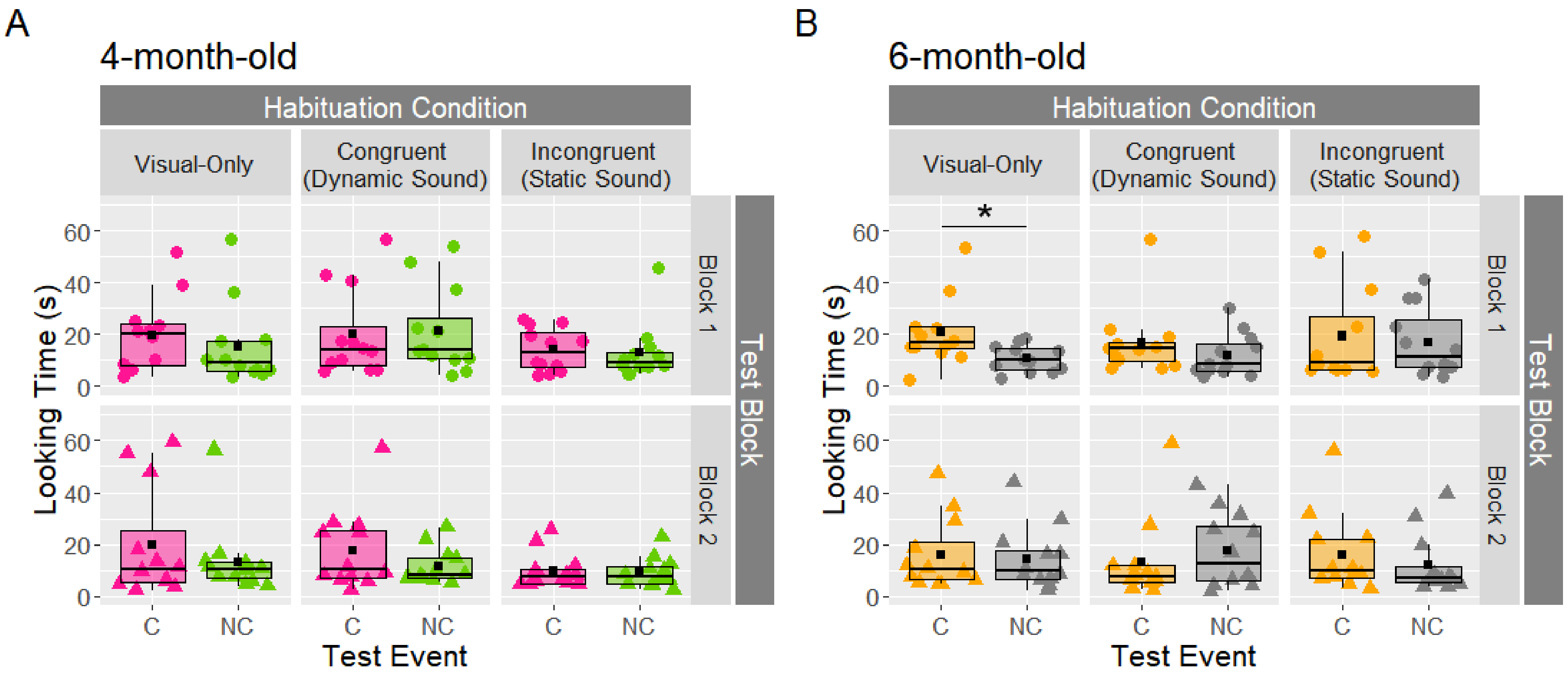

Figure 4.3. Individual looking times (in seconds) at the test events presented separately for each age group.

Change event (C), the ball changed its pattern during the occlusion. No Change event (NC), the ball kept its pattern during the occlusion. In the Visual-Only condition, the animation was presented in silence. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. *Note*: Black dots represent mean values. *$p < .05$, 2-tailed.

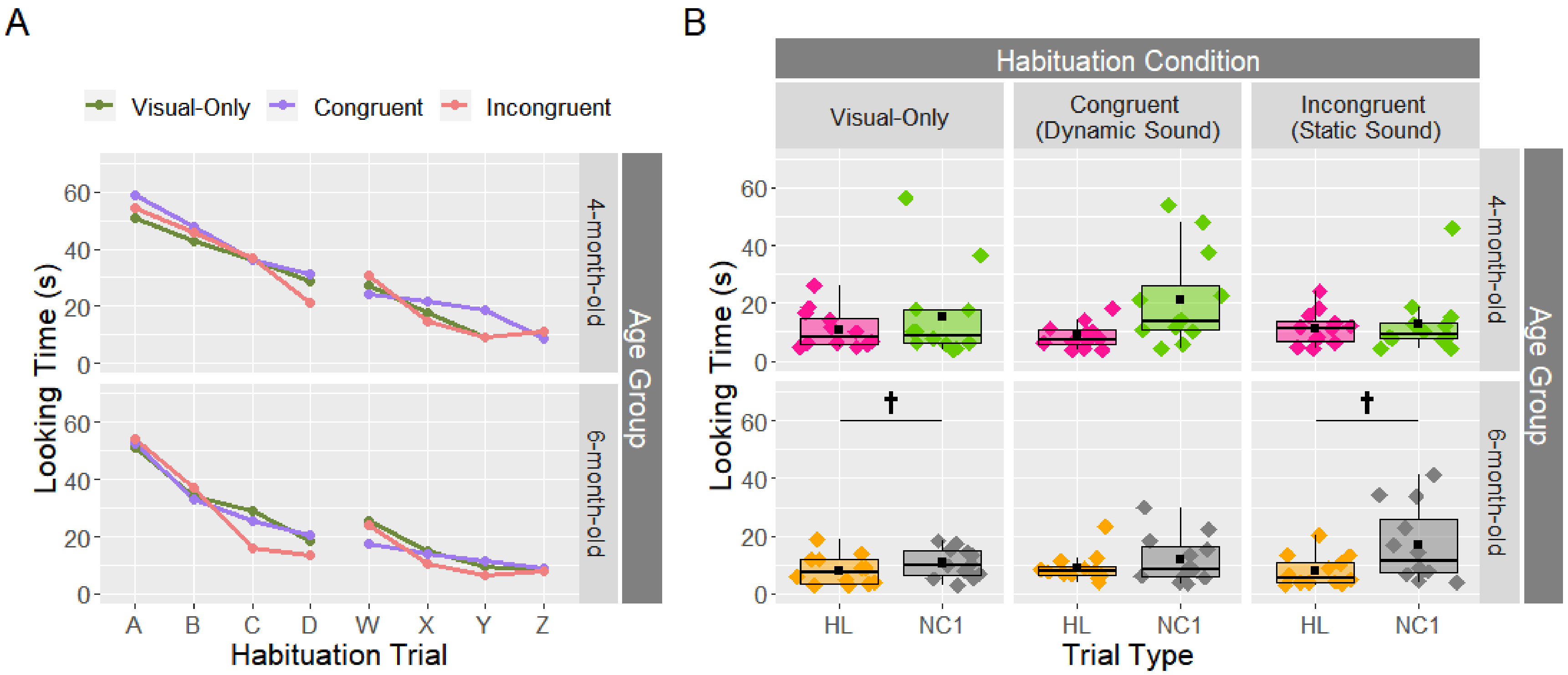

Figure 4.4. (A) Mean looking time (in seconds) in the first four and the last four habituation trials, and (B) individual looking times (in seconds) in the Last Habituation trial and the first No Change test trial.

(A) In the Visual-Only condition, the animation was presented in silence. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. (B) The last habituation trial (HL) and the first No Change test trial (NC1) were visually identical and depicted a ball that remained unchanged during the occlusion. *Note*: Black dots represent mean values. $^{\dagger}p < .10$, 2-tailed.

The analysis produced a main effect of Test Block, as the infants watched the stimuli for longer in *Test Block 1* (*M* = 1.10, *SD* = .24) than in *Test Block 2* (*M* = 1.00, *SD* = .28)*,* *F*(1, 66) = 8.94, *p* = .004, η_p^2 = .12. With the exception of Test Event, which was marginally significant - *F*(1, 66) = 3.08, *p* = .08 , η_p^2 = .05, no other main effects or interactions reached significance (all *F* < 1.65, all *p* > .20). Given that my participants were very young, and the test displays were very similar to the habituation display, I reasoned that any initial effects that may be observed in *Test Block 1*, could be reduced by fatigue in *Test Block 2*. Therefore, I decided to run a 3-way ANOVA between Age Group, Habituation Condition, and Test Event only on *Test Block 1*.

Test Block 1. There was a main effect of Test Event, as the infants looked longer at the *Change* event (*M* = 1.15, *SD* = .32) than the *No Change* event (*M* = 1.05, *SD* = .32)*,* *F*(1, 66) = 4.25, *p* = .04, η_p^2 = .06. No other main effects or interactions reached significance - Habituation Condition: *F*(2, 66) = .26, *p* = .77, *ns*; Age Group: *F*(1, 66) = .11, *p* = .74, *ns*; Test Event x Habituation Condition: *F*(2, 66) = 1.03, *p* = .36, *ns*; Test Event x Age Group: *F*(1, 66) = .84, *p* = .36, *ns*; Habituation Condition x Age Group: *F*(2, 66) = .99, *p* = .38, *ns*; Test Block x Habituation Condition x Age Group: *F*(2, 66) = .46, *p* = .63, *ns.*[39]

[39] When I included Test Event Order in the analysis, as a between-subjects factor, I found a significant main effect of Test Event, *F*(1, 60) = 4.89, *p* = .03, η_p^2 = .08, and a 2-way interaction between Test Event x Test Event Order, *F*(1, 60) = 12.66, *p* = .001, η_p^2 = .17. To follow-up the interaction, I split the data set between infants who watched the *Change* event first and infants who watched the *No Change* event first. In the *No Change First* group, there was no effect of Test Event, *t*(36) = .79, *p* = .43, *ns*. In the *Change First* group, there was a significant effect of Test Event, *t*(34) = 4.57, *p* < .001, Cohen d_z = .77, as the infants watched the *Change* event for longer. Such order effects are hard to interpret, for this reason I counterbalance the order of test events across participants.

To test the *a priori* prediction that only the 6-month-old infants would be able to detect the change in the ball's pattern, I followed-up the main effect of Test Event found in *Test Block 1* with two 2-way mixed ANOVAs between Test Event and Habituation Condition, one for each age group. In the *4-month-old* group, there was no significant main effect or interaction - Test Event: $F(1, 33) = .58$, $p = .45$, *ns*; Habituation Condition: $F(2, 33) = 1.13$, $p = .34$, *ns*; Test Event x Habituation Condition: $F(2, 33) = .46$, $p = .63$, *ns*. In the *6-month-old* group, there was a significant main effect of Test Event, $F(1, 33) = 5.12$, $p = .03$, $\eta_p^2 = .13$, as the infants preferred the *Change* event ($M = 1.16$, $SD = .31$) over the *No Change* event ($M = 1.02$, $SD = .30$). No other effects were found in this age group - Habituation Condition: $F(2, 33) = .12$, $p = .89$, *ns*; Test Event x Habituation Condition: $F(2, 33) = 1.11$, $p = .34$, *ns*. Planned comparisons revealed that only the 6-month-old infants in the *Visual-Only* condition looked significantly longer at the *Change* event, $t(11) = 3.27$, $p = .007$, Cohen $d_z = .95$ (see **Table 4.4**).

As with the previous study (see Chapter 3), I assessed whether the infants' looking behaviour changed between the *Last Habituation* trial and the first *No Change* test trial (*No Change 1*; see **Figure 4.4B**). The habituation display and the *No Change* test display were identical, the only difference being that for some infants the habituation trials were accompanied by a musical sound meanwhile the test trials were presented in silence. For this comparison, I conducted a 2 (Trial Type: *Last Habituation* vs. *No Change 1*) x 3 (Habituation Condition: *Visual-Only* vs. *Congruent* vs. *Incongruent*) x 2 (Age Group: *4-month-old* vs. *6-month-old*) mixed ANOVA on $\log_{10}$-transformed looking time data. Trial Type was manipulated within-subjects, and Habituation Condition and Age Group were manipulated between-subjects. The analysis yielded a main effect of Trial Type, as the infants' attention recovered from the *Last Habituation* trial (*M* = .90, *SD* = .25) to the *No Change 1* trial (*M* = 1.05, *SD* = .32), *F*(1, 66) = 10.31, *p* = .002, η_p^2 = .14. The recovery was marginally significant in the *4-month-old* group, *F*(1, 33) = 3.57, *p* = .07, η_p^2 = .10, and significant in the *6-month-old* group, *F*(1, 33) = 7.38, *p* = .01, η_p^2 = .18. Across the three habituation conditions, the 6-month-old infants looked longer at the *No Change 1* trial (*M* = 1.02, *SD* = .30) than they looked at *Last Habituation* trial (*M* = .85, *SD* = .25). However, when considered separately for each habituation conditions, the increase was: (1) marginally significant in the *Visual-Only* condition, *t*(11) = 2.19, *p* = .05, Cohen d_z = .63, and in the *Incongruent (Static Sound)* condition, *t*(11) = 2.24, *p* = .05, Cohen d_z = .65, and (2) non-significant in the *Congruent (Dynamic Sound)* condition, *t*(11) = .61, *p* = .56, *ns*. It is unclear why the infants in the *Congruent (Dynamic Sound)* condition failed to show a significant increase in their looking times at *No Change 1*. This partially contradicts what I found in the previous study with 10-month-old infants (see Chapter 3) and suggests that either the infants in the current study did not associate the musical sound with the visual display, or they did not consider the change important for the outcome of the event.

Table 4.4. Average looking time (log10-transformed data) at the two test events (Change vs. No Change) in Test Block 1 presented separately for each age group of infants, and for the 6-month-old infants in each habituation condition.

Age Group	**Habituation Condition**	**Test Event**		**T-test**			
		Change	**No Change**	***t*-value**	***df***	***p*-value** (2-tailed)	**Cohen d_z**
4-month-old	All	1.14 (.33)	1.08 (.33)	.77	35	.45	.13
6-month-old	All	1.16 (.31)	1.02 (.30)	2.26	35	.03*	.38
	Visual-Only	1.23 (.31)	.97 (.25)	3.27	11	.007*	.95
	Congruent (Dynamic Sound)	1.14 (.25)	.98 (.30)	1.35	11	.20	.39
	Incongruent (Static Sound)	1.12 (.38)	1.10 (.36)	.16	11	.88	.05

Note. Values represent mean (SD). Change test event, the ball changed pattern during the occlusion. No Change test event, the ball maintained its pattern during the occlusion. Visual-Only condition, the animation was presented in silence. Congruent (Dynamic Sound) condition, the animation was presented together with a musical sound that was spatiotemporally congruent with the movement of the ball. Incongruent (Static Sound) condition, the animation was presented together with a musical sound that was incongruent with the ball. All the *t*-tests were planned comparisons. **p* < .05, 2-tailed.

4.3.1. Exploratory Analysis

In the previous study (see Chapter 3), I found that the 10-month-old infants who had better visual-perceptual skills looked longer at the *Change* event. To assess whether this was the case with the infants who participated in this study, I correlated the infants' scores on the Mullen VR scale (Mullen, 1995) and the EMQ PA section (Libertus & Landa, 2013) with their looking preference score. The looking preference score was calculated by dividing the infants' looking time at the *Change* event by their accumulated looking time at both test events, $PTLT_{change} = LT_{change}/(LT_{change} + LT_{no\ change})$. Looking time data was pooled across the test blocks.[40]

As shown in **Table 4.5**, none of the correlations was statistically significant. These results contradict my findings with 10-month-old infants and suggest that there is no relationship between 4- and 6-month-old infants' ability to detect changes in the pattern of a briefly occluded object and their general visual-perceptual skills. This may be because, at this young age, infants often fail on manual search tasks which measure indirectly infants' ability to represent occluded objects. Instead, infants succeed on other behavioural tasks which may not be measuring the same perceptual skills that modulate the processing of occlusion events and allow infants to notice that the occluded object has changed pattern.

[40] The results would have been similar if the analysis had been conducted only on data from *Test Block 1*.

Table 4.5. Pearson correlation coefficients between infants' visual preference for the Change event ($PTLT_{change}$) and various participant characteristics.

Age Group	Measure	*N*	Correlation with $PTLT_{change}$
4-month-old			
	Age (days)	36	.25
	Mullen VR T-score	36	.05
	EMQ PA raw-score	34	-.02
6-month-old			
	Age (days)	36	.06
	Mullen VR T-score	36	.16
	EMQ PA raw-score	34	-.10

Note. No Change event, the ball kept its pattern unchanged during the occlusion. Change event, the ball changed its pattern during the occlusion. Infants' looking preference was calculated by dividing infants' looking time at the Change event by their accumulated looking time at both test event, $PTLT_{change} = LT_{change}/(LT_{change} + LT_{no\ change})$. Age, infants' age (in days) at test. Mullen VR T-score, standardized score on the Mullen Visual Reception Scale. EQM PA raw-score, raw score on the Early Motor Questionnaire Perception-Action section. Positive correlation coefficients indicate that the infants who had a stronger visual preference for the Change event (i.e., higher $PTLT_{change}$ score), were older or scored higher on a particular measure.

4.4. Discussion

This study sought to find out whether habituating 4- and 6-month-old infants with a cartoon in which a moving ball is specified simultaneously across both vision and audition interferes with infants' ability to encode the pattern on the ball. I found that, while the 4-month-old infants did not discriminate between a test event that depicted a change in the ball's pattern (*Change* event) and another test event that showed no change (*No Change* event), the 6-month-old infants looked longer at the *Change* event. The difference in 6-month-olds' looking at the test events was statistically significant in the *Visual-Only* habituation condition (looking time increased on average by ~7.5 s or 82% increase from the *No Change* to the *Change* event).[41] However, it was non-significant in the *Congruent (Dynamic Sound)* and the *Incongruent (Static Sound)* habituation conditions (looking time increased by 45% in the *Congruent* condition and by 5% in the *Incongruent* condition). Although these results suggest that only the 6-month-old infants who habituated to the cartoon in silence encoded the pattern on the ball, the difference between the test events was transient, emerging only in *Test Block 1* (i.e., between the first *Change* event and the first *No Change* event). It is unclear why the infants failed to display a consistent differential response to the test events across all the test trials. However, I reasoned that the visual similarity between the test events prompted the infants to habituate and to lose interest in the stimuli as the study progressed.[42] Equally intriguing was the fact that the 6-month-old infants in the *Visual-Only* and the *Incongruent* conditions looked marginally more at the first *No Change* test trial than the *Last Habituation* trial. I expected the infants in the two multisensory conditions to regain interest in the *No Change* test display because the test trials were silent. Nonetheless, I was surprised to find that the infants in the *Visual-Only* condition exhibited visual recovery as well. I do not have a plausible explanation for this finding, and it may be spurious. Finally, the exploratory analysis did not reveal any association between the infants' preference for the *Change* event and their scores on the EMQ PA (Libertus & Landa, 2013) and Mullen VR (Mullen, 1995) scales. The lack of

[41] I calculated the percentage change by transforming the Test Event means (calculated on log10-transformed data) into seconds via an exponential function. Then I subtracted the difference between the means and divided that by the *No Change* test event mean. Finally, I multiplied the quotient by 100.

[42] Looking time decreased significantly between *Test Block 1* and *Test Block 2*.

an association may be because the visual-perceptual skills that these instruments measure in 4- and 6-month-old infants may not underlie the infants' ability to abstract object pattern and use it to interpret occlusion events.

These results are consistent with the hypotheses and suggest that 6-month-old infants, but not 4-month-old infants, spontaneously use pattern information to interpret occlusion events. Although the interaction between *Test Event* and *Age Group* was non-significant, I concluded that there were age-group differences because of a power analysis which suggested that my study was underpowered. More specifically, to demonstrate a significant interaction between *Test Event* and *Age Group*, with effect size η_p^2 = .013 (alpha = .05) and power = .75, I should have tested *N* = 134 infants in each age group. Due to limited resources, I was able to test only *N* = 36 infants per group. Therefore, I had to draw conclusions based on the results of the *t*-tests conducted to address the *a priori* hypothesis and on previously reported differences between 4- and 6-month-old infants.

Previous research has shown that 4-month-old infants do not use colour and pattern information to connect the visible segments of partially occluded objects (Kellman & Spelke, 1983), segregate stationary objects in a visual display (Needham, 1999), and individuate objects (Wilcox, 1999). One possible explanation for these findings is that, at this age, the infants have not yet learned that the colour and pattern of objects can guide perception (see Needham, 1999; Wilcox, 1999). Two lines of research support this argument. Firstly, when 4-month-old infants explore an item visually and manually, they encode the shape of the object but not its colour and pattern (Hernandez-Reif & Bahrick, 2001). Secondly, 4-month-old infants do not need to be primed to detect changes in the shape of an occluded object (Bahrick, 1992; Wilcox, 1999). However, they need to see two differently patterned items, positioned side-by-side, to notice that the pattern of a third object has changed during the occlusion (Wilcox et al., 2011; Wilcox & Chapa, 2004). In the present study, the infants watched a ball that frequently moved in and out of sight. This visual event does not prompt 4-month-old infants to attend to the colour and pattern of a visually tracked object. Therefore, this group of infants may not have even noticed when the ball changed.

While the 4-month-old infants do not spontaneously attend to the visual pattern of objects, the 6-month-old infants encode this object feature when they look at the items that they are manually exploring (Hernandez-Reif & Bahrick, 2001). Besides, the results of the present study suggest that 6-month-old infants can detect when an object changes pattern during a brief occlusion interval. Previously, Wilcox (1999) reported that 7.5-month-old infants look longer at an occlusion event if it depicts an occluded object that repeatedly changes pattern. Furthermore, Wilcox & Chapa (2004) found that, if 5.5-month-old infants first watch an actor that performs various actions with differently patterned items, then the infants pay more attention to this object feature. In the present study, I did not prime the 6-month-old infants to attend to the pattern on the ball. Instead, I used an infant-controlled habituation procedure which ensured that each infant took as long as they needed to process the visual event. This procedure may have given the infants more time to encode the pattern on the ball.[43]

The results of this study speak to the predictions of the IRH (Bahrick & Lickliter, 2000, 2002, 2012). The theory argues that infants attend more to the modality-specific properties of objects (e.g., shape, colour, pattern, orientation, or pitch) when only one sensory modality specifies them but not when multiple sensory modalities define them simultaneously. In my study, nine out of twelve 6-month-old infants in the *Visual-Only* habituation condition looked longer at the *Change* event than the *No Change* event. The *Change* event was perceptually more novel as it depicted a change in the ball's pattern that had not occurred during the habituation. For infants to detect this change, they had to either encode the ball's pattern or learn that it remained unchanged during the occlusion. Irrespective of what the infants learned during the habituation this result is consistent with the IRH. It suggests that, when the infants received only visual stimulation, they processed at least one modality-specific object property (i.e., the pattern).

On the other hand, when the ball's movement was accompanied by a musical sound that was either spatiotemporally congruent (*Congruent* condition) or incongruent with the ball (*Incongruent* condition), the 6-month-old infants responded only slightly different to the two test events. However, the difference in looking times was not statistically significant. The IRH argues that, when an object is specified simultaneously by multiple sensory modalities (like in the *Congruent* condition), the infants

[43] Wilcox (1999) and Wilcox & Chapa (2004) employed a fixed-interval familiarization procedure. During this procedure, all the infants watch the stimuli for the same interval of time (set by the researcher) before they watch the test trials. While the pre-set interval may be enough for some infants to process the visual display, this may not be the case for other infants.

attend to those object properties that are common to both sensory modalities (i.e., amodal properties), such as the object's location, trajectory, tempo and rhythm of motion. I did not test these object properties, so I cannot say whether the infants in the present study encoded them. Nonetheless, 75% of the infants in the *Congruent* condition looked longer at the *Change* event, which suggests that they may have encoded the pattern on the ball. That said, it is unclear whether they may have learned this object property less well than the infants in the *Visual-Only* condition. The interaction between *Test Event* and *Habituation Condition* was non-significant. But the power analysis I conducted showed that for an interaction with an effect size of η_p^2 = .013 to be significant when power = .75, I should have tested *N* = 36 6-month-old infants per habituation condition. Given that my study was underpowered, I interpreted the results based on the *t*-tests I conducted to test my *a priori* hypotheses. Because the difference in looking time between test events in the *Congruent* condition did not reach statistical significance, I assumed that this finding is potentially consistent with the IRH and with Bahrick et al. (2006). Bahrick et al. found that 3- and 5-month-old infants learn the orientation of a hammer (i.e., a modality-specific property) when they see the hammer tapping, but not when they both see and hear it tapping.

The results in the *Visual-Only* and the *Congruent (Dynamic Sound)* conditions suggest that infants may be learning better the modality-specific properties of objects in unisensory than in multisensory-synchronous contexts. However, the data gathered in the *Incongruent (Static Sound)* condition indicates that the 6-month-old infants in this condition struggled the most to encode the pattern on the ball during the habituation.[44] Only half of the infants in this *Incongruent (Static Sound)* condition looked longer at the *Change* event. This finding is inconsistent with Bahrick et al. (2006), who reported that a group of 3-month-old infants encoded the orientation of a tapping hammer despite hearing unrelated impact sounds while they watched the tapping event. Based on this finding, Bahrick et al. argued that the congruency between the sensory cues is what drives the infants away from the modality-specific object/event properties, and not the additional auditory information. One methodological difference between the present study and that of Bahrick et al. is that the incongruent sound used in this study lasted throughout the trial, and it appeared to originate for another object in the display. Meanwhile, in Bahrick et al., the incongruent sound was periodic, and it was independent of the visual tapping event. Therefore, in the present study, the infants may have struggled to decide whether the sound was relevant for the occlusion event. In turn, this may have interfered with their visual processing (see also Barr, Shuck, Salerno, Atkinson, & Linebarger, 2010; Robinson & Sloutsky, 2007a, 2007b, 2008).

This study built on my previous research into the effects of multisensory stimulation on infants' processing of occlusion events (see Chapter 3) by testing two younger age groups of infants: 4- and 6-month-old infants. I found that 4-month-old infants do not use the pattern on an object to disambiguate occlusion events. Meanwhile, 6-month-old infants process the (visual) object pattern and use it to keep track of briefly occluded objects. I also found that, unlike 10-month-old infants, 6-month-old infants are

[44] Despite the apparent differences in how infants from different habituation conditions responded to the test events, caution is needed when interpreting these results given that the interaction between Test Event and Habituation Condition was non-significant.

distracted by auditory input when abstracting an object's pattern. The effect was more pronounced when the sound appeared to originate from another item in the display. Nonetheless, it was also apparent when the sound seemed to come from the visually tracked object (however, this finding may have been spurious). To investigate whether auditory stimulation interferes with the infants' visual processing in general or it affects only the processing of some object features, I planned another study for 6-month-old infants, in which I intended to manipulate another object property aside from the pattern (see Chapter 5).

CHAPTER 5

Effects of multisensory stimulation on 6-month-old infants' encoding of object pattern and trajectory

5.1. Introduction

The study reported in Chapter 4 suggests that 6-month-old infants may learn better the pattern on an object when they perceive the item unimodally. After habituating to a ball that moved silently across the display, the infants looked longer at the test stimuli when the ball changed its pattern than when it did not. In contrast, when the movement of the ball was accompanied by a sound, which was either spatiotemporally congruent or incongruent with the ball, the infants did not exhibit a differential looking behaviour between the test events. These results partially support the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012). However, they do not reveal whether the infants learned any object properties during the multisensory conditions. Equally possible is that the auditory stimulation may have placed additional demands on the infants' attentional resources, and it may have distracted them from processing the occlusion event. To answer this question, in the current study, I decided to include a new test event which depicted a change in the ball's trajectory.

The IRH predicts that infants learn better the trajectory of an object when they both see and hear the object/event. For example, when infants see a car passing in front of them, they also hear the sound of its engine. Based on the difference in the arrival time of the sound between the two ears (i.e., the interaural time difference), the infants can locate the sound in space and track it as it is moving. Since both vision and audition specify the trajectory of the car, the IRH argues that this is an amodal object property which infants learn at the expense of the modality-specific car properties, such as its colour and pattern. Consistent with the IRH, various studies have shown that infants encode better the amodal properties of objects/events when they receive congruent audiovisual stimulation than incongruent or unimodal stimulation (Bahrick et al., 2002; Bahrick & Lickliter, 2000; J. G. Bremner et al., 2012; Hernandez-Reif & Bahrick, 2001; Kirkham, Wagner, et al., 2012).

One such amodal property is the rhythm. The rhythm is a regular succession of elements and breaks that is apparent in a dynamic event (e.g., when clapping our hands, we can clap them at equal intervals of time either once or twice in quick succession). When objects move, their movement pattern or rhythm is specified unimodally or bimodally. To study the effect of unimodal and bimodal object presentation on infants' learning of rhythm, Bahrick & Lickliter (2000) showed 5-month-old infants a video of a hammer repeatedly striking a surface. Some infants watched the video in silence (*Unimodal* condition), while other infants watched the video accompanied by a tapping sound (*Bimodal* condition). The sound was either synchronous with the visual event or asynchronous. After the infants habituated to the movement of the hammer, the authors changed the rhythm of tapping. Bahrick & Lickliter found that only the infants in the *Bimodal-Synchronous* habituation condition displayed visual recovery during the test trials. More specifically, the infants looked longer at the two test trials (which depicted a novel tapping rhythm) than at the last two habituation trials. These results suggest that receiving congruent audiovisual information about the tapping hammer helped the infants to learn the rhythm of the tapping (see also Lewkowicz & Marcovitch, 2006).

Another property that infants encode better when they experience an audiovisual object/event is the tempo. The tempo describes the speed or the pacing of elements within a dynamic event (e.g., when capping our hands, we can clap them either fast or slow - the average number of capping-bursts per minute defines the tempo of clapping). Using a similar testing paradigm as Bahrick & Lickliter (2000), Bahrick et al. (2002) found that 3-month-old infants learn better the tempo of a tapping hammer when both vision and audition specify the event (*Bimodal* condition), but not when either vision or audition does (*Unimodal* condition). While 3-month-old infants need to see and hear an object to detect its tempo, 5-month-old infants rely less on bimodal information, and they can discriminate this object property even when the item is presented only visually (Bahrick & Lickliter, 2004).[45] It is unclear what drives this change in infants' attention to different object/event properties or whether the presence of multisensory stimulation early in infancy is essential for the typical development of tempo and rhythm perception. However, training adults with temporally congruent audiovisual stimuli improves adults' perception of visual rhythm, whereas visual-only training does not (Barakat et al., 2015).

Arguably, discriminating the visual rhythm and tempo of an object would not be possible without continuously tracking the object's location in space. Providing redundant auditory and visual information about the location of an item helps infants remember where they last saw it. One

[45] A similar developmental change is apparent in infants' perception of rhythm. Five-month-old infants detect changes in the rhythm of a hammer tapping only after bimodal-synchronous habituation. By comparison, 8-month-old infants notice such changes both after bimodal-synchronous and unimodal habituation (Bahrick & Lickliter, 2004).

testing paradigm that has shed light into infants' object-location memory has been the manual search task. During the manual search task, infants first witness as the experimenter hides an item, and then they are encouraged to retrieve it. Employing this paradigm with both silent and sounding objects, Moore & Meltzoff (2008) found that 10-month-old infants were more successful at finding sounding items than silent ones.[46]

Further evidence that congruent audiovisual input aids infants' representation of an object's location comes from Shinskey (2017). In her studies, Shinskey employed a manual search task in which 10-month-old infants had to retrieve a hidden item from one of two adjacent containers. To succeed, the infants first had to remove the object from location A, for a few times, and then from location B. Infants aged between 9 and 12 months of age, who retrieve an object from container A more than once, continue to search for the item in container A even after seeing the experimenter placing the item in container B (Piaget, 1954). The exact mechanism behind infants' failure to adjust their searching behaviour is still under debate (Marcovitch & Zelazo, 2009; Ruffman et al., 2005; L. B. Smith et al., 1999). However, Shinskey found that 10-month-old infants searched more often in location B when the hidden object was a sounding object than a silent one.[47] These results and those of Moore & Meltzoff (2008) provide evidence that infants

[46] However, this was not the case in the group of 8.75-month-old infants that Moore & Meltzoff (2008) tested. A possible explanation being that manual search tasks require both a robust representation of the object (Munakata et al., 1997) and the ability to plan and execute an action (Bertenthal, 1996) which may not have fully developed by this age.

[47] The change in infants' behaviour in the sounding object condition may have been simply because the sound provided a perceptual cue to the object's location. That said, other studies (Kirkham, Richardson, Wu, & Johnson, 2012; Richardson & Kirkham, 2004) provide more compelling evidence that congruent audiovisual cues help infants to spatially index objects/events.

learn better the location of multisensory objects/events (see also Kirkham, Richardson, Wu, & Johnson, 2012; Richardson & Kirkham, 2004).

Considering these findings, I decided to assess whether, in the previous study (see Chapter 4), the multisensory presentation of the ball facilitated 6-month-old infants' learning of the ball's trajectory at the expense of the ball's pattern. Arguably, the ball's trajectory was specified by both vision and audition in the *Congruent (Dynamic Sound)* habituation condition. Meanwhile, in the *Visual-Only* and the *Incongruent (Static Sound)* habituation conditions, it was defined only by vision. To answer this question, in the present study, I habituated three groups of 6-month-old infants with a ball that moved back and forth behind a centrally located box. Some infants watched the event in silence (*Visual-Only* condition). Other infants heard a musical sound during the occlusion event. In the *Congruent (Dynamic Sound)* condition, the sound appeared to come from the ball. By contrast, in the *Incongruent (Static Sound)* condition, the sound seemed to originate from the box. After the habituation, the infants watched three test events in silence. One of the test events was visually identical to the habituation event (*No Change* event). The other two test events were perceptually more novel. In one test event, the ball changed its pattern during the occlusion (*Change* event), and in the other, the ball changed its trajectory during the occlusion (*Trajectory Change* event).

Based on the empirical evidence reviewed above and the IRH, I hypothesised that the infants in the *Congruent (Dynamic Sound)* habituation condition would learn the trajectory of the ball and would look longer at the *Trajectory* Change event than the *No Change* event. Furthermore, I expected infants in the *Visual-Only* habituation condition to learn the pattern on the ball and look longer at the *Change* event than the *No Change* event (see Chapter 4). Finally, in the *Incongruent (Static Sound)* habituation condition, I did not expect to find any difference between the test events (see Chapter 4; Bahrick & Lickliter, 2000).

5.2. Methods

5.2.1. Design

An infant-controlled habituation paradigm was used in this study. The infants were randomly assigned to one of the following habituation conditions: *Visual-Only*, *Congruent (Dynamic Sound)*, and *Incongruent (Static Sound)*. After the habituation, the infants watched 3 test events: *Change*, *No* Change, and *Trajectory Change*. The events were presented only once, in random order, for a total of 3 test trials.[48] This resulted in a 3 x 3 mixed study design with *Habituation Condition* (Visual-Only vs. Congruent vs. Incongruent) as a between-subjects factor and *Test Event* (Change vs. No Change vs. Trajectory Change) as a within-subjects factor. The dependent variable was the infants' looking time at the stimuli during each test trial.

5.2.2. Participants

The final sample had *N* = 42 six-month-old infants (*M* = 181.98 days, range = 164 - 197 days, 21 females). Five additional infants were tested (i.e., 11.90 % of the total *N* = 47), but they were not included in the analysis

[48] I opted for 3 test trials because the infants completed a preferential looking experiment before this study (see Chapter 6), and I was worried that the infants would get very tired if the study lasted too long.

because they were too fussy during the study (*n* = 2), did not meet the minimum gestational age of 37 weeks (*n* = 2), and experimenter error (*n* = 1). The infants were full-term babies (i.e., gestational age: 37 weeks or more), did not have any sight or hearing problems, as reported by the caregivers, and the majority came from middle-class, White-Caucasian families. The participants were recruited in the same way as described in Chapter 2 (i.e., via invitation letter and follow-up phone calls). The infants were randomly assigned to one of the three habituation conditions. This resulted in *n* = 14 infants per habituation condition. No significant differences were found between the groups in infants' age, accumulated looking time during the habituation, and level of attention at the beginning and the end of the study (see **Table 5.1**).

As with the previous studies, I carried out a power analysis in G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) to estimate the number of participants needed in each habituation condition. The analysis was based on the looking time data of the *6-month-olds* infants in the *Visual-Only* habituation condition, reported in Chapter 4. I based the power analysis on this group because it was the only one that showed a significant difference in looking behaviour between the *Change* and *No Change* test events (in *Test Block 1*), an effect which I aimed to reproduce in the present study. This difference had an effect size of Cohen d_z = .95. The projected sample size for a paired samples t-test (2-tailed), with effect size = .95, alpha = .05, and power = .75, was minimum *N* = 10 infants. I stopped testing when I had *n* = 14 infants in each habituation condition.

Table 5.1. Participant characteristics by habituation condition.

	Habituation Condition			**One-Way ANOVA**		
6-month-olds	**Visual-Only** (*n* = 14)	**Congruent (Dynamic Sound)** (*n* = 14)	**Incongruent (Static Sound)** (*n* = 14)	***F*-value**	***df1, df2***	***p*-value** (2-tailed)
Age (days)	184.43 (9.57)	179.86 (8.77)	181.64 (11.96)	.72	2, 41	.50
Habituation LT (s)	176.08 (51.02)	172.59 (43.34)	157.41 (40.33)	.68	2, 41	.51
Pre-test LT (s)	59.04 (2.59)	51.05 (16.53)	53.49 (11.54)	1.70	2, 41	.20
Post-test LT (s)	56.92 (10.61)	47.82 (15.53)	49.22 (16.30)	1.63	2, 41	.21

Note. Values represent mean (SD). Age, infants' age (in days) at test. Habituation LT (s), accumulated looking time during the habituation (seconds). Pre-test LT (s), infants' looking time at a 60 s control video that indicated the infants' level of attention before the study (seconds). Post-test LT (s), infants' looking time at the same control video that indicated the infants' level of attention at the end of the study (seconds). There were no significant differences between conditions.

5.2.3. Apparatus and Stimuli

The testing set-up and the stimuli were like those used in Chapters 3 (see **Figure 5.1**). The habituation display depicted a ball that moved horizontally behind a centrally located box. The cartoon either was silent (*Visual-Only* condition), or it was accompanied by a musical sound. In the *Congruent (Dynamic Sound)* condition, the musical sound was spatiotemporally congruent with the movement of the ball. In the *Incongruent (Static Sound)* condition, the sound was incongruent with the ball.

During the test trials, the infants watched in silence three occlusion events: *Change*, *No Change* and *Trajectory Change* (see **Figure 5.2**). In the *Change* event, the ball changed its pattern during the occlusion. In the *No Change* event, the display remained unchanged relative to the habituation display, but the sound stopped. Finally, in the *Trajectory Change* event, the ball changed its trajectory during the occlusion. Instead of moving Left - Right - Left behind the box (as in the habituation display), the ball translated half-way until it was behind the box then returned to its starting point and only after that it translated to the other side of the box. In other words, the new trajectory of the ball was: Left - Left - Right - Right - Left. In the *Trajectory Change* event, the ball appeared to the left and the right side of the box for the same number of times as it had done in the habituation display.

The infants' looking behaviour was coded online, and approximately a third of the video-recordings (*n* = 15) were re-coded offline by a research assistant to establish reliability.[49] Both the experimenter and the research assistant were blind to the habituation condition and the test events that the infants watched. Furthermore, the research assistant was unaware of the study hypotheses. According to a two-way mixed intra-class correlation analysis with absolute-agreement (Trevethan, 2017), there was an excellent inter-rater agreement, $ICC_{2,1} = .99$, on infants' total looking times at the test trials.

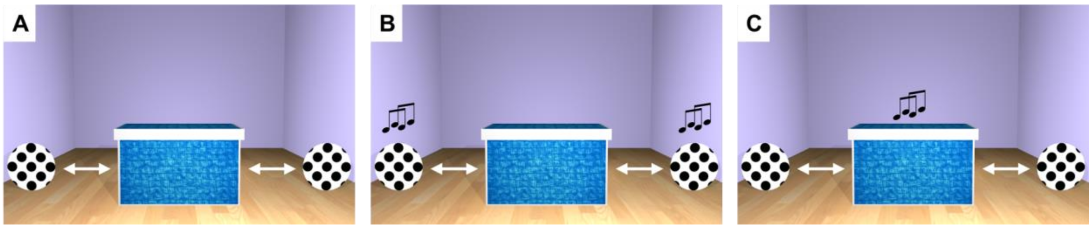

Figure 5.1. Habituation conditions employed in the study.

(A) Visual-Only condition. The habituation display was presented in silence (B) Congruent (Dynamic Sound) condition. The habituation display was accompanied by a musical sound that appeared to originate from the ball. (C) Incongruent (Static Sound) condition. The habituation display was accompanied by a musical sound that appeared to originate from the box.

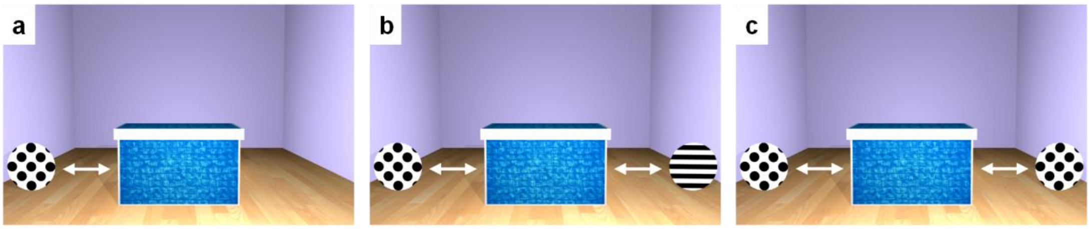

Figure 5.2. Test events used in the study.

(a) Trajectory Change event. The ball changed its trajectory during the occlusion. It translated up to half-point behind the box, then returned to its starting point, and only after that it translated to the other side of the box. (b) Change event. The ball changed its pattern during the occlusion - it went in dotted, and it remerged stripped and the reverse. (c) No Change event. The ball kept its trajectory and pattern unchanged during the occlusion. *Note:* The test events were presented in silence.

[49] The *n* = 15 videos coded offline were: *n* = 7 from the *Visual-Only* condition, *n* = 4 from the *Congruent* condition, and *n* = 4 from the *Incongruent* condition.

5.2.4. Procedure

The study procedure was similar to that described in Chapter 3 (see **Figure 5.3**). There was a *Pre-test* trial, followed by 5 to 12 habituation trials, then 3 test trials, and finally a *Post-test* trial. About a third of the infants watched the *Change* event first (*n* = 15), another third of the infants watched the *No Change* event first (*n* = 15), and the rest watched the *Trajectory Change* event first (*n* = 12). As per my previous studies, I used a minimum looking time at Post-test (i.e., 5 s or 8% of the entire video duration) as a criterion to include the infants' data in the statistical analysis. This was to ensure that the infants were not too tired during the test phase.

The infants took ~7 minutes to complete the study. Before the study, the infants completed a 4-minutes preferential looking study (see Chapter 6), in which one of two female faces spoke in synchrony with a voice recording. Given that the stimuli and the testing paradigms were different, I judged that any carry-over effects from one study to the other would be minimal. Upon completing both studies, the families were debriefed and received a baby t-shirt and a certificate for participating in the studies.

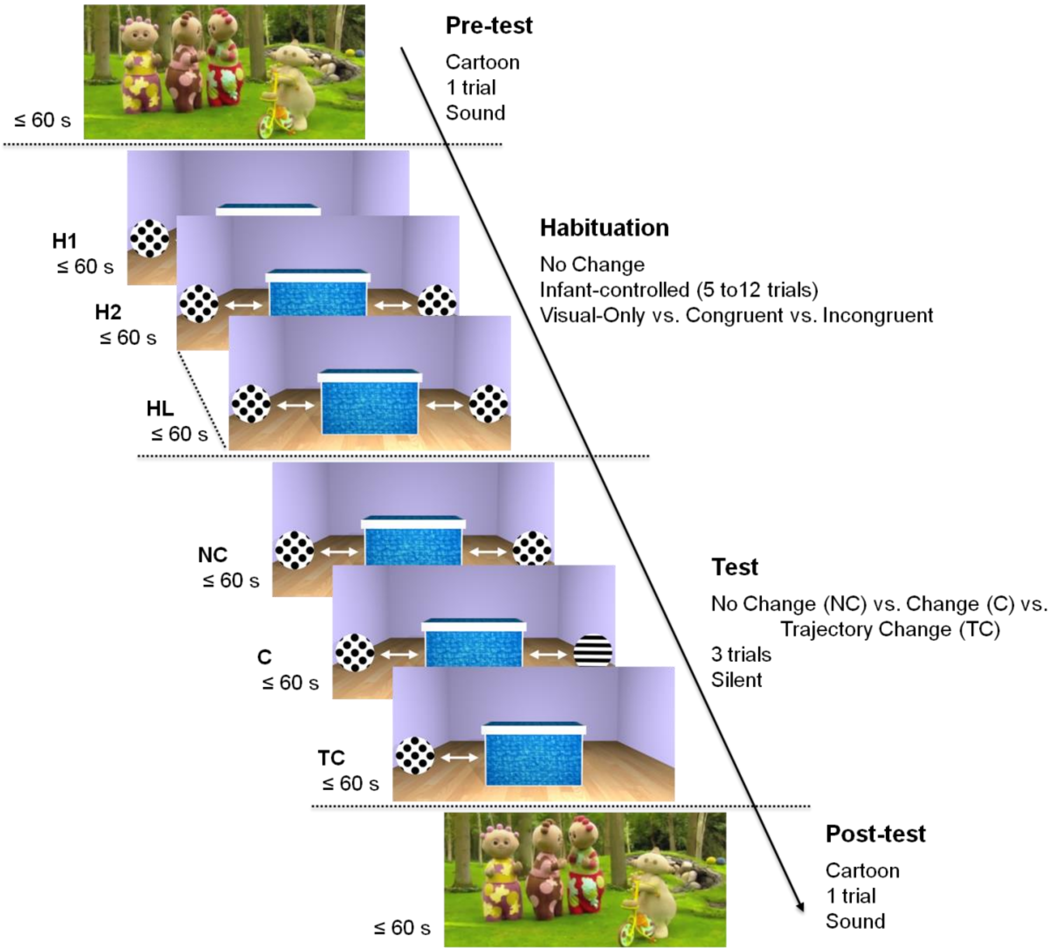

Figure 5.3. Timeline detailing the study procedure.

During Pre-test and Post-test, the infants watched a video from the TV series "In the Night Garden". During the habituation, the infants watched an animation of a dotted ball that moved horizontally behind a box. The habituation display was either presented in silence (Visual-Only), or it was accompanied by a musical sound. The sound was either spatiotemporally congruent with the ball, and "moved" left-right together with the ball (Congruent condition), or it was independent of the ball, and it remained located in the centre of the display throughout the trial (Incongruent condition). The habituation criterion was reached when the infants' accumulated looking time in the last 4 habituation trials was less than half of their accumulated looking time in the first 4 habituation trials. The infants completed between 5 and 12 habituation trials. At test, the infants watched three occlusion events in silence: the familiar event, in which the ball remained unchanged during the occlusion (No Change event), a novel event in which the ball changed its pattern during the occlusion (Change event), and another novel event in which the ball changed trajectory during the occlusion (Trajectory Change event). The three test events were presented in alternating order and about a third of the infants watched either of the test events first. Each trial lasted 60 s or until the infants looked away from the screen for more than 2 s. In between the trials, an audiovisual animation was presented at the centre of the screen.

5.3. Results

Infants' looking times (in seconds) at the three test events, *Change*, *No Change* and *Trajectory Change*, are displayed in **Figure 5.4**. Pooled across the habituation conditions, 28 out of 42 infants looked longer at the *Change* event vs. *No Change* event (Wilcoxon signed ranks test z = 2.77, p = .01) and 25 out of 42 infants looked longer at the *Trajectory Change* vs. *No Change* event (z = 2.22, p = .03). As it can be seen in **Table 5.2**, the number of infants who looked longer at either of the two novel events (*Change* or *Trajectory Change*) was higher in the *Visual-Only* and *Congruent (Dynamic Sound)* conditions.

To confirm these observations, I conducted a 3 (Habituation Condition: *Visual-Only* vs. *Congruent* vs. *Incongruent*) x 3 (Test Event: *Change* vs. *No Change* vs. *Trajectory Change*) mixed ANOVA. Habituation Condition was manipulated between-subjects and Test Event was manipulated within-subjects. The analysis was conducted on $\log_{10}$-transformed data because the raw data was positively-skewed (see also Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016).

The analysis yielded a main effect of Test Event, $F(2, 78) = 6.56$, p = .002, η_p^2 = .14, as the infants looked longer at the *Change* event (M = 1.04, SD = .27) vs. *No Change* event (M = .88, SD = .27), $t(41) = 3.30$, p = .002, Cohen d_z = .51, and at the *Trajectory Change* event (M = .99, SD = .30) vs. *No Change* event, $t(41) = 2.54$, p = .015, Cohen d_z = .39. The main effect of Test Event was qualified by a significant Test Event x Habituation Condition interaction, $F(4, 78) = 2.75$, $p = .03$, $\eta_p^2 = .12$.[50] This interaction was followed up with three one-way repeated measures ANOVA (one for each habituation condition) with Test Event as a within-subjects factor. The main effect of Test Event was marginally significant in the *Visual-Only* condition, $F(2, 26) = 3.14$, $p = .06$, $\eta_p^2 = .19$, highly significant in the *Congruent (Dynamic Sound)* condition, $F(2, 26) = 11.01$, $p < .001$, $\eta_p^2 = .46$, and non-significant in the *Incongruent (Static Sound)* condition, $F(2, 26) = .82$, $p = .45$, *ns*.

Since I hypothesized (1) that the infants in the *Visual-Only* condition would look longer at the *Change* event vs. *No Change* event, and (2) the infants in the *Congruent (Dynamic Sound)* condition would look longer at the *Trajectory Change* event vs. *No Change* event, I conducted planned comparisons between the different test events in each habituation condition (see **Table 5.3**). The analysis confirmed the predictions and it showed that the infants in the *Congruent (Dynamic Sound)* condition looked longer at the *Change* event than the *No Change* event, which was unexpected.

[50] I found similar results when Test Event Order (*Change*, *No Change* or *Trajectory Change* was presented first) was included in the analysis as a between-subjects factor. The main effect of Test Event Order was non-significant, $F(2, 33) = .94$, $p = .40$, *ns*, and neither was any interaction between Test Event Order and other factors, all $F < 1.12$, all $p > .36$, *ns*.

Table 5.2. The number of participants who looked longer at the Change vs. No Change event, and at the Trajectory Change vs. No Change event, in each habituation condition and Wilcoxon signed ranks test results.

Comparison	Habituation Condition	Group size	Preferred Change/Trajectory Change Event	Wilcoxon Test	
		N	*n*	*z*-value	*p*-value (2-tailed)
Change vs. No Change					
	Visual-Only	14	11	2.35	.02*
	Congruent (Dynamic Sound)	14	11	2.79	.01*
	Incongruent (Static Sound)	14	6	.35	.73
Trajectory Change vs. No Change					
	Visual-Only	14	9	1.60	.11
	Congruent (Dynamic Sound)	14	10	2.17	.03*
	Incongruent (Static Sound)	14	6	.16	.88

Note. Change test event, the ball changed its pattern during the occlusion. No Change test event, the ball maintained its pattern during the occlusion. Trajectory Change test event, the ball changed its trajectory during the occlusion. Visual-Only condition, the animation was presented in silence. Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. * *p* < .05, 2-tailed.

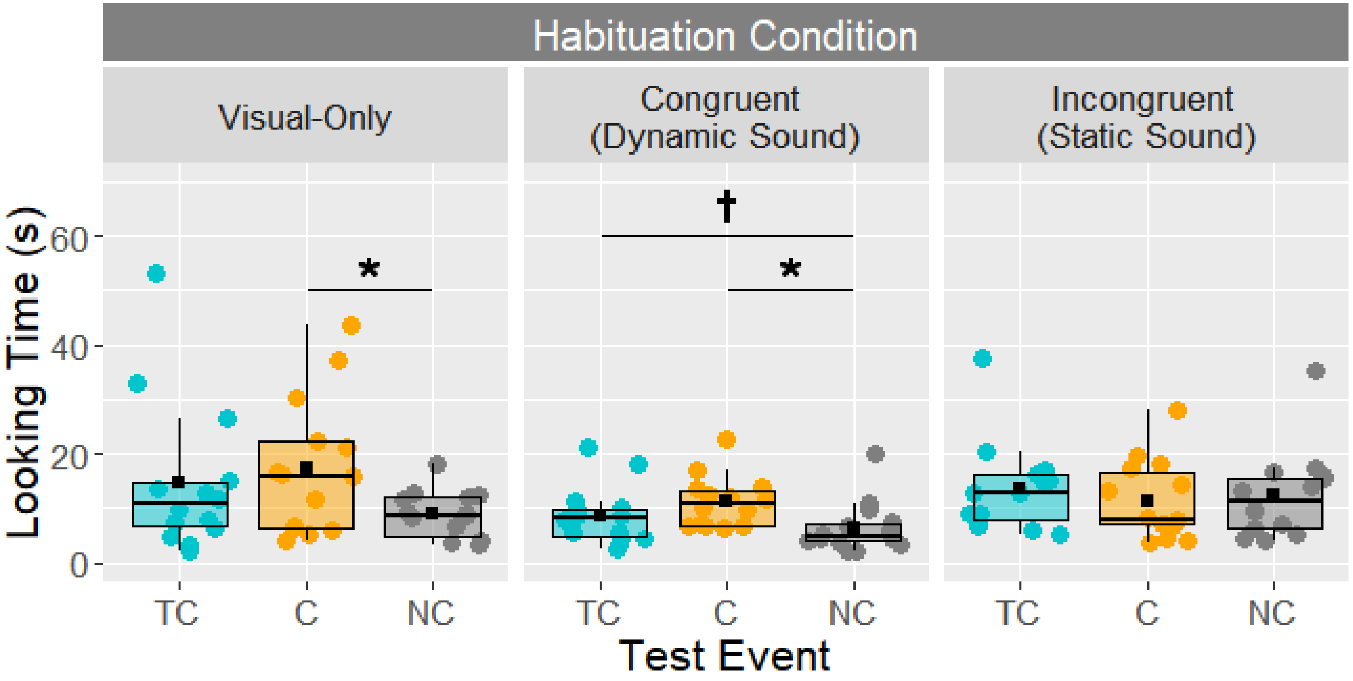

Figure 5.4. Individual looking times (in seconds) at the test events.

Trajectory Change event (TC), the ball changed its trajectory during the occlusion. Change event (C), the ball changed its pattern during the occlusion. No Change event (NC), the ball maintained its pattern and trajectory during the occlusion. In the Visual-Only condition, the animation was presented in silence. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. *Note*: Black dots represent mean values. *p < .05, $^{\dagger}p$ < .10, 2-tailed.

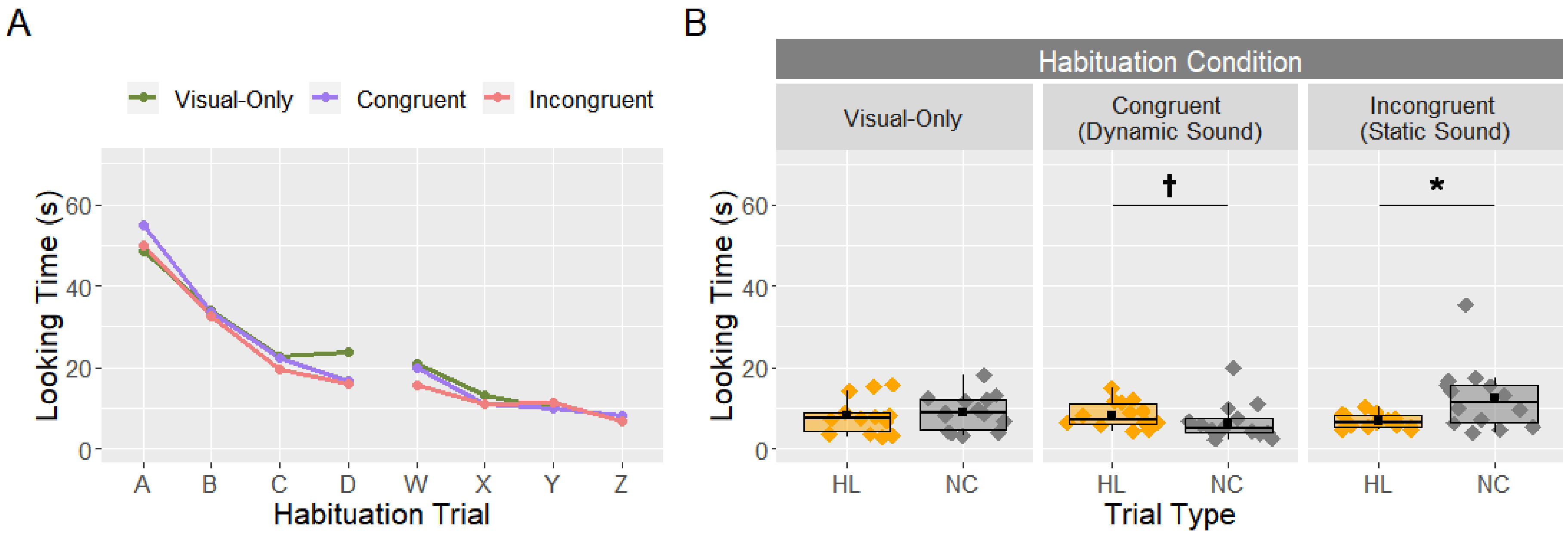

Figure 5.5. (A) Mean looking time (in seconds) in the first four and the last four habituation trials, and (B) individual looking times (in seconds) in the Last Habituation trial and the No Change test trial.

(A) In the Visual-Only condition, the animation was presented in silence. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. (B) The last habituation trial (HL) and the first No Change test trial (NC1) were visually identical and depicted a ball that remained unchanged during the occlusion. *Note*: Black dots represent mean values. *p < .05, †p < .10, 2-tailed.

Table 5.3. Average looking time (log10-transformed data) at the three test events (Trajectory Change vs. Change vs. No Change) presented separately for in each habituation condition.

Comparison	Habituation Condition	Test Event			T-test			
		Trajectory Change	Change	No Change	*t*-value	*df*	*p*-value (2-tailed)	Cohen d_z
Change vs. No Change								
	Visual-Only		1.13 (.33)	.89 (.23)	2.64	13	.02*	.71
	Congruent (Dynamic Sound)		1.02 (.17)	.74 (.26)	4.32	13	.001*	1.15
	Incongruent (Static Sound)		.98	1.02 (.27)	.56	13	.59	.15
Trajectory Change vs. No Change								
	Visual-Only	1.02 (.38)		.89 (.23)	1.62	13	.13	.43
	Congruent (Dynamic Sound)	.87 (.26)		.74 (.26)	2.09	13	.06†	.56
	Incongruent (Static Sound)	1.07 (.23)		1.02 (.27)	.73	13	.48	.20

Note. Values represent mean (SD). Trajectory Change test event, the ball changed its trajectory during the occlusion. Change test event, the ball changed its pattern during the occlusion. No Change test event, the ball maintained its trajectory and pattern during the occlusion. Visual-Only condition, the animation was presented in silence. Congruent (Dynamic Sound) condition, the animation was presented together with a musical sound that was spatiotemporally congruent with the movement of the ball. Incongruent (Static Sound) condition, the animation was presented together with a musical sound that was incongruent with the ball. *p < .05, †p < .10, 2-tailed.

To evaluate whether the infants' looking behaviour changed between the *Last Habituation* trial and the *No Change* test trial, I conducted a 2 (Trial Type: *Last Habituation* vs. *No Change*) x 3 (Habituation Condition: *Visual-Only* vs. *Congruent* vs. *Incongruent*) mixed ANOVA on $\log_{10}$-transformed. Trial Type was manipulated within-subjects and Habituation Condition was manipulated between-subjects. Since the habituation trials and the *No Change* test trial were visually identical, the only difference being that the test trials were presented in silence, I expected the infants in the multisensory habituation conditions (*Congruent* and *Incongruent*) to look longer at the *No Change* test trial. The analysis yielded a significant interaction between Trial Type x Habituation Condition, $F(2, 39) = 6.45$, $p = .004$, $\eta_p^2 = .25$. No other effects reached significance - Trial Type, $F(1, 39) = .60$, $p = .44$, *ns*; Habituation Condition, $F(2, 39) = 1.26$, $p = .30$, *ns*. The difference in looking times between the *Last Habituation* trial and the *No Change* trial was: (1) non-significant in the *Visual-Only* condition, $t(13) = .75$, $p = .47$, *ns*, (2) marginally significant in the *Congruent* condition, $t(13) = 2.07$, $p = .06$, Cohen $d_z = .55$, and (3) significant in the *Incongruent* condition, $t(13) = 2.90$, $p = .01$, Cohen $d_z = .81$. As expected, the infants in the two multisensory habituation conditions differentiated between the *Last Habituation* trial and the *No Change* trial (see **Figure 5.5B**). The infants in the *Congruent (Dynamic Sound)* condition watched the *Last Habituation* trial ($M = .89$; $SD = .16$) for longer than the *No Change* trial ($M = .74$; $SD = .26$), while infants in the *Incongruent (Static Sound)* condition looked longer at the stimuli in the *No Change* test trial ($M = 1.02$; $SD = .27$) than in the *Last Habituation* trial ($M = .82$; $SD = .11$).

5.3.1. Exploratory Analysis

Because I employed similar stimuli and testing procedure, and I tested 6-month-old infants in both the current study and the one reported in Chapter 4, I decided to check whether the effects found across experiments were robust. Furthermore, in the previous experiment, 75% of the infants in the *Visual-Only* and *Congruent (Dynamic Sound)* conditions looked longer at the *Change* than the *No Change* event. However, the difference was statistically significant only in the *Visual-Only* habituation condition. To assess whether this was because the previous study was underpowered or the effect achieved in the *Congruent* condition was too small (Cohen d_z = .39), I conducted a power analysis. To demonstrate a statistically significant difference between test events, in the *Congruent* condition, with a 2-tailed *t*-test (alpha = .05) with Cohen d_z = .39 and power =.75, I would have had to test *N* = 48 infants. However, due to limited resources, I could include only *N* = 12 infants in this condition in the previous study.

For these reasons, I compared how the 6-month-old infants in both the previous and the current study responded to the *Change* event vs. *No Change* event in the three habituation conditions (see **Figure 5.6**). To do this, I conducted a 2 (Test Event: *Change* vs. *No Change*) x 3 (Habituation Condition: *Visual-Only* vs. *Congruent* vs. *Incongruent*) x 2 (Experiment: *Previous* vs. *Current*) mixed ANOVA on $\log_{10}$-transformed looking time data. Test Event was manipulated within-subjects, and Habituation Condition and Experiment were manipulated between-subjects. To maintain the signal-to-noise ratio across experiments, I analysed only the *Test Block 1* data from the previous study.

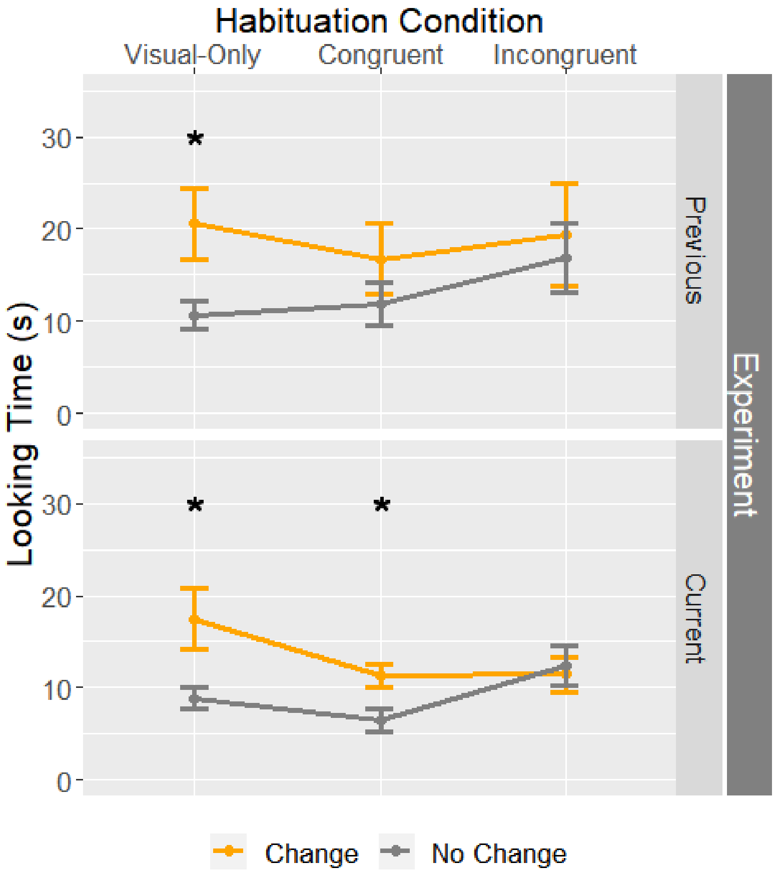

Figure 5.6. Mean looking times (in seconds) at the test events in the previous and the current studies.

Change event (in yellow), the ball changed its pattern during the occlusion. No Change event (in grey), the ball maintained its pattern during the occlusion. In the Visual-Only condition, the animation was presented in silence. In the Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. In the Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. Previous Experiment, mean looking times that the 6-month-old infants in the previous study (see Chapter 4) exhibited in Test Block 1. Current Experiment, mean looking times that the 6-month-old in the present study exhibited at test. *Note*: Error bars represent standard error of the mean. *$p < .05$, 2-tailed.

This analysis showed a main effect of Experiment, $F(1, 72) = 5.85$, $p = .02$, $\eta_p^2 = .08$, as the infants in the *Previous* study looked longer at the test stimuli ($M = 1.09$, $SD = .23$) than the infants in the *Current* study ($M = .96$, $SD = .23$). Furthermore, there was a main effect of Test Event, $F(1, 72) = 16.01$, $p < .001$, $\eta_p^2 = .18$, as the infants looked longer at the *Change* event

(*M* = 1.10, *SD* = .20) than the *No Change* event (*M* = .95, *SD* = .21). Finally, there was a significant interaction between Test Event x Habituation Condition, *F*(2, 72) = 4.56, *p* = .01, η_p^2 = .11. No other main effects or interactions reached significance - Habituation Condition, *F*(2, 72) = 1.15, *p* = .32, *ns*; Habituation Condition x Experiment, *F*(2, 72) = .32, *p* = .73, *ns*; Experiment x Test Event, *F*(1, 72) = .03, *p* = .87, *ns*; Habituation Condition x Test Event x Experiment, *F*(2, 72) = .52, *p* = .60, *ns*. As it can be seen in **Table 5.4**, only the infants in the *Visual-Only* and *Congruent (Dynamic Sound)* conditions looked significantly longer at the *Change* event vs. *No Change* event, which suggests that they detected the change in the ball's pattern.

Table 5.4. Average looking time (log10-transformed data) at the two test events (Change vs. No Change), pooled across the 6-month-old infants in the current study and those in the previous study (Test Block 1; see Chapter 4).

	Test Event		**T-test**			
	Change	**No Change**	***t*-value**	***df***	***p*-value** (2-tailed)	**Cohen d_z**
Visual-Only	1.17 (.32)	.93 (.24)	4.16	25	<.001*	.82
Congruent (Dynamic Sound)	1.08 (.22)	.85 (.24)	3.48	25	.002*	.68
Incongruent (Static Sound)	1.04 (.33)	1.06 (.31)	.17	25	.87	.03

Note. Values represent mean (SD). Change test event, the ball changed its pattern during the occlusion. No Change test event, the ball maintained its pattern during the occlusion. Visual-Only condition, the animation was presented in silence. Congruent (Dynamic Sound) condition, the animation was accompanied by a musical sound that was spatiotemporally congruent with the movement of the ball. Incongruent (Static Sound) condition, the animation was accompanied by a musical sound that was incongruent with the ball. Looking times were pooled across the current and the previous study (see Chapter 4), and from the previous study only the Test Block 1 data was used. **p* < .05, 2-tailed.

5.4. Discussion

This study tried to find whether 6-month-old infants learn different object properties (e.g., pattern, trajectory) when they only see an object compared to when they both see and hear it. I found that the infants in the *Visual-Only* condition looked longer at a test event that depicted a ball that changed pattern during the occlusion (*Change* event) than at a test event that showed no change at all (*No Change* event). Similarly, the infants in the *Congruent (Dynamic Sound)* condition looked longer at the *Change* event than the *No Change* event. Furthermore, this group of infants looked longer at a test event that depicted a change in the ball's trajectory (*Trajectory Change* event) than the *No Change* event. By contrast, the infants in the *Incongruent (Static Sound)* condition did not differentiate between the three test events. This result suggests that they processed the ball differently from the other two groups. Besides, I found that the infants' looking times changed between the *Last Habituation* trial and the *No Change* test event in the *Congruent (Dynamic Sound)* condition and in the *Incongruent (Static Sound)* condition. Specifically, in the *Congruent (Dynamic Sound)* condition, the infants' looking times decreased from the *Last Habituation* trial to the No *Change* test event. In the *Incongruent (Static Sound)* condition, the looking times increased. It is unclear why the results differ between habituation conditions. However, the fact that the infants' looking behaviour varied between the two trials suggests that they detected whether the musical sound was on/off. Finally, the exploratory analysis showed that the 6-month-old infants in the current experiment responded to the *Change* and the *No Change* test events in the same way as the 6-month-old infants in the previous study, in *Test Block 1* (see Chapter 4). The fact that I reproduced the results across two experiments provides evidence that (1) the effects reported are stable, and (2) the sensory context in which infants encounter an object affects their encoding of object pattern.

These results are partially consistent with the IRH (Bahrick & Lickliter, 2000, 2002, 2012). The theory argues that infants learn the modality-specific properties of objects (e.g., shape, colour, pattern, orientation, or pitch) when the items are perceived unimodally. Furthermore, the IRH predicts that infants learn the amodal properties of objects (e.g., tempo, rhythm, location, trajectory) when the items are specified bimodally. I found the infants that watched the ball moving silently across the display (*Visual-Only* condition), learned the ball's pattern during the habituation and discriminated between the test trials (see also Bahrick, Lickliter, & Flom, 2006). However, in addition to the visual pattern, the infants may have learned the ball's trajectory as well. Nine out of 14 infants in the *Visual-Only* condition also looked longer at the *Trajectory Change* event than the *No Change* event. The effect did not reach statistical significance, but it suggests that some of the infants encoded both features. Previously, Bahrick & Lickliter (2000) and Bahrick et al. (2002) found that two groups of 3-month-old infants failed to learn the rhythm and the tempo (two amodal properties) of unimodally presented objects. Consequently, the authors interpreted the results as evidence that the unimodal presentation of objects/events facilitates young infants' encoding of the modality-specific object/event properties at the expense of the amodal properties. The present results suggest that this interference does not occur when infants learn the object trajectory - another amodal property.[51]

In and of themselves, the findings I obtained in the *Visual-Only* condition are not problematic for the IRH. Nonetheless, the fact that I got a similar pattern of results in the *Congruent (Dynamic Sound)* habituation condition is inconsistent with the predictions of the IRH. Specifically, I found that when the ball was specified concurrently by vision and audition, the infants learned both the pattern and the trajectory of the ball and looked longer at the test events that depicted changes in these object properties than at the *No Change* event. I had expected the infants in this habituation condition to learn the ball trajectory, but not the pattern (see Bahrick et al., 2002; Bahrick & Lickliter, 2000). The fact that the infants encoded both object properties within a single episode of exploration suggests that the spatiotemporally congruent sound facilitated the learning of the object/event amodal properties without interfering with the processing of modality-specific properties.

Further support that spatiotemporally congruent audiovisual stimulation does not interfere with the encoding of modality-specific object/event properties comes from the analysis I conducted between the current study and the previous one (see Chapter 4). In the former experiment, I used similar stimuli and testing procedure, and I studied the same age group of infants (i.e., 6-month-olds). I found that the difference in looking times between test events was not statistically significant in the *Congruent (Dynamic Sound)* condition even though 75% of the infants

[51] The infants in the present study were 6-months-old. The fact that they were older than the infants tested by Bahrick & Lickliter (2000) and Bahrick et al. (2002) could also explain the different results.

watched for longer the *Change* event. Because, in the current study, I found that this difference was statistically significant, I decided to compare the infants' test-related looking times across studies and habituation conditions. This analysis revealed a statistically significant interaction only between *Habituation Condition* and *Test Event*. Based on these results, I concluded that 6-month-old infants learn the visual pattern (i.e., a modality-specific object property) depicted on an audiovisual object. The results of my studies are inconsistent with those of Bahrick et al. (2006), who found that congruent audiovisual stimulation prevented two groups of 3- and 5-month-old infants from learning the orientation of an object (which is also a modality-specific property). Therefore, I argue that either the redundant stimulation has different effects on different visual properties of an object/event, or the effects observed are specific to the testing paradigm used. Either way, the findings do not support the IRH.

In contrast to the infants in the *Visual-Only* and *Congruent (Dynamic Sound)* conditions, the infants in the *Incongruent (Static Sound)* habituation condition failed to learn both the pattern and the trajectory of the ball. Since the results in this condition are so different from those in the other two habituation conditions and mirror those in my previous study, I argue that this is the most significant finding of the current study. In essence, what this finding suggests is that a musical sound does not only have an alerting effect on young infants but, if that sound is spatiotemporally incongruent with the explored object/event, it impairs infants' visual processing of the object/event. One potential mechanism for this effect may be that the incongruent sound disrupted the infants' visual tracking during the habituation. As a result, the infants did not fixate the ball for long enough to encode its pattern and trajectory. However, without eye-tracking data, I cannot confirm whether the interval of time that the infants looked at the ball during the habituation was related to their looking behaviour during the test.

Irrespective of the cognitive mechanism that may have led to the infants' impaired processing of the ball's pattern and trajectory in the *Incongruent (Static Sound)* condition, these results are in line with those reported by other studies. Previously, Barr, Shuck, Salerno, Atkinson, & Linebarger (2010) found that a spatiotemporally incongruent sound affects infants' learning of object-directed actions, while J. G. Bremner et al. (2012) found that it interfered with infants' processing of the trajectory of a briefly occluded object (see also Kirkham, Wagner, et al., 2012). However, the results stand in contrast with those of Bahrick et al. (2006). Bahrick et al. found that an incongruent tapping sound did not interfere with young infants' encoding of the orientation of a tapping hammer. Given that the object orientation is a modality-specific property just like the pattern, it is surprising that the results are different. One explanation might be that Bahrick et al. used a periodic tapping sound during the habituation, which may have allowed the infants to determine faster that the sound was irrelevant to the tapping event observed and to ignore. By contrast, in the present study, the spatiotemporal incongruent sound was continuous, it lasted throughout the trial, and it appeared to originate from another object in the display. Therefore, the 6-month-old infants in this study may have struggled to decide whether they should link the sound to the ball. To address this research question, I could familiarise the infants with the incongruent sound before conducting the study. This way, the infants learn that the sound is irrelevant to the occlusion event. In this case, I would expect the infants to both the ball pattern and the trajectory during the habituation (see also Robinson & Sloutsky, 2007b).

The current study built on my previous findings that multisensory stimulation affects 6-month-old infants' learning of object pattern (see Chapter 4). Specifically, I investigated whether 6-month-olds learn different properties of a briefly occluded object (e.g., pattern, trajectory) when the infants watch the occlusion event in silence or accompanied by a musical sound. The sound either appeared to originate from the occluded item (i.e., spatiotemporally congruent sound) or a different object in the display (i.e., incongruent sound). I found that the infants learned the pattern and the trajectory of the briefly occluded object when the item was silent or appeared to produce the musical sound. However, they failed to do so when the sound was spatiotemporally incongruent with the object. These results partially support the IRH, as they show that young infants benefit from congruent audiovisual information when learning the trajectory of an object. However, the infants in this study did not prioritise the learning of the object trajectory (an amodal property) over that of the object pattern (a modality-specific property) when the audiovisual information was congruent, which is inconsistent with the IRH. Given these findings, I argue that further research is needed to understand whether young infants indeed prioritise the learning of amodal over modality-specific properties when they process multisensory objects/events. With regards to this, I planned another study (see Chapter 6), in which I intended to assess whether arbitrary associations of modality-specific properties (e.g., face-voice gender correspondences) affect infants' perception of amodal properties (e.g., audiovisual speech synchrony).

CHAPTER 6

Interactions between audiovisual gender correspondences and speech synchrony perception in 6-month-old infants

6.1. Introduction

In the study reported in Chapter 5, I tested two predictions of the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012): the *intersensory facilitation* and the *unimodal facilitation*. The results were inconsistent with the *unimodal facilitation* prediction because the 6-month-old infants who participated in the study learned the pattern on a ball (a modality-specific property) irrespective whether they received only visual stimulation or congruent (i.e., redundant) audiovisual stimulation. On the other hand, the results supported the *intersensory facilitation* prediction, because only the infants who received congruent audiovisual stimulation encoded the trajectory of the ball (an amodal property). Notably, when the visual and the auditory inputs were incongruent, the infants learned neither the object's pattern nor its trajectory. These findings (together with those reported in Chapters 2 to 4) raise the question as to whether infants prioritise the processing of the amodal properties over that of the modality-specific properties when they perceive an object/event concurrently across multiple sensory modalities. In other words, do infants perceive the amodal properties of these objects/events directly or unmediated as the IRH assumes (see Chapter 1)? To answer this question, in the present study, I looked at how audiovisual gender correspondences (i.e., arbitrary associations of modality-specific properties) influence the perception of speech synchrony (an amodal speech property).

Speech signals contain both spatiotemporal and semantic correlations between the auditory and the visual speech stimuli. When a person speaks, the auditory and the visual cues have the same onset, tempo, rhythm, and they originate from the same location in space. The IRH argues that these relations, which define speech synchrony, form the amodal properties of the speech signal, and they are detected automatically by infants (Bahrick & Lickliter, 2002, 2012). Besides, some visual features correspond to some acoustic characteristics of the speech signal, as is the case with face-voice gender correspondences. For example, women tend to have smaller faces, more arched eyebrows, narrower chins and less prominent noses than men (Bruce et al., 1993; Fellous, 1997). At the same time, women have high-pitched voices and a different timbre than men (Pernet & Belin, 2012). The IRH defines these gender correspondences as arbitrary associations and argues that, during an episode of exploration, infants first detect the amodal properties of the speech signal and then they identify the arbitrary face-voice relations (Bahrick & Lickliter, 2002, 2012).[52]

Some support for the IRH comes from the fact that, if 2-month-old infants habituate to a video of a person speaking in synchrony with a voice, they look longer at a test event which depicts the facial movements of the speaker lagging behind the voice by 400 ms or more (Lewkowicz, 2010). However, if equally young infants habituate to two videos of two people speaking, each uttering a monologue, they fail to respond when the speakers swap voices at test (Bahrick et al., 2005). While these findings provide some evidence that young infants may prioritise the detection of speech synchrony over that of face-voice correspondences, differences in task difficulty could explain the different results. Specifically, Lewkowicz

[52] There is inconsistency in how researchers classify face-voice gender correspondences. For example, Bahrick et al. (2005, p. 543) classify them as amodal relations. By contrast, Walker-Andrews (1994, p. 48) and Bahrick, Netto, & Hernandez-Reif (1998, p.1263) state that gender relations are neither exclusively amodal nor arbitrary. However, Bahrick & Lickliter (2002, 2012) argue that face-voice associations are arbitrary.

habituated infants with a video of one speaker, whereas Bahrick et al. showed infants alternating videos of two speakers. Habituating infants with two speakers increased the cognitive load and the task difficulty (in addition, it is unclear how easy it was to discriminate between the face-voice pairs) and may have resulted in the infants failing to notice the face-voice swap.[53]

Other research conducted on infants' processing of audiovisual speech stimuli has revealed that infants are sensitive to both speech synchrony and face-voice gender correspondences from a young age. In the case of speech synchrony perception, there is corroborating evidence that already at four months of age infants can detect which facial movement corresponds to the vowel sound heard. For example, when 4.5-month-old infants see two side-by-side speakers, one that says /a/ and the other one /i/, they look longer at the person who articulates the audible vowel (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999). Similarly, when 5- to 15-month-old infants watch two adjacent videos of a woman articulating two trisyllabic pseudo-words, infants prefer to look at the video that is synchronous with the sound (Baart et al., 2014). Finally, the infants' preference for synchronous speech is not limited to short, repetitive audiovisual stimuli. When shown two videos of a woman reciting a monologue in two different languages, 6-month-old infants spend more time looking at the video that corresponds to the sound (Kubicek et al., 2014; but see Shaw, Baart, Depowski, & Bortfeld, 2015). However, if the monologues belong to the same language, infants display a preference for the audible monologue only after 12 months of age

[53] It is unclear whether 2-month-old infants can detect changes in the acoustic characteristics of a person's voice if they habituate to only one speaker. Some evidence that young infants can discriminate auditory modifications in an audiovisual pair comes from Bahrick (1992; Exp. 2). Bahrick found that 3.5-month-old infants looked longer at a video of an object striking a surface when its impact sound changed from habituation to test.

(Lewkowicz et al., 2015). Altogether, these findings reveal a development in infants' ability to detect speech synchrony. Younger infants can identify synchrony when the speech stimuli are short and repetitive, whereas older infants can do this with more complex, multisyllabic stimuli. This developmental pattern is at odds with the IRH because, if speech synchrony detection were automatic, infants of different ages would perform similarly.

As with speech synchrony perception, infants' ability to identify gender-specific face-voice correspondences improves with age. For example, when 3.5-month-old infants listen to a voice recording of one of their parents uttering some sentences, they look more often at the parent whose voice they hear even if both parents are sitting silently in front of them (Spelke & Owsley, 1979). Similarly, when 6-month-old infants watch a video of two unfamiliar adults - a man and a woman - speaking in synchrony, side-by-side, they look more at the speaker who matches the gender of the voice heard. Interestingly, this preference for the gender-matched speaker is apparent irrespective of whether the lip movements are synchronous with the voice modulations or asynchronous (Richoz et al., 2017; Walker-Andrews et al., 1991). The effect is harder to detect when the stimuli are short, repetitive vowels (Patterson & Werker, 2002; Exp. 5), and seems to be more robust when infants match female voices and faces (Hillairet de Boisferon et al., 2015; Poulin-Dubois et al., 1994; Richoz et al., 2017). This asymmetry in detecting face-voice gender correspondences may reflect the fact that infants typically have more perceptual experience with female speakers. However, it is unclear whether this perceptual experience has to be with synchronous audiovisual speech, as the IRH proposes. It is equally possible that infants learn that there is a higher probability of hearing a high-pitched voice when they see a person with more feminine facial features. This alternative explanation does not require infants to perceive speech synchrony to identify face-voice gender correspondences. The recurrent exposure to specific acoustic characteristics alongside particular facial features would be enough to learn the statistical probabilities.

To find out whether listeners prioritise audiovisual speech synchrony over face-voice gender correspondences, some researchers have put the two types of relations into conflict. For example, Patterson & Werker (2002; Exp. 3) showed 4.5-month-old infants two side-by-side videos of a woman who repeatedly uttered the vowel /a/ in one video, and /i/ in the other. While the infants watched the videos, they heard a male voice that said either /a/ or /i/. Patterson & Werker reported that the infants looked longer at the synchronous video even though the voice belonged to the opposite gender. Although these results support the IRH, it is unclear whether the infants showed this preference because they prioritised the perception of speech synchrony or because the auditory stimuli were too short. By listening only to vowel sounds, the infants may have struggled to extract the gender-specific voice characteristics.[54] Some support for this latter explanation comes from a follow-up study by Patterson & Werker (Exp. 4). In this study, the infants saw simultaneously a video of a man saying /a/, and another one of a woman saying /i/. During the presentation of the videos, the infants heard a voice recording of a man articulating /i/. Patterson & Werker found that the infants

[54] Patterson & Werker (2002) found that 8- but not 6-month-old infants can make audiovisual gender matches when the stimuli are short, repetitive vowel sounds. Instead, with more naturalistic speech stimuli, infants can detect face-voice gender correspondences from 6 months of age (Walker-Andrews et al., 1991; Richoz et al., 2017).

looked equally long at both videos. The fact that the infants failed to show a preference for the synchronous video in this follow-up study suggests that they noticed the gender-relevant audiovisual cues and this information interfered with their ability to detect speech synchrony.

Adults too are susceptible to interference from audiovisual gender correspondences when processing speech synchrony. In a series of experiments, Vatakis & Spence (2007) showed participants audiovisual speech stimuli that had varying degrees of temporal asynchrony between them (i.e., either the video led the sound by 0 to 300 ms or the reverse). Furthermore, the stimuli were gender-matched (i.e., a female face uttered some words concurrently with a female voice) or they were gender-mismatched. On each trial, the participants had to judge if either the auditory or the visual stimulus came first. Vatakis & Spence found that adults need more time between the auditory and visual speech cues to indicate their order correctly when the stimuli are gender-matched than gender-mismatched (see also Vatakis, Ghazanfar, & Spence, 2008). These findings, together with those of Patterson & Werker (2002), suggest that listeners (naive or experts) do not always prioritise the processing of audiovisual speech synchrony over that of gender correspondences as the IRH proposes.

This study aimed to investigate further whether 6-month-old infants prioritise the processing of the amodal properties of audiovisual speech (i.e., speech synchrony) over that of the arbitrary face-voice correspondences, as the IRH argues. The experiment was like Patterson & Werker’s (2002) Exp. 3. However, it employed two longer voice recordings which were potentially more informative with regards to the gender of the voice than the vowel sounds used by Patterson & Werker. More specifically, during the experiment, the infants listened to a person who repeatedly uttered either "Hello baby!" or "Good job!". At the same time, the infants watched two side-by-side silent videos of a woman speaking. The videos originated from the same video recording, but I played one video forwards (in synchrony with the voice recording), and the other one backwards. In the *Congruent (Gender-Match)* condition, the infants heard a woman speaking (different from the person in the video). While in the *Incongruent (Gender-Mismatch)* condition, the infants listened to a voice recording of a man. I conducted the study with 6-month-old infants because at this age infants can match unfamiliar faces and voices by gender (Richoz et al., 2017; Walker-Andrews et al., 1991) and they can also detect speech synchrony in multisyllabic words (Baart et al., 2014; Kubicek et al., 2014). Based on the empirical findings reviewed above, I predicted that the infants in the *Congruent (Gender-Match)* condition would look longer at the synchronous video. Furthermore, I expected that the infants in the *Incongruent (Gender-Mismatch)* condition would display no preference for either video.

6.2. Methods

6.2.1. Design

In the present study, I used an intermodal preferential looking paradigm. The infants watched two side-by-side silent videos of a woman repeatedly uttering some phrases. One video was synchronous with a voice recording that played alongside the videos, while the other one was asynchronous. In one condition - *Congruent (Gender-Match)* - the voice was female, and in the other – *Incongruent (Gender-Mismatch)* - the voice was male. For the female voice recording, I asked a different woman than the one in the videos to record her voice. During the study, I used two phrases: "*Hello baby!*" and "*Good job!*" to ensure that my results do not just reflect one set of stimuli. The study had a 2 x 2 x 2 mixed design with *Condition* (Congruent vs. Incongruent) as a between-subjects factor and *Video* (Synchronous vs. Asynchronous) and *Phrase* (Hello baby! vs. Good job!) as within-subjects factors. The dependent variable was the infants' proportional total looking time to the synchronous video ($\text{PTLT}_{\text{sync}}$). This was the time that the infants spent looking at the synchronous video divided by their accumulated looking time at both videos ($\text{PTLT}_{\text{sync}} = \text{LT}_{\text{sync}} / (\text{LT}_{\text{sync}} + \text{LT}_{\text{async}})$).

6.2.2. Participants

The final sample consisted of *N* = 31 six-month-old infants (*M* = 180.84 days, range = 164 - 197 days, 13 females). Fifteen other infants were tested (i.e., 32.60 % of the total *N* = 46), but their data was not included in the analysis. This was because either the infants did not meet the minimum gestational age of 37 weeks (*n* = 2), they spent more than 80% of the time looking at one side of the screen (*n* = 6), they looked at the stimuli for less than 4 s in one or more trials (*n* = 5), or the equipment failed (*n* = 2). The infants were full-term babies (i.e., gestational age: 37 weeks or more), did not have any sight or hearing problems, and the majority came from middle-class, White-Caucasian families. Based on parental report, the infants heard English in proportions varying between 50% and 100% (*M* = 92.58; *SD* =

15.10) of the total language input they received. The participants were recruited in the same way as described in Chapter 2 (i.e., via invitation letter and follow-up phone calls), after I obtained ethical approval from the Ethics Committee board of Goldsmiths, University of London. At the beginning of the study, the caregivers provided informed consent for their children to participate. The infants were randomly assigned to one of the two voice conditions, which resulted in *n* = 16 infants in the *Congruent (Gender-Match)* condition and *n* = 15 infants in the *Incongruent (Gender-Mismatch)* condition. There were no differences between groups in the infants' age, but the infants in the *Congruent* condition had slightly more exposure to English than the infants in the *Incongruent* condition. This difference was marginally significant (see **Table 6.1**).

To estimate the number of participants needed in each condition, I carried out a power analysis in G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). The analysis was based on the preference scores (i.e., $\text{PTLT}_{\text{sync}}$) reported by Kubicek et al. (2014; Exp. 2) for the 6-month-old infants in their study.[55] I limited the analysis to the test trials, which were similar to the trials in the *Congruent (Gender-Match)* condition. Furthermore, I averaged the preference scores for the German ($\text{PTLT}_{\text{sync}}$ = 54.6%) and French ($\text{PTLT}_{\text{sync}}$ = 61.1%) audio conditions in Kubicek et al.'s study because I did not have any reason to base my analysis on either language. The calculated preference score was $\text{PTLT}_{\text{sync}}$ = 57.85 (*SD* = 10.2), which I compared to 50% chance level. The comparison yielded an effect size of Cohen d_z = .77. The projected sample size for a one-sample *t*-test (2-tailed)

[55] I chose to carry the power analysis on Kubicek et al. (2014; Exp. 2) because out of all the studies I reviewed on speech synchrony perception, it was the only one that tested a group of 6-month-old infants and had similar stimuli as the present study.

against constant = 50 (i.e., chance level), with effect size = .77, alpha = .05, and power = .75, was minimum *N* = 14 infants. I stopped testing when I had *n* = 16 infants in the *Congruent* voice condition and *n* = 15 infants in the *Incongruent* condition.

Table 6.1. Participant characteristics by voice condition.

	Voice Condition		**T-Test**		
	Congruent (Gender-Match) (*n* = 16)	**Incongruent (Gender-Mismatch)** (*n* = 15)	***t*-value**	***df***	***p*-value** (2-tailed)
Age (days)	180.94 (11.30)	180.73 (9.66)	.05	29	.96
English (%)	97.81 (7.52)	87.00 (19.07)	2.05	18.02#	.06†

Note. Values represent mean (SD). Age, infants' age (in days) at test. English (%), infants' daily level of English exposure expressed as a proportion of the total amount of speech heard (based on parental report). In the Congruent (Gender-Match) condition, the infants listened to a voice recording of a woman. In the Incongruent (Gender-Mismatch) condition, the infants heard the voice recording of a man. #For the English (%) comparison, the degrees of freedom were adjusted because the Levene's test for equality of variance was significant, F = 17.85, p < .001. The infants in the Congruent condition had marginally more exposure to English than the infants in the Incongruent condition. †p < .10, 2-tailed.

6.2.3. Apparatus and Stimuli

The equipment used in this study was: a PC, a 24" BenQ video screen (resolution 1920 x 1080), and two loud-speakers placed under the screen (at ~2 cm below the screen) and 50 cm from each other. The infants' looking behaviour during the study was video recorded using a surveillance video-camera which was positioned under the screen, out of the infants' sight. The video recording was presented live on a second screen, located outside the testing booth, and was used by the researcher to judge when to show the stimuli. An in-house computer script controlled the presentation of the stimuli. The script used MATLAB 2017b and Psychtoolbox 3.0.13 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

The researcher coded the infants' looking behaviour after the testing session was completed. Another in-house computer script allowed the researcher to go through the video recording frame-by-frame and mark the frames where the infant looked to the left or the right side of the screen. To minimize coding bias, the researcher was unaware of the voice condition (*Congruent* vs. *Incongruent*) or the left/right position of the *Synchronous* video. For n = 18 (*n* = 9 from each condition) randomly selected video recordings, the infants' looking behaviour was coded by a second person who was naive to the study hypothesis, the voice condition, and the position of the *Synchronous* video. A two-way mixed intra-class correlation analysis with absolute-agreement (Trevethan, 2017) revealed an excellent inter-rater agreement on infants' accumulated looking time at the videos, $ICC_{2,1} = .94$, and $PTLT_{sync}$, $ICC_{2,1} = .96$.

During the study, the infants listened to either a female (*Congruent* condition) or a male (*Incongruent* condition) voice recording that repeatedly said either "*Hello baby!*" or "*Good job!*". At the same time, the infants watched two side-by-side videos of a woman silently uttering one of the two phrases (see **Figure 6.1**). One of the videos was played forwards, in synchrony with the voice (*Synchronous* video), and the other backwards (*Asynchronous* video). The videos originated from a video recording in which the speaker uttered: "*Hello baby! You are doing very well. Good job!*". During the filming, the woman had her hair tied back, did not wear any pieces of jewellery or make-up, and appeared against a black background. Besides, she used Infant-Directed Speech and smiled at the camera, as if she was talking to an infant.

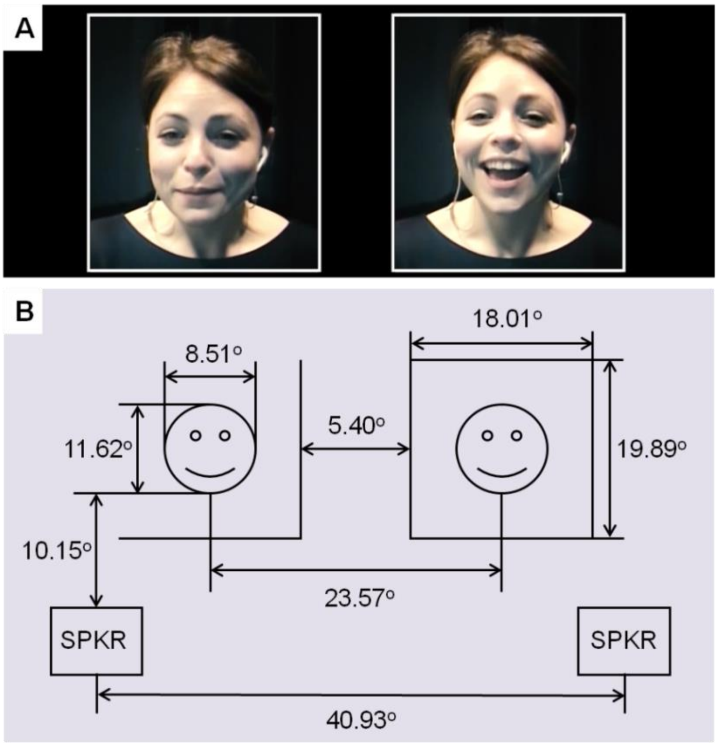

Figure 6.1. Example of videos shown to infants and stimuli dimensions.

(A) The infants listened to a voice recording that repeatedly said either "Hello baby!" or "Good job!". At the same time, they watched two side-by-side video clips of a woman uttering one of the two phrases. One of the video clips was played forwards, in synchrony with the voice, and the other backwards. During the study, the side of the synchronous video alternated left-right across the trials. (B) Schematic drawing of the video clips. Numbers represent the stimuli dimensions in degrees of visual angle, as seen from 70 cm distance (i.e., infants' viewing distance). The model has given informed consent for her picture to be published.

To prevent the infants from using idiosyncratic cues to detect the audiovisual synchrony, I replaced the audio of the original video. For this, I asked two different adults - a man and a woman - to record themselves (separately) uttering the same phrases as the woman in the video. The two adults were both native English speakers and were of a similar age as the woman I had filmed speaking. To ensure that the new voice recordings had the same rhythm, tempo, and speech register as the original, I instructed the adults to listen to the original video via headphones while they produced their recordings. Once I received the new voice recordings, I used Audacity (www.audacityteam.org) to process them further. I aligned the onsets and offsets of the words with those in the original video, and I normalized the voice recordings using root-mean-square amplitude. As a result, the amplitude of the voice recordings varied between 54 and 58 dB at 70 cm distance from the screen and loudspeakers (i.e., the infants' viewing distance). Upon completing the editing of the voice recordings, I used Praat 4.2.1 (Boersma, 2001) to analyse the audio characteristics of the male and female voices. Lastly, I asked a group of 10 adults to rate the voice recordings on a scale from 1 to 5, where 1 represented "masculine" and 5 represented "feminine" voice. I did this to check that the audio editing did not distort the gender specifications of the voices. **Table 6.2** summarises the acoustic characteristics of the voice recordings and the average voice-ratings**.**

I used Adobe Premiere Pro to overlay each voice recording onto the original video. This process resulted in two full-colour films (25 frames/ second): one in which the woman in the video had a feminine voice, and another one in which she had a masculine voice. From each one of these films, I extracted the segments that corresponded to "*Hello baby!*" and "*Good job!*". Each video segment had a total duration of ~1000 ms (25 frames). The audio component started ~280 ms (7 frames) after the video onset and lasted for ~720 ms for "*Hello baby!*", and for ~560 ms for "*Good job!*". I duplicated each segment 25 times, leaving a blank screen of ~200 ms (5 frames) between each copy. The duplication allowed me to create two 30 s long films in which the model repeatedly said either "*Hello baby!*" or "*Good job!*". I decided to use these two phrases and repeat them because young infants are better at detecting speech synchrony when the stimuli are repetitive (Patterson & Werker, 1999, 2003).

Table 6.2. Fundamental frequency characteristics (in Hertz) of the voice recordings and average voice-ratings.

Voice	Phrase	Ave	Min	Max	Ran	Voice-Ratings Ave
Female	Hello baby	220.99	135.92	302.97	167.10	3.80 (0.79)
Female	Good job	211.41	127.04	281.45	154.40	4.20 (0.42)
Male	Hello baby	114.65	84.79	147.01	62.22	1.40 (0.52)
Male	Good job	144.49	79.66	199.25	119.6	1.70 (0.48)

Note. Values represent mean (SD). Voice, voice recording. Phrase, phrase uttered. Ave, average pitch level used in the utterances. Min, minimum pitch level. Max, maximum pitch level. Ran, pitch range. Voice-Ratings Ave, average voice-ratings. The acoustic characteristics of the voice recordings were extracted using Praat 4.2.1. The average voice-ratings were provided by ten adults who scored each voice recording on a scale from 1 to 5, where 1 = masculine, 2 = quite masculine, 3 = neither masculine nor feminine (robotic), 4 = quite feminine, 5 = feminine.

For each 30 s film, I rendered two versions: one version advanced forwards, in synchrony with the voice, and the other backwards (i.e., from the end of the video towards the beginning; see **Figure 6.2**). I decided to use the backwards playing video as the *Asynchronous* video to make sure that it had the same number of social cues and amount of facial movement as the *Synchronous video*. To check whether the *Asynchronous* videos stood out from the rest, I asked four naive adults to watch the stimuli in silence. When asked whether they noticed anything about the videos, one person reported that, in one of the videos, the woman said: "*Hello baby!*". The other adults reported that they could not figure out what the woman was saying. Crucially, none of them noticed that the *Asynchronous* videos were advancing backwards.

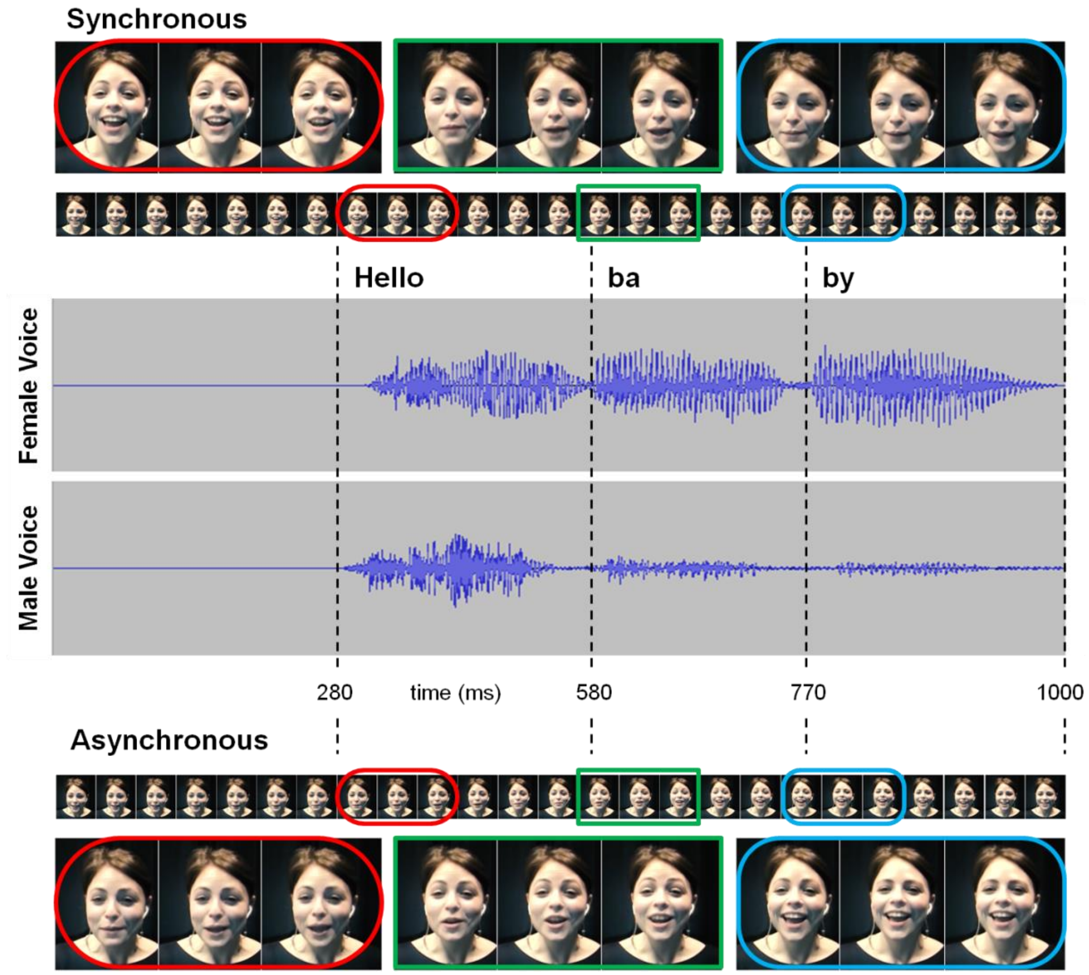

Figure 6.2. Schematic overview of the Synchronous and Asynchronous videos and waveforms of the female and male voice recordings.

The visual input for the "Hello baby!" phrase is depicted by individual video frames (25 frames; 1000 ms). The upper frames correspond to the Synchronous video, and the lower ones to the Asynchronous video. The enlarged sections correspond to the onset of the syllables. The middle panels display the waveforms of the female (Congruent condition) and male (Incongruent condition) voice recordings, relative to the timing of the visual speech (see the timeline on the X-axis). The dashed line marks the onset and offset of the syllables. The onset of the first syllable occurred ~280 ms (7 frames) after the video onset. During the trial, the "Hello baby!" phrase was repeated 25 times, with ~200 ms (5 frames) blank screen between repetitions. The model has given informed consent for her picture to be published.

6.2.4. Procedure

The infants were brought to the lab by their caregivers, who also gave informed consent. The study took place in a dimly illuminated room, and the infants were seated on the caregivers' lap, at ~70 cm from the stimuli video screen. I instructed the caregivers to look at the infant's head and to avoid interacting with the infants during the experiment. The experiment consisted of four 30 s trials during which the infants saw two side-by-side silent videos of a woman speaking (see **Figure 6.3**). At the same time, the infants listened to a female (*Congruent* condition) or a male (*Incongruent* condition) voice recording in which the phrases "*Hello baby!*" and "*Good job!*" were repeatedly uttered (n = 15 infants heard the male voice recording). For each infant, the gender of the voice remained unchanged during the study. However, the left/right position of the *Synchronous* video on the screen, and the phrase that the infants heard alternated throughout the experiment. All the infants listened to the "*Hello baby!*" utterance first, but whether they saw the *Synchronous* video on the left or right side of the screen in the first trial was counterbalanced across participants (for n = 15 infants, the *Synchronous* video was on the left-hand side of the screen in trial 1). If the infants looked at the screen for less than 2 s (i.e., 6.67% of the trial duration), the trial was stopped and repeated. I embedded this minimum looking interval to minimise data loss and to ensure that the infants heard each phrase at least twice during a trial. At the beginning of each trial, the infants saw a 4^{o} x 4^{o} audiovisual looming animation at the centre of the screen. The attention-getter had a minimum duration of 1.5 s and lasted until the infant looked at the screen. The infants completed the study in ~3 minutes, without a break between the trials. The infants then took a 2-minutes break and then participated in the habituation study described in Chapter 5. After the experiments, I asked the parents to estimate what percentage of their child's language input was English. I took this percentage as an estimate of the infants' exposure to English. At the end of the visit, I debriefed the families and rewarded them with a baby t-shirt and a certificate.

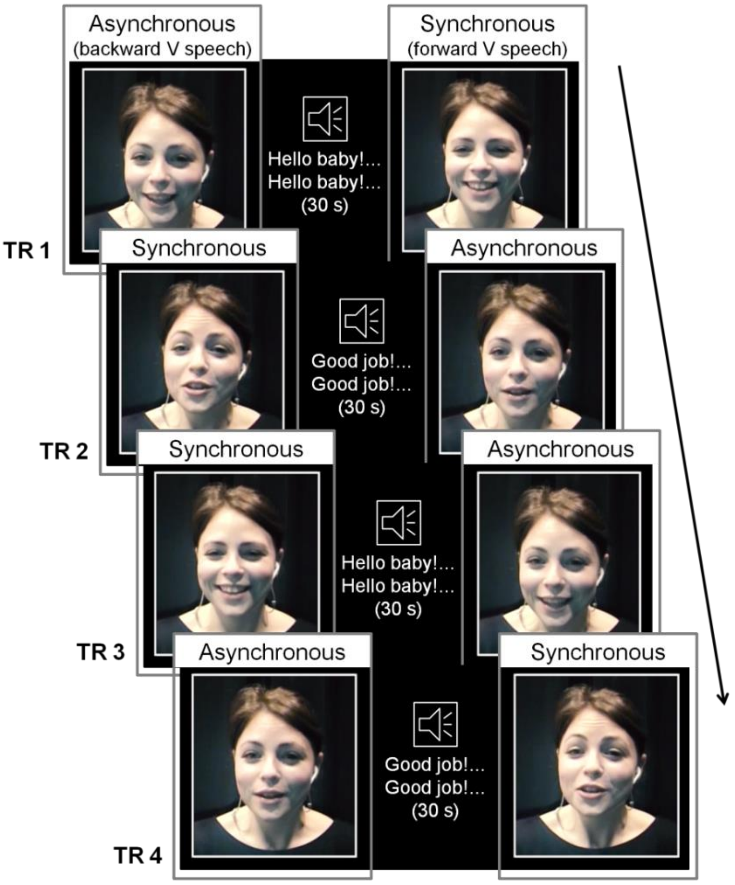

Figure 6.3. Timeline detailing the study procedure.

During the study, the infants listened to a female (Congruent condition) or a male (Incongruent condition) voice recording that said either "Hello baby!" or "Good job!". At the same time, they watched two silent side-by-side videos of a woman. In the Synchronous videos, the lip-movements of the model were synchronous with the voice. In the Asynchronous video, the lip-movements were played backwards (from the end of the video to the beginning). The videos corresponding to each phrase were repeated twice throughout the four trials, with left/right reversal for the positioning of the Synchronous video. Each trial lasted 30 s (in this interval, the phrase was repeated 25 times and there were 200 ms of blank screen between repetitions), and was repeated if the infant looked at the stimuli for less than 2 s. In between the trials, an audiovisual animation was presented at the centre of the screen for a minimum of 1.5 s or until the infants looked at the screen.

6.3. Results

The infants spent, on average, 23.02 s (i.e., 77% of the trial duration) looking at both the *Synchronous* and *Asynchronous* videos. In the *Congruent*

(Gender-Match) voice condition, the infants looked at the stimuli for 23.16 s (*SD* = 4.04), and they spent 51% of this time looking at the left-hand side of the screen. In the *Incongruent (Gender-Mismatch)* condition, the infants viewed the videos for 22.86 s (*SD* = 3.14), and they looked 51% of this interval towards the left video. There were no significant differences between voice conditions in the infants' total looking time to the stimuli, *t*(29) = .23, *p* = .82, *ns*, or proportion of time spent looking to left, *t*(29) = .13, *p* = .90, *ns*. To determine whether the infants detected the audiovisual speech synchrony, I calculated the proportion of total looking time that the infants spent looking at the synchronous video ($\text{PTLT}_{\text{sync}}$) in each trial. If the $\text{PTLT}_{\text{sync}}$ was higher than 50% chance level (i.e., the infants spent more than half of the time looking at the *Synchronous* video), I assumed that they perceived the speech synchrony.

Infants' $\text{PTLT}_{\text{sync}}$ scores are displayed in **Figure 6.4**. The infants in the *Congruent* condition had a marginally higher $\text{PTLT}_{\text{sync}}$ score than the infants in the *Incongruent* condition (Mann-Whitney U = 75, *p* = .08). This observation was confirmed with a 2 (Voice Condition: *Congruent* vs. *Incongruent*) x 2 (Phrase: *Hello baby* vs. *Good job*) mixed ANOVA with Voice Condition as a between-subjects factor, and Phrase as a within-subjects factors. The analysis showed a main effect of Voice Condition, *F*(1, 29) = 4.22, *p* = .05, η_p^2 = .13. The $\text{PTLT}_{\text{sync}}$ score was higher in the *Congruent* condition (*M* = 54.72, *SD* = 7.99) than in the *Incongruent* condition (*M* = 49.60, *SD* = 5.57). Neither the main effect of Phrase nor the interaction Voice Condition x Phrase reached significance - Phrase: *F*(1, 29) = 2.45, *p* = .13, *ns*; Voice Condition x Phrase, *F*(1, 29) = .30, *p* = .59, *ns*.

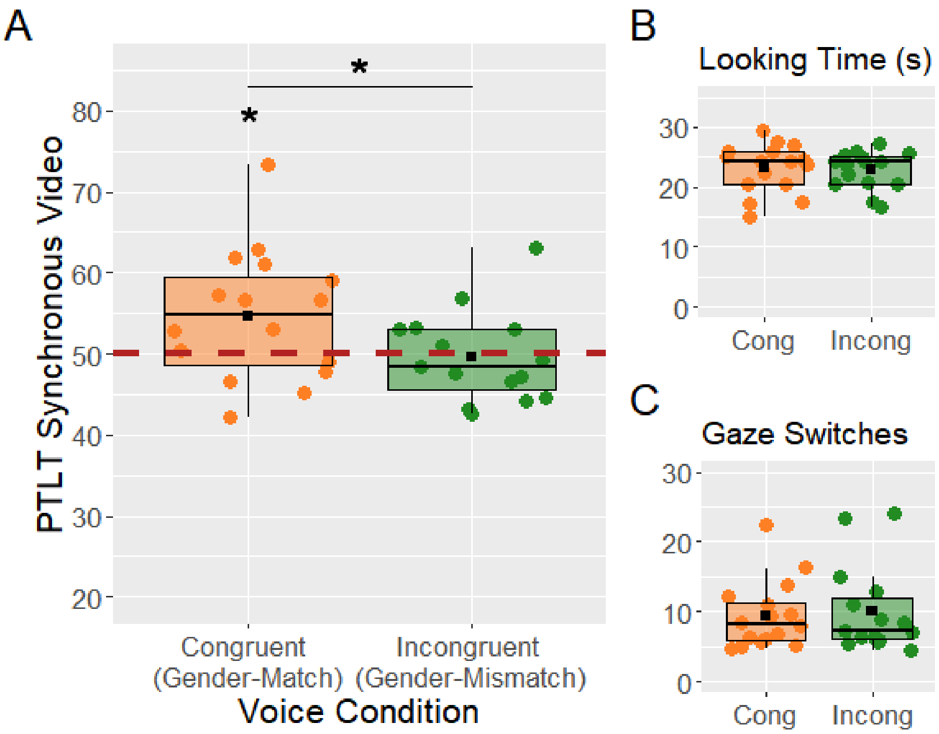

Figure 6.4. Infants' scores on three different measures split by voice condition.

(A) Proportion of total looking time directed to the Synchronous video (PTLTsync) averaged across trials. In the Congruent (Gender-Match) condition, the infants listened to the voice recording of a woman who repeatedly said either "Hello baby!" or "Good job!". In the Incongruent (Gender-Mismatch) condition, the infants heard the same phrases recited by a man. While the infants listened to the voice recordings, they watched two side-by-side silent videos of a woman speaking. In one video, the woman's lip movements were synchronous with the voice recording (Synchronous video), in the other they were asynchronous (to achieve this, I played the video backwards; Asynchronous video). The PTLTsync was calculated by dividing infants' looking time to the synchronous video by their accumulated looking time at both videos, PTLTsync = LTsync / (LTsync + LTasync). Higher PTLTsync indicates greater preference for the Synchronous video. (B) Total looking time (in seconds) at both videos averaged across trials. (C) The number of gaze switches that the infants made between the Synchronous and the Asynchronous videos averaged across trials. *Note*: Black dots represent mean values, and the dashed red line in panel A marks the 50% chance level. $^{*}p < .05$, 2-tailed.

To test the *a priori* prediction that only the infants in the *Congruent (Gender-Match)* condition would prefer the *Synchronous* video over the Asynchronous one, I conducted two one-sample *t*-tests, one for each voice condition. For this analysis, I compared $\text{PTLT}_{\text{sync}}$ to 50% chance level (because the infants watched two side-by-side videos). In the *Congruent*

(Gender-Match) condition, PTLT$_{sync}$ was significantly higher than chance, *t*(15) = 2.36, *p* = .03, Cohen d$_z$ = .59. At the individual level, I found that 11 out of 16 infants (69%) spent more than half of the time looking at the *Synchronous* video (binomial test, *p* = .21, *ns*, 2-tailed). In the *Incongruent (Gender-Mismatch)* condition, PTLT$_{sync}$ was not significantly different from chance, *t*(14) = 2.36, *p* = .79, *ns*. In this condition, 6 out of 15 infants (38%) looked longer at the *Synchronous* video (binomial test, *p* = .61, *ns*, 2-tailed).

Given that the infants' looking at the videos could be made-up of either a large number of short glimpses or a small number of long gazes, I also coded the number of gaze switches that the infants made between the two videos. I found that the infants completed, on average, 9.76 gaze switches (*SD* = 5.49) per trial. The number of gaze switches in the *Congruent* condition (*M* = 9.41, *SD* = 4.86) was similar to that in the *Incongruent* condition (*M* = 10.13, *SD* = 6.24), *t*(29) = .36, *p* = .72, *ns*. Therefore, the difference in PTLT$_{sync}$ between voice conditions was due to how long the infants spent looking at the *Synchronous* video rather than the infants employing two different visual scanning strategies.[56]

Finally, I looked at the relationship between infants' exposure to English and PTLT$_{sync}$. Previously, Mercure et al. (2019) found that 6.5- to 8-month-old monolingual-English infants, but not bilingual infants, look more at the mouth area of a speaker when the lip movement does not correspond with the English syllable heard than when it does. An explanation that the

[56] The visual scanning strategy could have been different if, in one condition, the infants had switched only a few times between the *Synchronous* and the *Asynchronous* videos, and they had fixated the *Synchronous* video for much longer. However, the results suggest otherwise. In both voice conditions, infants inspected the two videos for an equally large number of times but, in the *Congruent* condition, infants looked slightly longer at the *Synchronous* than the *Asynchronous* video.

authors provided for this finding was that those infants who learned other languages in addition to English may have extracted the phonemes from different phoneme populations than the monolingual infants (and that is why the bilingual infants did not look longer at the mouth in the mismatch condition). Considering these findings and the fact that the infants in the *Congruent (Gender-Match)* condition had marginally higher exposure to English, I decided to investigate the relationship between infants' English exposure and $\text{PTLT}_{\text{sync}}$. The Pearson correlation analysis showed that there was no association between the infants' English exposure and their $\text{PTLT}_{\text{sync}}$, $r = .08$, $n = 31$, $p = .67$.[57] Consequently, the differences I found in $\text{PTLT}_{\text{sync}}$ between the voice conditions cannot be explained by the infants' daily English input.

6.3.1. Exploratory Analysis

Previous research suggests that different "attention-getting" and "attention-holding" mechanisms underlie infants' cognitive processing of visual displays (L. B. Cohen, 1972). For example, L. B. Cohen showed that the stimulus size modulates how fast infants orient to a checkerboard pattern, while the number of checkers present on the board affects how long infants look at it. Other research has found that, in complex visual displays, infants direct a significantly larger proportion of first looks to human faces

[57] I also conducted two independent samples *t*-tests on $\text{PTLT}_{\text{sync}}$ between Voice Conditions – one for the Monolingual ($N = 23$) and the other for the Bilingual infants ($N = 8$). The Monolingual infants looked slightly more at the *Synchronous* video in the *Congruent* ($M = 54.90$, $SD = 8.48$) than in the *Incongruent* condition ($M = 49.13$, $SD = 4.52$). However, the difference was only marginally significant, $t(21) = 1.87$, $p = .08$. By contrast, the Bilingual infants looked equally long at the Synchronous video irrespective whether they were in the *Congruent* ($M = 53.49$, $SD = 4.33$) or the *Incongruent* condition ($M = 50.31$, $SD = 7.29$), $t(6) = .57$, $p = .59$. Although these results suggest that the effect of Voice Condition was smaller in the group of Bilingual infants, these results should be interpreted with caution given the small sample size.

and they spend more time looking at faces during a study (Gliga et al., 2009; Gluckman & Johnson, 2013; Kwon et al., 2016). These findings indicate that socially relevant stimuli both capture infants' attention (measured via first looks to the target) and hold it (measured via looking time at the target).

Arguably, in the present study, the *Synchronous* video is the more socially relevant stimulus. While the $\text{PTLT}_{\text{sync}}$ analysis shows that, in the *Congruent (Gender-Match)* conditions, the *Synchronous* video held the infants' attention for longer, it is unclear whether it also attracted it more often at the beginning of the trial. To investigate how audiovisual synchrony and face-voice gender correspondences affect the infants' attention-orienting behaviour, I looked at the infants' proportion of first looks to the *Synchronous* video. On average, the infants directed 61.29% (*SD* = 21.25) of their first looks to the *Synchronous* video. This proportion was significantly higher than 50% chance, *t*(30) = 2.96, *p* = .006, Cohen d_z = .53. The proportion of first looks to the *Synchronous* video was slightly higher in the *Congruent* condition (*M* = 65.63, *SD* = 23.94) than in the *Incongruent* condition (*M* = 56.67, *SD* = 17.59), but the difference was not statistically significant, *t*(29) = 1.18, *p* = .25, *ns*. Therefore, at the beginning of the trial, the *Synchronous* video captured more often the infants' attention, but it did so in similar proportions across the two voice conditions. Although revealing, this analysis should be interpreted with caution because the experiment had only four trials, and the variance within each voice condition is large.

6.4. Discussion

The present study tried to find out whether face-voice gender correspondences affect 6-month-old infants' perception of speech synchrony. During the experiment, the infants listened to a voice recording while they watched two side-by-side videos of a woman speaking. One video was synchronous with the voice recording and the other one asynchronous. I found that, in the *Congruent (Gender-Match)* condition, when the infants listened to the voice recording of a woman, they looked more at the *Synchronous* video (measured through $\mathrm{PTLT}_{\mathrm{sync}}$). In contrast, in the *Incongruent (Gender-Mismatch)* condition, when the infants heard a man speaking, they watched the *Synchronous* and *Asynchronous* videos for equally long. There were no differences between voice conditions in the infants' total looking time at the stimuli, or the number of gaze switches made between the videos. Finally, the infants' level of daily exposure to English was unrelated to their preference for the *Synchronous* video. These results support my hypotheses and suggest that the *Synchronous* video held the infants' attention for longer than the *Asynchronous* video but only when the voice matched the gender of the face.

Previous studies have shown that young infants can detect the synchrony between auditory and visual speech stimuli when they are simple, and repetitive stimuli (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; 2003). However, when two adjacent speakers utter different multisyllabic words at the same time, the infants can match the voice modulations with the synchronous speaker when careful consideration is given to the opening and closing of the speakers' mouth (Baart et al., 2014), as well as to the rhythm of speech (Kubicek et al., 2014). In the present study, I controlled for these aspects by playing forward one of the two videos, in synchrony with the voice recording, and the other one backwards. I replicated previous findings that 6-month-old infants can detect speech synchrony in multisyllabic words. Furthermore, I found that the face-voice gender correspondences modulate this effect.

Around 6-months of age, infants begin to use the gender-related characteristics of a voice to guide their visual exploration of potential speakers. Walker-Andrews et al. (1991) and Richoz et al. (2017) found that 6-month-old infants looked long at the video of a woman than that of a man when they heard a voice recording of a woman playing alongside the videos. This preference for the same-gender speaker was apparent both when the videos were synchronous with the voice recording and asynchronous, which suggests that detecting face-voice gender correspondences is independent of speech synchrony. By contrast, the perception of speech synchrony seems to be affected by audiovisual gender associations. Patterson & Werker (2002; Exp. 4) found that, when the gender of the voice conflicted with that of the speaker, 4.5-month-old infants failed to show a preference for the synchronous video. This interference occurred when the infants watched two side-by-side videos of a man and a woman, and they listened to a male voice recording that played in synchrony with the video of the women. However, in a separate experiment, when Patterson & Werker (Exp. 3) played the same male voice recording alongside two videos of a woman, the infants looked longer at the synchronous video despite the gender mismatch.

My results are inconsistent with this latter experiment, which could be because I used longer voice recordings than Patterson and Werker.

The voice recordings used by Patterson & Werker (2002) consisted of repetitive vowel sounds (i.e., /a/ and /i/), which proved difficult even for 6-month-old infants to match to the speaker of the same gender (Patterson & Werker; Exp. 5). The acoustic characteristics of the voice recordings reveal that, although the female voice had a higher pitch, its pitch range was equally broad as that of the male voice. By comparison, in Richoz et al. (2017), who found that 6-month-old infants can match faces and voices by gender, the pitch range of the female voice was double that of the male voice. The difference in pitch range may have been because Richoz et al. used short phrases, which captured the pitch and timbre of the speakers, as well as possible gender-related differences in intonation.[58] The female voice recording in the present study had a higher pitch and pitch range than the male voice recording, which may have allowed the infants to extract the gender of the voices and use this information to guide their looking behaviour during the study.

The results of this study are inconsistent with the IRH (Bahrick & Lickliter, 2000, 2002, 2012), which argues that amodal object/event properties such as speech synchrony are detected automatically, and they inform subsequent looking behaviour. If that had been the case, the infants in both the *Congruent (Gender-Match)* and *Incongruent (Gender-Mismatch)* conditions should have looked longer at the *Synchronous* video. I found that the infants preferred the *Synchronous* video only when the gender of the

[58] Haan & Van Heuven (1999) found that women have a broader pitch range than men both when they read and spontaneously asked questions in patient-doctor interviews (see also Simpson, 2009, for a discussion).

voice matched that of the face. This pattern of results suggests that 6-month-old infants extract both speech synchrony and face-voice gender correspondences when they look at potential speakers and use both types of information to decide for how long they keep looking at a speaker. Processing in parallel speech synchrony and gender associations might be an adaptive strategy for identifying speakers in a busy environment where multiple people could be speaking concurrently. When the speech is fluent, the boundaries between words are not always clear, and the pauses in the speech signal are hard to detect. Therefore, infants have to use the structure and the statistical regularities between syllables to discover candidate words and word boundaries (Saffran, 2001; Saffran et al., 1996). Such a process requires infants to pay attention to a speaker for a short interval of time, which can be demanding when infants must inspect multiple speakers. By adjusting their looking behaviour based on the face-voice gender correspondences, infants can be more successful in identifying the synchronous speaker.

An alternative explanation for the results could be that the 6-month-old infants have lost the ability to detect synchrony in gender-incongruent speech stimuli due to their extensive exposure to synchronous, gender-matched stimuli. This interpretation would be consistent with the multisensory perceptual narrowing phenomenon reported by Lewkowicz & Ghazanfar (2006). Lewkowicz & Ghazanfar found that 4- and 6-month-old infants successfully matched monkey vocalizations with their corresponding video, whereas 8- and 10-month-old infants failed to do so. The authors argued that the results reflect a pruning down of less relevant intersensory connections in favour of more relevant ones which infants encounter more often. Although this might be an explanation for the present results, it should be treated with caution because the study does not have cross-sectional data to capture developmental changes in infants' behaviour. Besides, in both the visual and the auditory domains, the perceptual narrowing phenomenon has been reported in older infants, aged between 9 and 12 month olds (Pascalis et al., 2002; Werker & Tees, 1984).

Using male speakers with gender-matched or gender-mismatched voices might be another way of investigating the role of perceptual experience on infants' detection of speech synchrony. Various studies have shown that infants display more robust matching of female faces and voices (Hillairet de Boisferon et al., 2015; Poulin-Dubois et al., 1994; Richoz et al., 2017), possibly because infants tend to spend more time with women than men during their first year of life. Therefore, 6-month-old infants may not have accumulated enough experience with male speakers to show the same effects that I found with female speakers. Furthermore, testing the same age-group of infants might be a more viable solution than conducting the study with younger infants, who might fail to detect speech synchrony because the stimuli are too complex to process.

The present study tried to find out whether 6-month-old infants can detect speech synchrony both when the speaker's face and voice are gender-matched (i.e., a female face with a female voice) and gender-mismatched (i.e., a female face with a male voice). I found that the infants preferred to look at the visual speech stream that was synchronous with the auditory speech stream when the stimuli were gender-matched. These results suggest that the detection of audiovisual speech synchrony is not automatic, as the IRH proposes, and that it is affected by face-voice gender correspondences. The study employed only female faces. Given that infants display an asymmetry in making face-voice gender matches between female and male speakers, future studies could use male faces.

CHAPTER 7

Discussion

The studies reported in this thesis aimed to find out how multisensory stimulation affects infants' visual perception and learning. Prior research has shown that adults do benefit from correlated multisensory stimulation when they process objects and events. For example, adults respond faster to congruent audiovisual stimuli (Harrington & Peck, 1998; Hughes et al., 1994; J. Miller, 1982; Mordkoff & Yantis, 1991). They make more accurate perceptual judgements about such stimuli (Chen & Spence, 2010; Kim et al., 2012; Rohe & Noppeney, 2015b), and even remember them better (Lehmann & Murray, 2005; Murray et al., 2004). By contrast, incongruent audiovisual stimulation seems to hinder adults' cognitive processing relative to unisensory stimulation (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Lehmann & Murray, 2005; Thomas, Nardini, & Mareschal, 2017; but see Harrington & Peck, 1998; cf. Innes-Brown & Crewther, 2009). However, this latter effect is less robust and seems to be task-dependent. Determining whether similar congruency effects are detectable in infants might shed some light on the development of cross-modal binding and could inform the existing models of multisensory integration (e.g., Ernst & Bülthoff, 2004; Rohe & Noppeney, 2015b). I will discuss this issue later in the chapter.

As outlined in Chapter 1, a theory that addresses the role of multimodal stimulation on infants' cognitive processing is the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000, 2002, 2012). This account states that spatiotemporally congruent audiovisual stimulation guides infants' attention to some object/event properties and away from other properties. Furthermore, it proposes that the effect of multisensory stimulation changes with age. Lastly, the IRH argues that infants detect the amodal relations between multisensory cues irrespective of the arbitrary associations between them. The IRH defines the spatiotemporal relations between stimuli as "amodal" and the semantic relations as "arbitrary". Building on the predictions of the IRH, in this thesis, I set out to address the following research questions: (1) Does multisensory stimulation affect infants' visual processing and learning of social and non-social stimuli? (2) Does this change across the first postnatal year? (3) Do semantic correspondences (e.g., face-voice gender matches) affect infants' ability to process multisensory events (e.g., speech synchrony)?

This chapter summarizes the results of the experiments reported in this thesis. The findings are then related to the predictions of the IRH and the existing research on multisensory processing in infants and adults. A robust effect across the studies is that (spatiotemporal and semantic) incongruent audiovisual stimulation hinders 6-month-old infants' visual processing and learning. Potential explanations for this effect are considered, alongside the implications that it might have for infants' cognitive development. Finally, in this chapter, I will discuss some of the limitations of the studies presented and indicate directions for future research.

7.1. Summary of findings

7.1.1. Ten-month-old infants' encoding of object pattern. Part 1 and 2

This thesis begins by reporting an investigation into the role of spatiotemporally congruent and incongruent audiovisual stimulation on 10-month-old infants' encoding of object pattern. The experiment in Chapter 2 built on J. G. Bremner, Slater, Johnson, Mason, & Spring (2012) and

Kirkham, Wagner, Swan, & Johnson (2012). These studies found that 4-month-old infants benefit from spatiotemporally congruent audiovisual stimulation when processing occlusion events. More specifically, infants represent an occluded object for longer and are better at anticipating when and where it will reappear if it is specified concurrently across vision and audition. The study I conducted aimed to find out whether the encoding of other object properties, such as the object's visual pattern, is similarly facilitated by congruent multisensory stimulation.

The study found that irrespective whether the infants received spatiotemporally congruent or incongruent information about an object they processed the pattern in the same way. More specifically, after the habituation, the infants looked equally long at two test events: one that depicted a change in the object's pattern during the occlusion and the other one which showed no change. Given that the former event was perceptually more novel, I had expected the infants to prefer the *Change* event. However, the results showed that the infants did not differentiate between the two test events, which suggests that either they did not encode the pattern on the object during the habituation, or the change employed was not salient enough.

Therefore, Chapter 3 describes a second study that I conducted to look at the effects of multisensory stimulation on 10-month-old infants' encoding of object pattern. The new experiment used a shorter period of occlusion (i.e., the occlusion lasted 634 ms) because previously S. P. Johnson, Bremner, et al. (2003) and J. G. Bremner et al. (2012) found that infants younger than ten months can represent the trajectory of an occluded object when the interval of occlusion is ~600 ms long. Besides, the pattern changed from dots to stripes because Wilcox (1999) found that 7.5-month-old infants detect such a visual change. Finally, the experiment included a visual-only habituation condition alongside the congruent and incongruent audiovisual conditions. The decision to add a unimodal condition was motivated by the fact that background (auditory) noise hinders infants' ability to: detect changes in a visual display, form visual categories, differentiate between the objects they touch, and learn object-directed actions (Barr et al., 2010; Lejeune et al., 2016; Robinson & Sloutsky, 2007a, 2007b).

This second experiment found that the infants looked significantly longer at the test event that depicted a change in the object's pattern than at the test event that showed no change. The results are in line with previous studies (e.g., Wilcox, 1999) and suggest that 10-month-old infants can encode the pattern on a briefly occluded object. Furthermore, the study did not find any differential effects of habituation condition. In other words, the type of sensory stimulation (visual-only, audiovisual congruent, or audiovisual incongruent) that the infants received did not affect their encoding of object pattern. The lack of a differential response between condition was surprising given that previous studies have found that infants younger than ten months old benefit from congruent audiovisual stimulation when performing various cognitive tasks (J. G. Bremner et al., 2012; Kirkham, Wagner, et al., 2012; Lawson, 1980).

7.1.2. Four- and six-month-old infants' encoding of object pattern

The study presented in Chapter 4 investigated whether the effects of multisensory stimulation are more pronounced in younger infants. Previous studies have reported that younger infants are affected differently by audiovisual simulation than older infants. For example, Bahrick (1994) found that, out of three age groups of infants habituated to two pairs of auditory and visual stimuli, only the 7-month-old infants detected swaps between these pairs but not the 3- and the 5-month-old infants. Similarly, other studies have revealed that the older, but not the younger, infants respond to changes in the orientation (Bahrick et al., 2006) and visual order (Lewkowicz, 2004a, 2004b) of some objects that are both seen and heard striking a surface. These findings indicate that concurrent audiovisual stimulation hinders the visual processing and learning of young infants. To examine whether 4- and 6-month-old infants' processing of object pattern is affected by multisensory stimulation, the study reported in Chapter 4 employed the same stimuli as in Chapter 3.

The results showed that the 6-month-old infants encoded the pattern on the briefly occluded object, but the 4-month-old infants did not. These findings are in line with Needham (1999) and Wilcox (1999), studies which have reported that 4-month-old infants do not spontaneously use pattern information to segment objects in a visual display and to disambiguate occlusion events. This age difference may be because, by four months, the infants have not learned to attend to the pattern of objects (Needham, 1999; Wilcox, 1999), or the occluded item may have moved too fast for them to process this feature (Burnham, 1987). The data also revealed that, in the group of 6-month-old infants, the response to the change in object pattern was more robust in the unisensory condition. Therefore, audiovisual stimulation had an impact on 6-month-old infants' learning of object pattern.

However, the effect seemed to be more pronounced in the incongruent condition. The results of this study are inconsistent with Bahrick et al. (2006), who found that only congruent audiovisual stimulation affected young infants' encoding of object orientation (another visually-specified object property). Methodological differences between the two studies (e.g., stimuli used and study procedure) could explain the different results.

7.1.3. Six-month-old infants' encoding of object pattern and trajectory

Chapter 5 presented another study which investigated the effects of multisensory stimulation on 6-month-old infants' cognitive processing. More specifically, the study tried to find out whether receiving audiovisual stimulation affects only the infants' encoding of object pattern, or it interferes with their learning of object trajectory as well. As discussed before, there are empirical findings which suggest that congruent audiovisual stimulation helps young infants to represent an occluded object for longer (J. G. Bremner et al., 2012) and learn the objects' trajectory (Kirkham, Wagner, et al., 2012). Additionally, infants learn the rhythm and tempo of a striking hammer when they both see and hear it (Bahrick et al., 2002; Bahrick & Lickliter, 2000, 2004). Finally, infants seem to benefit from congruent audiovisual stimulation when they learn the location of multisensory objects/events (Kirkham, Richardson, et al., 2012; Moore & Meltzoff, 2008; Shinskey, 2017).

This fourth study revealed that the infants learned both the pattern and the trajectory of an object when they received congruent audiovisual information about it. By contrast, when they had only visual information, the infants encoded only the object pattern. Lastly, when the audiovisual information was spatiotemporally incongruent, the infants failed to learn either of the two object properties. Given the methodological similarities between Chapter 4 and Chapter 5, I pooled the data together and analyzed the infants' response to the change in object pattern. This analysis showed that the 6-month-old infants in the unisensory and the congruent conditions detected the pattern change. Altogether, these findings suggest that spatiotemporally incongruent audiovisual stimulation hinders infants' visual processing and learning. These results are reminiscent of those reported by J. G. Bremner et al. (2012) and Kirkham, Wagner, et al. (2012) and provide some support for the idea that young infants find it difficult to ignore irrelevant auditory information.

7.1.4. Six-month-old infants' detection of speech synchrony

The study reported in Chapter 6 looked at how semantic congruency influences 6-month-old infants' perception of audiovisual synchrony. Semantic congruency refers to the arbitrary relations that exist between certain visual and auditory stimuli which are, nonetheless, associated because of their relevance to action and cognition (e.g., dogs bark, birds chirp, women have a high pitch voice). These associations are stable and predictive in the environment, and infants may use them to decide on how to combine cross-modal stimulation. The study examined this aspect by looking at the effect that face-voice gender correspondences have on infants' ability to detect speech synchrony. Numerous studies have shown that infants learn to match the lip movements of a speaker with their voice modulations during their first year of life (Baart et al., 2014; Kubicek et al., 2014; Kuhl & Meltzoff, 1982; Lewkowicz et al., 2015; Patterson & Werker, 1999). In parallel, infants learn to associate voices and faces by gender (Hillairet de

Boisferon et al., 2015; Richoz et al., 2017; Spelke & Owsley, 1979; Walker-Andrews et al., 1991). The ability to detect gender correspondences drives infants' looking behaviour and, as a result, infants spend more time fixating a woman than a man if they hear a female voice. This behaviour is not affected by speech synchrony (it occurs with both synchronous and asynchronous audiovisual speech recordings). However, it is unclear whether the perception of speech synchrony is influence by gender associations.

The study revealed that, when the infants heard a female voice recording and watched two side-by-side videos of a woman speaking, the infants looked longer at the video that was synchronous with the voice recording. This preference for the synchronous video was not apparent when the infants heard a male voice recording (i.e., when the stimuli were gender-mismatched). These results suggest that 6-month-old infants process in parallel both the face and voice characteristics of speakers and the synchrony between them. Furthermore, it appears that the ability of 6-month-old infants to detect speech synchrony is affected by gender congruency. Vatakis & Spence (2007) found a similar effect in a group of adults. They reported that when faces and voices are gender-congruent, adults find it harder to judge whether they hear a person uttering something first or whether they see the person's lips moving first. This difficulty in judging which sensory cue occurred first could be because the adults perceive the voice and the facial movements as being synchronous (see also Vatakis, Ghazanfar, & Spence, 2008).

7.2. Theoretical implications

The studies reported in this thesis show and confirm that multisensory stimulation affects infants' visual processing and learning (see also A. J. Bremner, Lewkowicz, & Spence, 2012; Lewkowicz & Lickliter, 1994). Infants' ability to encode different object properties, such as visual pattern and trajectory, seems to be hindered by incongruent audiovisual stimulation. It is unclear whether this is because infants cannot segregate incongruent sensory stimulation, or they find it harder to ignore auditory distractions. However, in the occlusion task used, the effect seemed to be more robust in younger than in older infants. The final study in the thesis provides an insight into how young infants process audiovisual speech, and that they use both the semantic and the spatiotemporal relations between the speech cues to identify specific speakers in a busy environment. More specifically, the infants detected which one of two people uttered something in synchrony with a voice recording when the speaker's face and voice were gender-congruent. These findings speak to the predictions of the IRH (Bahrick & Lickliter, 2000, 2002, 2012) and offer a glimpse into the development of multisensory processing in humans.

7.2.1. Implications for the IRH

As described in Chapter 1, the IRH is a theoretical framework which attempts to explain multisensory development in early life (see Bahrick & Lickliter, 2012). It proposes that infants can automatically detect congruent multisensory cues. Furthermore, it argues that these cues guide infants' attention towards the object/event properties that are specified redundantly by multiple sensory modalities (i.e., amodal properties). The assumption that infants spontaneously detect which sensory cues are related overlooks the computational issues that multisensory integration poses: the cross-modal binding problem and the reliability-weighted integration problem (Ernst & Bülthoff, 2004; Rohe & Noppeney, 2015b). Furthermore, although the infants are sensitive to the spatiotemporal relations between the sensory cues, they associate auditory and visual stimuli that are more distant in space and time, and which adults would not typically bind (Fenwick & Morrongiello, 1998; Lewkowicz, 1996, 2010). In a complex environment where unrelated stimuli are often concurrent, associating sensory cues that occur over a broader spatiotemporal window increases the likelihood of abstracting incorrect cross-modal relations.

Putting aside the computational problems that the IRH fails to address, the question that remains is whether infants attend to and encode the amodal properties of objects/events when they receive congruent audiovisual stimulation. One of the studies reported in this thesis found that 6-month-old infants learned an object's trajectory when this was specified concurrently by vision and audition. Although these findings are consistent with the IRH, it is unclear whether the infants learned the object's trajectory or a particular sequence of ocular motions (e.g., the congruent audiovisual cues may have helped the infants learn to look alternatively to the left and the right side of the screen). By arguing that infants preferentially process the amodal properties of objects/events, the IRH implies that infants know that the sensory cues have a common origin and reflect the same underlying property. While infants may engage in these computations, it is equally possible that they watched for longer the *Trajectory Change* test event because they needed to readjust their visual scanning pattern.

The congruent stimulation received during the habituation may have entrained the infants' looking behaviour to the motion of the object (see Kirkham, Wagner, et al., 2012). This entrainment may have been adaptive because it allowed the infants to track the item better across the display. However, during the test event, when the trajectory of the object changed, the infants had to adjust their visual scanning pattern. Instead of looking alternatively to the left and the right side of the screen, the infants had to learn to look to one side for longer because the object appeared twice, in succession, on the same side of the screen. This adjustment may have resulted in the infants looking longer at the screen. The same may have happened in Bahrick & Lickliter (2000) and Bahrick et al. (2002), who found that the infants responded to changes in the tempo and rhythm of a tapping hammer only after they habituated to a congruent audiovisual tapping event. Eye-tracking technology may provide an insight into whether congruent multisensory stimulation entrains infants' eye movements. For example, in the occlusion task presented in this thesis, the infants in the congruent audiovisual condition may have executed more anticipatory looks to the wrong side of the screen in the *Trajectory Change* test event. Alternatively, they could have taken longer to fixate the briefly occluded object after it re-emerged from behind the occluder because they would not be able to predict where the object would reappear (i.e., the reactive saccades had longer latencies).

The second prediction that the IRH makes is that the infants encode the modality-specific object/event properties (e.g., pattern, colour, pitch) when they perceive the object/event unimodally or when the multimodal information is incongruent. As discussed in Chapter 1, disentangling multisensory stimulation is a complex process that requires perceptual experience, precision, and sustained attention. Establishing whether the cues are related and deciding whether to combine or segregate them might affect unimodal processing. Some evidence in this regard comes from Robinson & Sloutsky (2007b). Robinson & Sloutsky showed 14-month-old infants two side-by-side visual streams, one in which the visual stimulus changed, and the other in which it remained unchanged. The authors reported that, when the infants heard a computer-generated sound alongside the two visual streams, they took longer to display a looking preference for the changing visual stream (which was more interesting), than when the infants watched the streams in silence. Infants also benefit from unimodal stimulation when they perform visual categorization and object individuation tasks (Robinson & Sloutsky, 2007a, 2008). However, in these studies, the sound may have disrupted the infants' visual processing because it was incongruent with the visual input. The IRH, by contrast, argues that, at this age, the incongruent stimulation supports infants' processing of modality-specific object properties.

The results of two of the studies reported in this thesis are consistent with Robinson & Sloutsky (2007a, 2007b, 2008). In both experiments, when 6-month-old infants received auditory input that was spatiotemporally incongruent with the motion of an object across the display, they learned neither its pattern nor its trajectory. However, when the sound was spatiotemporally congruent, the infants learned both object properties. Other studies have also reported that, compared to incongruent audiovisual stimulation, congruent stimulation helps infants' cognitive processing (J. G. Bremner et al., 2012; Kirkham, Wagner, et al., 2012; Lawson, 1980; Bahrick & Lickliter, 2000; Bahrick et al., 2002; but see Bahrick et al., 2006). Even more significant, is the fact that in the studies reported in this thesis the 6-month-old infants encoded the pattern on the object in both the unimodal and the congruent multimodal condition (there were no significant differences between these conditions).[59] These results are inconsistent with the IRH, which argues that modality-specific properties such as (visual) object pattern are encoded better when the stimulation is unimodal or incongruent.

The IRH predicts that the effects of multisensory stimulation should be more pronounced in younger than older infants. Consistent with this prediction, the studies reported in this thesis found that incongruent multisensory stimulation affects 6- but not 10-month-old infants' visual processing and learning. I did not find any significant differences between conditions in the 4-month-old age group, possibly because at this age infants do not spontaneously use the pattern on objects to segment and individuate objects in a visual display (Needham, 1999; Wilcox, 1999). Around six months of age, infants begin to look in anticipation to where they expect an

[59] In one of the studies (see Chapter 4), I found that the 6-month-old infants in the Congruent habituation condition did not differentiate between the Change and No Change test events. However, 75% of the infants in this condition looked longer at the Change event, and the increase in looking time averaged 45%. Furthermore, when I pooled the infants' looking time data to the Change and No Changed test event across Chapter 4 and Chapter 5, I found that only the infants in the Visual-Only and Congruent (Dynamic Sound) conditions differentiated between the test events and they learned the pattern on the ball.

occluded object to re-emerge (S. P. Johnson, Amso, et al., 2003), and they also start to keep better track of object pattern (Hernandez-Reif & Bahrick, 2001; Wilcox, 1999; Wilcox & Chapa, 2004). It is at this stage in infants' cognitive development that the effect of multisensory stimulation seems to be more pronounced. When infants reach ten months old, they encode the pattern on a briefly occluded object irrespective whether they see it in silence or accompanied by a spatiotemporally congruent or incongruent sound (see Chapter 3).

The explanation that the IRH provides for the developmental change in infants' response to multisensory stimulation is that infants' processing capacity and speed improves with age. As a result, older infants can process multiple object properties in parallel. Although this is possible, as infants grow up, they also gain more experience with cross-modal input. They become more precise in their spatiotemporal estimates (Fenwick & Morrongiello, 1998), and they learn that some sensory cues have a higher probability of occurring together (Hillairet de Boisferon et al., 2015; Richoz et al., 2017; Walker-Andrews et al., 1991). These developments probably help infants process multisensory information faster (see also Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006). Furthermore, infants' attentional control improves significantly during the first year of life (Atkinson & Braddick, 2012; Colombo, 2001), which may help infants deal better with noise and be less distracted by irrelevant cues.

The final study reported in this thesis addressed another prediction of the IRH, namely that infants attend first to the amodal properties of objects/events (e.g., tempo, rhythm, trajectory, onset) and then to the arbitrary relations between their visual and auditory characteristics. These arbitrary relations are, in fact, semantic correlations (e.g., object-sound or face-voice associations). More specifically, the study investigated whether 6-month-old infants detect speech synchrony irrespective whether the voice they hear matches the gender of the person they see speaking. The results showed that the infants' perception of audiovisual speech synchrony was affected by the face-voice gender correspondences. The 6-month-olds showed a preference for the actress who spoke in synchrony with a female voice recording, but not when the same actress uttered the phrases in synchrony a male voice recording. This finding suggests that infants use their prior knowledge about face-voice associations to process audiovisual speech.

Kirkham, Richardson, et al. (2012) is another study which found that infants process the arbitrary relations between the visual and the auditory features of a multisensory object/event and use these associations to index the spatial location (an amodal property) of the object/event. Kirkham, Richardson, et al. familiarized infants with two cartoon characters that occupied a specific square on the screen and moved in synchrony with one of two musical excerpts. After the familiarization, the infants saw only the two empty locations and heard either of the two musical pieces. The authors reported that 3- and 6-month-old infants preferred to look at the square previously associated with the musical sound heard. However, this was true only when the two characters used during the familiarization had different visual features. When they were identical, the infants did not learn the spatial location of the multisensory objects/events. According to the IRH, the infants should have responded similarly irrespective whether the characters were identical or not because, in both conditions, the audiovisual cues specified the location of the multisensory objects/events in the same manner. Aside from being inconsistent with the IRH, Kirkham, Richardson, et al.'s findings suggest that infants used their prior knowledge about cross-modal associations to guide their subsequent visual exploration of objects/events (see also Richardson & Kirkham, 2004).

7.2.2. Implications for multisensory development

The findings reported in this thesis are relevant not only to the IRH but also to the broader discussion about multisensory development in infants. The results indicate that, at six months, infants can differentiate between congruent and incongruent multisensory stimulation. Infants seem to use both the spatiotemporal and the semantic relations between cross-modal cues to solve the cross-modal binding problem (see also J.G. Bremner et al., 2012; Lawson, 1980; Bahrick & Lickliter, 2000; Walker-Andrews et al., 1991; Patterson & Werker, 1999). The exact mechanism through which infants learn these relations is unclear. One possibility is that infants map both the visual cues and the auditory cues onto the same reference frame, and the co-occurrence of some stimuli attracts infants' attention. But for this to occur, infants would have to decipher the different sensory inputs into the same code and then map them onto the same reference frame. The alternative possibility is that infants have two separate frames of reference - a visual frame and an auditory one. Dividing attention between these frames may allow the infants to learn the probabilistic relations between the cross-modal cues (see Pouget, Deneve, & Duhamel, 2004). But the puzzle in this situation is how do infants align these frames of reference. Studying neural entrainment and brain connectivity between different sensory areas may provide an answer (see Bauer, Debener, & Nobre, 2020).

Neurophysiological findings from animal studies provide more support to the first explanation. As mentioned in Chapter 1, the superior colliculus is a brain area involved in multisensory processing (Spence & Driver, 2004; Stein, 2012b). The neurons in this area have overlapping receptive fields for various sensory modalities, and they respond to the stimulus location rather than its presentation modality (Stein et al., 2004). These neurons use the animal's gaze direction as their frame of reference, and they have an enhanced firing rate in response to bimodal stimulation (coming from the same object/event) than unimodal stimulation. Research in juvenile animals has revealed that these "multisensory" neurons develop with age and are experience-dependent. More specifically, the neurons of newborn cats respond similarly to unimodal and bimodal stimulation. By comparison, the neurons of one-month-old kittens have a higher discharge rate in response to spatiotemporally overlapping cross-modal stimuli (Wallace & Stein, 1997). Besides, the number of multisensory neurons increases with age, while the size of their receptive fields decreases. Lastly, the activity of these neurons becomes increasingly anchored to the animal's gaze direction (King, 2004; Stein, 2012a).

Evidence that the visual cues drive the alignment of sensory maps in the superior colliculus comes from experiments in which researchers have altered the spatial relations between the maps. For example, King et al. (1988) surgically deviated the gaze direction of juvenile ferrets and then recorded the neuronal activity in the superior colliculus in response to auditory stimuli. The researchers found that the preferred sound location of the auditory neurons shifted laterally by a similar degree as the gaze deviation. However, this reorganisation of the auditory map did not occur when the ferrets had their eyes orientation deviated and eyelids sutured so they could not see (King & Carlile, 1995). These experiments (see also Knudsen & Knudsen, 1990) suggest that the repeated exposure to spatiotemporally overlapping cross-modal stimulation drives the alignment of the sensory maps in this brain area. Similar processes may occur in the developing human brain. However, due to ethical considerations, such neurophysiological studies have not been conducted on human infants.

Although infants may have some rudimentary mechanisms through which they untangle cross-modal stimulation, this does not necessarily mean that the multisensory processes are identical in infants and adults. As discussed in Chapter 1, a recurring finding in the literature on multisensory processing in adults is the so-called "congruency effect". Adults orient and respond faster to spatiotemporally congruent audiovisual cues than to unimodal cues (Harrington & Peck, 1998; Hughes et al., 1994; J. Miller, 1982; Mordkoff & Yantis, 1991). Furthermore, they discriminate and encode better semantically congruent audiovisual stimuli than purely visual stimuli (Chen & Spence, 2010; Lehmann & Murray, 2005; Murray et al., 2004). Only two of the studies reported in this thesis found a congruency effect in infants. Specifically, 6-month-old infants learned the trajectory of an object when the object was specified concurrently by both vision and audition than when it appeared only visually. However, this congruency effect did not extend to the object's pattern. It is unclear why the infants did not display a more robust congruency effect. Either the congruency effect found reflects a change in infants' visual scanning pattern when the object trajectory changed (see Section 7.2.1), or it is task-dependent. The second study that found a congruency effect was the one that investigated audiovisual speech processing in 6-month-old infants. There, the infants detected speech synchrony only when the audiovisual cues were semantically congruent.

A more robust effect that the studies reported in this thesis found was that incongruent audiovisual stimulation hindered infants' visual processing and learning. In two separate studies, 6-month-old infants failed to encode the pattern on an object when they received incongruent compared to unimodal stimulation. Furthermore, this hindering effect was not specific to the object pattern. The infants did not learn the object's trajectory, either, when the auditory and the visual input was spatiotemporally incongruent. The fact that the infants responded differently to congruent than incongruent audiovisual stimulation suggests that they did not preferentially process auditory information over visual information.

Various studies that have investigated multisensory processing in children have found that children rely more on auditory cues than adults do (Innes-Brown et al., 2011; Massaro et al., 1986; Napolitano & Sloutsky, 2004; Nava & Pavani, 2013; Thomas et al., 2017). For example, Thomas et al. asked different age groups of children and a group of adults to listen to and identify some animal sounds. While the participants listened to the sounds, they saw pictures of animals that were either congruent (e.g., the sound of a dog barking paired with a picture of a dog) or incongruent with the animal sounds. As a control condition, the researchers played the animal sound alongside a checkerboard image. Thomas et al. found that all the participants identified the animal sound faster in the congruent condition than in the control condition. However, the 8 to 9-year-olds and the adults responded slower in the incongruent condition than in the control condition. This incongruency effect was not apparent in the group of 6 to 7-year-olds, who were equally fast in both the incongruent and the control conditions. These results indicate that younger children give more weight to auditory input than older children and adults do. While it is possible that infants too attend more to auditory input than adults, the studies reported in this thesis found that auditory information is disruptive only when it is incongruent with the visual event (see also Barr et al., 2010). Notably, as infants grow up and their perceptual experience broadens, they learn to deal with auditory noise and are no longer affected by it (at least in the occlusion task I employed).

Even though infants seem to learn to overcome the effect of incongruent stimulation, the question as to why 6-month-old infants' visual processing and learning was affected by incongruent audiovisual cues remains. There are several explanations for this phenomenon, but I will focus on only three of them. The first explanation is that young infants are less familiar with the type of incongruent stimulation used in the studies presented in this thesis. Typically, in the environment, the incongruent cues are transient, and they do not occur in close spatial proximity. However, in the studies reported here, the infants heard the incongruent sound throughout the habituation period. As a result, the infants may have struggled to decide whether to associate or to segregate the cues. In turn, this process may have distracted them from the visual event. Robinson & Sloutsky (2007b) found that familiarizing infants with a computer-generated sound before completing a task benefited the infants, whose visual processing is hindered less by the sound. In this case, the fact that the infants heard the sound before the study may have allowed them to segregate the visual and the auditory inputs faster.

The second explanation is that the incongruent sound directed the infants' attention to the centre of the screen, while the movement of the object made the infants shift attention from one side of the screen to the other. Previously, Spence & McDonald (2004) reported that, in adults, presenting a cue in one sensory modality leads to covert spatial orienting of attention to the cued spatial location. This cross-modal spatial orienting effect results in participants responding faster to the stimuli presented in a cued location than in another place. The infants in the incongruent multisensory condition may have used the sound as a cue to look more to the centre of the screen. Alternatively, they may have followed the ball across the display, but they attended covertly to the centre of the screen. Using eye-tracking technology to record infants' looking pattern during the study could have offered an answer.

Finally, the static sound used in the incongruent condition may have been less alerting for infants than the dynamic sound used in the congruent condition. To avoid cueing the infants through a particular rhythmic variation, the musical excerpt created and used in the studies reported here had few modulations and was quite repetitive. In a previous study, Wada et al. (2009) found that a rare sound embedded in a sequence of frequent sounds helped infants detect an illusory contour figure in a succession of images. More specifically, the infants looked longer at the side of the screen that displayed the illusory contour figure when they heard a rare sound but not a frequent sound. The explanation that the authors provided for this finding was that the infrequent sound was more alerting for the infants. As a result, the infants processed better the visual display. In light of this finding, it may well be the case that the sound used in the incongruent condition was not alerting or engaging enough. Across the studies conducted, the infants' total looking duration during the habituation period did not differ significantly between the congruent and incongruent conditions. Therefore, the infants engaged similarly with the stimuli, but maybe they attended to different aspects.

In sum, this thesis paints a complex picture of infants' multisensory processing. The studies conducted suggest that infants use the spatiotemporal and semantic relations between sensory cues to disentangle stimulation. Furthermore, infants seem to benefit from congruent audiovisual stimulation in some tasks (e.g., object occlusion, audiovisual speech processing), but the effect is not as robust as in adults. By contrast, incongruent multisensory stimulation confuses infants and interferes with their visual processing and learning. These findings are inconsistent with the IRH, which argues that infants attend to different object/event properties when the audiovisual cues are congruent compared to when they are incongruent. However, consistent with the IRH, the effects observed were more robust in younger than in older infants.

7.3. Limitations

In the empirical chapters of this thesis, I discussed some of the shortcomings of the studies reported. Therefore, in this section, I would like to reflect on three more general limitations which I have not addressed previously. The first one is that, in three studies, I used only one set of visual and auditory stimuli. More specifically, I habituated the infants to a dotted ball that moved across the display. Furthermore, in the congruent and incongruent multisensory conditions, the infants heard a musical sound while they watched the ball rolling. Again, I used only one sound across the studies. Although it is not uncommon for infant habituation studies to use only one set of stimuli (e.g., Bahrick et al., 2006; J. G. Bremner et al., 2012; Lewkowicz, 1996), one can question whether the results can be generalized.

In the first empirical chapter, I reported a habituation study in which I used a different set of visual stimuli. In that study, I habituated two groups of 10-month-old infants with a half-red and half-green ball. I found that the infants did not differentiate between a test event which showed no change in the pattern on the ball and another test event in which the ball changed from half red and half green to chequered red and green. As discussed in Chapter 2, this pattern change was probably not salient enough for infants to notice it. As a result, in the subsequent studies, I employed a different pattern change - the ball changed from dots to stripes. Other studies have used a similar pattern change with different age-groups of infants (e.g., Wilcox, 1999; Wilcox & Chapa, 2004; Wilcox, Smith, & Woods, 2011). Therefore, it seemed reasonable to follow what other researchers have done. Habituating some infants to the dotted ball, and other infants to the stripy ball could have partially addressed the issue of using one set of stimuli. Similarly, editing a second musical excerpt and having some infants listen to one sound and other infants listen to the second sound would have been beneficial.

The second issue is that I used the same task (i.e., an occlusion task) across four studies. Although this approach allowed me to detect age-related effects of multisensory stimulation, it prevented me from getting an insight into how audiovisual cues affect 4-month-old infants' processing of visual object pattern. The 4-month-old infants I studied did not differentiate between the test events (irrespective of which habituation condition they were in), which suggests that the occlusion task was too difficult for them. To address this limitation, I could have conducted a new study with this age group of infants, in which the ball remained fully visible during the habituation. In this study, the *Change* event could have displayed a ball that changed its pattern abruptly during the crossing of the display. This way, the memory demand would have been smaller, and the infants may have processed the object pattern (see Burnham, 1987; Hartlep, 1983; Morton & Johnson, 1991).

Thirdly, the studies conducted did not investigate whether the infants learned anything about the other object in the display. As I mentioned earlier in the thesis, in the incongruent condition, the sound appeared to originate from the box located in the centre of the screen. Therefore, the 6-month-old infants in the incongruent condition could have looked more at and encoded better the pattern on the box. If that was the case, then the incongruent stimulation did not hinder infants' visual processing and learning more broadly. Instead, it affected only their attention to the ball. To mitigate this limitation, I could have conducted another habituation study that could have measured infants' response to changes in either the ball's pattern or the box's pattern. This study could have clarified whether the sound cued the infants' attention to a particular spatial location.

Lastly, some of the results reported in this thesis could be due to floor and ceiling effects. More specifically, the occlusion interval used in three of the studies reported here was 634 ms. This interval may have been too long for the 4-month-old infants (in Chapter 4) to recall the pattern on the ball. Previous research by S. P. Johnson, Bremner, et al. (2003) found that the 4-month-old infants can represent an occluded object for about 400 ms, while the 6-month-old infants can do that for roughly 600 ms. Therefore, the results reported in the 4-months-old group may reflect a floor effect rather than an inability to use pattern information to individuate objects involved in occlusion events. At the same time, this interval of 634 ms may have been too short for the 10-month-old infants (in Chapter 3), who detected the ball pattern change irrespective of their habituation condition. This lack of variability in infants' responses could suggest a ceiling effect rather than the fact that older infants are better at ignoring spatiotemporally incongruent stimulation. To deal with floor and ceiling effects, I could have adjusted the length of the occlusion interval depending on the age of the infants.

7.4. Future research

In this thesis, I have tried to find out whether infants differentiate between congruent and incongruent audiovisual cues, and whether different kinds of stimulation affect their cognitive processing. Inevitably, the studies conducted have raised other questions about multisensory processing in infants. In this section, I will propose three areas of further investigation, which would build on the findings of this thesis. The first one concerns whether linguistic labels affect how infants process visual objects/events and whether the effect is similar to that of spatiotemporally congruent or incongruent sounds. Parents often label the toys that infants manipulate or look at (see Bornstein, Tamis-LeMonda, Hahn, & Haynes, 2008; Tamis-LeMonda, Kuchirko, & Song, 2014). Therefore, infants must learn to segregate the linguistic input from the visual information. Robinson & Sloutsky (2007a, 2007b, 2008) have already looked into this problem, and they have found that object labels are less distracting than computer-generated sounds. However, the sounds they used were incongruent with the objects shown. It remains to be seen whether the effect of linguistic labels is like that of spatiotemporally congruent sounds. For example, in the occlusion task presented in this thesis, hearing the same object label every time the object re-emerges from behind the occluder could help infants learn the pattern. Previous research found that, if infants hear the same object label when they watch a succession of objects, they look more at the shared features of the items (Althaus & Mareschal, 2014). Therefore, linguistic input may help infants interpret occlusion events.

Secondly, it would also be interesting to look at whether social cues help infants overcome the effect of incongruent audiovisual stimulation. Incongruent sounds (or acoustic noise) are encountered frequently in everyday activities, and sometimes the sources of those sounds are not apparent. Yet, infants appear to deal successfully with the noise and learn about the objects they encounter. It is unclear how infants do that, but the linguistic input and social cues they receive from parents may guide infants' attention. For example, infants are more likely to look at the object that an adult is gazing at if the adult first makes eye-contact or addresses the infant using infants directed speech (Parsons et al., 2019; Senju & Csibra, 2008). Whether these cues are equally engaging in a noisy environment, and the infants continue to follow the adult's gaze to the referenced object despite the acoustic noise is an empirical question that needs answering.

Finally, in Chapter 6, I discussed the fact that infants display more robust matching of female faces and voices (Hillairet de Boisferon et al., 2015; Poulin-Dubois et al., 1994; Richoz et al., 2017). This imbalance in matching faces and voices of different genders may be because infants spend more time with women than men during their first year of life. If having experience with speakers of different genders is essential for forming face-voice gender associations, then infants may struggle to detect gender correspondences in speakers coming from other cultures/races. Cross-cultural studies with infants have found that, at three months of age, infants can differentiate between male and female faces that belong to their own-race but not to other-races (Liu et al., 2015; Quinn et al., 2008). This finding suggests that infants may have difficulties deciding whom to attribute the voice they hear. In turn, this difficulty may impact how they process audiovisual speech in foreign speakers. I have already started investigating this research question, but I had to pause the research project due to time constraints and limited resources.

7.5. Conclusions

Infants grow up in complex environments where they are recipients of continuous streams of sensory information across multiple sensory modalities. The multisensory stimulation is either related, such as when both the auditory and the visual inputs originate from the same object/event, or unrelated. The focus of this thesis concerned addressing how infants navigate this complex environment and whether they benefit from multisensory stimulation when processing objects/events. The studies conducted show that 6-month-old infants use the spatiotemporal and semantic relations between auditory and visual cues to interpret sensory input. Evidence supporting this conclusion comes from the fact that the infants learned the pattern and the trajectory of a moving object when a spatiotemporally congruent sound accompanied the object, but not when an incongruent sound did. Besides, 6-month-old infants were able to detect audiovisual speech synchrony when they saw and heard faces and voices of the same gender but not of different genders. The effect of incongruent audiovisual stimulation was more robust in 6- than in 10-month-old infants, which suggests that infants learn to deal with incongruent cues more efficiently as they accumulate more experience with multisensory stimulation. The studies reported in this thesis offer an insight into how infants process social and object-related audiovisual information and reveal that incongruent cues hinder young infants visual processing and learning.

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Althaus, N., & Mareschal, D. (2014). Labels direct infants' attention to commonalities during novel category learning. *PLoS ONE*, *9*(7): e99670. https://doi.org/10.1371/journal.pone.0099670

Atkinson, J., & Braddick, O. (2012). Visual attention in the first years: typical development and developmental disorders. *Developmental Medicine & Child Neurology*, *54*(7), 589–595. https://doi.org/10.1111/j.1469-8749.2012.04294.x

Baart, M., Vroomen, J., Shaw, K., & Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition*, *130*(1), 31–43. https://doi.org/10.1016/j.cognition.2013.09.006

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development*, *10*(4), 387–416. https://doi.org/10.1016/0163-6383(87)90039-7

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, *59*(1), 197–209. https://doi.org/10.1111/j.1467-8624.1988.tb03208.x

Bahrick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Child Psychology*, *53*(2), 180–199. https://doi.org/10.1016/0022-0965(92)90048-B

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology*, *6*(2), 111–123. https://doi.org/10.1207/s15326969eco0602_2

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, *41*(4), 352–363. https://doi.org/10.1002/dev.10049

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face-voice relations. *Developmental Psychology*, *41*(3), 541–552. https://doi.org/10.1037/0012-1649.41.3.541

Bahrick, L. E., Krogh-Jespersen, S., Argumosa, M. A., & Lopez, H. (2014). Intersensory redundancy hinders face discrimination in preschool children: Evidence for visual facilitation. *Developmental Psychology*, *50*(2), 414–421. https://doi.org/10.1037/a0033476

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190–201. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 30, pp. 153–187). Academic Press. https://doi.org/10.1016/S0065-2407(02)80041-6

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 137–147. https://doi.org/10.3758/CABN.4.2.137

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–205). Oxford University Press.

Bahrick, L. E., Lickliter, R., & Flom, R. (2006). Up versus down: The role of intersensory redundancy in the development of infants' sensitivity to the orientation of moving objects. *Infancy*, *9*(1), 73–96. https://doi.org/10.1207/s15327078in0901_4

Bahrick, L. E., McNew, M. E., Pruden, S. M., & Castellanos, I. (2019). Intersensory redundancy promotes infant detection of prosody in infant-directed speech. *Journal of Experimental Child Psychology*, *183*, 295–309. https://doi.org/10.1016/j.jecp.2019.02.008

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*(5), 1263–1275. https://doi.org/10.2307/1132264

Bahrick, L. E., Walker, A. S., & Neisser, U. (1981). Selective looking by infants. *Cognitive Psychology*, *13*(3), 377–390. https://doi.org/10.1016/0010-0285(81)90014-1

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science*, *13*(3), 89–94. https://doi.org/10.1111/j.0963-7214.2004.00281.x

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208. https://doi.org/10.1016/0010-0277(85)90008-3

Baillargeon, R., & Wang, S. (2002). Event categorization in infancy. *Trends in Cognitive Sciences*, *6*(2), 85–93. https://doi.org/10.1016/S1364-6613(00)01836-2

Barakat, B., Seitz, A. R., & Shams, L. (2015). Visual rhythm perception improves through auditory but not visual training. *Current Biology*, *25*(2), R60–R61. https://doi.org/10.1016/j.cub.2014.12.011

Barr, R., Shuck, L., Salerno, K., Atkinson, E., & Linebarger, D. L. (2010). Music interferes with learning from television during infancy. *Infant and Child Development*, *19*, 313–331. https://doi.org/10.1002/icd.666

Bauer, A. K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of neural oscillations and cross-modal influences. *Trends in Cognitive Sciences*, *24*(6), 481–495. https://doi.org/10.1016/j.tics.2020.03.003

Begum Ali, J., Spence, C., & Bremner, A. J. (2015). Human infants' ability to perceive touch in external space develops postnatally. *Current Biology*, *25*(20), R978–R979. https://doi.org/10.1016/j.cub.2015.08.055

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*(6), 578–584. https://doi.org/10.3758/BF03207374

Bertenthal, B. I. (1996). Origins and early development of perception, action, and representation. *Annual Review of Psychology*, *47*(1), 431–459. https://doi.org/10.1146/annurev.psych.47.1.431

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9–10), 341–347.

Bogartz, R. S., Shinskey, J. L., & Schilling, T. H. (2000). Object permanence in five-and-a-half-month-old infants? *Infancy*, *1*(4), 403–428. https://doi.org/10.1207/S15327078IN0104_3

Bornstein, M. H., Ferdinandsen, K., & Gross, C. G. (1981). Perception of symmetry in infancy. *Developmental Psychology*, *17*(1), 82–86. https://doi.org/10.1037/0012-1649.17.1.82

Bornstein, M. H., & Krinsky, S. J. (1985). Perception of symmetry in infancy: The salience of vertical symmetry and the perception of pattern wholes. *Journal of Experimental Child Psychology*, *39*(1), 1–19. https://doi.org/10.1016/0022-0965(85)90026-8

Bornstein, M. H., Tamis-LeMonda, C. S., Hahn, C.-S., & Haynes, O. M. (2008). Maternal responsiveness to young children at three ages: Longitudinal analysis of a multidimensional, modular, and specific parenting construct. *Developmental Psychology*, *44*(3), 867–874. https://doi.org/10.1037/0012-1649.44.3.867

Bourguignon, M., Baart, M., Kapnoula, E. C., & Molinaro, N. (2020). Lip-reading enables the brain to synthesize auditory features of unknown silent speech.

The Journal of Neuroscience, *40*(5), 1053–1065. https://doi.org/10.1523/JNEUROSCI.1101-19.2019

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (Eds.) (2012). *Multisensory development*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.001.0001

Bremner, A. J., & Mareschal, D. (2004). Reasoning . . . what reasoning? *Developmental Science*, *7*(4), 419–421. https://doi.org/10.1111/j.1467-7687.2004.00360.x

Bremner, J. G., Slater, A. M., Johnson, S. P., Mason, U. C., & Spring, J. (2012). The effects of auditory information on 4-month-old infants' perception of trajectory continuity. *Child Development*, *83*(3), 954–964. https://doi.org/10.1111/j.1467-8624.2012.01739.x

Bremner, J. G., Slater, A. M., Johnson, S. P., Mason, U. C., Spring, J., & Bremner, M. E. (2011a). Two- to eight-month-old infants' perception of dynamic auditory-visual spatial colocation. *Child Development*, *82*(4), 1210–1223. https://doi.org/10.1111/j.1467-8624.2011.01593.x

Bremner, J. G., Slater, A. M., Mason, U. C., Spring, J., & Johnson, S. P. (2013). Trajectory perception and object continuity: Effects of shape and color change on 4-month-olds' perception of object identity. *Developmental Psychology*, *49*(6), 1021–1026. https://doi.org/10.1037/a0029398

Bremner, J. G., & Wachs, T. D. (Eds.) (2010). *The Wiley-Blackwell handbook of infant development*. Wiley-Blackwell. https://doi.org/10.1002/9781444327564

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception*, *22*(2), 131–152. https://doi.org/10.1068/p220131

Burnham, D. K., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220. https://doi.org/10.1002/dev.20032

Burnham, D. K. (1987). The role of movement in object perception by infants. In B. E. McKenzie & R. H. Day (Eds.), *Perceptual development in early infancy: Problems and issues* (pp. 143–172). Erlbaum.

Calvert, G. A., Spence, C., & Stein, B. E. (Eds.) (2004). *The handbook of multisensory processes*. MIT Press. https://doi.org/10.5860/choice.42-4014

Cashon, C. H., & Cohen, L. B. (2000). Eight-month-old infants' perception of possible and impossible events. *Infancy*, *1*(4), 429–446. https://doi.org/10.1207/S15327078IN0104_4

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. https://doi.org/10.1016/j.cognition.2009.10.012

Chen, Y.-C., & Westermann, G. (2018). Different novelties revealed by infants' pupillary responses. *Scientific Reports*, 8: 9533, 1-8. https://doi.org/10.1038/s41598-018-27736-z

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Inc. http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, *43*(3), 869–879. https://doi.org/10.2307/1127638

Cohen, L. B. (2004). Uses and misuses of habituation and related preference paradigms. *Infant and Child Development*, *13*(4), 349–352. https://doi.org/10.1002/icd.355

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*(4), 495–506. https://doi.org/10.1016/S1388-2457(02)00024-X

Colin, C., Radeau, M., Deltenre, P., & Morais, J. (2001). Rules of intersensory integration in spatial scene analysis and speechreading. *Psychologica Belgica*, *41*(3), 131–144. https://doi.org/10.5334/pb.977

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225–234. https://doi.org/10.1016/j.nlm.2008.06.002

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536. https://doi.org/10.1037/dev0000083

Cumming, G. (2012). *Understanding the new statistics*. Routledge. https://doi.org/10.4324/9780203807002

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony.

Perception, *9*(6), 719–721. https://doi.org/10.1068/p090719

Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. *Current Directions in Psychological Science*, *26*(5), 451–457. https://doi.org/10.1177/0963721417709087

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5): 7. https://doi.org/10.1167/7.5.7

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Fantz, R. L. (1958). Pattern vision in young infants. *The Psychological Record*, *8*(2), 43–47. https://doi.org/10.1007/BF03393306

Fantz, R. L. (1961). The origin of form perception. *Scientific American*, *204*(5), 66–73. https://doi.org/10.1038/scientificamerican0561-66

Fantz, R. L. (1963). Pattern vision in newborn infants. *Science*, *140*(3564), 296–297. https://doi.org/10.1126/science.140.3564.296

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fellous, J.-M. (1997). Gender discrimination and prediction on the basis of facial metric information. *Vision Research*, *37*(14), 1961–1973. https://doi.org/10.1016/S0042-6989(97)00010-2

Fenwick, K. D., & Morrongiello, B. A. (1998). Spatial co-location and infants' learning of auditory-visual associations. *Infant Behavior and Development*, *21*(4), 745–759. https://doi.org/10.1016/S0163-6383(98)90042-X

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*(1), 238–252. https://doi.org/10.1037/0012-1649.43.1.238

Fulkerson, M. (2014). Rethinking the senses and their interactions: The case for sensory pluralism. *Frontiers in Psychology*, *5*, 1–14. https://doi.org/10.3389/fpsyg.2014.01426

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. H. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*(5), 550–562. https://doi.org/10.1080/15250000903144199

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*: 527. https://doi.org/10.3389/fpsyg.2013.00527

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–149. https://doi.org/10.1006/jecp.1998.2438

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition. *Perspectives on Psychological Science*, *8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Haan, J., & Van Heuven, V. J. (1999). Male Vs. female pitch range in Dutch questions. *14th International Conference of Phonetic Sciences*, *January 1999*, 1581–1584. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1581.pdf

Hadad, B., Schwartz, S., Maurer, D., & Lewis, T. L. (2015). Motion perception: A review of developmental changes and the role of early visual experience. *Frontiers in Integrative Neuroscience*, *9*, 1–18. https://doi.org/10.3389/fnint.2015.00049

Hammond-Kenny, A., Bajo, V. M., King, A. J., & Nodal, F. R. (2017). Behavioural benefits of multisensory processing in ferrets. *European Journal of Neuroscience*, *45*(2), 278–289. https://doi.org/10.1111/ejn.13440

Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research*, *122*(2), 247–252. https://doi.org/10.1007/s002210050512

Hartlep, K. (1983). Simultaneous presentation of moving objects in an infant tracking task. *Infant Behavior and Development*, *6*(1), 79–84. https://doi.org/10.1016/S0163-6383(83)80010-1

Hellier, J. L. (Ed.) (2016). *The five senses and beyond: The encyclopedia of perception*. Greenwood.

Hernandez-Reif, M., & Bahrick, L. E. (2001). The development of visual-tactual perception of objects: Amodal relations provide the basis for learning arbitrary relations. *Infancy*, *2*(1), 51–72. https://doi.org/10.1207/S15327078IN0201_4

Hillairet de Boisferon, A., Dupierrix, E., Quinn, P. C., Lœvenbruck, H., Lewkowicz, D. J., Lee, K., & Pascalis, O. (2015). Perception of multisensory gender coherence in 6- and 9-month-old infants. *Infancy*, *20*(6), 661–674. https://doi.org/10.1111/infa.12088

Howard, I. P., & Templeton, W. B. (1966). *Human Spatial Orientation*. Wiley.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 131–153. https://doi.org/10.1037/0096-1523.20.1.131

Humphrey, G. K., Humphrey, D. E., Muir, D. W., & Dodwell, P. C. (1986). Pattern perception in infants: Effects of structure and transformation. *Journal of Experimental Child Psychology*, *41*(1), 128–148. https://doi.org/10.1016/0022-0965(86)90055-X

Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. G. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. *Developmental Science*, *14*(5), 1089–1099. https://doi.org/10.1111/j.1467-7687.2011.01059.x

Innes-Brown, H., & Crewther, D. (2009). The impact of spatial incongruence on an auditory-visual illusion. *PLoS ONE*, *4*(7), e6450. https://doi.org/10.1371/journal.pone.0006450

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, *5*(2), 52–65. https://doi.org/10.1080/17470215308416626

Jay, M. F., & Sparks, D. L. (1987a). Sensorimotor integration in the primate superior colliculus. I. Motor convergence. *Journal of Neurophysiology*, *57*(1), 22–34. https://doi.org/10.1152/jn.1987.57.1.22

Jay, M. F., & Sparks, D. L. (1987b). Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *Journal of Neurophysiology*, *57*(1), 35–55. https://doi.org/10.1152/jn.1987.57.1.35

Johnson, M. H., & de Haan, M. (2010). *Developmental cognitive neuroscience* (3rd ed.). Wiley-Blackwell.

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences*, *100*(18), 10568–10573. https://doi.org/10.1073/pnas.1630655100

Johnson, S. P., & Aslin, R. N. (1995). Perception of object unity in 2-month-old infants. *Developmental Psychology*, *31*(5), 739–745. https://doi.org/10.1037/0012-1649.31.5.739

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development*, *74*(1), 94–108. https://doi.org/10.1111/1467-8624.00523

Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research*, *174*(3), 588–594. https://doi.org/10.1007/s00221-006-0634-0

Jones, J. A., & Munhall, K. G. (1997). The effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics*, *25*(4), 13–19.

Jordan, K. E., & Brannon, E. M. (2006). The multisensory representation of number in infancy. *Proceedings of the National Academy of Sciences*, *103*(9), 3486–3489. https://doi.org/10.1073/pnas.0508107103

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*(4), 483–524. https://doi.org/10.1016/0010-0285(83)90017-8

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology*, *104*(1), 105–114. https://doi.org/10.1016/j.jecp.2009.01.006

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kim, R., Peters, M. A. K., & Shams, L. (2012). 0 + 1 > 1: How adding noninformative sound improves performance on a visual task. *Psychological Science*, *23*(1), 6–12. https://doi.org/10.1177/0956797611420662

King, A. J. (1993). A map of auditory space in the mammalian brain: Neural computation and development. *Experimental Physiology*, *78*(5), 559–590. https://doi.org/10.1113/expphysiol.1993.sp003708

King, A. J. (2004). Development of multisensory spatial integration. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 1–24). Oxford University Press.

King, A. J., & Carlile, S. (1995). Neural coding for auditory space. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 279–293). MIT Press.

King, A. J., Doubell, T. P., & Skaliora, I. (2004). Epigenetic factors that align visual and auditory mas in the ferret midbrain. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 613–624). MIT Press.

King, A. J., Hutchings, M. E., Moore, D. R., & Blakemore, C. (1988). Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. *Nature*, *332*(6159), 73–76. https://doi.org/10.1038/332073a0

Kirkham, N. Z., Richardson, D. C., Wu, R., & Johnson, S. P. (2012a). The importance of " what": Infants use featural information to index events. *Journal of Experimental Child Psychology*, *113*(3), 430–439. https://doi.org/10.1016/j.jecp.2012.07.001

Kirkham, N. Z., Wagner, J. B., Swan, K. A., & Johnson, S. P. (2012). Sound support: Intermodal information facilitates infants' perception of an occluded trajectory. *Infant Behavior and Development*, *35*(1), 174–178. https://doi.org/10.1016/j.infbeh.2011.09.001

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(14), 1–16. https://doi.org/10.1068/v070821

Knudsen, E. I., & Knudsen, P. F. (1990). Sensitive and critical periods for visual calibration of sound localization by Barn owls. *Journal of Neuroscience*, *10*(1), 222–232. https://doi.org/10.1523/jneurosci.10-01-00222.1990

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9): e943. https://doi.org/10.1371/journal.pone.0000943

Kubicek, C., Hillairet de Boisferon, A., Dupierrix, E., Pascalis, O., Lœvenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PLoS ONE*, *9*(2): e89275. https://doi.org/10.1371/journal.pone.0089275

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, *105*(32), 11442–11445. https://doi.org/10.1073/pnas.0804275105

Kwon, M.-K., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, *52*(4), 537–555.

https://doi.org/10.1037/dev0000080

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*: 863. https://doi.org/10.3389/fpsyg.2013.00863

Lartillot, O., & Toiviainen, P. (2007). MIR in Matlab (II): A toolbox for musical feature extraction from audio. *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007*, 127–130.

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*(4), 405–414. https://doi.org/10.1007/s00221-004-1913-2

Lawson, K. R. (1980). Spatial and temporal congruity and auditory-visual integration in infants. *Developmental Psychology*, *16*(3), 185–192. https://doi.org/10.1037/0012-1649.16.3.185

Lecanuet, J.-P., & Schaal, B. (1996). Fetal sensory competencies. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *68*(1–2), 1–23. https://doi.org/10.1016/0301-2115(96)02509-2

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, *24*(2), 326–334. https://doi.org/10.1016/j.cogbrainres.2005.02.005

Lejeune, F., Parra, J., Berne-Audéoud, F., Marcus, L., Barisnikov, K., Gentaz, E., & Debillon, T. (2016). Sound interferes with the early tactile manual abilities of preterm infants. *Scientific Reports*, *6*: 23329. https://doi.org/10.1038/srep23329

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1094–1106. https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J. (2004a). Perception of serial order in infants. *Developmental Science*, *7*(2), 175–184. https://doi.org/10.1111/j.1467-7687.2004.00336.x

Lewkowicz, D. J. (2004b). Serial order processing in human infants and the role of multisensory redundancy. *Cognitive Processing*, *5*(2), 113–122. https://doi.org/10.1007/s10339-004-0015-1

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, *46*(1), 66–77. https://doi.org/10.1037/a0015579

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences*, *103*(17), 6771–6774. https://doi.org/10.1073/pnas.0602027103

Lewkowicz, D. J., & Kraebel, K. (2004). The value of multisensory redundancy in the development of intersensory perception. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 655–678). MIT Press.

Lewkowicz, D. J., & Lickliter, R. (Eds.) (1994). *The development of intersensory perception: Comparative perspectives*. Lawrence Erlbaum Associates, Inc.

Lewkowicz, D. J., & Marcovitch, S. (2006). Perception of audiovisual rhythm and its invariance in 4- to 10-month-old infants. *Developmental Psychobiology*, *48*(4), 288–300. https://doi.org/10.1002/dev.20140

Lewkowicz, D. J., Minar, N. J., Tift, A. H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147–162. https://doi.org/10.1016/j.jecp.2014.10.006

Libertus, K., & Landa, R. J. (2013). The Early Motor Questionnaire (EMQ): A parental report measure of early motor development. *Infant Behavior and Development*, *36*(4), 833–842. https://doi.org/10.1016/j.infbeh.2013.09.007

Liu, S., Xiao, N. G., Quinn, P. C., Zhu, D., Ge, L., Pascalis, O., & Lee, K. (2015). Asian infants show preference for own-race but not other-race female faces: the role of infant caregiving arrangements. *Frontiers in Psychology*, *6*: 593. https://doi.org/10.3389/fpsyg.2015.00593

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131–141. https://doi.org/10.3109/03005368709077786

Marcovitch, S., & Zelazo, P. D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Developmental Science*, *12*(1), 1–18. https://doi.org/10.1111/j.1467-7687.2008.00754.x

Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M., Thomas, M. S. C., & Westermann, G. (2007). *Neuroconstructivism, Vol. I: How the brain constructs cognition*. Oxford University Press.

Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cognitive Science*, *26*(3), 377–389. https://doi.org/10.1016/S0364-0213(02)00062-9

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, *41*(1), 93–113. https://doi.org/10.1016/0022-0965(86)90053-6

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, *22*: e12701. https://doi.org/10.1111/desc.12701

Meredith, M. A., & Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, *365*(2), 350–354. https://doi.org/10.1016/0006-8993(86)91648-3

Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development*, *5*, 143–159. https://doi.org/10.1016/S0163-6383(82)80024-6

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Minar, N. J., & Lewkowicz, D. J. (2018). Overcoming the other-race effect in infancy with multisensory redundancy: 10-12-month-olds discriminate dynamic other-race faces producing speech. *Developmental Science*, *21*(4), e12604. https://doi.org/10.1111/desc.12604

Moore, M. K., & Meltzoff, A. N. (2008). Factors affecting infants’ manual search for occluded objects and the genesis of object permanence. *Infant Behavior and Development*, *31*(2), 168–180. https://doi.org/10.1016/j.infbeh.2007.10.006

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 520–538. https://doi.org/10.1037/0096-1523.17.2.520

Morrongiello, B. A. (1988). Infants’ localization of sounds along the horizontal axis: Estimates of minimum audible angle. *Developmental Psychology*, *24*(1), 8–13. https://doi.org/10.1037/0012-1649.24.1.8

Morrongiello, B. A., Fenwick, K. D., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior and Development*, *21*(4), 543–553. https://doi.org/10.1016/S0163-6383(98)90028-5

Morrongiello, B. A., Fenwick, K. D., & Nutley, T. (1998). Developmental changes in associations between auditory-visual events. *Infant Behavior and Development*, *21*(4), 613–626. https://doi.org/10.1016/S0163-6383(98)90033-9

Morrongiello, B. A., & Rocca, P. T. (1987). Infants’ localization of sounds in the horizontal plane: Effects of auditory and visual cues. *Child Development*, *58*(4), 918–927. https://doi.org/10.1111/j.1467-8624.1987.tb01429.x

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, *98*(2), 164–181. https://doi.org/10.1037/0033-295X.98.2.164

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (AGS Editio). American Guidance Services, Inc.

Munakata, Y. (2000). Challenges to the Violation-of-Expectation paradigm: Throwing the conceptual baby out with the perceptual processing bathwater? *Infancy*, *1*(4), 471–477. https://doi.org/10.1207/S15327078IN0104_7

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, *104*(4), 686–713. https://doi.org/10.1037/0033-295X.104.4.686

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, *58*(3), 351–362. https://doi.org/10.3758/BF03206811

Murray, M. M., Foxe, J. J., & Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage*, *27*(2), 473–478. https://doi.org/10.1016/j.neuroimage.2005.04.016

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*(1), 125–135. https://doi.org/10.1016/j.neuroimage.2003.09.035

Murray, M. M., & Sperdin, H. F. (2010). Single-trial multisensory learning and memory retrieval. In M. J. Naumer & J. Kaiser (Eds.), *Multisensory Object Perception in the Primate Brain* (pp. 191–208). Springer New York. https://doi.org/10.1007/978-1-4419-5615-6_11

Napolitano, A. C., & Sloutsky, V. M. (2004). Is a Picture Worth a Thousand Words? The Flexible Nature of Modality Dominance in Young Children. *Child Development*, *75*(6), 1850–1870. https://doi.org/10.1111/j.1467-8624.2004.00821.x

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, *18*(9), 689–693. https://doi.org/10.1016/j.cub.2008.04.021

Naumer, M. J., & Kaiser, J. (2010). *Multisensory object perception in the primate brain*. Springer New York. https://doi.org/10.1007/978-1-4419-5615-6

Nava, E., Grassi, M., Brenna, V., Croci, E., & Turati, C. (2017). Multisensory motion perception in 3–4 month-old infants. *Frontiers in Psychology*, *8*: 1994.

https://doi.org/10.3389/fpsyg.2017.01994

Nava, E., & Pavani, F. (2013). Changes in sensory dominance during childhood: converging evidence from the Colavita effect and the sound-induced flash illusion. *Child Development*, *84*(2), 604–616. https://doi.org/10.1111/j.1467-8624.2012.01856.x

Needham, A. (1999). The role of shape in 4-month-old infants' object segregation. *Infant Behavior and Development*, *22*(2), 161–178. https://doi.org/10.1016/S0163-6383(99)00008-9

Needham, A., & Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition*, *62*(2), 121–149. https://doi.org/10.1016/S0010-0277(96)00727-5

Needham, A., & Kaufman, J. (1997). Infants' integration of information from different sources in object segregation. *Early Development and Parenting*, *6*(3–4), 137–148. https://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<137::AID-EDP153>3.0.CO;2-C

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, *9*(5), 454–464. https://doi.org/10.1111/j.1467-7687.2006.00512.x

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, *11*(3), 255–268. https://doi.org/10.1080/15248371003699977

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early Human Development*, *18*(1), 27–36. https://doi.org/10.1016/0378-3782(88)90040-0

Orioli, G., Bremner, A. J., & Farroni, T. (2018). Multisensory perception of looming and receding objects in human newborns. *Current Biology*, *28*(22), R1294–R1295. https://doi.org/10.1016/j.cub.2018.10.004

Otsuka, Y., Kanazawa, S., & Yamaguchi, M. K. (2004). The effect of support ratio on infants' perception of illusory contours. *Perception*, *33*(7), 807–816. https://doi.org/10.1068/p5129

Otsuka, Y., & Yamaguchi, M. K. (2003). Infants' perception of illusory contours in static and moving figures. *Journal of Experimental Child Psychology*, *86*(3), 244–251. https://doi.org/10.1016/S0022-0965(03)00126-7

Parsons, J. P., Bedford, R., Jones, E. J. H., Charman, T., Johnson, M. H., & Gliga, T. (2019). Gaze following and attention to objects in infants at familial risk for ASD. *Frontiers in Psychology*, *10*: 1799. https://doi.org/10.3389/fpsyg.2019.01799

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*(5406), 1272–1273. https://doi.org/10.1126/science.283.5406.1272

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *296*(5571), 1321–1323. https://doi.org/10.1126/science.1070223

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237–247. https://doi.org/10.1016/S0163-6383(99)00003-X

Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, *81*(1), 93–115. https://doi.org/10.1006/jecp.2001.2644

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196. https://doi.org/10.1111/1467-7687.00271

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, *3*: 23. https://doi.org/10.3389/fpsyg.2012.00023

Piaget, J. (1954). *The construction of reality in the child*. Basic Books. https://doi.org/10.1037/11168-000

Pouget, A., Deneve, S., & Duhamel, J.-R. (2004). A computational neural theory of multisensory spatial representations. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 122–140). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524861.003.0006

Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology*, *30*(3), 436–442. https://doi.org/10.1037/0012-1649.30.3.436

Quinn, P. C., Slater, A. M., Brown, E., & Hayes, R. A. (2001). Developmental change in form categorization in early infancy. *British Journal of Developmental Psychology*, *19*(2), 207–218. https://doi.org/10.1348/026151001166038

Quinn, P. C., Uttley, L., Lee, K., Gibson, A., Smith, M., Slater, A. M., & Pascalis, O. (2008). Infant preference for female faces occurs for same- but not other-race faces. *Journal of Neuropsychology*, *2*(1), 15–26. https://doi.org/10.1348/174866407X231029

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*(5), 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

Radeau, M., & Colin, C. (1999). The role of spatial separation on ventriloquism and McGurk illusions. *Eurospeech*, 1–4. http://www.ulb.ac.be/rech/inventaire/chercheurs/7/CH1297.html

Rand, K., & Lahav, A. (2014). Maternal sounds elicit lower heart rate in preterm newborns in the first month of life. *Early Human Development*, *90*(10), 679–683. https://doi.org/10.1016/j.earlhumdev.2014.07.016

Reich, S. (1978). Section I. In *Music for 18 Musicians*. ECM.

Richardson, D. C., & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General*, *133*(1), 46–62. https://doi.org/10.1037/0096-3445.133.1.46

Richoz, A.-R., Quinn, P. C., Hillairet de Boisferon, A., Berger, C., Loevenbruck, H., Lewkowicz, D. J., Lee, K., Dole, M., Caldara, R., & Pascalis, O. (2017). Audio-visual perception of gender by infants emerges earlier for adult-directed speech. *PLoS ONE*, *12*(1): e0169325. https://doi.org/10.1371/journal.pone.0169325

Robinson, C. W., & Sloutsky, V. M. (2007a). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, *11*(3), 233–253. https://doi.org/10.1111/j.1532-7078.2007.tb00225.x

Robinson, C. W., & Sloutsky, V. M. (2007b). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*, *10*(6), 734–740. https://doi.org/10.1111/j.1467-7687.2007.00627.x

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. *Developmental Science*, *11*(6), 869–881. https://doi.org/10.1111/j.1467-7687.2008.00751.x

Rohe, T., Ehlis, A.-C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications*, *10*: 1907. https://doi.org/10.1038/s41467-019-09664-2

Rohe, T., & Noppeney, U. (2015a). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, *15*(5), 1–16. https://doi.org/10.1167/15.5.22

Rohe, T., & Noppeney, U. (2015b). Cortical hierarchies perform bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073. https://doi.org/10.1371/journal.pbio.1002073

Ruffman, T., Slade, L., Sandino, J. C., & Fletcher, A. (2005). Are A-not-B errors caused by a belief about object location? *Child Development*, *76*(1), 122–136. https://doi.org/10.1111/j.1467-8624.2005.00834.x

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*(1), 141–145. https://doi.org/10.1016/0304-3940(91)90914-F

Sann, C., & Streri, A. (2007). Perception of object shape and texture in human newborns: Evidence from cross-modal transfer tasks. *Developmental Science*, *10*(3), 399–410. https://doi.org/10.1111/j.1467-7687.2007.00593.x

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science*, *6*(3), 233–241. https://doi.org/10.1111/1467-7687.00276

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755–759. https://doi.org/10.1017/S0048577298980714

Schröter, H., Frei, L. S., Ulrich, R., & Miller, J. (2009). The auditory redundant signals effect: An influence of number of stimuli or number of percepts? *Attention, Perception, & Psychophysics*, *71*(6), 1375–1384. https://doi.org/10.3758/APP.71.6.1375

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*(6614), 308–308. https://doi.org/10.1038/385308a0

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Shams, L., Kamitani, Y., & Shimojo, S. (2004). Modulations of visual perception by sound. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 27–33). MIT Press.

Shaw, K., Baart, M., Depowski, N., & Bortfeld, H. (2015). Infants' preference for native audiovisual speech dissociated from congruency preference. *PLoS*

ONES ONE, *10*(4): e0126059. https://doi.org/10.1371/journal.pone.0126059

Shimojo, S., Scheier, C., Nijhawan, R., Shams, L., Kamitani, Y., & Watanabe, K. (2001). Beyond perceptual modality: Auditory effects on visual perception. *Acoustical Science and Technology*, *22*(2), 61–67. https://doi.org/10.1250/ast.22.61

Shinskey, J. L. (2017). Sound effects: Multimodal input helps infants find displaced objects. *British Journal of Developmental Psychology*, *35*(3), 317–333. https://doi.org/10.1111/bjdp.12165

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Linguistics and Language Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Slater, A. (2003). Bouncing or streaming? A commentary on Scheier, Lewkowicz and Shimojo. *Developmental Science*, *6*(3), 242–242. https://doi.org/10.1111/1467-7687.00277

Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory−visual pairings. *Developmental Science*, *2*(3), 333–338. https://doi.org/10.1111/1467-7687.00079

Slaughter, V., & Suddendorf, T. (2007). Participant loss due to “fussiness” in infant visual paradigms: A review of the last 20 years. *Infant Behavior and Development*, *30*(3), 505–514. https://doi.org/10.1016/j.infbeh.2006.12.006

Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review*, *106*(2), 235–260. https://doi.org/10.1037/0033-295X.106.2.235

Smith, N. A., Folland, N. A., Martinez, D. M., & Trainor, L. J. (2017). Multisensory object perception in infancy: 4-month-olds perceive a mistuned harmonic as a separate auditory and visual object. *Cognition*, *164*, 1–7. https://doi.org/10.1016/j.cognition.2017.01.016

Song, J.-H., Rafal, R. D., & McPeek, R. M. (2011). Deficits in reach target selection during inactivation of the midbrain superior colliculus. *Proceedings of the National Academy of Sciences*, *108*(51), E1433–E1440. https://doi.org/10.1073/pnas.1109656108

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research*, *14*(1), 139–146. https://doi.org/10.1016/S0926-6410(02)00068-X

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605–632. https://doi.org/10.1037/0033-295X.99.4.605

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, *13*(2), 113–142. https://doi.org/10.1111/j.2044-835X.1995.tb00669.x

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Spelke, E. S., & Owsley, C. J. (1979). Intermodal exploration and knowledge in infancy. *Infant Behavior and Development*, *2*(1), 13–27. https://doi.org/10.1016/S0163-6383(79)80004-1

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, *1296*(1), 1–19. https://doi.org/10.1111/nyas.12121

Spence, C., & Driver, J. (Eds.) (2004). *Crossmodal space and crossmodal attention*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524861.001.0001

Spence, C., & McDonald, J. (2004). The cross-modal consequences of the exogenous spatial orienting of attention. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 3–25). MIT Press.

Stein, B. E. (2012a). Early experience affects the development of multisensory integration in single neurons of the superior colliculus. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 589–606). MIT Press.

Stein, B. E. (Ed.) (2012b). *The new handbook of multisensory processing*. MIT Press.

Stein, B. E., Burr, D., Constantinidis, C., Laurienti, P. J., Meredith, M. A., Perrault Jr, T. J., Ramachandran, R., Röder, B., Rowland, B. A., Sathian, K., Schroeder, C. E., Shams, L., Stanford, T. R., Wallace, M. T., Yu, L., & Lewkowicz, D. J. (2010). Semantic confusion regarding the development of multisensory integration: a practical solution. *European Journal of Neuroscience*, *31*(10), 1713–1720. https://doi.org/10.1111/j.1460-9568.2010.07206.x

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. MIT Press.

Stein, B. E., Stanford, T. R., Wallace, M. T., Vaughan, J. W., & Jiang, W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. In C.

Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 25–50). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524861.003.0002

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why Is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126. https://doi.org/10.1177/0963721414522813

Tham, D. S. Y., Rees, A., Bremner, J. G., Slater, A., & Johnson, S. (2019). Auditory information for spatial location and pitch–height correspondence support young infants' perception of object persistence. *Journal of Experimental Child Psychology*, *178*, 341–351. https://doi.org/10.1016/j.jecp.2018.05.017

Thomas, R. L., 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. https://doi.org/10.1016/j.jecp.2017.04.020

Townsend, J. T., & Nozawa, G. (1997). Serial exhaustive models can violate the race model inequality: Implications for architecture and capacity. *Psychological Review*, *104*(3), 595–602. https://doi.org/10.1037/0033-295X.104.3.595

Trevethan, R. (2017). Intraclass correlation coefficients: clearing the air, extending some cautions, and making some requests. *Health Services and Outcomes Research Methodology*, *17*(2), 127–143. https://doi.org/10.1007/s10742-016-0156-6

Ujiie, Y., Kanazawa, S., & Yamaguchi, M. K. (2020). Development of the multisensory perception of water in infancy. *Journal of Vision*, *20*(8): 5. https://doi.org/10.1167/jov.20.8.5

Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision*, *8*(9): 14. https://doi.org/10.1167/8.9.14

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, *69*(5), 744–756. https://doi.org/10.3758/BF03193776

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*(1), 12–23. https://doi.org/10.1016/j.actpsy.2006.12.002

Vatakis, A., & Spence, C. (2010). Audiovisual temporal integration for complex speech, object-action, animal call, and musical stimuli. In M. J. Naumer & J. Kaiser (Eds.), *Multisensory Object Perception in the Primate Brain* (pp. 95–121). Springer New York. https://doi.org/10.1007/978-1-4419-5615-6_7

Wachs, T. D., & Smitherman, C. H. (1985). Infant temperament and subject loss in a habituation procedure. *Child Development*, *56*(4), 861–867. https://doi.org/10.2307/1130098

Wada, Y., Shirai, N., Otsuka, Y., Midorikawa, A., Kanazawa, S., Dan, I., & Yamaguchi, M. K. (2009). Sound enhances detection of visual target during infancy: A study using illusory contours. *Journal of Experimental Child Psychology*, *102*(3), 315–322. https://doi.org/10.1016/j.jecp.2008.07.002

Walker-Andrews, A. S. (1994). Taxonomy for intermodal relations. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 39–56). Lawrence Erlbaum Associates, Inc.

Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, *3*(2), 55–75. https://doi.org/10.1207/s15326969eco0302_1

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*(1), 21–25. https://doi.org/10.1177/0956797609354734

Wallace, M. T., & Stein, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *Journal of Neuroscience*, *21*(22), 8886–8894. https://doi.org/10.1523/jneurosci.21-22-08886.2001

Wallace, Mark T., & Stein, B. E. (1996). Sensory organization of the superior colliculus in cat and monkey. *Progress in Brain Research*, *112*, 301–311. https://doi.org/10.1016/S0079-6123(08)63337-3

Wallace, Mark T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience*, *17*(7), 2429–2444. https://doi.org/10.1523/jneurosci.17-07-02429.1997

Wang, S., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition*, *93*(3), 167–198. https://doi.org/10.1016/j.cognition.2003.09.012

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wheeler, A., Anzures, G., Quinn, P. C., Pascalis, O., Omrin, D. S., & Lee, K. (2011). Caucasian infants scan own- and other-race faces differently. *PLoS ONE*, *6*(4): e18621. https://doi.org/10.1371/journal.pone.0018621

Whitchurch, E. A., & Takahashi, T. T. (2006). Combined auditory and visual stimuli facilitate head saccades in the barn owl (tyto alba). *Journal of Neurophysiology*, *96*(2), 730–745. https://doi.org/10.1152/jn.00072.2006

Wilcox, T. (1999). Object individuation: infants' use of shape, size, pattern, and color. *Cognition*, *72*(2), 125–166. https://doi.org/10.1016/S0010-0277(99)00035-9

Wilcox, T., & Baillargeon, R. (1998a). Object individuation in young infants: Further evidence with an event-monitoring paradigm. *Developmental Science*, *1*(1), 127–142. https://doi.org/10.1111/1467-7687.00019

Wilcox, T., & Baillargeon, R. (1998b). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology*, *37*(2), 97–155. https://doi.org/10.1006/cogp.1998.0690

Wilcox, T., & Chapa, C. (2004). Priming infants to attend to color and pattern information in an individuation task. *Cognition*, *90*(3), 265–302. https://doi.org/10.1016/S0010-0277(03)00147-1

Wilcox, T., Smith, T., & Woods, R. (2011). Priming infants to use pattern information in an object individuation task: The role of comparison. *Developmental Psychology*, *47*(3), 886–897. https://doi.org/10.1037/a0021792

Xiao, W. S., Xiao, N. G., Quinn, P. C., Anzures, G., & Lee, K. (2013). Development of face scanning for own- and other-race faces in infancy. *International Journal of Behavioral Development*, *37*(2), 100–105. https://doi.org/10.1177/0165025412467584

Xu, F., & Carey, S. (1996). Infants' Metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*(2), 111–153. https://doi.org/10.1006/cogp.1996.0005

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*(6), 1574–1583. https://doi.org/10.2307/1130476