Understanding the multimodal integration of emotional information during development

Peiwen, Yeh, M.Sc.

This thesis is submitted for the degree of Doctor of Philosophy

Department of Psychology

Lancaster University

July 2017

Declaration

This thesis is my own work and no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification at this or any other institute of learning.

Peiwen Yeh 20th, July, 2017

Acknowledgment

The work presented in the thesis cannot be accomplished without support from numerous people including my supervisors, friends, colleagues, and my families. I would like to express many thanks for all the people who made the thesis possible and an unforgettable experience for me.

I would first like to wholeheartedly thank my supervisor Professor Vincent Reid, who kept me to study my Ph.D. at Lancaster University. He gave me the opportunities to development my research, obtain experimental skills and present my work. He also ensured that I balance my life between research and others. I also acknowledge my co-supervisor, Dr. Elena Geauga for contributing her expertise to my work and gaining my knowledge of research skills.

I would also like to thank others for supporting my research and my life in Lancaster. Thanks members at Lancaster University Babylab: Katharina, Kirsty, Uschi, Aine, Barrie, Diana, and Alison, for assisting my work including recruiting participants and doing experiments. Thanks to Katharina for precise guidance and comments that allows me to accurately learn about infant experiments. My time in Lancaster was also enriched by kindly friends in the Department of Psychology, Claire, Steven, Liam, Francescaki, Christian, Irina, James, Han and others. I am particularly indebted to Claire as she was always willing to help my difficulties in life.

I extend my appreciation to my Taiwanese friends, Wenju, Ming, Meiyin, Claude and Ya-Ning, for their companionship and encouragement that kept me going strong through the Ph.D. in Lancaster. I am also thankful to my flatmate, Alison for her understanding and caring when I was in the difficult timing of writing thesis.

Many thanks to the financial support from the William Ritchie and Friend's

Programme and Graduate School for giving me opportunities for attending workshops and conferences.

Finally, I would express my deepest gratitude to my parents and sisters for their continuous encouragement and love throughout the process of applying, starting until completing my Ph.D. Also special thanks goes to my previous supervisor, Dr. Chia-Ying, Lee, who inspired me to peruse my goal of researching. It was all of them to help me to accomplish my dream of studying my Ph.D. in UK.

Peiwen
Lancaster University
July, 2017

Abstract

In recent years, body expressions have been demonstrated to be effective visual cues for conveying emotional information. We usually perceive others' emotions from multiple modal sources in our daily lives, such as via the face, sounds and touch. As such, it is an important issue that we seek to understand how we perceive emotional cues as a coherent percept rather than separate percepts. With behavioral measurements, previous studies have provided evidence that a combination of multiple emotional cues can assist in making a more accurate and rapid discrimination of emotional contents. However, little to no research has focused on the integration of emotion perception from body expressions combined with other modal information, especially during development. As a consequence, the aim of this thesis was to investigate developmental changes in neural activity underlying the integration of emotion perception via body expressions and the voice. In Chapter 1, literature on multisensory processing in infants and children was reviewed, and the objectives of the thesis were described. In Chapter 2, processing for unisensory (sounds or body expressions) compared to audiovisual conditions (body expressions with sounds) was measured in adults. In Chapter 3, influences of types of body presentations (dynamic/static) with emotions were examined in an audiovisual paradigm with adults. In Chapter 4, 6.5-month-old infants processed emotional information in a paradigm derived from Chapter 3. In Chapter 5, audiovisual emotion perception was examined in 5-6 year-old children. These studies showed separate processing for interactions between visual and auditory perceptual sources, and for the assessment of combined emotional content across the three ages groups. This series of studies also revealed maturational changes in neural correlates to audiovisual emotion processing in ERP components indexing perception and cognition. A final chapter explores the implications of the findings for understanding the audiovisual emotion processing from a developmental perspective

Contents

List of Figures

List of Tables

Chapter 1 General Introduction

1.1. Theories of multisensory perception in early development

In daily life, we are often exposed to a wide array of information that simultaneously arrives to our different senses. The information from one sensory modality often automatically influences, or interacts with, processing of relevant information from other sensory modalities. An illustration of this is the McGurk effect, which is an illusion that reflects the binding of visual and auditory speech elements (McGurk & MacDonald, 1976). While attending to an auditory syllable /ba/, the listeners usually perceive it as /ga/ when the visual /da/ is presented simultaneously. The multisensory experience can sensitize our perceptions and allows us to react more quickly when compared to information from unimodal sources. In essence, it is challenging to process information via distinct senses, but this is easily taken for granted by experienced perceivers (Bahrick & Lickliter, 2012). It is because we selectively attend information that is relevant to our requirements and expectations, and simultaneously, ignore those that are relatively irrelevant. In order to perceive fluently, our brain also need to cohere the selected information into a unified percept rather than process separate percepts. An increasing number of studies with adults showed the integration or interaction of multisensory processing at the neural level; however, questions about how this is associated with multisensory information in early life largely remain unexplored. Convergent evidence from behavioural and neuropsychological studies highlights the importance of multisensory perception in language, attention, emotion, and other cognitive functions (see Murray, Lewkowicz, Amedi, & Wallace, 2016, for a review). Therefore, it is essential to explore how the brain determines which aspects of multisensory information are required and unified in a representation that then contribute to early development.

The development of multisensory perception begins before birth (Lewkowicz, 2014), but studies in humans have almost exclusively examined the postnatal years. A number of behavioural studies have shown rapid emergence of the ability to detect relationships among multisensory information during infancy (e.g. Bahrick & Lickliter, 2000; Flom & Bahrick, 2007). With little experience of the world, infants at birth are supposed to perceive multisensory information based on simple cues from each sensory system, rather than specific multisensory attributes of objects or events. Therefore, it remains to be investigated how multisensory relationships are constructed and perceived into definite multisensory attributes of objects or events in early development. Here, we review two theories that specify the current knowledge about the development of multisensory perception. One is the ***multisensory perceptual narrowing*** (Lewkowicz, 2014; Lewkowicz & Ghazanfar, 2009) which describes the process of increasing perceptual differentiation and tuning into specific patterns for objects or events as a result of experiencing statistical regularities in the environment. This process boosts the occurrence of expertise for multisensory information that is frequently present in one's environment, and may degrade sensitivities to ones that is less present. Recently, Murray et al. (2016) proposed a schema of three developmental stages of perceptual narrowing for multisensory systems: immature, broadly tuned, and narrowly tuned (see **Figure 1**). In these schematic representations, auditory and visual stimulus parameter are symbolized as red and blue geometric shapes, respectively. The corresponding shapes refer to features of the same object. The curves illustrate the tuning profiles of neural populations, and the turning function for an exemplar stimulus parameter is highlighted. The right side of the schemas describes putative responses to concurrent presentation of a given auditory and visual parameter. (A) At an immature age, neural tuning is very broad and widely responds to physical stimulus. The neural circuits for each modal source of information contains shared statistical features (e.g., spatial or temporal relationships) which are likely to converge. However, at this age multisensory integration (interaction) does not happen. (B) During an intermediate stage, neural tuning becomes narrow and multisensory interactions may be observed. Perceptual interaction can occur for a broader range of stimulus attributes than seen at later stages. In the first two stages, multisensory representations for low-level physical stimuli, experienced with statistical multisensory inputs and constructed from general-category, become more restrictive and specific. (C) During the last stage, the neural tuning increasingly becomes narrow and specialized for complex behaviours and experience. The integration takes place when stimulus attributes shared across the modalities are present concurrently; however, no integration occurs with unshared attributes are paired. Overall, this schematic illustration describes the process of neural tuning for multisensory perception. It is important to note, though, that a dynamic shift between stages (B) and (C) can occur with learning experience and task contingencies. In addition, perceptual narrowing occurs in each multisensory circuit depending on the attributes of the information, the modality involved and the rate of neural maturation.

Evidence for the *multisensory perceptual narrowing* has been provided by studies in different domains of perception. For example, Lewkowicz and Ghazanfar (2006) examined the ability to recognize non-human species' faces in 4- to 10-month-old infants by presenting two side-by-side rhesus monkeys' facial gestures producing a coo call and a grunt call. Each face was also presented with or without sounds while the looking times were measured. Infants younger than 6-months preferred to look at the face that corresponded with a congruent matched sound; however, there was no difference in the looking time for any of the faces when the

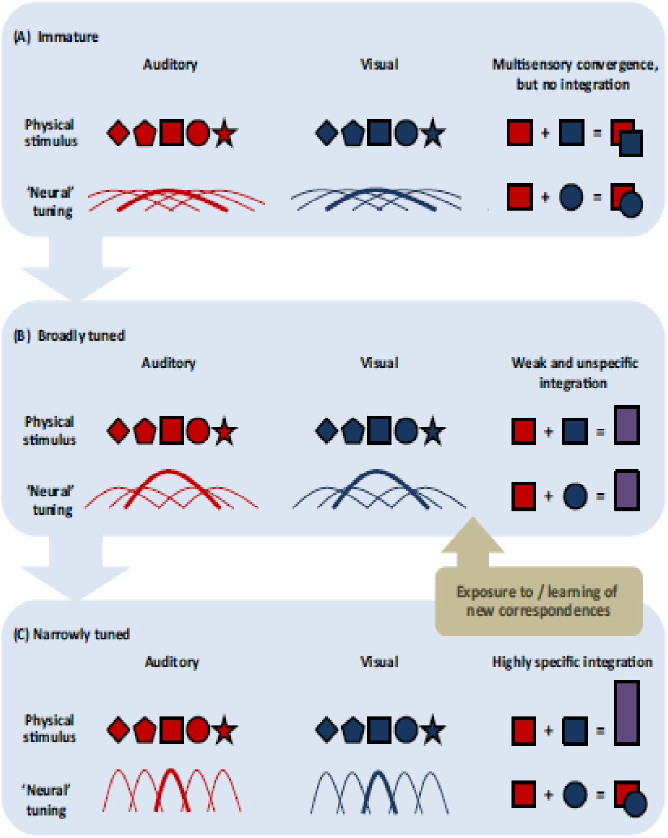


Figure 1.1. Three developmental stages of perceptual narrowing for multisensory integration (from Murray, et. al., 2016)

faces were presented without sounds. In contrast, 8- and 10-month-old infants did not show a preference for the face paired with the congruent sounds. As these older infants can still easily discriminate between auditory stimuli and between visual stimuli (Lewkowicz, Sowinski, & Place, 2008), the decline in detecting the multisensory matches in another specie was unlikely to be related to deficits in unisensory perception. Perceptual narrowing also can be seen in the language domain. Pons, Lewkowicz, Soto-Faraco, and Sebastian-Galles (2009) observed that both English and Spanish-learning infants at 6-months can successfully match visual speech to auditory English-syllables (/ba/ versus /va/) that are not present in Spanish speech. At 11-month-olds, the ability can still be observed in the English-learning infants but not in the Spanish-learning infants. Consequently, the perceptual system is likely to be a domain-general tuning process. The system arguably tunes widely to the presentation of multisensory information, which is both typical and non-typical at birth, but this then narrows as one is exposed more to their environment towards the end of the first postnatal year.

Another hypothesis, the **Intersensory Redundancy Hypothesis (IRH)**, emphasizes attentional processes underlying the development of multisensory perception (Bahrick & Lickliter, 2012; Bahrick & Lickliter, 2000). *Intersensory redundancy* refers to the fact that certain types of amodal information present in spatially and temporally synchronous events (e.g. rhythm, tempo, duration) is perceived simultaneously across different senses. According to the IRH, the redundant information is quickly built into salient amodal attributes that guides attention and other cognitive processes during early development. This process can also be adjusted by the perception of non-redundant and modality-specific information (information specific to one particular sense) (Bahrick, Lickliter, Castellanos, & Todd, 2015). For example, the sights and sounds of a ball bouncing are simultaneously detected through their temporal rhythm and tempo across visual and auditory modalities. The synchronized stimulation across the two modalities is considered the same amodal information, separating it from other events that do not share the same properties (duration, tempo, rhythm).

The most fundamental principle of the IRH is **intersensory facilitation**, which states that the amodal properties are more salient and detectable in bimodal synchronous stimulation when contrasted with the same amodal properties presented through just one sense. For example, young infants were likely to detect changes in the tempo of a toy hammer tapping when videos and sounds were synchronously presented, but not when they experienced rhythm in one modality alone or accompanied with temporally asynchronous visual information (Bahrick, Flom, & Lickliter, 2002; Bahrick & Lickliter, 2000). Intersensory redundancy also facilitates the perception of socially related events. Flom and Bahrick (2007) habituated 3- to 7-month-old infants to dynamic films of females portraying happy, sad, or angry faces with/without emotionally matched speech. When the stimulation was presented bi-modally in the habituation phase, the ability to detect emotional changes was evident even in 4-months-old infants. In contrast, when only unisensory information was presented, the ability to discriminate between emotional expressions was present above 5-months-olds for auditory information and only in 7-months-olds for visual-only presentations. Taken together, intersensory redundancy promotes the saliency of amodal properties presented across multiple modalities when compared with the same properties presented in one modality. This has importance for the guiding of attention during multisensory processing that contributes to social, emotional, and language learning in early development, including face discrimination, sequence detection and word comprehension (see Bahrick & Lickliter, 2012, for a review).

As noted above, the two prominent hypotheses, with a wealth of behavioural evidence, have provided frameworks on how multisensory information guides and shapes perceptual and cognitive learning in early development. While *multisensory perceptual narrowing* states that the process for binding occurs via neural narrowing and tuning to the native environment, the *IRH* emphasizes the importance of attentional allocation to amodal properties. Despite the different perspectives, both theories highlight the importance of multisensory learning during infancy. Through experiencing statistical regularities in the environment, infants develop effective patterns of perceiving multisensory relationships. These patterns gradually allow them to be experienced perceivers, preferentially processing the unified multimodal attributes. However, a number of crucial questions about how this development takes place remain largely unanswered. For example, how does the process underlying each modal attribute bind into amodal attributes in the first postnatal year? How do other factors, such as attention, influence the neural tuning to multisensory perception with specific attributes? In addition, questions about development beyond infancy also remain largely unexplored but worthy of further investigation. Both theories agree that multisensory processes are plastic and dynamic, thereby explaining the improvement in perception with experience. It is plausible that a developmental reweighting may occur during maturational progresses, like low-level physical features which initially weight more, while later increasingly more complicate and unified attributes are prioritized. This idea is supported by studies which have shown that behaviors benefit from multisensory cues, but that the advantage changed over childhood (e.g. Bair, Kiemel, Jeka, & Clark, 2007; Gil, Hattouti, & Laval, 2016), suggesting changes in the processing strategies or reweighting to multisensory information in children. It is also little understood how these patterns turn into the expert processing seen during adulthood. In order to address these highly relevant questions, evidence about processing at neural level across development is also needed. In the following section, we review neurophysiological findings which help us understand the mechanism underlying multisensory integration.

1.1.2. Neural evidence for multisensory perceptions

A traditional view of multisensory perceptions is that they take place at high-level associative cortical regions such as the premotor cortices and sensorimotor subcortical regions. However, this perspective has been recently reformulated since an increasing number of neuroimaging and electrophysiological studies indicate that sensory systems have the capacity to influence one another, even at very early processing stages (Murray et al., 2016). The neural circuits for multisensory interactions are not entirely determined by low-level factors, like physical features of the stimuli themselves (e.g. intensity, shape, location). The higher-level processes related to task demands, attention or semantic could also influence multisensory processing. In that case, multisensory perception is processed by the dynamic interplay between low-level and high-level factors.

To comprehend how the brain selects and integrates relevant information across time and space into a coherent percept, electroencephalography (EEG) and event-related potentials (ERPs) are optimal ways of investigating the neural processing for the multisensory perception. Due to high temporal resolution, these techniques, particularly ERPs, can rapidly record changes in the neural activity, improving our understanding about the time course of processing stages between a stimulus and a response (Steven, 2005). Moreover, EEG/ERPs do not necessarily involve complex tasks or a covert behavioural response to stimulus or instructions, so they are well suited to research on developing and clinical populations (DeBoer, Scott, & Nelson, 2007). With the same methodology, the EEG/ERP applied to a wide age range of participants can explore the maturation of the neural processing strategies. Below we briefly summarize how recent work utilising EEG/ERPs investigating multisensory processing in adults, infants and young children, have illuminated our understanding of the parameters of multisensory processes.

ERP research with adults has shown that that the interactions between multisensory information occur during the early processing stages. The integration effects are typically measured by quantifying the responses to multisensory stimulation compared to the sum of the responses to each unisensory condition (e.g.

Besle, Fort, Delpuech, & Giard, 2004; Giard & Peronnet, 1999; Molholm et al., 2002). According to the assumption that the unisensory information is processed independently, the bimodal responses are supposed to equal the sum of the unisensory responses. If the multisensory response differs, being either reduced (supra-additive effect) or increased (super-additive effect), from the sum of the unisensory responses, this may reflect the timing for the integration or interaction between multisensory processing. To date, many audiovisual studies have shown that the auditory N1 (or N100, a negative peak occurring after onset of sounds) is often attenuated and speeded up in an audiovisual condition compared to the sum of each unisensory condition in speech (Besle et al., 2004), emotion (Jessen & Kotz, 2011) and other perceptual domains (Stekelenburg & Vroomen, 2007). This implies that visual and auditory perception interact at an early stage of auditory sensory processing. The attenuation may occur as the visual stimulus provides accurate predictions for evaluating the following auditory information, so less attentional or other cognitive resources are demanded for auditory processing (van Wassenhove, Grant, & Poeppel, 2005).

In terms of developmental evidence, studies with infants prefer to observe their neural activities for congruency in certain features of objects or events across modalities, such as emotional expressions, synchronized timing or speech content (Grossmann, Striano, & Friederici, 2006; Hyde, Jones, Flom, & Porter, 2011; Reynolds, Bahrick, Lickliter, & Guy, 2014). The ERP component, Nc (the negative central component), a negative deflection peaking around 400 ms post-stimulus, is often taken as an index for congruency effects in infants. The Nc is considered to reflect visual attention allocation in infants, recording greater negative responses to salient or infrequent stimuli (Ackles & Cook, 1998; de Hann & Nelson, 1999). Thus, the Nc elicited during the multisensory processing may reflect abilities to detect violation of known concepts or unexpected information across modalities. However, the Nc is associated with attention processing, which belongs to cognitive processing rather than a sensory processing. As such, the congruency paradigm may not be the most suitable for exploring whether the integration occurs at an early stage of processing.

Comparatively, a few studies have assessed the occurrence of multisensory integration by comparing responses to unimodal and multiple modalities during infancy (Reynolds et al., 2014) and early childhood (Brandwein et al., 2011; Knowland, Mercure, Karmiloff-Smith, Dick, & Thomas, 2014). Although these studies provide valuable information which indicates that perceptual interactions occur at the sensory levels, different important aspects related to the recording and the analysis of the developmental ERP data need to be considered. Since the synaptic density, neuronal alignment and other maturational processes change throughout infancy, childhood and adolescence, the auditory ERPs waveforms greatly vary across development until adulthood. For example, adults' auditory ERPs typically show a negative peak (N1, ~ 100 ms after onset of sounds) and then transit to a positive response (P2, ~ 200 ms). In contrast, early in infancy, the auditory ERPs mostly show a broad positive deflection by 250 ms from stimulus onset, followed by a negativity (**Figure 1.2.**) (see Coch & Gullick, 2011, for a review). After the age of 4 years, an adult-like N1 gradually emerges, and other surrounding components, such as P1 and P2, decrease in latency and amplitude. Nevertheless, these maturational changes are nonlinear (**Figure 1.3.**) (e.g. Ponton, Eggermont, Kwong, & Don, 2000). In addition, great differences in individual ERPs can be observed in terms of latencies, distributions and polarities. For instance, while some show a positive deflection, a negative waveform is found in others during the same period of time (Trainor, 2007). As such, the grand average patterns might become flat which might make the evaluation of statistically significant ERP regions difficult.

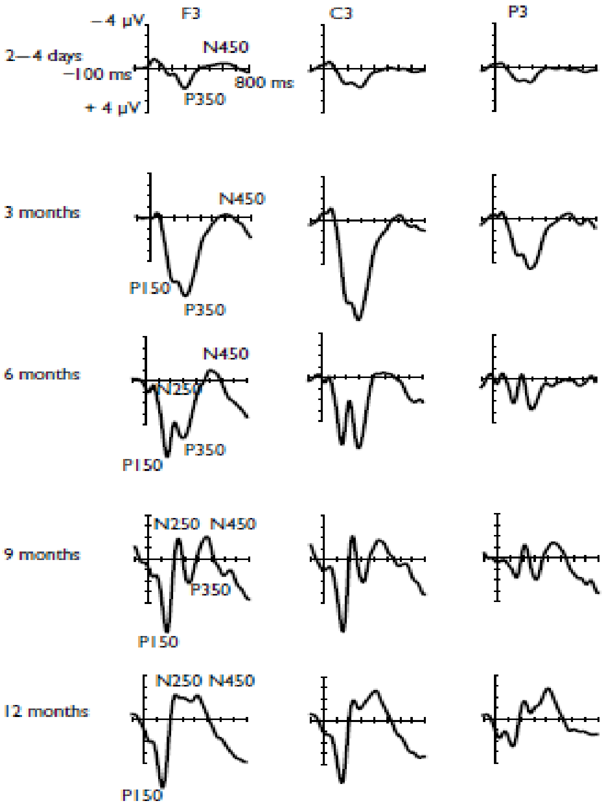


Figure 1.2. The grand average ERP to tones at birth and 3, 6, 9 and 12 months of age (from Kushnerenko et al., 2002)

Taken together, electrophysiological methods have provided new insights into multisensory perception by showing processes at perceptual and cognitive levels. Particularly, EEG/ERPs do not require behavioural responses, so it is a practical tool to explore the neural mechanism underlying behaviour in infants and children. Despite of great variation in brain maturation, an increasing number of studies using cross-sectional and longitudinal methods revealed changes in the morphology of the auditory ERPs from infancy across childhood (e.g. Kushnerenko et al., 2002; Ponton et al., 2000; Shafer, Yu, & Wagner, 2015; Sussman, Stemschneider, Gumenyuk, Grushko, & Lawson, 2008). These findings could allow us to more precisely identify the responses to specific information, and allow us to understand the developmental changes that occur for certain processed.

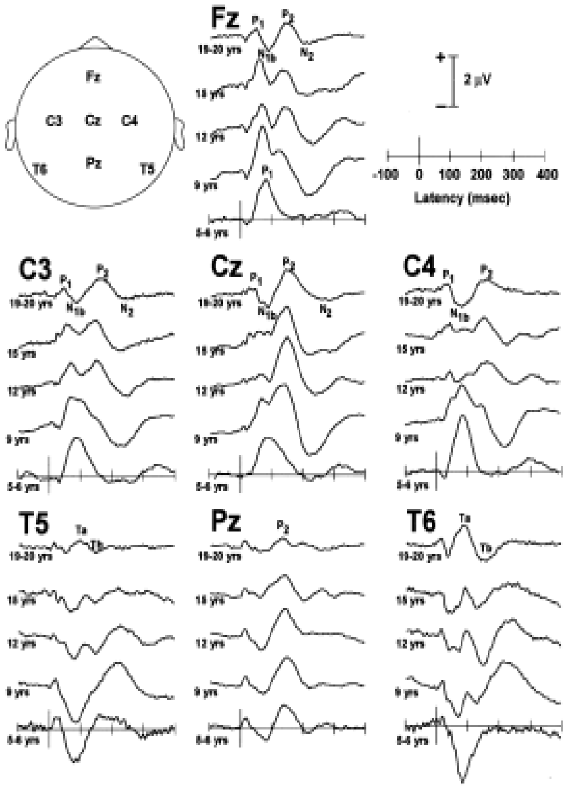


Figure 1.3. The grand average of ERP responses to sounds from 5- to 20-year-old (from Ponton et. al, 2000)

1.2. Emotion

1.2.1. Current studies on multisensory processing of Emotion Perception

Neuroimaging studies with adults have identified the processing routes for audiovisual emotional perception. Based on these findings, Symons, El-Deredy, Schwartze, and Kotz (2016) proposed three stages of specialized emotion processing across modalities: detection, integration and evaluation. During the detection stage, the salient emotion signals are processed, including an early perceptual level, which is the traditional view of modality-specific processing. For example, the visual detection of emotion occurs in the region of the occipito-temporal cortex, fusiform gyrus, with other sub-cortical regions specifically related to facial or body expressions. The specialized processing of auditory stimuli was found in the primary auditory cortex and temporal lobe. During the next stage, the extracted low-level visual or other physical features are precisely processed and might converge within the superior temporal sulcus (STS). This can be evidenced by studies (Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009; Robins, Hunyadi, & Schultz, 2009) showing that the STS and the surrounding structures are sensitive to vocal or facial emotions expressions, with overlapped brains areas for audiovisual emotional information. In the last stage, the motivational value of the current content might be evaluated within the inferior frontal gyrus (IFG) and orbitofrontal cortex (OFC) as these areas have been functionally related to the processing of reward and punishment (Kringelbach & Rolls, 2004). In addition to these cortical regions, subcortical structures, such as the amygdala and basal ganglia, are also involved in emotion perception from facial, body and vocal expressions at early and late stages.

Although neuroimaging studies from adults have advanced our understanding of process in audiovisual emotional perception, it is difficult to apply these methods with young populations to explore their neural circuits for multisensory processing of emotion. Despite this, behavioural studies have indicated that understanding the association between facial and auditory emotional expressions has emerged by the age of 7-months (e.g. Soken & Pick, 1992; Walker-Andrews, 1986). Additionally, multisensory experience benefits emotion understanding in early development (e.g. Flom & Bahrick, 2007; Walker-andrews & Lennon, 1991). For example, 5-month-old infants prolonged their looking time to changes in vocal expressions when facial expressions were presented, whereas no changes for looking time were found when a checkerboard was presented with vocal expression during habituation phase (Walker-andrews & Lennon, 1991). Further, several studies using ERPs (e.g. Grossmann et al., 2006; Otte, Donkers, Braeken, & Van den Bergh, 2015) found the effects to emotional congruency across auditory and visual modalities in infants, indicating the timing of the neural processing for detecting the relationships of emotion content between the two modalities. However, infant studies usually utilise simple designs with relatively few conditions due to the infants' limited attention span. Therefore, many other relevant variables usually remain unexplored within the same individuals. Due to the great variation in brain development, the results are likely to vary as a function of task difficulties, emotion types, modal information and other factors. Thus, more research that addresses the complex interaction between these factors is needed. For example, the majority of studies so far targeted contrasts between emotions with opposite valence emotions (e.g. happiness versus anger), typically expressed visually (i.e., faces). Although such contrasts are informative, the results might be influenced by other confounding factors, such as the degree of familiarity. Compared to other types of emotion, happiness is more common in infants' environments (Walker-Andrews, 2008). Thus, infants can more easily discriminate another emotion that is greatly different from happiness in terms of valence. The following section outlines several variables have been discussed in the studies with developmental populations on audiovisual emotional perception from facial expressions and sounds.

1.2.2. Emotional Differentiation

Each emotional expression has its unique cognitive and physiological functions, which specifically supports communication and behavioural adaption to the social world. Therefore, the expression and perception of different emotions may be underpinned by distinct patterns of neural activities and connectivity. Valence is one of the widest standard way to classify emotions, with the categorization of emotions into a positive (pleasant) and negative (unpleasant) emotion (Symons et al., 2016). To date, a large body of studies from healthy adults (e.g. Canli, Desmond, Zhao, Glover, & Gabrieli, 1998; Killgore & Yurgelun-Todd, 2004) and lesion-brain patients (e.g., Adolphs, Jansari, & Tranel, 2001; Borod et al., 1998) have demonstrated that there are hemispheric asymmetries for emotion processing. Several theories have attempted to account for the lateralization of emotion perception. *Right Hemisphere* Hypothesis proposed that the right hemisphere is dominant for processing all emotions (Borod et al., 1998). Considering the valence of emotion, the *Valence-Specific* Hypothesis states that the left hemisphere is specialized for processing positive emotion, whereas the right hemisphere is specialized for processing negative emotions (Ahern & Schwartz, 1985). This hypothesis is similar to the *approach-withdrawal* hypothesis, which suggests that emotions can be categorized into approach (e.g. happy face) and withdrawal (e.g. sad face) behaviour, and are processed within left and right hemisphere, respectively (Davidson, 1992b). Since anger drives the individual to fight, it is classified into the same category as happiness and surprise (approach emotion). However, this is incongruent with the *Valence hypothesis* where anger is the opposite emotion to happiness. Despite opposing perspectives, both hypotheses are supported by large empirical evidence. At the same time, the hemispheric asymmetry in emotion processing could be influenced by tasks demands (Kotz et al., 2003; Kotz, Meyer, & Paulmann, 2006). Another perspective is that it might be partially overlapping neural connectivity that accounts for the processing of different emotions (LeDoux, 2000). This however requires more investigation in order to differentiate the subtle changes in the neural connections (Symons et al., 2016).

1.2.3. Different developmental trajectories across emotions

Humans differentiate between emotions at different levels across development, which might be associated with discrepant developmental curves for perception to each emotion. Infants by 10 months of age might rely on emotional valence or perceptual features to detect or discriminate emotions from one another (Widen & Russell, 2008). Until late in the first postnatal year, infants probably learn to extract emotional meaning from faces and voices and this aids in guiding their own behaviour and predict others behaviour. At this age, they understand other people's emotional expressions and also link the expression to external events, such as reward, punishment, whom to approach and whom to avoid. By the age of 24 months, infants and toddlers modify their behavior based on others emotional expressions and acquire emotional meaning in the events they experience. This differentiation of understanding emotions starts with discrimination and proceeding toward its meanings, or more complicated emotion type (e.g. surprising) appears in later life

Studies with young children further demonstrated that the maturational course of perception to each emotion is inconsistent. The differences in maturation are also related to modal resource. For facial expressions, happiness is easily detected in comparison to the other basic emotions (happiness, anger, fear, disgust, sadness) in early childhood, whereas sadness is the least accurately recognized (Herba, Landau, Russell, Ecker, & Phillips, 2006; Montirosso, Peverelli, Frigerio, Crespi, & Borgatti, 2010). With increasing age, a slower improvement in accuracy has also been observed for sadness and anger relative to happiness and fear. As for emotional sounds, children at the age of 5 appear to understand positive emotions, fear and anger from non-verbal vocalization; however, sadness can be well recognized if sounds contain linguistic elements (e.g. speech prosody; Sauter, Panattoni, & Happe, 2013). Further, several studies have directly compared responses to auditory and visual information, showing that visual emotions are more easily detected than auditory cues. Nelson and Russell (2011) found that pre-schoolers have the ability to recognize anger, happiness and sadness from facial expressions and body postures (> 70%), whereas their performance for vocal recognition was only high for sadness in contrast to the other emotions (< 51%). Another study by Chronaki, Hadwin, Garner, Maurage, and Sonuga-Barke (2015) observed that performance for emotion recognition from faces achieved an adult-like state by 11-years of age, but extracting emotion from voice still developed till adolescence. In addition, accuracy for both facial and vocal recognition was largely lower for sadness than for anger and happiness. Taking into account the above studies, the ability to recognise each emotion from facial and vocal expressions improves with age, but at differing speeds. Moreover, some types of emotions can be perceived earlier, particularly from the face (e.g. happiness), while others are recognized with age at a slower speed (e.g. sadness). These findings could also be influenced by other factors, such as task, stimuli characteristic (e.g. verbal and nonverbal voice), and emotional intensity.

1.2.4. Modality Dominance

Precisely, some characteristics of stimuli appear to be perceived accurately through one modality when contrasted with others, which has been termed *modality dominance* (Spence & Squire, 2003; Welch, DuttonHurt, & Warren, 1986). For instance, vision is more sensitive to spatial changes than hearing, whereas hearing has a greater influence by temporal synchronization than vision. This can be found in adult studies where modality dominance is divergent across each emotion. Paulmann and Pell (2011) have shown that the accuracies of emotion recognition were higher for anger, happiness and disgust from facial expressions, than that from voices, which implies visual expressions dominate the three emotions compared to auditory expressions. Later, Takagi, Hiramatsu, Tabei, and Tanaka (2015) discovered that attentional instruction differently modulates modality dominance for each emotion (anger, disgust, fear, happiness, sadness and surprise). Accuracies were higher for emotions of anger, disgust, happiness and surprise when attention was directed to the face than when it was directed to one's voice in audiovisual conditions, whereas voice dominance was shown for fear. The study further divided congruency effects into facilitation (unisensory versus emotionally congruent condition) and interference effects (unisensory versus incongruent condition). For anger, only a facilitation effect was found when attention was directed to the voice. Comparatively, only an interference effect achieved significance when attention was given to the face. As for fear, no facilitation or interference effects were found when either faces or sounds were attended to. Both facilitation and interference effects were observed for happiness when voice was attended to. The findings suggest that the benefits and costs from multiple modal cues disproportionally affected each type of emotion during multisensory processing. Moreover, the more dominant a modality was, the more difficult it was to ignore the modality. However, either the facilitation or interference of the modality dominance can be modulated by attention.

ERP research further provided evidence of neural processing for modality dominance during multisensory perception. For example, Ho, Schroger, and Kotz (2014) presented angry/neutral face-voice pairs and found opposite directions of congruency effects within the P200 (P2, a positive response at approximately 200ms after onset of sounds). When an angry face was presented and preceded a neutral sound (incongruent pair), the P2 amplitude was reduced compared to when both the face and corresponding sound was neutral (congruent). The congruency effect was reversed when an angry sound was presented, that is, an increased P2 was observed for an incongruent pair compared to a congruent pair. It appears that the P2 amplitude was reduced when the angry face was presented beforehand. In multisensory literature, the P2 has been explained as a competition across auditory and visual information (Knowland et al., 2014; Stekelenburg & Vroomen, 2007) or a processing for the combined emotion information (van Wassenhove et al., 2005). Based on these assumptions, the emotion of anger expressed by face is likely to carry a much stronger message than by voice, that influences processing for the holistic emotion. Moreover, the P2 patterns for the congruency effects were not significantly modulated by attention, which was opposite to the findings for auditory N1. As such, N1 and P2 might reflect functionally dissociated processes during the emotionally audiovisual perception.

In a similar vein, the modality dominance can affect multisensory processing during infancy. Grossmann et al. (2006) presented either a happy or angry face with emotionally (in)congruent prosody to 7-month-old infants. Further analyzing incongruent effects, a response was more negative for an angry face with a happy prosody contrasted when emotional information reversed across auditory and visual modalities. The author inferred that infants are usually more exposed to happy than angry expressions in their daily social interactions. Therefore, it is not surprising for infants to expect that a happy face presented with sounds correspond to the same emotion. However, presentation of an angry sound violated their expectancy, which triggered a larger changing response.

As discussed above, the modality dominance for each emotion can be modulated by attention and familiarity of emotion source. However, we have less understanding related to whether there is a developmental transition of emotion perception by emotion category and modalities. Although studies have shown that most emotions (e.g. disgust, happiness) are more easily perceived by the face or by the voice (e.g. fear) in children, these results were obtained from indirect comparisons in unisensory presentations (Chronaki et al., 2015; Nelson & Russell, 2011). It is unknown whether modality dominance for emotions alters across development, that is, the use of emotional cues might shift, with other cues predominating in childhood. A behavioral study by Gil et al. (2016) showed a degree of facial ambiguity (a continuum of an extremely happy face to an extremely sad face) with prosody to 5- 9-year-old children and adults. The response curves to categorize the emotion pairs were different between 5- to 7- and 9-year-old children. However, similar performance patterns were found in 9-year-old children and adults. As such, the authors inferred that the processes for audiovisual emotional effects develop nonlinearly between infancy and adulthood. At a certain age, there might be a developmental turning point toward adult-like strategies for using emotional cues, which leads to effective cognitive processes and social interaction.

1.2.5. Emotion perception from body expression (with) other modal information

Prior studies mostly focused on visual emotion from facial expressions. In recent decades, behavioral and neuroimaging research have demonstrated that body expressions are also crucial visual cues of conveying affective information (see de Gelder, 2009; de Gelder, de Borst, & Watson, 2015, for reviews). These studies showed similar findings between facial and body perceptions, including behavioral performance (Atkinson, Dittrich, Gemmell, & Young, 2004) and the brain regions that are involved in the recognition of body and facial expressions (van de Riet, Grezes, & de Gelder, 2009). However, processing body expressions partially differed from facial expressions. Directly compared to responses of faces, van de Riet et al. (2009) found that the perception of body expressions elicits a common (e.g. superior temporal sulcus, and fusiform gyrus) and a specific network of brain areas (e.g. intraparietal sulcus, parietal-occiptial gyrus). In addition, body expressions can more effectively convey emotional information than faces in some circumstances, for example, when people are at a long distance from each other (de Gelder, 2009).

Several studies further extended the issues towards the perception of body expressions combined with other sensory information in adults. This is a more important and naturalistic issue as we are often exposed to a combination of emotional cues in daily life. From behavioral data, Van den Stock, Righart, and de Gelder (2007) observed that a degree of emotional congruency of body-face and body-voice pairs influenced both judgment of facial and vocal emotions. This result implied that the interaction of emotion perception inevitably occurred when relevant information, such as facial, vocal or body expression, is presented from different modalities simultaneously. ERP studies advanced the understanding about the time course underlying emotion processing on body expressions and voice. Consistent with other domain multisensory studies, Jessen (Jessen & Kotz, 2011; Jessen, Obleser, & Kotz, 2012) demonstrated that the auditory N1 was reduced for an audiovisual condition compared to the sum of unisensory conditions, suggesting that the interactions of affective perception from body expressions and vocalization occur at an early sensory processing stage.

From a developmental perspective, the ability to distinguish emotions from body expressions is likely to develop by 8-months of age (Missana, Atkinson, & Grossmann, 2015; Missana, Rajhans, Atkinson, & Grossmann, 2014) Recently, a behavioral study by Zieber, Kangas, Hock, and Bhatt (2014b) found via an intermodal preference paradigm that 6.5-month-old infants looked longer at a video portraying a person with an angry body expression compared to a happy body expression when paired with an angry voice. By contrast, the infants were more likely to watch a happy video when it corresponded with a happy sound. This pioneering study disclosed that the ability to associate anger and happiness from body expressions to vocal sounds has emerged early in life. Despite this, little is known about neural activities for the integration of emotion perception related to body during early development. For example, whether the processing for audiovisual emotional perception in the developing populations is similar to the findings in adults where it occurs at a sensory processing stage? Whether there are distinct processes for each emotion across modalities? Also, it is unknown the strategy of processing emotion perception changes between infancy and adulthood. Additional work is required to understand the developmental changes in the neural mechanisms underlying audiovisual emotional perception.

Thesis objectives

The goal of the current thesis is to understand the developmental change in the integration of emotion perception from body expressions and affective sounds. In a different way to infant studies that typically use preferential looking time, we conducted ERPs studies to record changes in brain responses in infants as well as young children and adults. ERPs can non-invasively record neural activities without any required behavioural responses, which makes the technique particularly suitable for studies with developing populations. The measurement can also advance our understanding of emotion processing at an early and a late processing stage (or a sensory and a cognitive level) underlying infant or child behaviours. To observe the changes in the ERP waveform for emotion perception across development, the same paradigm was presented to the three ages groups (6.5 month-old infants, 5-6 year-old children and adults). However, the paradigm was modified due to different requirements for each age group.

There are many ways of exploring the integration of emotion processing from body expressions and sounds. For instance, a priming paradigm (Otte et al., 2015) or synchronized/ asynchronized presentations of audiovisual pairs (Hyde et al., 2011) where middle latency components are expected to reflect the process of audiovisual perception. However, studies with adults have shown that the integration of audiovisual perception occurs at an early period of perceptual processing whereby the auditory responses differed between auditory-only and audiovisual conditions at 100 ms (Besle et al., 2004; Giard & Peronnet, 1999; Jessen & Kotz, 2011; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). It would be a question as to whether audiovisual emotion perception also emerges at an early stage of perceptual processing in infants and children; therefore, we determined the paradigm comparing auditory responses between auditory-only and audiovisual contexts in this thesis. Moreover, the paradigm that was employed allows us to separately observe audiovisual perception during perceptual and cognitive processing in young populations. Most important of all, the current work from typically developing populations could provide a reference for future work on atypically developing individuals, understanding their deficits related to processing stages of audiovisual emotional perception.

In the first study (see Chapter 2), the investigation into the interaction of emotion perception from body expressions and sounds was explored in an adult sample. Conducting typical analyses in multisensory perception (e.g. Giard & Peronnet, 1999), we examined the integration (or interaction) between the auditory and visual perception by comparing auditory ERP responses in auditory-only to audiovisual conditions. In the study, 600-ms body expressions preceding the sounds reflects that the visual context is a predictor for the following auditory processing. In order to observe the influence of the visual context (Ho et al., 2014), we also considered the effect of emotional congruency across the two modalities. As such, there were four conditions in the study: auditory-only, visual-only, emotionally congruent and incongruent audiovisual conditions. We manipulated other factors as well, such as selective attention, emotion intensity and emotion types (anger and fear). The study also extended Jessen's findings (Jessen & Kotz, 2011) which showed a shorter peak latency for anger than for fear in both auditory-only and audiovisual conditions, suggesting divergent processing for the two negative emotions. Moreover, other studies indicated the influence of attention on modality dominancy for each type of emotion (Ho et al., 2014; Takagi et al., 2015). Due to many variables in the study, we used a mixed-factors design, with between subjects for selective attention, with the other factors being within. In that case, one group was instructed to attend to the people's dressing (implicit task), and another one was to attend to the vocal emotion (explicit task). The congruency effects in addition to modality dominancy were observed within two auditory ERP components, N1 and P2. The modulation by other factors was also examined during the emotion processing.

The second study (see Chapter 3) extended the issue of the first study, aiming to understand whether the integration of emotion perception is influenced by presentation of body types in adults. The idea of the study was that there might be different processing for emotion recognition from moving and static body expressions. Compared to static stimulus, dynamic stimulus contain elicit movements that may provide more information related to emotion perception. Behavioral data have shown the dynamic expressions enhanced accuracy for body emotion recognition (Atkinson et al., 2004). From neuroimaging evidence, more prominent activation areas are activated to the dynamic stimulus (Grezes, Pichon, & de Gelder, 2007; Pichon, de Gelder, & Grezes, 2008) and have accounted for understanding action as well as emotions (Iacoboni, 2005). Nevertheless, some studies suggest that each emotion might be optionally specialized to be recognized (Coulson, 2004). While sadness is usually exhibited with less movement or is even motionless, anger is characterized by a higher velocity of movements (Roether, Omlor, Christensen, & Giese, 2009b; Volkova, Mohler, Dodds, Tesch, & Bulthoff, 2014). To investigate the modulation of motion on perceptual integration, the visual stimulus displayed body expressions with (dynamic type) and without movements (static type). The auditory stimulus were the same across the blocks with the both the two types of body expressions. The intensity of both the visual stimulus and auditory stimulus were controlled. Likewise, the modality and congruency effects were expected to occur during the timing of auditory

N1 and P2. In the same latency of the two components, we also examined the modulation of visual types for anger and fear, respectively.

To understand the early development in the integration of emotion perception, the third study (see Chapter 4) focused on 6.5-month-old infants. Considering infants have a short attentional span, the paradigm in the study was shortened compared to the one presented in the adult studies (the first and second studies). Based on the adults' data, the modality and congruency effects were more pronounced for angry than fearful expressions. Therefore, infant participants were only presented to angry expressions in auditory-only, emotionally congruent and incongruent conditions. Another point we considered is about paradigm. Prior infant studies tended to compare responses to emotionally congruent and incongruent pairs for the multisensory emotion processing (e.g. Grossmann et al., 2006). However, the congruency effects elicited after 400 ms of the stimulus, which might belong to a higher cognitive processing. It seems that the congruency paradigm can reveal how the visual content modulates the sounds at a later processing stage, but cannot show whether the visual processing interacts with auditory perception at an early processing stage. Consequently, we followed the methodology in adults' studies that compared auditory responses in auditory-only to audiovisual conditions. There is also another challenge that is to define infants' auditory components. During infancy, ERPs to sounds dramatically change in terms of peak latency and response phases. The maturational changes in auditory ERP waveform are also influenced by tasks and stimulus types (e.g. Kushnerenko et al., 2002). Despite this, a growing number of longitudinal and cross-sectional studies have shown that infants' auditory components are usually dominant by a broad positive response by 250 ms, followed by a broad negative response (see Trainor, 2007, for a review). According to Kushnerenko et al.

(2002) and visual inspection, we observed the modality and congruency effects within three ERP components: two positive peaks at 150 (P150) and 350 ms after onset of sounds (P350), and a negative response at 450 ms (N450). The factor of lateralization was also calculated to understand if there is any hemisphere specialized in processing for emotional integration during infancy.

In the final study (see Chapter 5), we extended the issue of audiovisual emotion perception in typically developing 5- to 6-year-old children. The age we observed is based on Nelson and Russell (2011) which revealed that children above the age of 5-years performed with high accuracy when they labelled angry and happy body expressions (70 - 80 %). Children at this age are also likely to stably show auditory evoked potentials (P1-N2; Ponton et al., 2000; Shafer et al., 2015). Identical to infant studies, we presented angry expressions in three conditions (auditory-only condition, emotionally congruent and incongruent audiovisual conditions). To keep children participants' attention, they were also instructed to response to a non-emotional picture randomly presented after auditory stimuli. The incongruent pairs in the study were also modified into angry sounds paired with happy instead of fearful body expressions. This is because the congruency effects were expected to be more salient to the opposite valence of emotional expressions. As discussed earlier, it is difficult to define statistically significant ERPs regions in young children (Trainor, 2007). In addition, none existing studies have explored the multisensory perception with the paradigm that compares modalities in children at this age. As we primarily focused on the auditory components at a sensory processing, the data were analyzed by 400 ms after onset of the sounds. Each component of the time window was segmented based on polarities transition in auditory-only condition. Likewise, we also included the factor of lateralization for the developmental change in processing strategies.

Taken together, the series studies in the thesis aim to explore the maturational changes in integration of emotion perception from body expressions combined with sounds. Through electrophysiological approach, we could explore changes in brain activities underlying the perceptual integration in infants and children. We compared auditory ERPs in auditory-only to audiovisual modalities across three age groups, allowing us to observe the processing at a sensory and a cognitive level. This is in contrast to previous developmental studies, which use a congruency paradigm but only examined the later processing stages for multisensory perception. The current data showed the responses differed between auditory-only and audiovisual conditions as early as 200 ms in infants, implying the interaction of auditory and visual perception on emotion have occurred at a sensory processing stage early in life. Due to neural changes in early development, more developmental studies are required to assess the results and establish reliable analyses. Despite this, the present studies could provide pioneering work for future research on multisensory emotion processing in typically and atypically developing populations.

Chapter 2 Study1

A primary investigation on the integration of emotion perception s in adults

Abstract

The body is an important cue for efficiently understanding others' emotional states in the social world. However, little work has discussed multisensory emotional perception related to body expressions. The study therefore attempted to examine emotion perception of body expressions and how this combines with sounds in adults. To examine the time course of interactions between visual and auditory conditions, the ERP responses were recorded in four conditions: auditory-only, visual-only emotionally congruent and incongruent audiovisual conditions. Results (N=18) showed that the auditory N1 amplitudes were reduced in audiovisual compared to auditory-only conditions, implying that there was an interaction of the auditory and visual perceptual mechanisms which occurred at an early stage of sensory processing. Another component, the P2, was sensitive to emotional congruency across visual and auditory information. Either increased or reduced P2 responses in congruent compared to incongruent conditions, depended on how the composite pairs present across emotion type and modalities. We also observed the influence of attention and visual emotional intensity within both components. Results from another group (N=18) showed that the N1 amplitudes were more affected by attentional instruction in contrast to the P2. Overall, the study indicates that two functionally dissociated processing mechanisms are underlying N1 and P2 components. In addition, attention and emotional intensity differently modulate angry and fearful expressions, which may well be related to divergent dominance for specific modalities for each of the two emotions.

2.1. Introduction

In our daily life, the perception of others' emotions gives us a good insight into their dispositions. Reading the emotions of others allows us to anticipate suitable responses in complex dynamic social interactions. In the natural environment, we usually simultaneously detect emotional information through multisensory (e.g., faces and bodies, or via vocalizations) instead of via unisensory modal processing. In fact, a combination of multiple emotional cues can assist in making a more accurate and rapid detection and discrimination of emotional content (e.g. Collignon et al., 2008). In order to comprehend how we perceive emotions, it is consequently essential to understanding how emotional information from multiple modalities unifies into a coherent percept.

Over the last ten years, body postures have been demonstrated to be important visual cues that convey reliable emotional content, whereas relatively few studies have discussed the perception of visual cues combined with other modal information, such as via an emotional voice (de Gelder, 2006). Instead, a number of studies have focused on the audiovisual emotion perception on facial expressions with sounds (e.g., Collignon et al., 2008; Ho et al., 2014; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007). Although recognition performance of body expressions and facial expression processing are alike, the neural networks involved are still different between the two visual cues (van de Riet et al., 2009). Electrophysiological studies (EEG/ERP) also provided evidence for the different processing between emotional faces and body expressions. The processing in emotional body expressions is detected at approximately 100 ms, which is similar to the initial processing during the observation of facial expressions; however, a significant sustained frontal-central ERP response to fear compared to neutral stimuli was only found in body processing but not in face processing, suggesting a prolonged attention to bodily expressed emotions compared to facial stimuli (e.g., van Heijnsbergen, Meeren, Grezes, & de Gelder, 2007). It may even be the case that emotion processing derived from body expressions offer more information than that from face expressions in some cases. For instance, emotion signals are more readable from body movements than from faces when viewing people at a distance (de Gelder, 2006). Therefore, the processing of body expressions is an important element in understanding others' emotions in social relationships, particularly when investigating with other modal cues.

In agreement with prior studies observing multisensory effects, Van den Stock et al. (2007) demonstrated that the reaction time for emotion discrimination sped up when the emotion signals from body and voice were more emotionally congruent. Using ERP measurements, Jessen and Kotz (2011) explored the perceptual integration effects from processing information in unisensory modalities compared to audiovisual modalities. They compared responses in audiovisual conditions to the sum of other two unisensory condition, which has been typically used to examine perceptual integration (Besle et al., 2004; Giard & Peronnet, 1999; Molholm et al., 2002; Stekelenburg & Vroomen, 2007). This analysis is based on the assumption that the information to each modality is processed independently; therefore, the bimodal response is supposed to equal to the sum of unisensory responses (i.e., Audiovisual (AV) = Audio (A) + Visual (V)). If the bimodal response differs from the sum of unimodal responses, either in a supra-additive manner (AV < A+V) or in a sub-additive way (AV > A+V), this points towards interactions occurring between the two modalities (Giard & Peronnet, 1999). In line with prior audiovisual studies (e.g., Besle et al., 2004; Stekelenburg & Vroomen, 2007), Jessen's results showed a reduced N1 (peaking around 100 ms after onset of voices) for the audiovisual condition compared to unisensory auditory condition. Further, Jessen et al. (2012) showed that the latency of N1 peak was reduced for the audiovisual stimuli when compared with an auditory-only condition at a high noise background, whereas the N1 reduction effect was not found at a low noise level. This implies that the requirement for processing bodily derived signals improves performance of emotional discrimination in a noisy environment. The reduced N1 to audiovisual stimuli suggests that perceptual interaction has already occurred during an early stage of sensory processing (Giard & Peronnet, 1999), with the deactivation considered a means to minimize the processing of redundant information for multiple modalities (van Wassenhove et al., 2005).

The integration of emotion perception might, however, be different for specific emotions. Jessen and Kotz (2011) found a reduced N1 latency in response to anger when contrasted with a fearful stimulus in auditory and in audiovisual conditions. Although the authors did not provide a conclusive explanation for this effect, brain imaging studies have provided evidence for common and specific neural circuits during the perception of anger and fear derived from body expressions (Grezes et al., 2007; Pichon et al., 2008; Pichon, de Gelder, & Grezes, 2009). On top of this, some types of emotion are more easily recognized from one modality than other modalities (Collignon et al., 2008; Paulmann & Pell, 2011; Takagi et al., 2015). This can be accounted for via modality appropriateness (Welch & Warren, 1980) or modality dominance (Spence & Squire, 2003) whereby some characteristics of stimuli are perceived accurately through one modality when contrasted with others. For instance, vision is more sensitive to spatial change than hearing, whereas hearing has a greater influence on temporal processing than vision. Modality dominance has also been found to be divergent across each emotion. A behavioural study by Takagi et al. (2015)

examined the modality dominance for six types of emotions perceived via face and vocalizations. In their study, participants were required to make judgments for emotions on the basis of facial or vocal expressions in auditory-only, visual-only or audiovisual conditions. Comparing accuracy of emotion judgments in the unisensory situation (voice-only versus face-only), the accuracy for facial emotions was higher than that of the voice for anger, happiness, disgust and surprise. This was consistent with other findings (Collignon et al., 2008; Paulmann & Pell, 2011) where facial cues dominate vocal cues. On the other hand, the higher performance was observed for voices than for faces when judging emotions of fear. To further understand modality dominance, the study examined congruency effects in two ways: facilitation effects (congruent audiovisual versus auditory-only condition) and interference effects (incongruent audiovisual condition versus. auditory-only condition). The performance for recognition of anger was improved when emotionally congruent faces were presented in a voice-attended condition. Conversely, the accuracy was decreased for the emotionally incongruent voice when the face was attended. As for fear, the congruent effects, either as a facilitation or interference effect, were absent when facial or vocal expressions were attended to. These findings imply that it is difficult to ignore the information from dominant modalities, but such modality dominance can be modulated by selective attention.

In fact, multisensory integration in pre-attentional processing can be influenced by bottom-up and top-down attention (see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010, for a review). Bottom-up attention is an automatic process driven by salient objects or events relative to the environment. This stimulus-driven process is not related to high-level processing and the observers' expectation. In contrast, top-down attention is a selectively biased process for the events that aligned with the observers' goals. For multisensory perception, a neural response to a relatively salient stimulus presented in one modality will possibly automatically elicit a weaker response to a stimulus in another modality. However, when multiple stimuli within each modality are competing for processing resources, top-down attention for the relevant property of stimuli may be required for unifying multisensory perception effectively.

Multisensory integration effects are also influenced by whether attention is fully involved across each modality. Without attentional instruction, the N1 in multisensory conditions was smaller than the sum of response to each unisensory condition (Giard & Peronnet, 1999). However, the results were reversed when both auditory and visual stimuli were attended, that is, a larger response was observed for the multisensory compared to the sum of each unimodal response (Talsma, Doty, & Woldorff, 2007; Talsma & Woldorff, 2005). The modulation of attention on multisensory perception can also occur at the emotional level. Recently, Ho et al. (2014) examined the relationship between attention and modal dominance for the perception of the face and voice. In their study, participants were required to discriminate either facial or vocal expressions (neutral versus angry), or emotional congruency between the face and voice when visual and auditory stimuli were presented simultaneously. The N1 amplitudes for congruency effects were significantly reduced for the attend-voice condition, but were not robust for the attend-face and attend-congruence conditions. The authors concluded that emotional information from visual modality might be difficult to ignore, thereby meaning that the congruency effects were substantially presented in the attend-voice task.

The comparisons between the congruency and incongruency of audiovisual information processing has also been discussed in the context of another ERP component, the P2 (P200). This is a positive deflection peaking at 200-ms post-auditory stimulus (Kokinous, Kotz, Tavano, & Schroger, 2014). The P2 has been linked to the processing of the emotional quality of a stimulus (Paulmann, Jessen, & Kotz, 2009) or a general classification category (Garcia-Larrea, Lukaszewicz, & Mauguiere, 1992). It is also interpreted as being modulated by the competition between incompatible information from multisensory sources (Knowland et al., 2014; Stekelenburg & Vroomen, 2007). However, the P2 is unlikely to be influenced by attention modulation. Ho et al. (2014) found that there was no interaction between the P2 and different attention demand tasks, but, rather, the component was modulated by how the emotional information is presented across the auditory and visual modalities. Their findings revealed a suppression of the P2 amplitude for a neutral sound with an angry face compared to that presented with a neutral face. In contrast, the P2 amplitude was increased when an incongruent neutral sound was presented than when both modal sources displayed anger. As such, the P2 is likely to reflect functionally separate processes to the N1 component during multisensory integration of the emotional percept. While the N1 reflects visual anticipation for the following auditory perception and is modulated by attention (Ho et al., 2014), the P2 is associated with processes of assessing emotional contents across modalities (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005).

On the basis of this prior literature, the objective of the current study was to understand the mechanisms underlying the interaction of emotion perceptions from body expressions combined with affective sounds. We observed the N1 and P2 to each emotion (anger vs. fear) in three conditions (auditory-only, visual-only and audiovisual). In addition, we included emotionally congruent/incongruent body-voice pairs to explore the influence of the preceding visual context for bimodal perceptual mechanisms. In order to investigate how attention interacts with modality dominance, we also compared the study with and without directing attention to the emotional characteristics of the stimuli. Referring to Jessen and Kotz (2011), the N1 is expected to be suppressed in amplitude and increased in speed in the audiovisual domain when compared with the auditory-only conditions. Due to varying modality dominance for anger and fear, the congruency effect within the N1 might be modulated by attention differently in the presentation of two emotions. As the voice is apparently dominant for fear (Takagi et al., 2015), emotionally incompatible body expressions are unlikely to interfere with attention for the fearful voice. As such, congruency effects were not expected to be observed within the N1. There might, however, be a great influence on congruency effects for anger if attentional instruction was given. With regard to later processing, the P2 is thought to be associated with competition among the modal information or the process for assessment of the binding content. Therefore, it is predicted that this will be influenced by the way of presentation of emotional contents across auditory and visual modalities instead of attention instruction.

2.2. Methods

2.2.1. Participants

Participants were Caucasian adults from Lancaster University, with normal hearing and normal or normal-corrected vision. They had no report for any neurological or psychiatric disorders. Twenty-five participants took part in the unattended tasks, but 6 were excluded from data analysis because of excessive artifact (3), fatigue (1) or poor signal-to-noise ratio (2) compared to other datasets. The remaining data comprised 18 participants (7 male) with a mean age of 23 years (*SD* = 5.0 years). In the voice-attended task, 4 out of 22 participants were excluded from further analysis due to excessive artifact contamination (2) or poor signal-to-noise ratio (2). The mean age of the remaining 18 participants (7 male) were 23.7 years (*SD* = 5.3 years). All participants provided written informed consent and were paid (£10) for their participation. The study was approved by Lancaster University Ethics Committee.

2.2.2. Stimuli

The visual stimuli were selected from the Beatrice's groups database and have been utilised in their studies (Kret, Pichon, Grezes, & de Gelder, 2011). The visual stimuli were video clips recorded from two male and two female actors expressing either angry or fearful movements. The body expressions for anger included shaking a clenching fist and raising the arm, while fear expressions involved bending the body backwards and defensive movements of the hands. The face area was blurred in all conditions involving the visual modality. The characters were all dressed in black and they performed the body movements against a green background. The luminance of each video clip was analyzed by taking into account each pixel within a frame (33 frames/clip, 480 × 854-pixel/frame). Each pixel was measured on a gray-scale using MATLAB, with values ranging from 0 to 255. The values of all pixels within a frame were the averaged to obtain a luminance score for that frame. This allowed us to explore any potential variations in luminance, which may appear with time due to the velocity and frequency of motion. Following the procedure described by Jessen and Kotz (2011), we found out that the average luminance of the individual frames in the dynamic stimuli ranges from 64 to 68, with differences of no more than 1 between two consecutive frames.

The auditory stimuli were represented by audio recordings of interjections spoken with an angry or fearful prosody. The sounds were selected from the Montreal Affective Voices database (Belin, Fillion-Bilodeau, & Gosselin, 2008), which were edited to a 700 ms epoch. Table 1. shows the key parameters for each emotional sound in the ERP study.

Table 2.1. The mean intensity (dB) and pitch (Hz) of angry and fearful sounds, with standard deviant in parentheses.

	anger		fear	
	male	female	male	female
Mean intensity	74.64 dB (9.44)	78.76 dB (9.19)	74.19 dB (8.88)	77.91 dB (9.12)
Mean pitch	241.63 Hz (59.83)	394.73 Hz (75.84)	298.45 Hz (38.02)	341.99 Hz (45.10)

In the current study, the auditory stimuli with or without the visual stimuli were presented in the following conditions: visual-only (V), auditory-only (A), emotionally congruent audio-visual (CAV), and emotionally incongruent audio-visual conditions (IAV). In the V condition, a video clip displayed a dynamic human body expressing emotions in the absence of sound. In the A condition, a video clip displayed a non-emotionally static human body posture with emotional sounds. The AV conditions played affective sounds with either emotionally congruent body expression (CAV) or incongruent ones (IAV). With the factor of emotion (anger, fear) and gender, there were a total of 16 conditions in the study. Based on the type of visual stimulus, the study were divided into 8 blocks, with V, A, CAV, IAV with anger and fear presented 8 times/condition/block. Each condition was presented 32 times, resulting in a total of 512 trials.

2.2.3. Procedure

Participants sat comfortably in a dimly lit/darkened room, and were asked to make their response by pressing a button. Each stimulus was presented using the

Psychtoolbox 3.0 in Matlab 2012a. The visual stimuli were presented on a monitor at a distance 90-100 cm from the participants, and the auditory stimuli were binaurally played via two speakers at 75 dB for all participants. Each trial started with a 800-ms white fixation on a black screen, followed by the presentation of a video clip for 1300 ms. An interval randomised between a fixation and a video clip (visual stimulus) from 800 to 1200 ms. The auditory stimuli were shown 600ms after the onset of the visual stimulus and ended synchronously with the video clips. A question mark was occasionally (< 60% of trials) presented in the center of the screen after the end of a trial. The goal of the judgment tasks was to maintain the participants' attention during the course of the stimulus presentation. Participants in the unattended task were instructed to indicate the gender of the person in the video by pressing the left or the right button (e.g., "Was the person a male or female?"). In the voice-attended task, participants were instructed to respond to the question for the emotional sounds (e.g., " Was it is angry or fearful sound?"). The question mark disappeared once the participants had made their responses. The testing started after a practice session consisting of 10 trials, and the participants were able to take a self-defined break between blocks if required. The study lasted approximately 50 minutes, including breaks.

2.2.4. EEG recording and analysis

The data were recorded by the EGI NetStation system (Geodesic Sensor Nets, Inc., Eugene, OR) with a 128-channel electrode net. The EEG signal was sampled at 500 Hz and the impedances were kept to 50 Hz or less during recording. All electrodes were on-line referenced to the vertex (Cz). For computing the ERPs, the data was filtered with a 0.3-30 Hz bandpass filter and segmented off-line from 100 ms before to 700ms after sound onset. Baseline correction was applied to 100 ms prior to each segment before artifact rejection. Trials were rejected with EGI software once the eye movement exceeded +/- 140 uV, and eye blinks exceeded +/- 100 uV. Any channels that exceeded +/- 200 uV were marked as bad. If more than 12 electrodes within a trial were marked as bad, the trial was automatically rejected. The Netstation bad channel interpolation algorithm was then applied to the accepted trials. The remaining data were re-referenced into an average reference before averaged waveforms for each participant with each condition. The analysis was focused on the two ERP components, N1 and P2, which have been indexed in audiovisual emotion perception literature. Based on previous studies (e.g. Jessen & Kotz, 2011), and visual inspection of the present data, two different analyses were conducted: the first involved the latency to the peak amplitude between 90-180 ms (N1) and 160-330 ms (P2) after sound onset, and the second involved the mean amplitude for the time window centered on the peak latency of each condition (+/- 30 ms).

As the distribution between frontal-central and central-parietal sites showed a reversed polarity of the potentials, the statistical analysis were consequently performed individually, taking the average of these electrode clusters for frontal (6, 11, 19, 4, 12, 5), central (Ref/Cz, 7, 106, 80, 31, 55) and central-parietal (62, 61, 78, 79, 54) regions of interest (ROI) (**Figure 2.1**.)

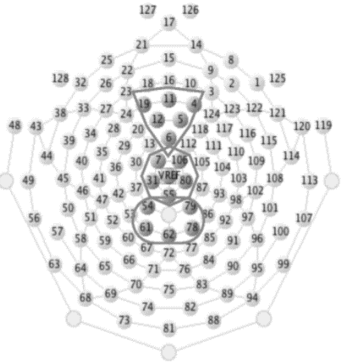


Figure 2.1. Averages were calculated based on electrode ROIs for frontal (6, 11, 19, 4, 12, 5), central (Ref/Cz, 7, 106, 80, 31, 55) and central-parietal (62, 61, 78, 79, 54) channels.

The emotional intensity of body expressions were divided into two levels, *high* and *low intensity* based on each participant's arousal rating of the body expressions. The participants were asked to judge the emotional intensity of the audiovisual stimuli after the EEG study. We used a 5-point Likert scale ranging from 1 (= very weak) to 5 (= very strong) for the rating of the emotional intensity of the stimuli. The accuracy rate for each emotion was above 90%. The results are displayed in Table 2. For the N1 and the P2, a mixed repeated-measures ANOVAs were conducted, with *group* (unattended versus voice-attended) as a between subjects factor, *conditions* (audio-only, visual-only, emotionally congruent audiovisual, and emotionally incongruent audiovisual), *emotion* (anger versus fear), *visual intensity* (high versus low intensity of body expressions) and *Site* (frontal, central, central-parietal sites) as within-subjects factors. Post-hoc analyses (least significant difference) were run where any significant (*p*-value < 0.05) interaction effects were reported. In order to better understand the effects of different conditions within unattended and voice-attended groups, we also separately ran ANOVAs with four within-subject factors (*condition, emotion, intensity and site*) for each group.

Table 2.2. Results of rating for the stimuli presented in the EEG study. Mean intensity (1 to 5 scale) for emotions of angry and fear in CAV (congruent audiovisual condition), with standard deviant in parentheses.

	Anger		Fear	
	High	low	high	low
intensity	4.05 (0.87)	3.54 (0.89)	4.34 (0.69)	3.51 (1.64)

2.3. Results

Figure 2.2. depicts the grand average for each condition (condition, emotional intensity of body expression and emotions) at Cz in the unattended and voice-attended groups. We only report key findings, any significant main effect or interactions including factors of *groups*, *condition* (comparison among A, CAV and IAV condition) and *emotion*. For brevity and coherence, we reported *post hoc* analyses for significant highest-order interactions involving *condition* or *emotion*.

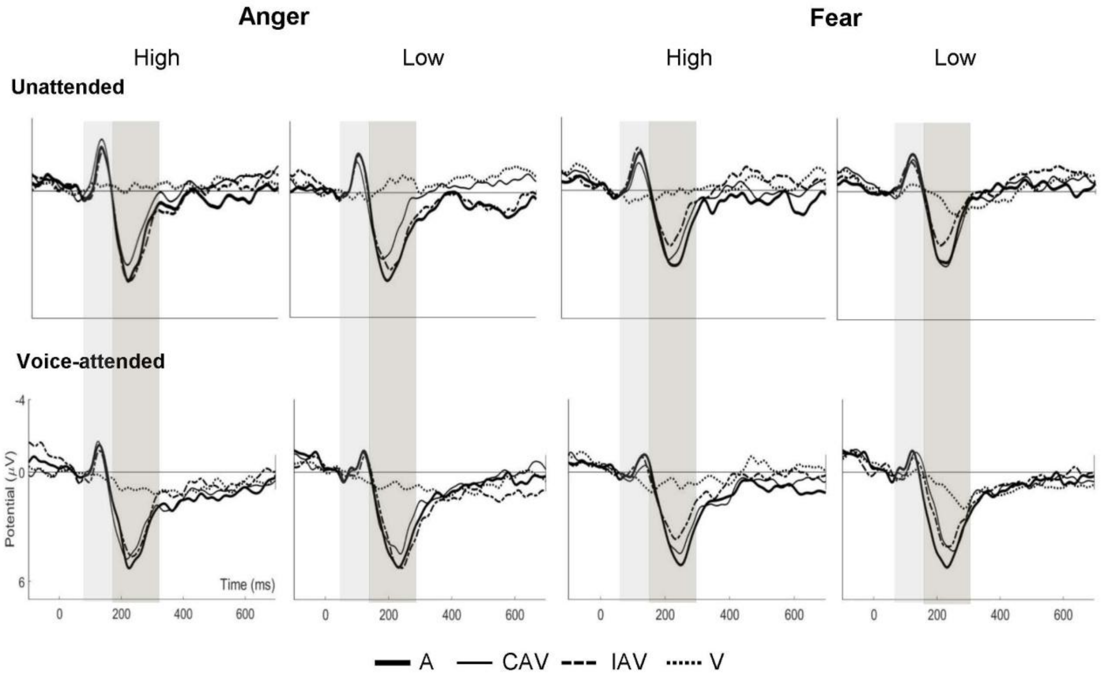


Figure 2.2. The grand average of N1 and P2 for each group for factors of condition, intensity of body expressions (high vs. low intensity), and emotions (anger vs. fear), which are indicative of effects in the region.

2.3.1. ERP Latency

Results of the N1 and the P2 peak latency for the comparisons between *group*, *conditio*n, *visual intensity*, and *sites* are listed in **Table 2.3.**

2.3.1a. N1

The main effect of *group* was significant (F(1,34) = 6.66, p =.014, η^2 =.164), with shorter latencies for the voice-attended group when contrasted with the unattended group. The *group* effect was also evidenced in the interactions with *condition* and *site* (F(6,204) = 5.36, p = .002, η^2 = .094), *emotion* and *site* (F(2,68) = 5.95, p = .004, η^2 =.149), *condition*, *emotion* and *site* (F(6,204) = 5.55, p < .0001, η^2 = .140). When angry sounds were presented, the latencies were sped up for the voice-attended compared to the unattended group in IAV conditions at frontal sites (p = .017). When the sounds were fearful, the group differences were found in CAV and IAV conditions at frontal sites (p = .012; p = .03, respectively) as well as in A condition at central sites (p = .014). We also examined 4 within-subjects factors (*condition, emotion, visual intensity,* and *site)* in each group (see **Table 2.4.**)

Table 2.3. Statistical Results for the N1 and P2 latency

		N1 latency			P2 latency		
	d*f*	F	*p*	*η2*	F	*p*	*η2*
condition	3,102	3.15	.028	.085	9.12	.000	.211
condition * group	3,103	.10	.959	.003	.30	.826	.009
emotion	1,34	29.49	.000	.464	8.60	.006	.202
emotion * group	1,34	.30	.587	.009	1.83	.185	.051
intensity	1,34	.47	.500	.014	5.23	.029	.133
intensity * group	1,34	.25	.623	.007	1.15	.291	.033
site	2,68	37.97	.000	.528	24.92	.000	.423
site * group	2,68	.48	.623	.014	4.42	.016	.115
condition * emotion	3,102	2.81	.043	.076	4.85	.003	.125
condition * emotion * group	3,102	1.95	.126	.054	.32	.814	.009
condition * intensity	3,102	1.81	.151	.050	3.29	.024	.088
condition * intensity * group	3,102	.93	.430	.027	.52	.672	.015
emotion * intensity	1,34	2.72	.108	.074	2.69	.110	.073
emotion * intensity * group	1,34	5.34	.027	.136	2.29	.139	.063
condition * emotion * intensity	3,102	.24	.867	.007	2.59	.057	.071
condition * emotion * intensity * group	3,102	.74	.534	.021	.43	.730	.013
condition * site	6,204	1.86	.089	.052	31.58	.000	.482
condition * site * group	6,204	3.54	.002	.094	2.06	.060	.057
emotion * site	2,68	1.15	.323	.033	.67	.517	.019
emotion * site * group	2,68	5.95	.004	.149	.52	.595	.015
condition * emotion * site	6,204	1.35	.238	.038	.22	.972	.006
condition * emotion * site * group	6,204	5.55	.000	.140	1.22	.297	.035
intensity * site	2,68	1.17	.317	.033	1.04	.359	.030
intensity * site * group	2,68	.04	.957	.001	.92	.403	.026
condition * intensity * site	6,204	2.80	.012	.076	3.36	.004	.090
condition * intensity * site * group	6,204	.15	.989	.004	1.33	.246	.038
emotion * intensity * site	2,68	7.92	.001	.189	1.41	.250	.040
emotion * intensity * site * group	2,68	1.06	.354	.030	.32	.728	.009
condition * emotion * intensity * site	6,204	3.23	.005	.087	1.01	.423	.029
condition * emotion * intensity * site * group	6,204	.76	.603	.022	.57	.756	.016

Table 2.4. Statistical Results of N1 latency in each group

	Unattended				*Voice-attended*		
	df	F	*p*	η^2	F	*p*	η^2
condition	3,51	1.12	.349	.062	2.23	.096	.116
emotion	1,17	13.91	.002	.450	16.68	.001	.495
Intensity	1,17	1.15	.299	.063	.01	.913	.001
site	2,34	11.95	.000	.413	31.43	.000	.649
*condition * emotion*	3,51	4.84	.005	.222	.05	.986	.003
*condition * intensity*	3,51	2.17	.103	.113	.51	.679	.029
*emotion * intensity*	1,17	.18	.675	.011	9.80	.006	.366
*condition * emotion * intensity*	3,51	.24	.870	.014	.95	.425	.053
*condition * site*	6,102	1.23	.297	.067	4.27	.001	.201
*emotion * site*	2,34	6.44	.004	.275	1.90	.348	.060
*condition * emotion * site*	6,102	2.38	.034	.123	4.20	.001	.198
*intensity * site*	2,34	.90	.417	.050	.42	.663	.024
*condition * intensity * site*	6,102	2.11	.058	.111	1.01	.425	.056
*emotion * intensity * site*	2,34	8.05	.001	.321	2.00	.151	.105
*condition * emotion * intensity * site*	6,102	4.62	.000	.214	.45	.842	.026

Unattended group

The main effect of *emotion* (F(1,17) = 13.91, *p* =.002, η^2 =.450) was found, indicating faster responses for angry (*M* = 128.59 (1.27) ms) compared to fearful expressions (*M* = 134.35 (.89) ms). This effect was qualified by 2-way interactions between *emotion* and *condition* (F(3,51) = 4.84, *p* =.005, η^2 =.222), *emotion* and *site* (F(2,34) = 6.44, *p* = .004, η^2 = .275), *emotion, condition, and site* (F(6,102) = 2.38, *p* =.034, η^2 =.123), and a 4-way interaction between *condition, emotion, intensity* and *site* (F(6,102) = 4.62, *p* < .0001, η^2 =.214). For the blocks presented with high-intensity body expressions, the differences in *emotion* were obvious in CAV at central (*d* = 8.00 (2.69) ms; *p* = .009) and central-parietal sites (*d* = 19.14 (4.02) ms; *p*

< .0001) as well as IAV conditions at central regions (*d* = 8.63 (3.101) ms; *p* = .013). As the emotional intensity of body expression became lower, the emotion effects were only observed in CAV condition at frontal (*d* = 19.14 (3.65) ms; *p* < .0001) and central regions (*d* = 12.15 (3.14) ms; *p* = .001)

Voice-attended group

Similar to the unattended group, only the main effect of *emotion* (F(1,17) = 16.68, *p* =.001, η^2 =.495) was observed, with faster responses to angry (*M* = 124.79 (1.57) ms) compared to fearful expressions (*M* = 129.49 (1.61) ms). This effect was qualified by a 2-way significant interaction between *emotion* and *intensity* (F(1,17) = 9.80, *p* =.006, η^2 = .366). The pairwise comparisons indicated that the difference in *emotion* was present when lower-intensity body expressions were presented (*d* = 8.85 (1.76) ms; *p* < .0001). A 3-way interaction between *condition*, *emotion* and *site* (F(6,102) = 4.20, *p* = .001, η^2 = .198) was also found. Follow up analysis indicated that the emotion effects were prominent in A condition at frontal sites (*d* = 7.57 (3.55) ms; *p* =.048) as well as in CAV condition at central (*d* = 6.74 (3.18) ms; *p* = .049) and central-parietal regions (*d* = 9.47 (3.64) ms*;* *p* = .019).

2.3.1b. P2

The main effect of *group* was also observed (F(1,34) = 4.27, *p* = .046, η^2 = .112), with a faster P2 peak latency in the unattended group when contrasted with the voice-attended group. The *group* marginally showed interactions with *condition* and *site* (F(6,204) = 2.06, *p* =.006, η^2 =.057). The group differences were found in CAV condition at frontal and central sites (*p* = .008, *p* = .037, respectively), IAV and A condition at central sites (*p* = .005; *p* = .047, respectively). Further analyses were conducted to investigate the emotion effects and condition effects by factors of *condition, emotion*, *intensity*, *site* in each group (see **Table 2.5**.).

Table 2.5. Statistical Results of P2 latency in each group

	Unattended				*Voice-attended*		
	df	F	*p*	η^2	F	*p*	η^2
condition	3,51	5.90	.002	.258	3.60	.020	.175
emotion	1,17	12.37	.003	.421	.99	.333	.055
Intensity	1,17	1.02	.326	.057	4.41	.051	.206
site	2,34	10.64	.000	.385	16.37	.000	.491
*condition * emotion*	3,51	1.43	.245	.078	4.00	.012	.191
*condition * intensity*	3,51	2.78	.051	.140	1.04	.381	.058
*emotion * intensity*	1,17	.07	.937	.000	6.57	.020	.279
*condition * emotion * intensity*	3,51	.86	.468	.048	2.18	.102	.114
*condition * site*	6,102	24.82	.000	.593	9.27	.000	.353
*emotion * site*	2,34	1.18	.318	.065	.11	.894	.007
*condition * emotion * site*	6,102	.91	.490	.051	.55	.770	.031
*intensity * site*	2,34	3.96	.028	.189	.18	.839	.010
*condition * intensity * site*	6,102	.51	.800	.029	4.60	.000	.213
*emotion * intensity * site*	2,34	1.06	.358	.059	.69	.511	.039
*condition * emotion * intensity * site*	6,102	.95	.462	.053	.64	.698	.036

Unattended group

The main effect of *condition* (F(3,51) = 5.90, *p* = .002, η^2 = .258) was observed, with shorter latencies in CAV than in A and IAV conditions (*p* = .003; *p* = .001, respectively). This effect was also qualified by an interaction between *condition* and *site* (F(6,102) = 24.82, *p* < .0001, η^2 = .593), with faster responses for CAV compared to A conditions at frontal (*d* = 5.01 (2.19) ms; *p* = .036) and central regions (*d* = 7.53 (1.79) ms; *p* =.001) as well as for CAV compared to IAV at central (*d* = 5.03 (1.62) ms; *p* = .007) and central-parietal regions (d = 9.72 (3.44) ms; *p* = .012).

In addition, the main effect of *emotion* (F(1,17) = 12.37, *p* = .003, η^2 =.421) reached significance, with shorter latencies for angry than for fearful expressions. However, no significant interactions involving *emotion* were found.

Attended group

The main effect of *condition* (F(3,51) = 3.60, *p* =.020, η^2 = .175) was observed, with longer latencies in V than in CAV and IAV conditions (*p* = .037; *p* = .040, respectively). *Condition* also showed a 2-way interaction with *site* (F(6,102) = 9.27, *p* <.0001, η^2 =.353), and a 3-way interaction among *condition*, *intensity* and *site* (F(6,102) = 4.60, *p* <.0001, η^2 =.213). However, further analysis showed no significant differences between A, CAV and IAV conditions.

We also found a 2-way interaction between *condition* and *emotion* (F(3,51) = 4.00, *p* =.012, η^2 =.191). Planned comparisons showed the differences were driven by angry expressions, with shorter latencies in CAV than in A (*d* = 7.01 (3.11) ms; *p*=.038) and IAV conditions (*d* = 10.01 (3.08) ms ; *p* = .005). Considering emotion effects, the latencies for angry expressions peaked earlier than fearful expressions in both CAV (*d* = 14.08 (6.30) ms, *p* = .039) and A contexts (*d* = 8.43 (3.41) ms; *p*=.024) .

2.3.2. ERP Amplitude

The time window for N1 and P2 peak amplitudes were based on their peak latency. **Figure 2.3.** and **Figure 2.4.** display the grand averages of the N1 and the P2, respectively, displaying mean amplitude (top) and topography (bottom) across *emotions*, *visual intensity* and *condition* for both groups. Statistical results of the N1 and the P2 amplitude for the comparisons between *group*, *condition*, *visual intensity*, and *sites* are listed in **Table 2.6**.

Table 2.6. Statistical Results for the N1 and P2 amplitudes

	N1 amplitude				P2 amplitude		
	df	F	*p*	η^2	F	*p*	η^2
condition	3,102	17.87	.000	.345	64.08	.000	.653
*condition * group*	3,103	1.09	.356	.031	.15	.930	.004
emotion	1,34	1.16	.289	.033	8.48	.006	.200
*emotion * group*	1,34	3.20	.083	.086	.610	.440	.018
intensity	1,34	.29	.595	.008	.00	.963	.000
*intensity * group*	1,34	.19	.670	.005	3.78	.060	.100
site	2,68	56.00	.000	.622	41.30	.000	.548
*site * group*	2,68	.49	.613	.014	1.53	.224	.043
*condition * emotion*	3,102	2.27	.085	.062	7.68	.000	.184
*condition * emotion * group*	3,102	1.83	.146	.051	1.20	.315	.034
*condition * intensity*	3,102	.09	.966	.003	.90	.446	.026
*condition * intensity * group*	3,102	.90	.446	.026	1.63	.188	.046
*emotion * intensity*	1,34	5.56	.024	.141	.52	.476	.015
*emotion * intensity * group*	1,34	.19	.667	.006	.40	.534	.012
*condition * emotion * intensity*	3,102	7.09	.000	.173	2.11	.104	058
*condition * emotion * intensity * group*	3,102	.36	.783	.010	.30	.829	.009
*condition * site*	6,204	9.36	.000	.216	10.71	.000	.240
*condition * site * group*	6,204	2.06	.060	.057	2.837	.011	.077
*emotion * site*	2,68	.58	.562	.017	.198	.821	.006
*emotion * site * group*	2,68	1.54	.221	.043	4.739	.012	.112
*condition * emotion * site*	6,204	5.40	.000	.137	3.509	.003	.094
*condition * emotion * site * group*	6,204	.68	.669	.019	.763	.600	.022
*intensity * site*	2,68	1.46	.240	.041	.933	.398	.027
*intensity * site * group*	2,68	.02	.981	.001	.010	.990	.000
*condition * intensity * site*	6,204	1.12	.352	.032	4.61	.000	.119
*condition * intensity * site * group*	6,204	.85	.534	.024	.40	.877	.012
*emotion * intensity * site*	2,68	1.18	.314	.033	8.09	.001	.192
*emotion * intensity * site * group*	2,68	.22	.801	.006	.41	.668	.012
*condition * emotion * intensity * site*	6,204	7.46	.000	.180	3.53	.002	.094
*condition * emotion * intensity * site * group*	6,204	.85	.530	.025	.63	.710	.018

2.3.2a. *N*1

The main effect of *group* (F(1,34) = 7.47, *p* = .010, η^2 = .180) reached significance, indicating larger N1 amplitudes in the unattended than in the voice-attended group. However, no interactions including group were significant. **Table 2.7.** shows the statistical results for four within subject factors (*condition, emotion, visual intensity, site*) in each group.

Table 2.7. Statistical Results for N1 amplitudes in each group

	Unattended				*Voice-attended*		
	df	F	*p*	η^2	F	*p*	η^2
condition	3,51	10.76	.000	.388	7.28	.000	.300
emotion	1,17	4.10	.059	.194	.25	.621	.015
Intensity	1,17	.49	.494	.028	.01	.942	.000
site	2,34	34.25	.000	.668	22.90	.000	.574
*condition * emotion*	3,51	3.46	.023	.169	.09	.964	.005
*condition * intensity*	3,51	.30	.824	.017	.76	.522	.043
*emotion * intensity*	1,17	5.96	.026	.260	1.38	.257	.075
*condition * emotion * intensity*	3,51	4.48	.007	.209	2.76	.052	.140
*condition * site*	6,102	4.35	.001	.204	7.51	.000	.306
*emotion * site*	2,34	1.82	.178	.096	.13	.877	.008
*condition * emotion * site*	6,102	3.66	.002	.177	2.22	.047	.116
*intensity * site*	2,34	.57	.571	.032	1.00	.379	.055
*condition * intensity * site*	6,102	1.74	.120	.093	.33	.921	.019
*emotion * intensity * site*	2,34	.16	.849	.010	1.47	.244	.080
*condition * emotion * intensity * site*	6,102	4.33	.001	.203	3.81	.002	.183

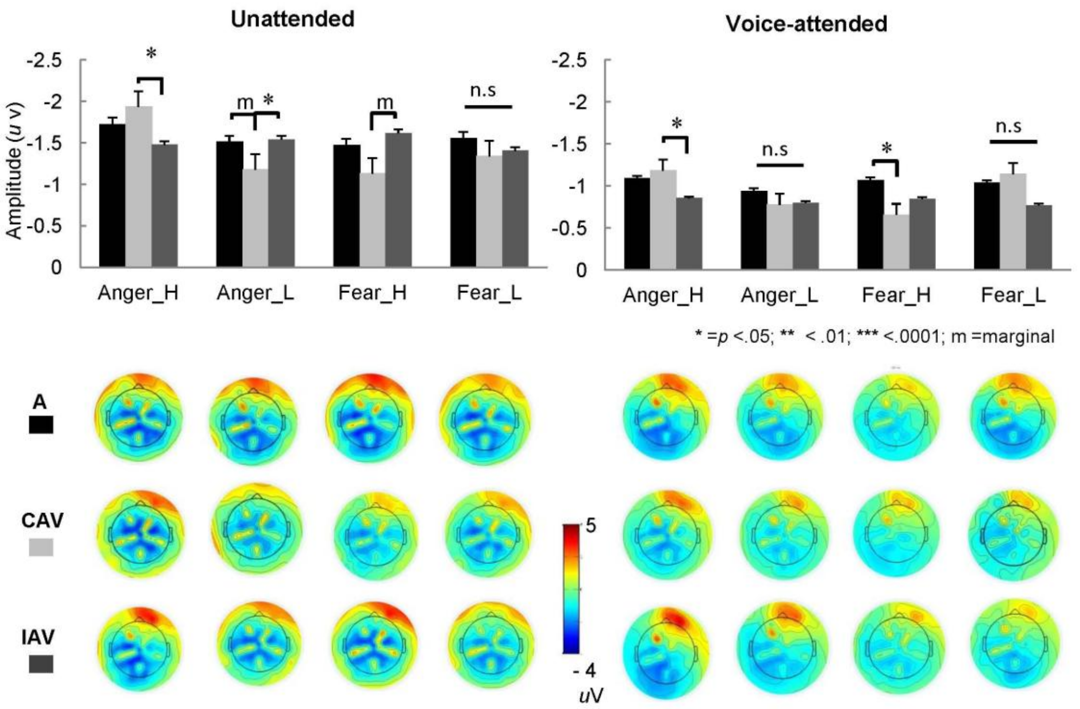


Figure 2.3. The ERPs displaying the grand average of N1 peak amplitude (top) at central electrode site and topography distributions (bottom) for angry and fearful information with intensity (H= high-intensity; L= low-intensity emotional body expression) in A, CAV, IAV conditions between 90 to 180 ms after onset of auditory stimulus

Unattended group

The main effect of *condition* was significant (F(3,51) = 10.76, *p* <.0001, η^2 =.388), with a trend for smaller amplitudes in CAV compared to A condition (*p* =.052). We also found significant interactions between *condition* and *emotion* (F(3,51) = 3.46, *p* =.023, η^2 = .169), between *condition* and *site* (F(6,102) = 4.35, *p* = .001, η^2 =.204), and between *condition*, *emotion* and *site* (F(6,102) = 3.66, *p* = .002, η^2 = .177). These effects were further qualified by a 4-way significant interaction among *condition, emotion, intensity and sit*e (F(6,102) = 4.33, *p* = .001, η^2 = .203). For the blocks with high-intensity body expressions, angry sounds elicited greater N1 amplitudes in CAV compared to the A condition at frontal regions (d_{CAV-A} = -.67 (.21) μV, *p* = .005) as well as in CAV compared to the IAV condition over frontal ($d_{CAV-IAV}$ = -1.10 (.23)

μV, p <.0001) to central regions (d= -.46 (.19) μV, p = .030). Further, the amplitudes were more negative for A than for IAV condition at frontal sites ($d_{A - IAV}$ = -.43 (.19) μV; p =.039). In contrast, the amplitude was reduced in CAV compared to the IAV condition at central regions ($d_{CAV - IAV}$ = .36 (.16) μV; p = .035) when the blocks presented low-intensity body expressions.

When fearful sounds were presented with high-intensity body expressions, smaller N1 amplitudes were observed for CAV compared to IAV conditions from frontal ($d_{CAV - IAV}$ = .78 (.25) μV; p = .007) to central regions (d = .49 (.25) μV; p = .063). There was also a trend for smaller amplitudes in A compared to IAV condition at frontal regions ($d_{A - IAV}$ = .48 (.27) μV; p = .091). Similar patterns were also observed for the blocks with low-intensity body expressions whereby N1 amplitudes were slightly smaller for A compared to IAV at frontal regions ($d_{A - IAV}$ = .31 (.17) μV; p = .084).

Voice-attended group

There was a main effect of *condition* (F(3,51) = 7.28, p < .0001, η^2 = .300), which was driven by smaller N1 amplitudes in V compared to the other three conditions (all p < .01). *Condition* also showed a 2-way significant interaction with *site* (F(6,102) = 7.51, p < .0001, η^2 = .306), a 3-way interaction with *emotion* and *site* (F(6,102) = 2.22, p = .047, η^2 = .116), and a 4-way interaction with *emotion, intensity* and *site* (F(6,102) = 3.81, p = .002, η^2 = .183). With high-intensity body expressions, angry sounds elicited larger N1 amplitudes in CAV compared to the IAV condition at frontal ($d_{CAV - IAV}$ = -.40 (.13) μV; p = .007) and central regions ($d_{CAV - IAV}$ = -.33 (.14) μV; p = .035). Conversely, fearful sounds elicited smaller amplitudes in CAV than in the A condition at central regions ($d_{CAV - A}$ = .41 (.19) μV; p = .041). For both emotional sounds, no significant difference between A, CAV and IAV conditions were found when low-intensity body expressions were presented.

2.3.2b. P2

No significant main effect of *group* was found (F(1,34) = 1.26, *p* =.270, η^2 =.036), but there was a significant interaction between *group*, *emotion* and *site* (F(2,68) = 4.74, *p* = .012, η^2 = .036). Further analysis showed that both responses to angry (*p* =.079) and fearful (*p* =.122) stimuli were slightly smaller in the unattended group than those elicited in the attended group over central-parietal regions. Further, the statistical results for four within-in subject factors in each group are listed in Table 8.

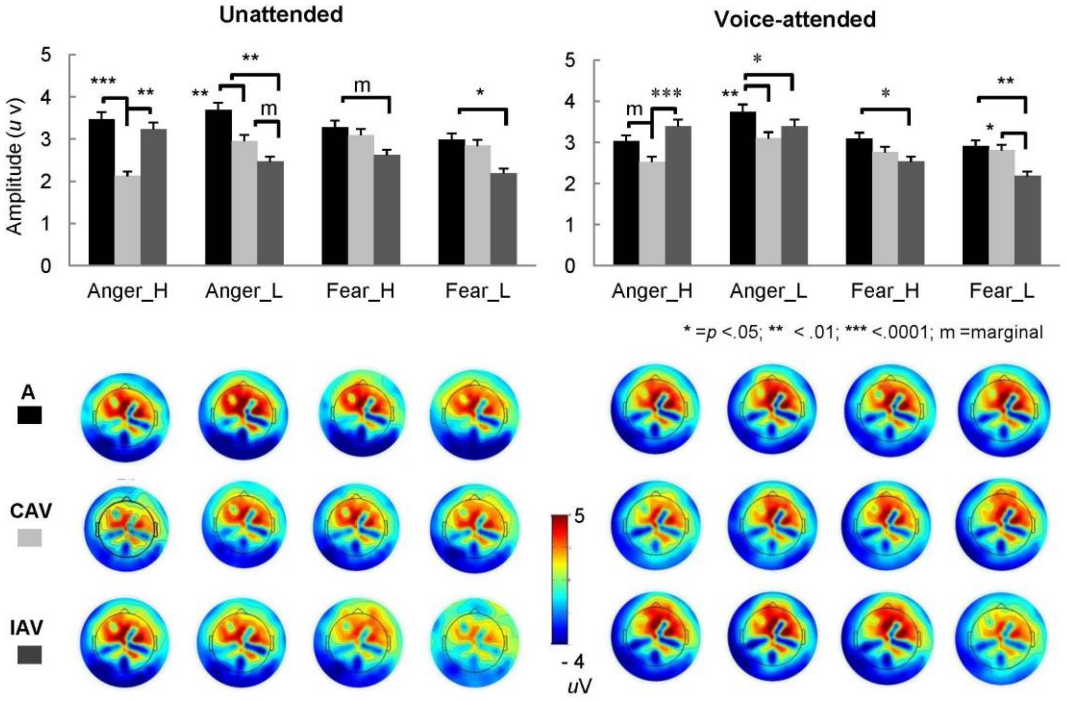


Figure 2.4. The ERPs displaying the grand average of P2 peak amplitude (top) at frontal-central electrode site and topography distributions (bottom) for angry and fearful information with intensity (H= high-intensity; L= low-intensity emotional body expression) in A, CAV, IAV conditions between 160 to 330 ms after onset of auditory stimulus

Table 2.8. Statistical Results for P2 amplitudes in each group

	Unattended				*Voice-attended*		
	df	F	*p*	η^2	F	*p*	η^2
condition	3,51	34.76	.000	.672	29.60	.000	.635
emotion	1,17	1.30	.270	.071	26.71	.000	.611
Intensity	1,17	1.62	.221	.087	2.23	.154	.116
site	2,34	38.11	.000	.692	10.99	.000	.393
*condition * emotion*	3,51	6.08	.001	.264	2.42	.077	.124
*condition * intensity*	3,51	1.70	.178	.091	.64	.592	.036
*emotion * intensity*	1,17	.60	.451	.034	.01	.926	.001
*condition * emotion * intensity*	3,51	.93	.435	.052	1.70	.180	.091
*condition * site*	6,102	8.52	.000	.334	5.21	.000	.235
*emotion * site*	2,34	2.72	.081	.138	2.15	.132	.112
*condition * emotion * site*	6,102	1.22	.302	.067	3.36	.005	.165
*intensity * site*	2,34	.34	.717	.019	.67	.517	.038
*condition * intensity * site*	6,102	1.66	.138	.089	4.27	.001	.201
*emotion * intensity * site*	2,34	2.18	.129	.113	10.14	.000	.374
*condition * emotion * intensity * site*	6,102	2.89	.012	.145	1.23	.299	.067

Unattended group

The main effect of *condition* was significant (F(3,51) = 34.76, *p* < .0001, η^2 =.672), with smaller P2 amplitudes in CAV and IAV compared to A conditions (*p* =.008; *p* =.029, respectively). Significant interactions between *condition* and *emotion* (F(3, 51) = 6.083, *p* = .001, η^2 = .264), between *condition* and *site* (F(6, 102) = 8.52, *p* < .0001, η^2 = .334), and between *condition* and *emotion*, *intensity and sit*e (F(6,102) = 2.89, *p* = .012, η^2 = .145) were identified. Subsequent analysis revealed that angrysounds elicited smaller P2 amplitudes in CAV compared to A condition at frontal (d_{CAV-A} = -1.34 (.20) μV; *p* < .0001) and central regions (d_{CAV-A} = -.47 (.20)

μV; *p* = .029) when high-intensity body expressions were presented. The P2 amplitudes were also reduced for CAV compared to IAV condition at frontal sites ($d_{CAV - IAV}$ = -1.11 (.28) μV; *p* = .001). For lower-intensity body expressions, the P2 amplitude was reduced in CAV compared to A at frontal ($d_{CAV - A}$ = -.73 (.24) μV; *p* =.007) and central regions ($d_{CAV - A}$ = -.76 (.20) μV; *p* =. 001). In addition, smaller amplitudes were found for IAV than for A at frontal electrode sites ($d_{IAV - A}$ = -1.22 (.29) μV; *p* = .001), and the difference slightly decreased at central sites ($d_{IAV - A}$ = -.67 (.38) μV; *p* = .095). There was also a trend-level effect for smaller amplitudes in IAV compared to the CAV condition at frontal regions ($d_{IAV - CAV}$ = -.48 (.23) μV; *p* =. 052)

When fearful sounds were presented with high-intensity body expressions, smaller amplitudes was marginally observed for IAV compared to A conditions at frontal ($d_{IAV - A}$ = -.66 (.32) μV; *p* = .055) and central sites ($d_{IAV - A}$ = -.52 (.26) μV; *p* = .062). With the presentation of lower-intensity body expressions, the reduced amplitudes were more prominent for IAV than for A over frontal ($d_{IAV - A}$ = -.80 (.34) μV; *p* = .03) to central electrode sites ($d_{IAV - A}$ = -.54 (.21) μV; *p* = .02). There was also a trend towards a smaller response to IAV than for the CAV conditions at central regions ($d_{IAV - CAV}$ = -.72 (.37) μV; *p* =. 069)

Voice-attended group

A main effect of *condition* was significant (F(3,51) = 29.60, *p* < .0001, η^2 = .635), which was mainly driven by smaller amplitudes in V compared to the other three conditions (all *p* < .0001). *Condition* also showed a 2-way significant interaction with *site* (F(6,102) = 5.21, *p* < .0001, η^2 = .235), and a 3-way interaction with *emotion and site* (F(6,102) = 3.36, *p* = .005, η^2 =.165). Further analysis showed that angry sounds elicited smaller P2 amplitudes in CAV compared to A ($d_{CAV - A}$ = -.57 (.17) μV; *p*

= .003) and IAV conditions at frontal regions ($d_{CAV-IAV}$ = -.58 (.14) μV; $p < .0001$). For fearful sounds, smaller amplitudes were found for IAV compared to A (d_{IAV-A} = -.64 (.18) μV; $p = .003$) and CAV conditions at frontal regions ($d_{IAV-CAV}$= -.43 (.09) μV; $p < .0001$), but both differences were attenuated at central regions ($p = .085$; p =.053, respectively).

A main effect of *emotion* was also found (F(1,17) = 26.709, $p < .0001$, η^2 =.611), indicating larger P2 amplitudes to angry than to fearful expressions. This effect was also qualified by a three-way interaction between *emotion, intensity* and *site* (F(2,34) = 10.14, $p = .005$, $\eta^2 = .374$). Subsequent analysis showed that the difference in *emotion* was observed for high-intensity body expressions at central (d = .23 (.07) μV; $p = .005$) and central-parietal sites (d = .36 (.12) μV; $p = .008$). When the visual stimulus were low-intensity, the emotion effects were found at frontal (d = .58 (.11) μV; $p < .0001$) and central sites (d = .20 (.08) μV; $p = .022$).

2.4. Discussion

The present study aimed to explore the neural processing underlying the integration (or interaction) of emotion perception for body expressions and sounds. This study also provided preliminary evidence for the modulation of attention, emotional types and intensity of body expressions on audiovisual emotion perception. With the manipulation of attentional instructions, we could further examine the influence on modality dominance for the emotions of anger and fear. The auditory N1 (90-180 ms) and the P2 (260-330 ms) were observed for the differences in responses to modalities and to emotional congruency across the audiovisual information. Both N1 and P2 peak latency were differentiated in response to angry and fearful information among unattended and voice-attended participants. This is congruent with Jessen's findings (Jessen & Kotz, 2011; Jessen et al., 2012) showing rapid N1 and P2 peak latencies for angry compared to the fearful stimuli in auditory-only and congruent audiovisual conditions. However, the responses to CAV, IAV and A conditions differed in terms of N1 and P2 amplitudes. **Table 2.9.** summarizes statistically significant comparisons between *conditions*, *emotions*, and *intensity of body expression* within N1 and P2 in each group. In unattended participants, the difference in N1 amplitudes were observed in auditory-only (A) compared to congruent (CAV) as well as incongruent (IAV) audiovisual conditions.

Table 2.9. Summary of the significant effects on N1 and P2 in each condition for unattended and voice-attended group

		Unattended		Attended	
		High intensity	Low intensity	High intensity	Low intensity
N1	Anger	CAV vs. A at F IAV vs A at F CAV vs IAV at F,C	CAV vs. A at C(m) IAV vs A at CP CAV vs IAV at C	CAV vs. IAV at F,C (m)	n.s
	Fear	CAV vs IAV at F, C(m) IAV vs A at C (m)	IAV vs A at F (m)	CAV vs A at C	CAV v.s IAV at F(m)
P2	Anger	CAV vs. A at F,C CAV vs. IAV at F,	CAV vs A at F,C CAV vs IAV at F(m)	CAV vs. A at F (m), CAV vs IAV at F	CAV vs A at F,C(m) IAV vs A at F CAV vs IAV at F(m)
	Fear	IAV vs A at F(m),C(m)	IAV vs A at F(m),C(m) CAV vs IAV at C(m)	IAV vs A at F CAV vs IAV at C	IAV vs. A at F CAV vs IAV at F

F = frontal-central site, C = central site, m = marginal significance

The N1 amplitudes also differed between congruent and incongruent pairs, and this effect was more prominent for angry sounds. However, the modalities and congruency effects were attenuated as attention was guided towards voices due to the instructions (voice-attended group). Particularly, when the blocks were presented with the low-intensity emotional body, these effects were attenuated nearly diminished. With regard to the P2 responses, the patterns across emotions and intensity in unattended participants were similar to voice-attended participants. Specifically, the comparisons between conditions were different between angry and fearful expressions. The P2 amplitudes to angry sounds were differentiated between CAV and IAV as well as A conditions. For fearful sounds, the difference was mainly found for the IAV compared to A conditions.

Overall, the data presented here support our hypothesis that the N1 amplitudes were modulated more by attention and emotional intensity of body expressions relative to the P2. Additionally, the congruency effects within N1, particularly for the fearful expressions, were attenuated with voice-attended instruction. In contrast, less attention-related changes were observed within the P2 component. Instead, the P2 was largely influenced by how the emotion content combined across the modalities. Therefore, the two ERP components are likely to reflect functionally different processes.

2.4.1. Attention Modulation and the N1

Compared to unattended participants, voice-attended participants showed decreased modality effects and congruency effects for both angry and fearful information. These results are consistent with Talsma et al. (2010) whereby the bottom-up, or stimulus-driven process from a certain modality automatically captures attention; nevertheless, this ‘pop-out’ effect would be attenuated when the multiple stimuli within each modality saliently compete for the resource. The involvement of selective attention can appropriately modify the interference and enhance the effectiveness of multisensory perceptual integration. However, if the features of the stimuli in the unattended modalities are intrinsically salient, particularly if it is emotionally incongruent for the information in attended modalities, then this can impede the processing of specific emotions in the attended modality. This interference could also possibly increase congruency effects and lower modality effects.

Our results show that the processing of angry and fearful information were both modulated by attention, but in different ways. This difference may be accounted for through concepts related to modality dominance (Spence & Squire, 2003). It has been reported that fearful information is advantageously recognized from auditory, rather than visual channels (Paulmann & Pell, 2011; Takagi et al., 2015). In this case, fearful sounds are likely to convey enough information to observers, particularly when the attention was focussed on the instructions related to emotional sounds. Whether the visual information was emotionally congruent or not was unlikely to become an interference or facilitation factor for the processing of fearful sounds. In line with our assumption, the current data showed little to no congruency and modality effects for fearful sounds in voice-attended participants ($p < .05$ or marginally significant) than those effects observed in the unattended group. We assume that the emotionally incongruent visual information for fearful sounds, (the angry body expression), could be a strong signal that influenced emotional recognition. Therefore, when the angry body displayed higher emotional intensity, the effects of emotional congruency between the auditory and visual information were activated in the unattended participants. Until attention was directed to the fearful voice during the audiovisual processing, the modulations from visual information, including modality and congruency effects, were marginally observed or were not seen at all. However, the modality dominance for anger is dependent on attentional instruction (Takagi et al., 2015). As such, when attention was not instructed, the fearful information from the visual modality was prone to impact upon the processing of angry sounds. Once attention was involved, the visual interference for the angry sounds possibly declined. This particularly occurred when low-intensity body expressions were presented. However, the current findings differed from Ho et al. (2014), where the congruency effects to angry sounds were more robust for the voice-attended task compared to when attention was instructed to visual or audiovisual stimuli. This difference may be related to the neutral faces that they used to pair with angry sounds, which were quite different to the stimuli that we used.

2.4.2. The Modality Effect and the N1

For fearful sounds, the N1 amplitudes were expected to reduce for the audiovisual conditions compared to auditory conditions (amplitude AV < A) (Jessen & Kotz, 2011). However, the assumption was consistent with angry stimuli only when the body expressions were a low-intensity emotion. Contrary to our expectation, the angry sounds with high-intensity angry bodies elicited larger amplitudes compared to sounds presented in isolation (amplitude AV > A). In the study, both blocks with high and low-intensity bodies presented the same sounds, and we also found no significant differences in the responses to auditory-only condition across the two blocks. Therefore, the different patterns of effects as a function of modality are not be caused by different levels of contrasts from auditory-only responses. Alternatively, we considered that the high-intensity angry body conveyed much stronger information when contrasted with the auditory stimuli. Consequently, it is not easy to ignore the visually angry expressions even if the instruction was given to attend to the sounds. This may be explained by Talsma et al. (2007) whereby the N1 patterns either increase or decrease amplitudes in audiovisual compared to the sum of the unisensory responses, depending on whether attention was fully directed to both auditory and visual modalities. It is plausible that attention was captured by both visual and auditory modalities when the high-intensity angry body was presented. However, there is evidence to suggest that fear is dominated by the auditory modality (Takagi et al., 2015). The fearful expressions from the visual modality might not strongly convey the signal of angry emotional content. Therefore, this reversed patterns of modality effects was not observed for fearful expressions, even if its emotion intensity was the same as that for the angry body.

2.4.3. The Modulation of the P2

Compared to the N1, attentional modulation was diminished within the P2. The patterns of modality and congruency effects in the unattended group were similar to voice-attended groups. For both of the two groups, the P2 amplitudes to angry sounds were reduced in CAV compared to the A and IAV conditions whereas the responses to fearful sounds were attenuated in the IAV condition compared to the other two conditions. These results are consistent with Ho et al. (2014), who reporting no significant interactions between the P2 and selective attention tasks. As such, the P2 reflects a process that is less affected by top-down attention.

Despite the finding being less related to top-down attention, angry and fearful information elicited opposite directions of congruency effects within the P2. The P2 amplitudes were decreased when angry body expressions preceded fearful sounds when contrasted with fearful body stimuli. In contrast, the P2 amplitudes were significantly larger when the fearful body expressions preceded the angry sounds compared with when both the sounds and body expressions presented anger. Overall, the present data showed that the P2 amplitude was attenuated when the angry body expression was presented beforehand. Prior work suggests that the P2 is related to assessing unifying perceptual content (van Wassenhove et al., 2005). This assessment, either in the direction toward suppression or facilitation for the congruency effects, relies on the preceding emotional context (Ho et al., 2014). We therefore assumed that the preceding angry body expression may convey a stronger signal which led to a greater expectation for the observers. This is strongly in conflict with the following auditory information, such as fearful sounds. The inconsistent expectation resulted in more attentional cost for reassessment in the following auditory information (Crowley & Colrain, 2004).

The P2 patterns for congruency effects might also be associated with modality dominance, as fear is more easily recognised from the auditory modality compared to visual modality when emotional face and sounds are presented simultaneously; however, the recognition for anger can be dominated either by visual or auditory modalities as it depends on attentional instruction (Takagi et al., 2015). As such, the preceding angry body might be effectively predictive for processing the following angry sound, which decreased the amplitude to minimize the redundant attention resource. In contrast, the fearful information was not a valid prediction through the visual channel even if the intensity of the fearful body expression was as the same as the angry body expression. It is plausible that the difference was more robust between IAV and A conditions than between CAV and A conditions.

2.5. Limitation

Although the modulation of attention on emotion perception was observed in the current study, several issues still remain. We presented the same emotional sounds to match with high and low intensity of body expressions, so the A condition were presented twice as much compared to the CAV, IAV and V conditions. The responses to the repetitive auditory-only stimuli might have become desensitized and attenuated as a result. However, we examined and found no difference in the auditory-only response between the blocks with high to low intensity of body expressions, implying that the modality effects were not due to attenuation in the auditory-only responses. Another consideration is that attention might not be balanced across the two modalities in the unattended task. Even if the participants were instructed to non-emotional visual properties of the stimuli, we cannot exclude the potential attention bias for emotional information from the visual rather than from the auditory modality. Moreover, more studies will be required to understand the modality dominance of emotional body expressions. Prior findings on modality dominance for emotions have been based on facial expressions with sounds. It is not known if the modality dominances are different for body expressions.

Finally, we utilised the same paradigm as Jessen and Kotz (2011) for the auditory-only condition, and it is possible that the static non-emotional (neutral) body images with emotional sounds, may have been interpreted as having emotional content by the participants. The visual information may elicit emotion-related responses that may have influenced auditory processing. To exclude the confound from visual emotion involving auditory processing (A condition), it might be better to present emotional sounds without body expressions .

2.6. Conclusion

The present study investigated the neural processes underlying the integration of emotion perception on body expressions and sounds, with the modulation of attention, emotional types and intensity of body expressions. Through the observation of the two ERP responses, the N1 and the P2, we could understand different processing stages during the integration of emotion perception. The modality effects within the N1 were associated with attentional instructions, with discrepant directions of congruency effects observed with the P2. However, both ERP components are likely to be modulated by modality dominance, which accounts for different results of modalities or emotional congruency effects across anger and fear. Overall, our data provide an important contribution to the integration of emotion processing in body expressions and sounds. However, future investigations are still required to improve a deeper understanding of the neural mechanisms underpinning the emotional perception across modalities, particularly when considering the role of attentional mechanisms.

Prelude to Chapter 3

How does dynamic information influence the integration of emotion perception?

Body expressions are crucial cues for conveying emotional information. In natural environments, body expressions often accompany other modal sources that convey emotional information to perceivers. ERP evidence allows us to understand the deeper integration of emotion processing from body expressions combined with sounds in adults appear to occur at an early sensory level (~ 100ms after onset of sounds) (Jessen & Kotz, 2011; Jessen et al., 2012). Our preceding work (the first study) also found that the auditory N1 and P2 change with variation in factors related to attention, emotional intensity of body expressions, and emotion types. Even though the N1 was largely modulated by attention and visual emotion intensity, the P2 was associated with the combination of emotional content across both visual and auditory modalities. Although the N1 and P2 reflected different functional processes underlying bimodal perceptual integration, both components were modulated by modality dominance which varies across each emotion. Consequently, these results suggest that each factor is linked to a specific processing stage, but some factors may more broadly influence early and late processing.

In addition to the factors we have already examined, this thesis concerned the influence of visual motion on body expressions during emotional audiovisual perception. It is the reason that there seems to be different emotional processing between dynamic and static body expressions during infancy. With presentation of static body expression, Missana (Missana et al., 2014) has shown that the more negative responses (Nc, 700-800 ms) were for fearful than happy expressions at both frontal and central regions. By contrast, the distinct responses between the two emotional expressions were more distributed over temporal and parietal regions at 700-1000 ms when dynamic body were displayed (Missana et al., 2015). The differences in the timing and topography distribution are likely to do with the fact that the expression in the static context is presented immediately, whereas the dynamic expression conveys emotional signals over time. Consequently, how the body expressions displayed is also critical to emotion processing as the brain mechanisms might process emotion differently for motion cues and posture cues

Studies with adults have implicated different neural networks for processing visual emotion in dynamic and static displays. Kilts, Egan, Gideon, Ely, and Hoffman (2003) investigated the neural processes for dynamic and static facial expressions in tasks that involved making an explicit emotional judgement. A greater activation of the superior temporal sulcus (STS) was observed for the emotion of anger in dynamic rather than static expressions. The STS is sensitive to biological motions that are related to a social signal, such as the moving eyes (e.g., Puce, Allison, Bentin, Gore, & McCarthy, 1998). Therefore, this brain area possibly supports how we understand social information, including intention and emotion. In contrast, both angry and happy static expressions were more associated with the activation of the premotor and motor cortices when compared with neutral expressions. The authors assumed that the sensory decoding of emotion shares the same neural activities as motor encoding for producing the same emotion. A link between sensory and motor systems is consistent with the concept of motor theory whereby motor systems simulate an agent's movement for understanding their motor intention (Decety & Grezes, 1999). In addition, activations in left primary sensory cortex were found for the two emotional static images, suggesting the decoding emotion from static images also involve a somatosensory as well as motor representation of the emotional state. Therefore, static expressions may covertly use motor systems to simulate the static percept to its dynamic mental representation. Comparatively, dynamic expressions rely on a lesser strategy of the simulation for understanding emotions. Although this is not a conclusive assumption, uncommon activations of brain regions to dynamic versus static expressions suggest different strategies in the processing of the two types of expressions even when the same emotion is expressed.

Differences in neural activity for dynamic and static displays of visual emotion can also be observed from body expression stimuli. For example, greater activation of the premotor cortex area was found via neuroimaging for dynamic compared to static images when an angry body was displayed (Pichon et al., 2008). The regions of the temporal-parietal junction (TPJ) were more engaged for a dynamic than for a static fearful body (Grezes et al., 2007). However, each emotion can be successfully recognized from body expressions in different ways, with body movements not required for recognition. Evidence by Atkinson et al. (2004) highlighted this, showing a higher accuracy score when identifying anger and fear from dynamic compared to static body expressions. When the body movements were exaggerated, performance improved for the two emotions. In contrast, the improvement did not occur for happiness. Moreover, the performance for sadness was worse when body movements were increasingly exaggerated. It could be interpreted that sadness or grief are often inferred from body postures in motionless natural environments. As such, when the speed of sad movements was artificially increased, observers tended to categorize them as other emotional expressions. Comparatively, anger is generally expressed by bodies with high velocity movements; therefore, faster angry movements were still classified as anger but rated with a higher level of intensity. This explanation also supports Roether et al. (2009b), which showed that anger and happiness were recognized more accurately with faster gait speeds compared to neutral ones, but fear and sadness were associated with smaller movements. Overall, these two findings suggest that kinetic cues play important roles in recognizing high arousal emotions (e.g. anger and happiness) relative to other emotions.

It should be highlighted that static bodies sometimes provide useful information for the recognition of emotions. Static expressions can be identified through critical features related to viewpoints and postures. For example, the elbow-flexion angle can be a key feature for the perception of anger and fear, whereas the perception of sadness can be dominated more by head inclination (Roether et al., 2009b). To address the effects of anatomical variables, body postures and viewpoint that contribute to the recognition of specific emotions, Coulson (2004) presented static body images which varied in weight transfer (backwards and forwards), viewpoint (front, side, rear) and joint rotations of the body. The results showed that angry expressions were characterized by a backwards head bend, arms raised forwards and upwards, and no abdominal twist. It was more likely to be perceived as anger when postures were observed from the front. For fear, head backwards and no abdominal twist were predictive features, but there was no effect of upper arm position. There was less attributed for fear when viewed from the front. Sadness was the only emotion characterised by a forwards head bend as well as forwards chest bend and no twisting. With regard to less well-recognized emotions, motion may be required or other situational cues may need to be present (e.g. surprise), or there may be no standard body expression for the emotion (e.g. disgust) or unrepresentative body posture (e.g. fear). Despite this, a static body is still able to offer a reliable source in which one can identify emotions with specific postures, even if the emotional repertoire is limited when contrasted with other sources of emotional content.

As above discussion, the factor of body expression types also plays an important role in visual emotion processing. Since the present work in the thesis aims to explore development in audiovisual emotion perception, we required an infant-friendly paradigm with an effective stimulus to investigate developing populations. As such, we did not further investigate the factors of attention, type of emotion, intensity and congruency (the first study). Instead, the new variable, type of body expression, was considered in another adults' study (the second study). We expected to observe strongly significant comparisons between auditory-only and audiovisual conditions in adults before investigating these issues with an infant population. In the next study, we presented two types of visual stimulus: angry or fearful body expressions with (dynamic type) and without (static type) movements. To rule out confounds related to the amount of motion between the angry and fearful body expressions, we referred to Jessen's work (Jessen & Kotz, 2011; Jessen et al., 2012) that controlled for pixel changes from frame to frame in each of the video clips. There were four conditions in the second study: auditory-only, visual-only, emotionally congruent and incongruent audiovisual conditions. However, the presentations of the stimulus were amended slightly when contrasted with the first study. For example, in the auditory-only condition affective vocalizations were only presented with a black screen (non-body images) rather than non-emotional body expressions. We reasoned that the body postures might elicit processing related to emotion perception; however, this could confound what we expect to observe when examining the processing of auditory information in isolation. In addition, the video clips were converted into gray scale in order to reduce the possibility that the visual emotions were discriminated between each other due to factors such as the background. The valence of the body expressions was also controlled across emotions (anger versus fear) and visual types (dynamic versus static). We nevertheless expected, the auditory N1 and P2 to reflect modality and congruency effects in the study. These effects are also likely to be modulated by types of body expressions. However, the modulation might be different depending on whether an angry and or fearful expression is displayed.

Chapter 3 Study2

Coherent emotional perception from body expressions and the voice in adults

Text as it appears in Yeh, P., Geangu, E., Reid, V. (2016). Coherent emotional perception from body expressions and the voice. Neuropsychologia, 91, 99-108.

Abstract

Perceiving emotion from multiple modalities enhances the perceptual sensitivity of an individual. This allows more accurate judgments of others' emotional states, which is crucial to appropriate social interactions. It is known that body expressions effectively convey emotional messages, although fewer studies have examined how this information is combined with the auditory cues. The present study used event-related potentials (ERP) to investigate the interaction between emotional body expressions and vocalizations. We also examined emotional congruency between auditory and visual information to determine how preceding visual context influences later auditory processing. Consistent with prior findings (N=18), a reduced N1 amplitude was observed in the audiovisual condition compared to an auditory-only condition. While this component was not sensitive to the modality congruency, the P2 was sensitive to the emotionally incompatible audiovisual pairs. Further, the direction of these congruency effects was different in terms of facilitation or suppression based on the preceding contexts. Overall, the results indicate a functionally dissociated mechanism underlying two stages of emotional processing whereby N1 is involved in cross-modal processing, whereas P2 is related to assessing a unifying perceptual content. These data also indicate that emotion integration can be affected by the specific emotion that is presented.

3.1. Introduction

In our daily life, the perception of others' emotions gives us a good insight into their dispositions and allows us to anticipate suitable responses during complex dynamic social interactions. Emotions are typically expressed through different sensory modalities (e.g., faces and bodies, or vocalization). The combination of multiple emotional cues can be particularly useful in making a more accurate and rapid detection and discrimination of emotional content (de Gelder & Vroomen, 2000; Massaro & Egan, 1996; Van den Stock et al., 2007). This is advantageous in life when information from one modality is unclear (Collignon et al., 2008). For instance, the affective prosody in someone's voice can help us disambiguate the emotional expression of their body posture when this is partially occluded in a crowded room. In order to understand how we process emotions, it is essential to elucidate how emotional information from multiple modalities can be unified into a coherent percept. It is for this reason that this study will investigate how auditory and visual information from voices and bodies are jointly processed.

Body postures are often essential visual cues that convey reliable emotional content (see de Gelder, 2006, for a review). One such circumstance is when attending to distal events, prior to the ability to see an emotional expression displayed on a face. Thus, body expressions provide important complementary emotional information in our daily life (de Gelder & Beatrice, 2009). Electrophysiological (EEG/ERP) data has provided evidence that the processing of emotional information from the body occurs at an early stage of visual processing at approximately 100 ms (Stekelenburg & de Gelder, 2004; van Heijnsbergen et al., 2007). How our bodily expressions interact with other social cues, such as those from the voice, illustrates a further challenging issue. To date, only a few studies on emotion perception have focused on the body and the voice (Jessen & Kotz, 2011; Jessen et al., 2012; Van den Stock et al., 2007). Recently, Jessen and colleagues (Jessen & Kotz, 2011; Jessen et al., 2012) used ERPs to examine neural mechanisms underlying the interaction of emotional perceptions from body expressions and affective interjections. This investigation reported a decrease in N1 amplitude in the bimodal condition compared to an auditory-only condition. The auditory N1 is usually reported at around 100 ms after the sound onset and it has shown sensitivity to sensory information such as intensity or frequency (e.g., Naatanen & Picton, 1987; Naatanen et al., 1988). Other multisensory studies (Besle et al., 2004; Stekelenburg & Vroomen, 2007) also observed the reduction in N1 amplitude to multisensory modalities compared to the sum of the unimodal modalities. If information from each modality was processed independently, the bimodal response is supposed to equal to the sum of unisensory response (AV = A+V). However, if the bimodal response differs from the sum of the unimodal responses in a sub-additive (AV < A+V) or supra-additive manner (AV > A+V), then this points towards interactions occurring between the two modalities (Giard & Peronnet, 1999). As such, the interaction of the auditory and visual information is likely to take place during an early stage of sensory processing. When interpreting how the visual stimuli modulate the auditory processing, van Wassenhove et al. (2005) proposed that the preceding visual stimulus acts as a predictor for the forthcoming information. There might be a deactivation mechanism that minimizes the processing of redundant information for multiple modalities, with the consequence that the auditory cortices decrease responses to the relevant information.

Nevertheless, it should be noted that information from multiple modalities is not always presented simultaneously to an observer. The information from one modality may well precede other modalities within the perceptual system, either in a way of suppression or facilitation (Ho et al., 2014; Takagi et al., 2015). Thus, the preceding one might be a prediction or a constraint to subsequent perceptual processing. However, Jessen's findings (Jessen & Kotz, 2011; Jessen et al., 2012) of the comparison between unimodal and bimodal information does not explore how the preceding visual context influences the processing of different emotions. Irrespective of this, studies on facial expression with voices have exploited the presentation of emotionally conflicted visual and auditory stimuli to reveal the contextual influence on emotional integration. Kokinous et al. (2014) provided evidence that N1 amplitudes were suppressed in both congruent and incongruent auditory-visual conditions with neutral sounds compared to neutral sound-only conditions. The N1 was only reduced in the congruent pairs with angry sounds when compared to the other two conditions. This emotion-specific suppression in N1 was interpreted in terms of the preceding angry visual stimulus being a stronger predictor compared to the neutral stimulus despite the presentation of incongruent information. In that case, the saliency of emotional contexts compared to non-emotional contexts is preferentially processed during early audiovisual integration.

Another component, the P2 (P200), was also reported in response to congruency and incongruency of audiovisual information (Kokinous et al., 2014). The P2 showing a positive deflection at 200-ms post-stimulus is modulated by the emotional quality of a stimulus (Paulmann et al., 2009). The component is also associated with attention to the competition between multisensory incompatible information (Knowland et al., 2014; Stekelenburg & Vroomen, 2007). More precisely, P2 is correlated to assessing a unifying perceptual content, dependent upon preceding contexts (van Wassenhove et al., 2005). Ho et al. (2014) have shown that a suppression of the P2 amplitude occurs for a neutral sound presented with an angry face compared to a neutral face. This could be interpreted as an effect of incongruency. However, the P2 amplitude increased when an angry sound was paired with a neutral face than when both sound and face were angry. The P2 implied the modulation of the previous emotional expression on the following neural responses. As such, it has been considered that the P2 is likely to be functionally separated processes to the N1 component during multisensory integration of the emotional percept. While the N1 is associated with visual anticipation for the following auditory processing, the P2 is considered to be content-dependent processing (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005).

In addition, there seems to be different cognitive processes from one emotion to another. A reduced N1 latency (Jessen & Kotz, 2011) in response to anger was observed when contrasted with a fearful stimulus either in auditory or in audiovisual conditions. Although the authors did not have a conclusive explanation for this effect, several brain imaging studies provided evidence for common and specific neural circuits during the perception of anger and fear derived from body expressions. For instance, the amygdala and temporal cortices were activated when participants recognized both angry and fearful behaviours compared to neutral (non-emotional) ones (Grezes et al., 2007; Pichon et al., 2008, 2009). More specifically, the perception of angry bodies particularly triggered activation within a wider array of the anterior temporal lobes whereas the perception of fearful bodies elicited responses in the right temporoparietal junction (TPJ) (Pichon et al., 2009). Based on these results, it is likely that there are particular neural routes for the perception of angry and fearful body expressions, respectively, which might modulate the integration of emotion perception information differently.

Moreover, moving stimuli and static stimuli may be processed differently.

Generally speaking, dynamic stimuli compared to static stimuli contain explicit movements, which arguably provide more information associated with emotion recognition. Behavioural findings indicate that accuracy rates of emotion recognition for dynamic body expressions are generally higher than for static expressions (Atkinson et al., 2004). Supported by fMRI data, responses to emotions were more pronounced when a body was presented with movement than when a still body was shown. For instance, the expression of fear elicited more activation of the TPJ when displayed in a dynamic compared to a static way (Grezes et al., 2007); and the regions of the premotor cortex were more engaged for the dynamic angry body (Pichon et al., 2008). These more pronounced activation areas for dynamic stimuli are linked to the understanding of actions during action observation; therefore, biological motion is likely to be contributing to emotion understanding (Gallese, Keysers, & Rizzolatti, 2004; Iacoboni, 2005).

This is not to say, however, that static body postures are not a reliable source of information for emotion recognition. With static postures of expressions displayed from three angles to different types of emotions, Coulson (2004) revealed that anger and happiness were accurately recognized for large numbers of postures whereas only a small number of postures were perceived for fear and surprise. Atkinson et al. (2004) also found that the classification accuracy for expressions of anger and fear was improved, but for sadness was impeded when increasing exaggeration presentation of moving body expressions. These results are in line with the natural differences in velocity between different emotional body expressions, with sadness featuring less movement or at times being even motionless, whereas anger is typically associated with a higher velocity movement (Roether, Omlor, Christensen, & Giese, 2009a; Volkova et al., 2014). Taken together, each type of emotion is likely to be optimized specifically, whether in a dynamic or static way, in order to be recognized successfully.

The aim of the current study was to investigate the mechanisms underlying the interaction of emotion perceptions presented in body expressions and affective sounds. We examined ERPs in order to compare both the N1 and P2 to emotions (anger vs. fear) and visual stimulus types (dynamic vs. static body expressions) in three conditions: auditory-only, visual-only and audiovisual. We also included emotionally congruent/incongruent body-voice pairs to explore the influence of the preceding visual context to the bimodal interaction. Since emotion processing is thought to be an automatic response (Mauss, Bunge, & Gross, 2007; Mauss, Cook, & Gross, 2007), we conducted the study without directing attention to the emotional characteristics of the stimuli. Based on previous work (Jessen & Kotz, 2011), the N1 is expected to be reduced in amplitude and increased in speed in the audiovisual when compared with the auditory conditions. This differentiation will particularly be observed with the presence of dynamic visual information. It is predicted that the N1 for the emotions of anger and fear will be different either in terms of latency and/or amplitude, and it will also be modulated by the emotional content within the audiovisual information. The P2 is hypothesized to reflect attention on incompatible information and process content of the binding perception; therefore, it is predicted that this will be influenced by emotional audiovisual congruency and visual type (body expression with/without movement)

3.2. Methods

3.2.1. Participants

Twenty-two students from Lancaster University (5 males) with a mean age of 21.5 years old (*SD* = 4.0 years) participated in this study. Three participants were excluded from the analysis because of fatigue and one further participant was excluded due to poor signal-to-noise ratio compared to other datasets. All participants had normal vision and hearing, and none reported any neurological or psychiatric disorders. Participants provided written informed consent and were paid (£10) for their participation. The study was approved by Lancaster University Ethics Committee.

3.2.2. Stimuli

All visual stimuli were obtained from the research group of Beatrice de Gelder. To compare the motion effect, there were two types of visual stimuli: a video depicting an actor expressing bodily emotions of anger or fear either with movements (i.e., dynamic condition) or with static postures only (i.e., static condition). The static visual stimuli were based on the results of the Bodily Expressive Action Stimulus Test (de Gelder & Van den Stock, 2011) whereas the dynamic stimuli were extracted from those used by Kret et al. (2011). The body expressions for anger included shaking a clenched fist and raising the arm, while fear expressions involved bending the body backwards and defensive movements of the hands. The face area was blurred in all conditions involving the visual modality. The characters were all male dressed in black and performed the body movements against a gray background. The luminance of each video clip was analyzed by taking into account each pixel within a frame (33 frames/clip, 480 × 854-pixel/frame). Each pixel was measured on a gray-scale using MATLAB, with values ranging from 0 to 255. The values of all pixels within a frame were the averaged to obtain a luminance score for that frame. This allowed us to explore any potential variations in luminance that may appear with time due to the velocity and frequency of motion. Following the procedure described by Jessen and Kotz (2011), we found out that the average luminance of the individual frames in the dynamic stimuli ranges from 64 to 68, with differences of no more than 1 between two consecutive frames. The luminance of the static stimuli was slightly lower than that of the dynamic ones, and varied between 30 to 44.

The auditory stimuli were audio recordings of interjections spoken with a fearful or angry prosody. The sounds were produced by male speakers as included in the Montreal Affective Voices database (Belin et al., 2008). All the voices were edited to last 700ms. The mean pitch (anger = 240.47 Hz (*SD* = 60.72); fear = 298.45 Hz (*SD* = 38.02)) and the mean intensity (anger = 71.66 db (*SD* = 9.60); fear = 73.19 db (*SD* = 8.88)) were not statistically different between the two emotional sounds.

In the study, the auditory stimuli with or without the visual stimuli were presented in the following conditions: visual-only (V), auditory-only (A), emotionally congruent audio-visual (CAV), and emotionally incongruent audio-visual conditions (IAV). In the V condition, a video clip displayed either a dynamic (dV) or static human (sV) body expressing emotions in the absence of sound. In the A condition, only a sound was played against a black background. The CAV and IAV conditions played affective sounds with either emotionally congruent dynamic (dCAV) or static (sCAV) body expression, or emotionally incongruent ones (dIAV and sIAV, respectively).

In order to account for the emotional properties of the stimuli, we asked two new groups of participants to judge the emotions and rate the intensity of the visual-only (N = 20) and the audiovisual stimuli (N = 20), respectively. For rating the intensity of the stimuli, we used a 5 point Likert scale ranging from 1 (= very weak) to 5 (= very strong). The **Table 3.1.** shows the mean accuracy in identifying the emotion and the mean intensity, with standard deviation in brackets (D = dynamic body; S= static body)

Table 3.1. Results of rating for the stimuli presented in the EEG study. Mean accuracies (%) and intensity (1 to 5 scale) for emotions of angry and fear in V (visual-only condition) and CAV (congruent audiovisual condition), with standard deviant in parentheses.

	V				CAV			
	Anger		Fear		Anger		Fear	
	D	S	D	S	D	S	D	S
Accuracy	95.24%	100%	100%	97.62%	100%	97.62%	100%	100%
	(0.15)	(0)	(0)	(0.11)	(0)	(0.11)	(0)	(0)
Intensity	3.60	3.35	4.50	3.35	3.65	3.62	4.38	2.73
	(0.90)	(0.80)	(0.51)	(0.49)	(0.69)	(0.92)	(0.58)	(0.62)

D = dynamic visual stimulus; S = static visual stimulus

3.2.3. Procedure

Participants sat comfortably in a dimly lit/darkened room, and were asked to make their response by pressing a button. Each stimulus was presented using the Psychtoolbox 3.0 in Matlab 2012a. The visual stimuli were presented on a monitor at a distance 90-100 cm from the participants, and the auditory stimuli were binaurally played via two speakers at a sound pressure of 70 dB for all participants. Each trial started with a 800-ms white fixation on a black screen, followed by the presentation of a video clip (CAV, IAV and V condition) or a black background (A condition) for 1300 ms. An interval randomised between a fixation and a video clip (visual stimulus) from 800 to 1200 ms. The auditory stimuli were shown 600ms after the onset of the visual stimulus and ended synchronously with the video clips. In V, CAV and IAV conditions, participants were required to indicate what the person in the video was wearing (e.g., "Did the person wear a jumper/belt?") by pressing the left or the right button. A question mark was also presented in the A condition, and participants also pressed the space bar as a response without any judgement. The question mark disappeared once the participants had made their response. Each block included 64 trials. In order to avoid learning the regularities of question marks presentation, in each block we randomly showed them after a trial in less than 60% of the cases (ranging from 20 to 33), by using a custom Matlab script. The presentation of a question mark after a trial was presented less than 5 consecutive times. The testing started after a practice session consisting of 10 trials, and the participants were able to take a self-defined break between blocks if required. The study consisted of 8 blocks, a total of 512 trials. In each of the 4 blocks, either the dynamic or static body expression (V) was presented (8 times/block) together with other factors of *condition* (A, CAV, IAV conditions) and *emotion* (anger and fear). The study lasted approximately 50 minutes, including breaks.

3.2.4. EEG recording and analysis

The data were recorded by EGI NetStation system (Geodesic Sensor Nets, Inc., Eugene, OR) with a 128-channel electrode net. The EEG signal was sampled at 500 Hz and the impedances were kept to 50 Hz or less during recording. All electrodes were on-line referenced to vertex (Cz). For computing the ERPs, the data was filtered with a 0.3-30 Hz bandpass filter and segmented off-line from 100 ms before to 700ms after sound onset. Baseline correction was applied to 100 ms prior to each segment before artifact rejections. Trials were rejected with EGI software once the eye movement exceeded +/- 140 uV, and eye blinks exceeded +/- 100 uV. Any channels that exceeded over +/- 200 uV for an electrode were marked as bad. If more than 12 electrodes within a trial were marked as bad, the trial was automatically discarded. The remaining trials were re-referenced into an average reference before averaged waveforms for each participant with each condition. The analysis was focused on the two ERP components, N1 and P2, which have been indexed in audiovisual emotion perception literature. Based on previous studies (e.g. Jessen & Kotz, 2011), and visual inspection of present data, two different analyses were conducted: the first involved the latency to the peak amplitude between 90-180 ms (N1) and 160-330 ms (P2) after sound onset, and the second involved the mean peak amplitude for the time window centered on the latency of each conditions (+/- 30 ms).

As the distribution between frontal-central and central-parietal sites showed a reversed polarity of the potentials, the statistical analysis were therefore performed individually, taking the average of these electrode clusters for frontal (6, 11, 19, 4, 12, 5), central (Ref/Cz, 7, 106, 80, 31, 55) and central-parietal (62, 61, 78, 79, 54) regions of interest (ROI) (**Figure 3.1.**). A 2 (*visual type:* dynamic, static body expression) x 4 (*conditions*: audio-only, visual-only, emotionally congruent audiovisual, and emotionally incongruent audiovisual) x 2 (*emotion*: anger, fear) x 3 (*ROI*: frontal-central, central, central-parietal sites) repeated-measures ANOVA was conducted on the two time windows. Post-hoc analyses (least significant difference) were run where any significant (*p*-value < 0.05) interaction effects were reported.

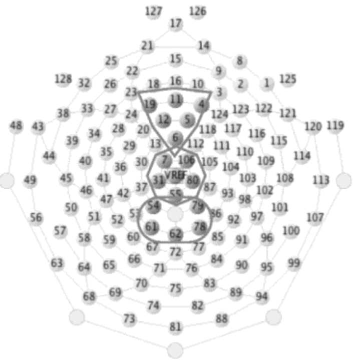


Figure 3.1. Averages were calculated based on electrode ROIs for frontal (6, 11, 19, 4, 12, 5), central (Ref/Cz, 7, 106, 80, 31, 55) and central-parietal (62, 61, 78, 79, 54) channels in study2

3.3. Results

The topography and the grand average of the N1 and the P2 at sequential time from 100 to 350 ms for each condition are presented separately for the dynamic (**Figure 3.2.**) and static (**Figure 3.3.**) visual stimuli. In the following sections, we only reported the key findings, particularly the comparison of condition for visual types (dynamic and static) and for emotional content (anger and fear) as we were interested in modality and congruency effects. A full list of all statistical comparisons can be found in the **Table 3.2.**

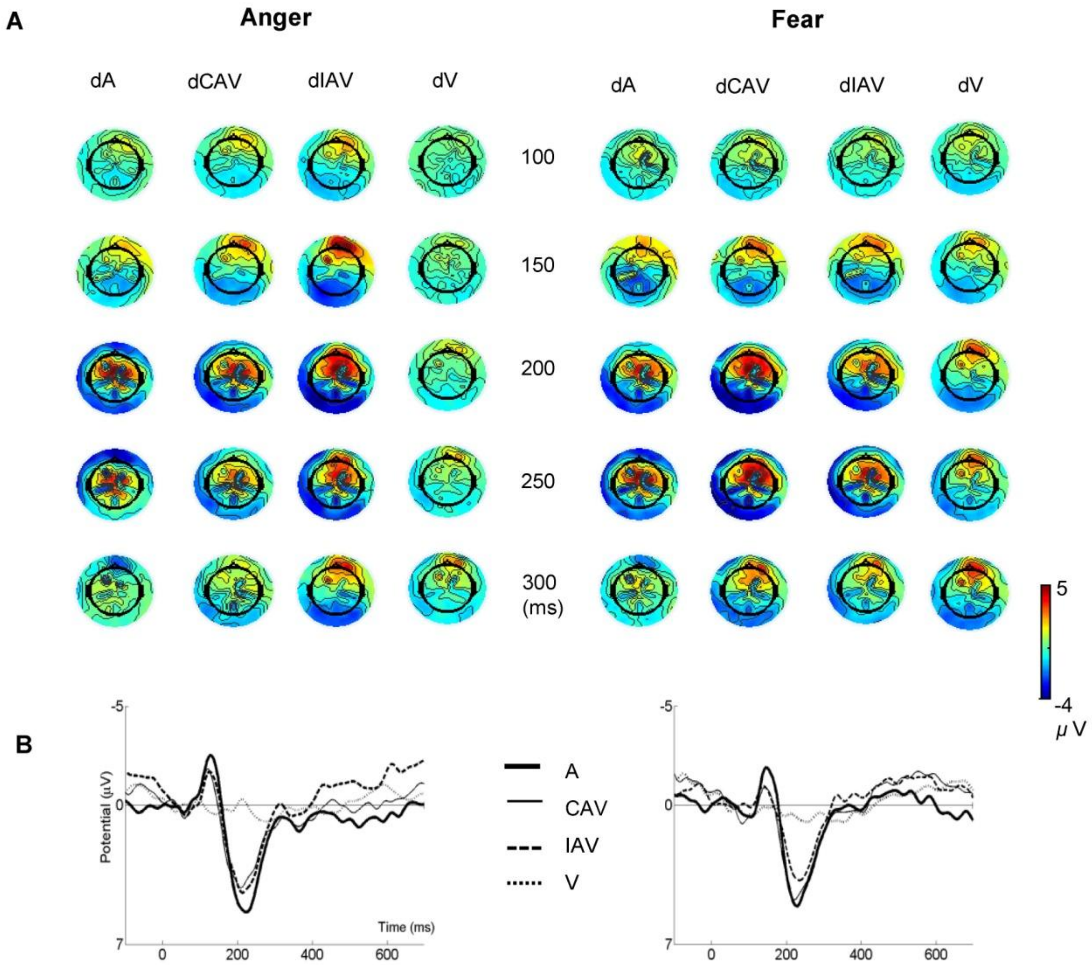


Figure 3.2. The ERPs displaying (A) the topography distributions for angry (4 left) and fearful (4 right) information in dA, dCAV, dIAV and dV conditions from 100 to 350 ms after onset auditory stimulus when the dynamic body expressions were presented. (B) The grand average for each condition at central electrode sites

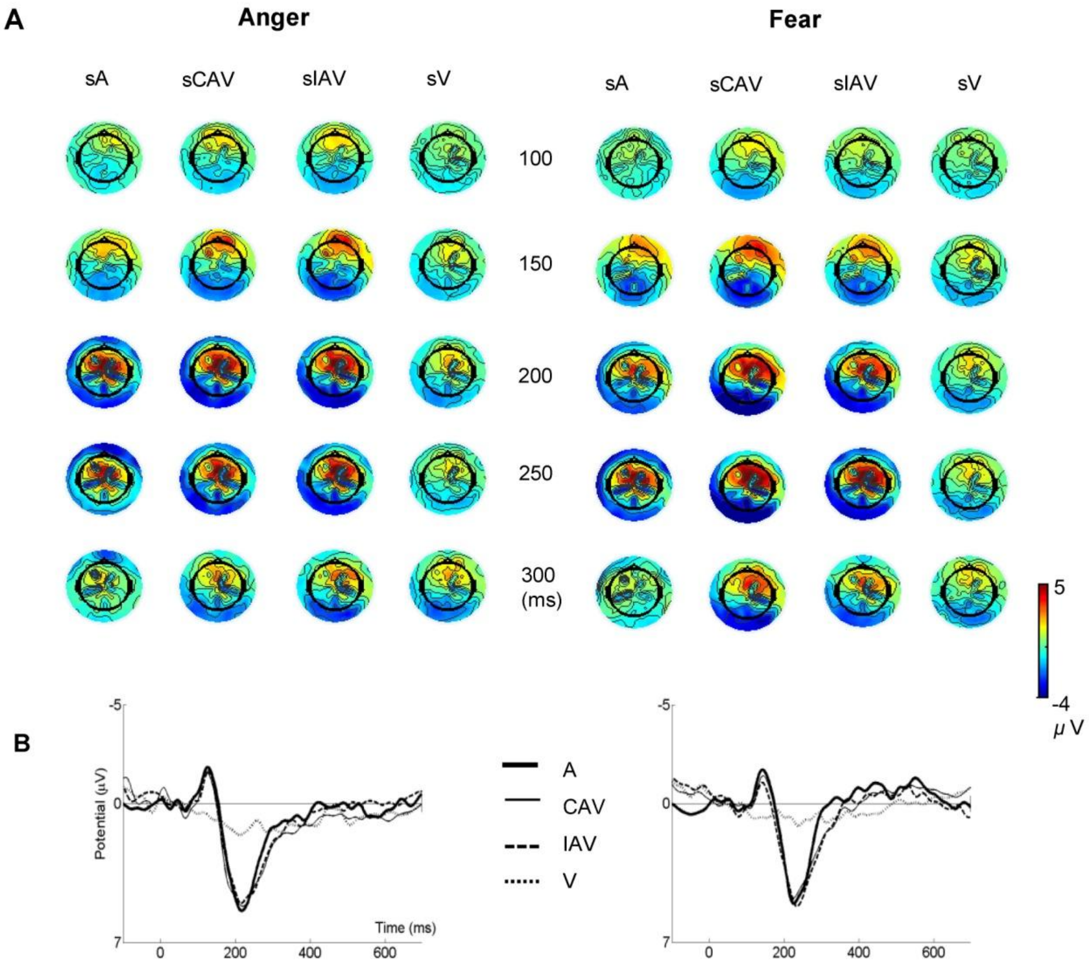


Figure 3.3. The ERPs displaying (A) the topography distributions for angry (4 left) and fearful (4 right) information in sA, sCAV, sIAV and sV conditions from 100 to 350 ms after onset auditory stimulus when the static body expressions were presented. (B) The grand average for each condition at central electrode site

Table 3.2. A summary of statistical analysis * *p* < .05; ** < .01; *** < .001

			N1 Amp		P2 Amp		N1 latency		P2 latency	
		df	F	*p*	F	*p*	F	*p*	F	*p*
Visual type		1,17	1.61		11.55	**	0.62		0.11	
Condition		3,51	15.98	***	38.42	***	0.83		0.66	
	CAV v.s A			*		.071				
	CAV v.s V			***		***				
	CAV v.s IAV									
	IAV v.s A			*		*				
	IAV v.s V			***		***		*		
	A v.s V			***		***				
Emotion		1,17	1.43		4.79	*	62.65	***	9.47	**
Site		2,34	50.80	***	56.75	***	14.21	***	12.43	***
Type*condition		3,51	1.71		0.81		0.88		3.67	*
Type* emotion		1,17	3.18	.092	5.45	*	0.60		5.77	*
Condition *emotion		3,51	0.37		0.26		3.26	*	3.41	*
Type*condition * emotion		3,51	0.86		1.45		0.44		4.12	*
Type *site		2,34	3,28	.050	6.28	**	0.10		1.41	
Condition *site		6,102	9.74	***	9.09	***	2.19	*	13.94	***
Type*condition * site		6,102	0.95		1.03		1.01		0.92	
Emotion *site		2,34	5.09	*	15.19	***	33.49	***	0.57	
Type*emotion *site		2,34	0.56		0.23		1.11		1.31	
Condition *emotion*site		6,102	5.35	***	10.36	***	2.74	*	0.45	
Type*condition *emotion*site		6,102	1.87	.093	2.45	*	1.33		1.24	

3.3.1. ERP latency

3.3.1a N1

Only the main effect of *emotion* (F(1,17) = 62.65, p <.0001, η^2 = .787) reached significance. A significant interaction between *emotion*, *condition* and *site* (F(6, 102) = 2.74, p =.017, η^2 = .139) (**Table 3.3**) was also found. *Post hoc* analysis of the interaction indicated that the N1 response to the angry stimuli peaked earlier than to the fearful stimuli, and the difference was most enhanced in both A and CAV conditions at central and central-parietal sites (all p < .0001).

Table 3.3. The mean in milliseconds of N1 peak latency for each condition at frontal-central (FC), central (C) and central-parietal (CP) sites (SD in parentheses)

		N1					
		anger			fear		
		FC	C	CP	FC	C	CP
Dynamic visual type	A	120.4 (22.54)	126.0 (12.34)	128.0 (14.26)	130.2 (27.72)	147.3 (15.44)	151.7 (10.81)
	CAV	113.2 (19.11)	122.1 (12.19)	126.2 (12.66)	116.30 (24.00)	128.2 (20.24)	155.6 (11.38)
	IAV	113.2 (22.00)	124.0 (15.46)	144.7 (15.69)	125.1 (22.83)	138.7 (20.21)	148.0 (16.57)
	V	137.7 (24.39)	125.1 (23.88)	127.5 (26.63)	120.8 (25.52)	142.0 (24.49)	144.7 (20.34)
Static visual type	A	117.4 (25.87)	127.6 (14.38)	123.0 (13.21)	124.0 (30.49)	140.60 (13.88)	153.6 (14.70)
	CAV	116.2 (21.74)	123.7 (14.37)	129.7 (13.21)	124.0 (30.49)	140.6 (13.88)	153.6 (14.69)
	IAV	115.4 (21.84)	123.2 (15.83)	134.4 (14.33)	124.1 (27.81)	135.4 (13.78)	145.5 (14.85)
	V	125.2 (24.47)	128.0 (21.37)	133.1 (24.17)	122.0 (24.87)	130.54 (27.72)	145.2 (25.19)

In addition to emotion effects, we also considered the comparison of the conditions. However, no significant effects were found when the three-way interactions (*emotion, condition and site*) were unpacked by the other two factors. The *condition* only showed a significant two-way interaction with *emotion* (F(3, 51) = 2.67, p =.029, η^2 = .151). Further analysis showed a shorter N1 latency was found for the angry stimulus in the CAV than in IAV condition (p = .022), whereas the latency was only reduced in the IAV compared to the A condition (p = .031) for the fearful stimulus.

3.3.1b P2

Only the main effect of *emotion* was significant (F(1,17) = 9.47, p = .007, η^2 = .358), revealing a rapid latency to the P2 peak for the angry compared to the fearful stimuli. The *emotion* also showed significantly interactions with *type* and *condition* (F(3,51) = 4.12, p = .011, η^2 = .195) (**Table 3.4**). Further analysis showed the different latencies between emotion were pronounced in the sounds-only condition (dA and sA: all p < .0001), and sounds with dynamic visual information (dCAV: p = .031; dIAV: p < .0001). However, the emotion effects were reduced when sounds were presented with static body expressions (sCAV: p = .042; sIAV: p = .024).

With regard to the condition effects, we only found the difference when the static body expressions were presented. Shorter latencies to angry sounds were observed in both sA and sCAV compared to sIAV conditions (p = .01; p < .0001, respectively). The peak was shorter for sCAV than for sIAV conditions when sounds were fearful (p = .049).

Table 3.4. The mean in milliseconds of P2 peak latency for each condition at frontal-central (FC), central (C) and central-parietal (CP) sites (SD in parentheses)

		P2					
		anger			fear		
		FC	C	CP	FC	C	CP
Dynamic visual type	A	203.7	222.8	245.0	224.6	232.8	263.3
		(19.64)	(16.41)	(24.45)	(23.83)	(15.77)	(26.58)
	CAV	212.0	222.8	243.9	231.4	233.8	252.5
		(30.69)	(28.01)	(35.37)	(21.26)	(18.53)	(24.27)
	IAV	214.4	213.0	240.11	228.3	239.5	256.4
		(16.52)	(18.11)	(26.19)	(29.65)	(23.92)	(19.91)
	V	248.3	248.0	234.87	255.3	237.8	229.4
		(36.22)	(42.06)	(36.10)	(26.85)	(35.56)	(41.65)
Static visual type	A	212.1	225.2	253.1	216.8	232.6	257.9
		(23.28)	(20.04)	(27.17)	(23.52)	(12.01)	(27.97)
	CAV	218.9	222.3	238.3	232.2	239.0	245.78
		(21.72)	(25.18)	(31.00)	(26.93)	(26.26)	(35.71)
	IAV	231.5	241.3	253.3	223.3	230.1	243.5
		(25.59)	(17.24)	(30.77)	(23.87)	(26.84)	(31.65)
	V	242.9	241.0	221.60	227.3	242.4	227.6
		(38.07)	(34.81)	(29.38)	(38.59)	(36.40)	(34.42)

3.3.2. ERP amplitude

Figure 3.4. shows the mean peak amplitude of the N1 (top) and of the P2 (bottom) components across emotions, visual type, conditions.

*3.3.2a N*1

A significant main effect of *condition* was found (F(3,51) = 15.98, p < .0001, η^2 = .485), with reduced N1 amplitudes in both CAV and IAV conditions compared to the A condition (p = .015 and p = .018, respectively). Of interest is the marginally significant four-way interactions between *condition, emotion, visual types* and *sites* (F(6,102) = 1.87, p = .093, η^2 = .099). When separated by *visual types, emotion,* and *sites*, smaller N1 amplitudes were observed for angry dCAV and dIAV conditions compared to the dA condition at frontal (*p* = .005; *p* = .001, respectively) and central sites (*p* = .014; *p* = .041, respectively). Conversely, no significant differences were found between conditions with static body expressions (sCAV vs. sA, *p* = .375; sIAV vs. sA, *p* = .282). In response to the fearful sounds, a reduced N1 amplitude for dCAV was found when contrasted with dA at central regions (*p* = .036). However, the reduced N1 was less significant for sCAV and sIAV compared to sA conditions at frontal sites (*p* = .018; *p* = 0.088, respectively)

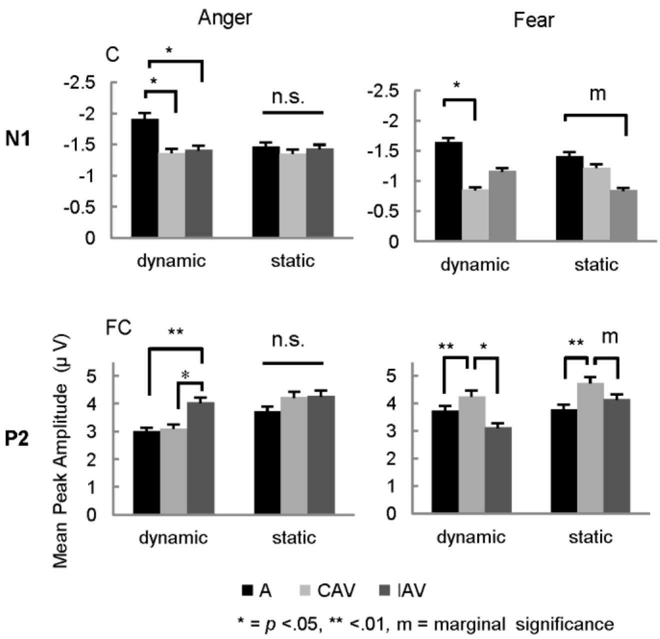


Figure 3.4. The N1 and P2 mean peak amplitudes for each factor (condition, visual types and emotions), which is indicative of effects in the region.

3.3.2b P2

We observed a significant main effect of *condition* (F(3,51) = 38.42, *p* < .0001, η^2 = .693). The *post ho*c analysis indicated that smaller P2 amplitudes were observed for IAV in comparison to A conditions (*p* = .017), but the reduction become less significant in CAV compared to A condition (*p* = .071). In addition, *type* (F(1,17) =

11.55, *p* = .003, η^2 = .405) as well as *emotion* showed significant main effects (F(1,17) = 4.79, *p* = .043, η^2 = .220). Planned comparison revealed a reduced P2 for dynamic compared to static visual stimuli, as well as for angry than for fearful expressions. Significant interactions between *type, condition*, *emotion* and *site* were also found (F(6,102) = 2.45, *p* = .030, η^2 = .13). Further analysis was separated by *visual types*, *emotion*, and *sites.* Generally, the differences between conditions were more pronounced with the presentation of dynamic contrasted with static body expressions. With presentation of the dynamic angry body, smaller P2 amplitudes were found in both dA and dCAV conditions compared to the dIAV condition at the frontal regions (*p* = .011; *p* = .001, respectively). In addition, smaller responses to dCAV compared to dA conditions nearly achieved significance at central sites (*p* = .085) but became robust at central-parietal sites (*p* =. 013). However, no significant differences were found in response to angry stimuli when the body expressions were static. In contrast, reduced P2 amplitudes were found for the fearful IAV condition compared to dA at frontal and central sites (*p* = .033, *p* = .005, respectively), and for the dIAV compared to dCAV conditions at frontal sites (*p* = .004). When static body expressions were presented, only larger P2 amplitudes were observed for the sCAV compared to sA condition at frontal regions (*p* = .003).

3.4. Discussion

In the current study, we used ERPs to measure the integration of emotion perception from body expressions and affective interjections. Both the emotion and the presence of dynamic visual information significantly modulated both the N1 and P2 components. However, the modality in which the emotional information was presented significantly affected the N1, whereas the effect of the congruency between visual and auditory information was only observed within the P2. These findings indicate that processing the interaction between visual information, related to body posture, and auditory information, specific to prosody, during emotion perception may occur at different stages, as reflected by the response of the N1 and P2 components. The influences of modality, visual type, emotion and audiovisual congruency within these two components will be discussed in more detail below.

3.4.1. Modality Effects

In agreement with the studies of Jessen and her colleagues (e.g. Jessen & Kotz, 2011; Jessen et al., 2012) on emotional integration from body postures and prosody, we found both reduced N1 latencies and amplitudes for the emotionally congruent and incongruent audiovisual compared to voice-only conditions. This observation is also consistent with other previous investigations of audio-visual integration outside the emotional domain (e.g. Stekelenburg & Vroomen, 2007), suggesting that the interaction of body posture and voice information occurs at a very early stage of perception. In addition, these modality effects in both amplitude and latency are likely to be activated by unspecific emotional information as the N1 was suppressed in both angry and fearful contexts.

3.4.2. Comparison between Emotions

The reduced N1 latency for anger was robustly found when compared with fearful stimuli in the auditory-only and audio-visual conditions. Since the N1 is interpreted as a sensory component, it shows that faster processing for anger than for fear at a very early stage, rather than a later stage of processing (Paulmann et al., 2009). With regard to the emotion component, both anger and fear are associated with high arousal and negative valence, yet they convey quite different social signals. In comparison to fear, anger often displays cues about the expressers' intentions to act, so it is an interactive message that requires observers to modify their behaviour in tune with the approaching interaction (Pichon et al., 2009). Neuroimaging studies also have demonstrated that the perceptions of the two emotions are different. For instance, the premotor area and temporal lobe, activate more when one perceives an angry rather than a fearful body (Pichon et al., 2009). The authors proposed that the function of the premotor area is to readjust our defensive behaviour in response to one monitoring a forthcoming threat, and the temporal area evaluates the emotional contexts by drawing from past experience. Consequently, this additional activation is crucial for one to be sensitive to the detection of anger, improving their social relationships.

However, the current differences between emotions could be the fact that the fearful stimuli we used do not evoke threat in the observer as efficiently as the angry stimuli. It has been indicated that if the expresser’s signals of emotions are directed to or successfully shared with the observer, these might become threats to the observer which require an adjustment in their behaviour (Adams & Kleck, 2003). Therefore, whether the emotional signals are clearly related to the observers is likely to influence the observer's emotion perception.

3.4.3. Congruency Effect

The differentiation between response to congruent and incongruent audiovisual conditions was not reflected by the N1, which is consistent with the findings of Stekelenburg and Vroomen (2007), but is in contrast to two other prior studies (Ho et al., 2014; Kokinous et al., 2014). Several reasons might influence the results. Firstly, Ho et al. (2014) and Kokinous et al. (2014) observed facial expressions whereas we presented body expressions as visual information. It is possible that perceiving emotions from bodily expressions is less sensitive than facial expressions at an early stage of processing; that is, faces compared to bodies appear to be better predictors of the auditory emotional information. Secondly, different combinations of emotions may modulate the congruency effects differently. Previous studies examined the audiovisual congruency effect by mismatching angry and neutral information, which is different from the present study, which paired the expressions of anger and fear. Both anger and fear are negative emotions conveying a message of threat, so it might be difficult to perceive the difference when the two emotions are displayed through separate modalities simultaneously. Differences could also arise due to distinctive methodology in analysis and instruction, and so further studies are required to demonstrate this assumption.

Although a congruency effect at the level of N1 was absent in our study, the same component reflected predominance of the information from bimodality than from unimodality. Conversely, only a significant congruency effect was observed for the P2 amplitude at frontal-central regions, which was specific to moving body expressions. These results suggest that the processing for the modal interaction emerges at an early sensory stage, but for conjunctions of emotional contents occurs at a later stage (Kokinous et al., 2014). The discrepancy within the two investigated components is also in line with the assumption that the AV effect on the N1 is modulated by visual anticipation but is independent of audiovisual coherence, whereas the P2 is driven by AV coherence and more dependent on specific contents (Ganesh, Berthommier, Vilain, Sato, & Schwartz, 2014).

More precisely, the direction for the congruency effects, either the suppression or facilitation within P2, might depend on the preceding emotional contexts (Ho et al., 2014). The current data has shown that the P2 amplitude reduced when the angry body presentation preceded the fearful sounds compared with when the fearful sounds were paired with fearful body expressions. Conversely, the P2 amplitude increased when the fearful body preceded the angry sounds compared with when both the sounds and body expressions presented anger. The preceding angry body expression may be considered to convey a strong signal and lead to a greater expectation by the participant. This is strongly in conflict with the following fearful sounds, leading to reassessing the stimulus with consequent processing costs and attention (Crowley & Colrain, 2004). However, the fearful image seems not to carry this message as strongly as the anger stimuli; therefore, the P2 was not suppressed to the incongruent combination with the angry sounds.

An alternative perspective for the reversed congruency effects may be related to the different dominance within separate modalities for the two emotions. The modality dominance might be different for each type of emotion when presenting audiovisual information (Takagi et al., 2015) as voice dominance was shown for fear, whereas anger was most linked to a visual modality. Considering the auditory-only condition as a baseline, we observed that the amplitude of the P2 was reduced whenever the angry body expression was displayed before the voices, whereas no significant effects appeared within the P2 when a fearful body was presented. In that case, the image of an angry body might serve as a very strong predictor, modulating the brain responses irrespective of the information provided subsequently by the emotional voices.

It has also previously been suggested that the P2 could represent a general stimulus classification process (Garcia-Larrea et al., 1992), and that the mismatched audiovisual pairs might yield new percepts. A noticeable example is the McGurk effect (McGurk & MacDonald, 1976), which comprises a speech sound (/ba/) overlaid with a face articulating another sound (/ga/) resulting in a fused percept (/da/), whereas a reverse combination of the auditory (/ga/) and visual (/ba/)is perceived as

/bga/. In that case, the combined information is likely to be perceived differently when the emotional information was reversed from the two modalities. Based on this assumption, in our study, the perception for the four types of combinations from two modalities during the processing of two emotions might be different, with consequent results found within the P2 in terms of latency or amplitude.

3.4.4. The Modulation of Motion

The current study showed that visual types of emotional body expressions are relevant for multisensory emotion processing. In particular, both modality and congruency effects were observed within N1 and P2, respectively when presenting dynamic materials; however, the results were not entirely extended to the static stimuli, especially for the angry stimulus. In support of the assumption that the kinetic cues from visual information fasten the process for the following auditory information (Stekelenburg & Vroomen, 2007). Some neuroimaging studies have provided evidence that specific brain areas are activated when one perceives a dynamic represented body compared to a static one (e.g. Pichon et al., 2008). The additional engagement of brain areas, such as the premotor cortex, are noted for the perception of biological motion, but have also been observed for the processing of understanding emotion (e.g. Iacoboni, 2005). Consequently, our results feed into a literature that indicate that viewing a dynamic angry body activates sensory regions as well as motor areas, which helps one to understand the emotion that is being portrayed.

On the other hand, the benefit of dynamic cues appears to partially apply to the recognition of fear. Although a larger activation of the premotor area has also been reported for a fearful body in dynamic compared to still states (Grezes et al., 2007), our data nonetheless indicated N1 suppression for the fearful audiovisual condition in static conditions. The more easily recognized fearful stimuli might be related to the angles within the postures for the current static body stimuli, whereas this may not be the case for anger (Coulson, 2004). In addition, a fearful body is often well recognized with fewer high velocity movements than angry expressions (Roether et al., 2009a). On this basis, we have assumed that a still body presentation is sufficient for the discrimination of fear.

3.5. Limitations

There are some potential limitations related to the present study. First, attention might not be balanced across the two modalities. We tried to divert participants' attention from the emotional information by asking them to make judgments about the non-emotional visual properties of the stimuli. However, this may not have fully removed attention from visual information, and attention away from auditory processes. Also, the auditory condition cannot be displayed in a dynamic and static way. We consequently presented twice the A compared to the other CAV, IAV and V conditions. To ensure the effects of the auditory-only were not attenuated, we examined the response in the blocks with each visual type and found no differences. Another limitation might be due to the fact that the emotional intensity of the fearful stimuli was higher for the dynamic than for the static presentations. However, the modality effect can be observed for the static fearful body expression but not for the angry static expression with the same intensity, which suggests that the emotional intensity and the biological motion per se are not the main contributors to the observed effects. Other factors might contribute more specifically to the perceptual integration of fear. Moreover, males and females are known to differ in processing emotional prosody (e.g. Schirmer & Kotz, 2003; Schirmer, Kotz, & Friederici, 2002) and this might be an important aspect to consider when investigating emotional information processing. As such, the lack of balance with respect to gender in the present study is a restriction for generalizing the findings to broader populations. However, the present study is more focused on understanding whether different types of emotions and body exhibitions influence the emotion perception from body expression and sounds. Given the number of variables in the current study, which currently features four factors (visual types, condition, emotions, and sites), the addition of another factor would dramatically increase the complexity of the study and the associated interpretations of results. In this case, we reasoned it is better to address gender issues across body expression and voice in future studies.

3.6. Conclusion

The present study reiterates the findings of Jessen and Kotz (2011) indicating a clear suppression of the N1 amplitude and latency for the emotionally congruent and incongruent audiovisual conditions than for auditory-only condition. Moreover, we have clearly shown that the availability of dynamic information about body expressions aids emotion processing, particularly at later stages as indexed by the P2. The N1 and the P2 were separately influenced by the presence of multimodal emotional information and their congruency, leading us to conclude that these components index different emotion processing functions. The current evidence supports the previous assumption that the N1 is affected by multisensory signals in a manner that is independent of congruency information, whereas the P2 is sensitive to the coherence of the integration of emotional content.

Prelude to Chapter 4

At what age that the capacity of the integration of audiovisual perception from body expression and sounds develop?

During the past ten years, a growing body of studies has indicated that body expressions are important visual social cues in order to understand others' emotions. When considering an ecological approach, we typically perceive emotion not only relying on a single modality, but also combining each individual modality with other modal information. It is therefore a crucial issue to understand how multisensory perceptions cohere into a unified perception rather than separated perceptions. Using ERP measurements, Jessen and Kotz (2011) provided a first step to explore emotion perception from body expression combined with affective sounds in adults. Their findings demonstrated that multimodal interactions during audiovisual perception occurred at an early sensory stage. However, the field is very sparse related to the integration of emotional information relevant to the body at the developmental level. To date, two ERP studies have shown that 8-month-old infants can discriminate emotions of happiness and fear from body expressions (Missana et al., 2015; Missana et al., 2014) Recently, a behavioural study by Zieber et al. (2014b) further found that the capacity for extracting body expressions to emotionally matched sounds emerges by 6.5-months. With preference looking measurements, infants tended to look longer to the corresponding emotional body expressions when angry/happy vocalizations were presented. Nevertheless, some questions about the integration of emotion perception in early development are still unanswered. For example, below infants' behavioural responses, it is unclear whether they automatically select the emotionally congruent pairs or whether attentional resources need to be allocated for this to occur; or how body expression with the presentation of emotional sounds sensitizes infants' perception compared to only when sounds are presented. ERP paradigms could help to resolve these issues. By examining infants' neurophysiological responses to multisensory information, we may be able to understand the neural mechanism underlying emotional audiovisual integration, including perceptual and cognitive processes. When contrasted with adults' data, a developmental change in the processing of perceptual integration can be observed. Most importantly, an ERP study does not require any overt behavioural responses, which is suitable for research with infant cohorts.

Reviewing previous research on multisensory perception in infants, including emotion, language or other domains of perception, two paradigms are regularly used to investigate the interaction across multisensory perception in infants. One is to compare congruency across multimodal information (Bristow et al., 2009; Grossmann et al., 2006; Otte et al., 2015). For instance, Grossmann et al. (2006) observed the response to emotionally (in)consistent pairs (happiness/fear) between facial expressions and sounds in 7-month-old infants. A greater negative response (Nc) peaking around 400-600 ms after sound stimuli was found for incongruent pairs compared to congruent pairs at frontal-central sites. The Nc is thought to reflect more attention involvement to salient or familiar visual stimuli for infants (Ackles & Cook, 1998; de Hann & Nelson, 1999). As a result, infants are likely to be aware of appropriate affective information between face and sounds at an early age. Another measurement is to involve manipulating the temporal synchrony across visual and auditory stimuli (Hyde et al., 2011; Reynolds et al., 2014). The Nc was observed to differentiate synchronized audiovisual stimuli from asynchronized stimuli. However, the Nc obtained from the two methods should be linked to attention processing in infants, which is a higher-level cognitive process (Csibra, Kushnerenko, &

Grossmann, 2008). Given that past research has focused on the Nc, it is unknown whether the interactions related to audiovisual integration have already taken place at the perceptual level, that is, occur earlier than 400 ms in infants.

In our adult's data (study1 and study2), the modality and congruency effects were found within N1 and P2, respectively. In addition, the specific congruency effects, that is, either an increase or a reduction in P2 amplitudes in the audiovisual condition compared to auditory-only responses, were different for anger and fear. Thus, we inferred that the N1 and P2 are two dissociated processes during emotional perceptual integration. While the N1 component reflects the interaction between multisensory perceptual systems, we have related P2 to the assessment of the combined audiovisual emotional content or the competition across the bimodal information. Consequently, we hypothesized that the function of the Nc might be similar to the P2, and is considered to be related to competition processing or the assessment of the content of multimodal information. There might be another process that occurs earlier than the Nc for multisensory processing during infancy.

To understand the integration of multisensory perception at the perceptual level, we should compare unisensory responses to multisensory processes (Giard & Peronnet, 1999; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). This method has been widely used in adult studies, observing that the N1 (a negative peak at 100ms) and P2 (a positive peak ~200 ms) amplitude reduce to multimodal stimuli than to the sum of unimodal stimuli. The N1 and P2 are typical cortical auditory evoked potentials (CAEPs) in adults that do not require the listener's attention (Trainor, 2007; Wunderlich & Cone-Wesson, 2006). Nevertheless, it is challenging to apply typical adult paradigms to infants. The definition of clear components of CAEPs in infancy is difficult. Due to an immature primary auditory cortex, infants' morphology of CAEPs are dissimilar to adults' responses. The waveform patterns also change dramatically across the first postnatal year, with varying transitions between positive and negative polarities (Kushnerenko et al., 2002). Furthermore, a large variance in peak latency and amplitude can be seen across individual infants, which looks to be flat and difficult to determine statically significant regions for effects. Further complexities involve the presentation of differential polarity waveforms that can be induced in different infants, with some infants responding a positivity whereas some show a negativity during the same time window (Trainor, 2007).

Despite this, several components can be recognized through the grand average in most studies with infants (see Coch & Gullick, 2011; Wunderlich & Cone-Wesson, 2006 for reviews). Unlike adults, infant's auditory ERPs are often reported as a predominantly positive peak (~100 to 300 ms) followed by a broad negative response. The positive response has been defined as an infantile P2, which peaks around 250 to 300 ms after the onset of the auditory stimulus at frontal-central sites. However, several studies have observed double positive peaks during the latency of 300 ms. For example, Kushnerenko et al. (2002) recorded infants' cortical responses to complex tones from birth until 12 months. Infants at 3-month showed peaks at 150 and 350 ms, labeled as P150 and P350, respectively. Furthermore, the P150 amplitude remained unchanged from 6-12 months of age whereas the P350 predominantly disappeared around 6-9 months. The author considered that the emergence of two components in infants may reflect separated neural processes already at birth. Instead, a negative peak around 250 ms (termed N250 by author) through the double peaks was discernible from 6-months of age and increased until 12 months. Another negative component (N450) peaking between 350-600 ms was also found to increase in amplitude but decrease in peak latency after 6 months of age. The author considered that the amplitude P350 decreased during the period due to the overlap with the increasing N250 and N450.

The following chapter investigates the audiovisual integration of emotional content in early development by comparing auditory ERP responses in unisensory and audiovisual conditions. The aim of the study is to examine whether emotion perception from body expressions and from sounds can be observed at the perceptual level in infants. This is different from the conventional methods of measuring congruency effects at a later stage of processing. Based on findings by Kushnerenko et al. (2002) in 6-month-old infants, the auditory ERP components, infantile P2 (P150 and P350) and N450 were indexed for the modality and congruency effects in the present study. Since it is difficult to maintain infants' attention, we modified a paradigm for adults' paradigm (study2) so that it was more infant-friendly. (1) As the adults' data showed significant modality effects and motion effects on emotions of anger relative to fear, we only focused on the perception of angry expressions from the body and via sounds in infants. (2) We conducted passive presentation of three conditions, auditory-only, emotionally congruent and incongruent audiovisual conditions for the infant paradigm. (3) Due to the variation of latency in infant’s responses, we also used different ERP analysis for infant data relative to adults'. The mean amplitudes were calculated to examine effects at frontal, central to parietal as well as in left and right topographical regions. Although the functions of each component of an infant's CAEPs are uncertain, the perceptual interaction from body and sounds is expected to appear during a period of sensory processing (P150 and P350) rather than during later processing (N450) in 6.5-month-old infants.

Chapter 4 Study3

Electrophysiological evidence of perceptual integration from emotional body expressions and sounds during infancy

Abstract

Perceiving emotions from multiple sensory systems often makes us respond effectively in daily life. However, our understanding of multimodal emotional processing during development is relatively limited. The current study observed the neural responses of 6.5-month-old infants (N=15) to the presentation of angry sounds paired with emotionally congruent and incongruent body expressions, as well as to angry sounds presented in isolation. The findings showed that responses were differentiated between audio-only and audiovisual conditions from approximately 100 ms after the onset of sounds. Emotional congruency effects were also lateralized across left frontal-central regions during the presentation of sounds. Taken together, the current findings indicate that the capacity to integrate body and sound information might already be present at an early stage during processing at the age of 6.5 months, with distinct processes for the interaction of multisensory perception and for the assessment of the combined emotional content.

4.1. Introduction

Body expressions are essential visual cues for understanding others' emotions, and sometimes provide stronger emotional information than facial expressions in our social life. One such circumstance is when attending to distal events, before the ability to see an emotional expression on a face is displayed (de Gelder, 2009). The existent electrophysiological evidence suggests that infants might have ability to extract emotions (happiness versus fear) from body postures by 8-months (Missana et al., 2015; Missana et al., 2014). However, in our daily life we are frequently exposed to information from different modalities. The multisensory experience usually sensitizes our perception and speeds up our responses (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). It is therefore worthwhile to explore how emotion from body expressions combined with other modal information is perceived as a unifying percept rather than separate percepts. The majority of audiovisual emotional research has focused on facial expressions and emotional sounds (Kokinous et al., 2014; Pourtois, Debatisse, Despland, & de Gelder, 2002); however, it has been indicated that processing facial expressions is different to processing body expressions (see de Gelder et al., 2010, for a review). As such, there is little insight into multisensory aspects related to interpreting body expressions, particularly across development.

Behavioural studies have determined that the capacity to match emotion information conveyed through body expressions to that presented through sounds appears in early infancy. Using a preferential looking measure, Zieber and colleagues (Zieber, Kangas, Hock, & Bhatt, 2014a; Zieber et al., 2014b) found that 6.5-month-old infants looked longer to a video presenting an angry body expression when hearing an angry vocalization, and preferred to watch a happy video linked with a sound that expressed happiness. Despite this, questions about development in the integration of emotion perception have not been completely disentangled through behavioral observations. For example, to what extent do infants rely on attentional resources to process the emotional information across the modalities? It is also unclear how the audiovisual stimulation (i.e., body plus voice) enhances infants' emotion perception when contrasted with a sound that is played without any accompanying visual information.

Event-related potentials (ERPs) measurements could compensate for the limitations of behavioral studies as they can trace the timing and sequence of neural processes present within behavioral observations. This is especially advantageous for developmental research as responses can be recorded without any behavioural requirements (Hoehl & Wahl, 2012 for a review). Studies with adults have shown that the integration of emotion perception from body expressions and vocalization occurs at an early stage of sensory processing. A negative polarized component (N1) at 100 ms after onset of auditory stimuli, was reduced in amplitude for a body-voice condition when contrasted with a voice-only condition (Jessen & Kotz, 2011; Jessen et al., 2012; Yeh, Geangu, & Reid, 2016), suggesting that the interaction between each modal perception emerges at this period of time (Giard & Peronnet, 1999; van Wassenhove et al., 2005). Moreover, the emotion consistency across visual and auditory information can modulate the auditory P2 amplitude (a positive peak ~ 200 ms) (Ho et al., 2014) , which is relevant to the type of emotion and modality (Yeh et al., 2016). Thus, the P2 is interpreted to reflect a competition across modalities (Knowland et al., 2014) or an early assessment of the combined audiovisual content (van Wassenhove et al., 2005). As the modality and the congruency can be separately recorded at the level of the N1 and the P2, this implies that at least two functionally distinct neural mechanisms are involved in the integration of emotion information..

To date, ERP studies have not yet explored how emotions are perceived from the body alongside vocalizations from a developmental perspective. Outside the emotional domain, relevant literature has revealed that the capacity to process multisensory information emerges in infancy. A component with negative deflection (labeled as the Nc) peaking around 400-600 ms post-stimulus at frontal-central regions, has been commonly indexed for temporally synchronizing (Hyde et al., 2011; Kopp, 2014; Reynolds et al., 2014) or detecting congruency for particular characteristics across modal information (Bristow et al., 2009; Grossmann et al., 2006). However, the Nc is primarily involved in the processing of salient or infrequent visual information during infancy (de Hann & Nelson, 1999; Nelson & de Haan, 1996). It is, consequently, suspected that the neural mechanisms underlying the negative deflection to audiovisual stimuli are similar to those reported in visual research. Additionally, the Nc is thought to be related to attention processing, which is a relatively higher level cognitive process when contrasted with perceptual mechanisms (Csibra et al., 2008). The question remains open as to whether the interaction of multisensory perception occurs during initial perceptual stages during infancy, or whether it requires the involvement of higher level cognitive processes.

In terms of understanding the issue of emotional integration at the perceptual level, an alternative approach is to compare responses across unimodal and bimodal conditions. The method has been widely applied to multisensory studies with adults for speech (van Wassenhove et al., 2005), emotion (Jessen & Kotz, 2011) and other cognitive aspects (Stekelenburg & Vroomen, 2007). Relatively few studies have discussed audiovisual perception in infants with a measurement of comparing ERP responses between unimodal and multimodal contexts (Hyde, Jones, Porter, & Flom, 2010). Such research usually has the challenge of clearly defining auditory ERP waveform changes in these young populations. For example, infants often show a dominant positive response preceding a broad negative response by 400 ms for sounds (see Coch & Gullick, 2011; Wunderlich & Cone-Wesson, 2006, for reviews). This is the reverse of the adults' ERP morphology that consists of a negatively polarized response (N1) following a positive-going (P2) deflection. Additionally, ERP responses vary dramatically between birth and 12 months of age, with multiple transitions between positive and negative polarities (e.g. Kushnerenko et al., 2002).

In spite of the above issues, several auditory ERP components have been robustly identified for specific processing during infancy. For example, a positive component is often elicited over frontal-central regions from approximately 300ms. (P350, Kushnerenko et al., 2002). The component has been observed for maternal versus stranger's voice (Purhonen, Kilpelainen-Lees, Valkonen-Korhonen, Karhu, & Lehtonen, 2004), and crying versus neutral sounds (Missana, Altvater-Mackensen, & Grossmann, 2017), implying the processing of attention shift to salient or novel stimuli. Moreover, a negative deflection is often prominently found after 400 ms (N450, Kushnerenko et al., 2002). which is equivalent to the Nc in audiovisual studies on infants (Grossmann et al., 2006; Hyde et al., 2011; Otte et al., 2015). Likewise, the N450 may reflect that the infant allocates attention to saliency of visual signals. Indeed, the N450 may be more linked to expectancy or processing combined content across audiovisual information. A study from Grossmann et al. (2006) demonstrated that 7-month-old infants responded more negatively when a happy face was presented with an angry voice than when an angry face appeared with a happy voice. The author argued that infants might have more exposure to happy as opposed to angry faces in everyday life. Thus, for infants, a happy face with an angry sound might have been more unexpected compared to the other incongruent pairs, resulting in a larger negative response.

The current study aims to investigate the neural mechanisms underlying bimodal emotional perception from body expressions and sounds in infants. The age we focused on is based on behavioural results that indicate emotionally compatible concepts between body and sounds appears to develop by 6.5-month (Zieber et al., 2014a, 2014b). The auditory ERP components were compared in auditory and emotionally congruent as well as incongruent audiovisual contexts. This allowed us to examine the integration of audiovisual perception in infants by analyzing those ERP components considered to index perceptual and cognitive processes. We only focused on angry expressions as Yeh et al. (2016) showed that congruency and modality effects were more salient for the emotion when contrasted with fearful stimuli.

4.2. Method

4.2.1. Participants

The final sample consisted of 15 6.5-month-old (mean = 195 days, *SD* = 8 days) infants (10 boys, 5 girls). Seven additional infants were tested but were not included in the final sample due to fussiness (n=3), excessive movement (n=2), or poor quality of the resulting recording (n=2). The study was approved by the University Ethics Committee. Parents or guardians provided written informed consent, were paid (£10) to cover travel expenses, and were given a children's book as a gift.

4.2.2. Stimuli

The current work was based on Yeh et al. (2016), with presentation of angry sounds with angry or fearful body expressions for congruent and incongruent pairs, respectively. The visual stimuli were obtained from the research group of Beatrice de

Gelder and were extracted from those used by Kret et al. (2011). The body expressions for anger included shaking a clenching fist and raising the arm, while fearful expressions involved bending the body backwards and defensive movements of the hands. The face area was blurred in all conditions involving the visual modality. The characters were all dressed in black and they performed the body movements against a grey background. The luminance of each video clip was analyzed by taking into account each pixel within a frame (33 frames/clip, 480 × 854-pixel/frame). Each pixel was measured on a gray-scale using MATLAB, with values ranging from 0 to 255. The values of all pixels within a frame were averaged to obtain a luminance score for that frame. This allowed us to explore any potential variations in luminance which may appear with time due to the velocity and frequency of motion. The average luminance of the individual frames in the dynamic stimuli ranges from 64 to 68, with differences of no more than 1 between two consecutive frames. The auditory stimuli were chosen from the Montreal Affective Voices (MAV) database (Belin et al., 2008), which were edited to last 700 ms. The sounds were two types of male interjections spoken with an angry prosody (mean pitch: 240.47 Hz (*SD* = 60.72); mean intensity: 71.66 db (*SD* = 9.60).

In the study, the auditory stimuli with or without the visual stimuli were presented in the following conditions: auditory-only (A), emotionally congruent audio-visual (CAV), and emotionally incongruent audio-visual conditions (IAV). In the A condition, only a sound was played against a black background. The CAV and IAV conditions played an affective sound with either an emotionally congruent (CAV) or incongruent (IAV) body expression. Each condition comprised 32 trials, amounting to an overall total of 96 trials.

4.2.3. Procedure

Infants were seated on their parents' lap, facing a 90-100 cm computer monitor with two loudspeakers next to the screen on each side. The auditory stimuli were bi-aurally played via two speakers at a sound pressure of 70 dB for all participants. Each stimulus was presented using the Psychtoolbox 3.0 in Matlab 2012a. Each trial started with an 800-ms white fixation on a black screen, followed by the presentation of a video clip (CAV and IAV condition) or a black background (A condition) for 1.3 s. An interval randomised between a fixation and a video clip (visual stimulus) from 800 to 1200 ms. The auditory stimuli were shown 600 ms after the onset of the visual stimulus and ended at the same time as the video clips. The study lasted approximately 7-10 minutes overall, including breaks.

4.2.4. EEG Recording and analysis

The data was recorded by EGI NetStation system (Geodesic Sensor Nets, Inc., Eugene, OR) with a 128-channel electrode net. The EEG signal was sampled at 500 Hz and the impedances were kept to 50 Hz or less during recording. All electrodes were referenced on-line to the vertex (Cz). For computing the ERPs, the data was filtered with a 0.3-30 Hz bandpass filter and segmented off-line from 100 ms before to 1300ms after the video clip onset. Baseline correction was applied 100 ms prior to each segment before artifact rejections. Trials were rejected with EGI software once the eye movement exceeded +/- 140 uV, and eye blinks exceeded +/- 100 uV during the presentation of auditory stimuli (600 ms after onset of visual stimuli). Any channels that exceeded over +/- 200 uV for an electrode were marked as bad. If more than 12 electrodes within a trial were marked as bad, the trial was automatically discarded. The remaining trials were re-referenced to an average reference before the creation of average waveforms for each participant with each condition. Based on existing literature (Kushnerenko et al., 2002) and visual inspection, three ERP components were observed: P150 (100-230 ms), P350 (250-400 ms) and N450 (350-480 ms). Since the peak of each component was not clearly defined, we analyzed the mean amplitude of each condition within certain time windows.

As the distribution between frontal-central and central-parietal sites showed a reversed polarity of the potentials, the statistical analysis were performed individually, taking the average of these electrode clusters for left frontal-central (13, 29, 20, 24, 19), mid frontal-central (6, 12, 5, 11), right frontal-central (112, 111, 118, 4, 124), left central-parietal (30, 37, 36, 42, 54), mid central-parietal (7, 106, 31, REF/Cz, 80, 55), and right central-parietal (105, 104, 93, 87, 79) regions (**Figure 4.1.**). Each infant with fewer than 7 trials per condition was removed from the final analysis. The average number of available trials for each infant was 10.9 for the auditory-condition (*SD* = 3.41), 13.4 for congruent audiovisual condition (*SD* = 4.21), and 11.7 for incongruent condition (*SD* = 4.04). A three-way repeated-measures ANOVA was conducted on the three time windows, with the factors of *condition* (A, CAV, and IAV), *laterality* (left, midline, right) and *region* (frontal-central, central-parietal). Post-hoc analyses (least significant difference) were run where any significant (*p*-value < .05) interaction effects were reported.

4.3. Results

Figure 4.2. shows the topography and **Figure 4.3.** displays the grand average ERP of each condition.

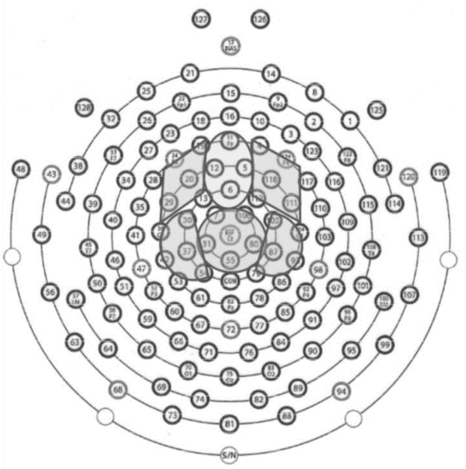
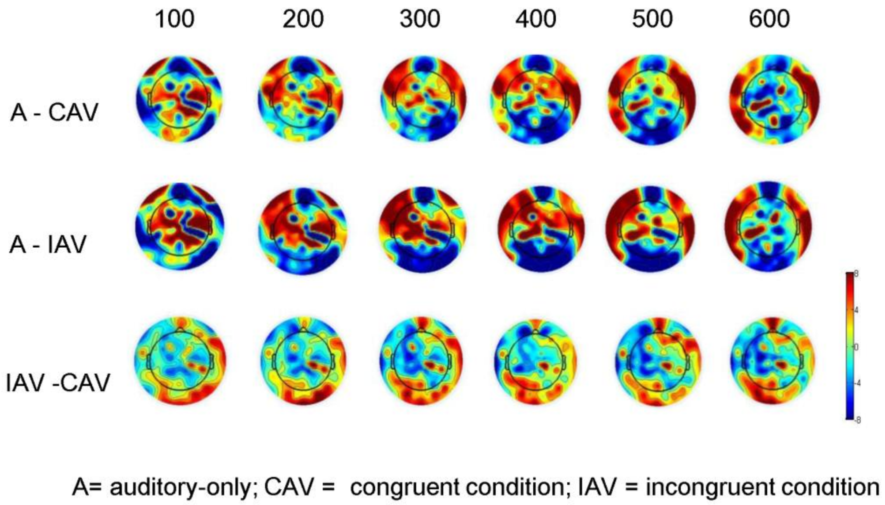


Figure 4.1. Averages were calculated based on electrode ROIs for left (13, 29, 20, 24, 19), midline (112, 111, 118, 4, 124) and right frontal-central (112, 111, 118, 4, 124), and left (30, 37, 36, 42, 54), mid (7, 106, 31, REF, 80, 55) and right central-parietal regions (105, 104, 93, 87, 79)



A= auditory-only; CAV = congruent condition; IAV = incongruent condition

Figure 4.2. The topography distributions for comparison between each condition from 100 to 600 ms after the onset of sounds in infants. (A = auditory-only; CAV = congruent audiovisual condition; IAV = incongruent audiovisual condition)

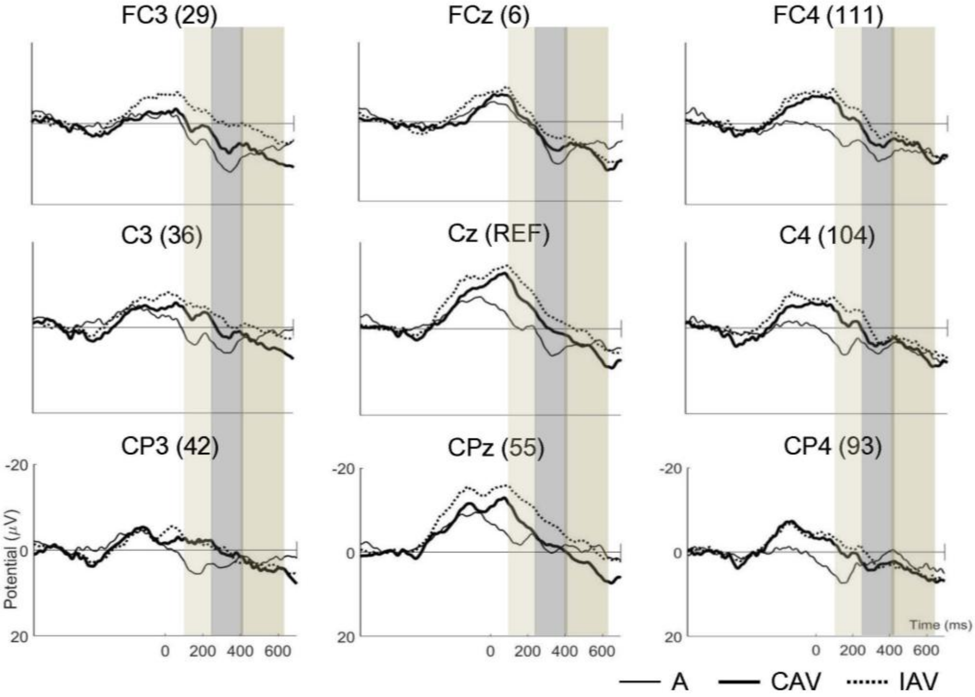


Figure 4.3. The grand average for each condition in infants. The shaded areas were the time windows for P150, P350 and N450, respectively (left to right).

4.3.1. P150 (100-230 ms)

There was a significant main effect of condition (F(2,28) = 7.391, p = .003, ɳ2 = .346), showing higher P150 amplitudes for the auditory-only condition (A) compared to the incongruent condition (IAV) (p = .001). In addition, there was a trend of larger responses for congruent (CAV) than for IAV pairs (p = .056). A main effect of laterality was also found (F(2,28) = 3.942, p = .031, η^2 = .220), with more positive deflections for left and right than for midline regions (p = .032; p =.008, respectively). A three-way interaction between condition, laterality and region was statistically significant (F(4,56) = 3.205, p = .019, η^2 = .186). Further analysis showed a larger P150 for A compared to CAV at mid and right central-parietal regions (p = .015; p =.029, respectively), as well as for A compared to IAV at left frontal-central regions (p = .040) and across all central-parietal regions (left: p = .004; mid: p < .001; right: p

= .002). The response was significantly more positive in CAV and IAV at left frontal-central (p = .032) and mid central-parietal electrodes (p = .023).

4.3.2. P350 (230-400 ms)

A significant main effect of condition (F(2,28) = 7.391, p = .003, η^2 = .346) was found, with higher positive amplitudes in A compared to IAV conditions (p = .008). Additionally, a marginally larger P350 response was found for the CAV than for the IAV condition (p = .067). The effect of condition also showed significant interactions with laterality and region (F(4,56) = 2.795, p = .035, η^2 = .166). Further analysis indicated different effects between A and IAV at left frontal-central (p = .009), left (p = .050) and mid central-parietal regions (p = .001). Significant comparisons between P350 in CAV and IAV conditions were observed at left frontal-central (p = .018) and mid central-parietal regions (p = .057).

4.3.3. N450 (400-650 ms)

No significant main effects were found; however, a three-way interaction between condition, laterality and region reached marginal significance (F(4,56) = 2.119, p = .091, η^2 = .131). Further analysis showed that a larger N450 for IAV than for A condition at left frontal-central (p = .015) and mid central-parietal sites (p = .035). There was also a trend for a larger N450 in IAV compared to CAV conditions at left frontal-central regions (p = .062)

4.4. Discussion

The current study investigated how 6.5-month-old infants perceive emotions from body expressions combined with auditory sounds at the level of perceptual and cognitive processing. We compared the two earlier sensory responses (P150 and P350) as well as the N450 between modalities and for emotional congruency across audiovisual information. Overall, the responses differed in the auditory-only condition to the response made in the audiovisual conditions, especially for emotionally incongruent conditions within three ERP components (P150, P350 and N450) at left frontal-central and central-parietal electrodes. Regarding the comparison between emotionally congruent and incongruent pairs, the effects were mainly elicited 100 ms after the onset of sounds at left frontal-central sites, diminishing at approximately 500 ms. These modality and congruency effects have specific implications for our understanding of infant emotion processing.

Examining the early processing stage, amplitudes of both the P150 and P350 were reduced for the presentation of the body with a sound than when the sound was presented in isolation. The reduced response for bimodal compared to unisensory information supported the notion that the interactions of multimodal perception occur at sensory processing stages (van Wassenhove et al., 2005). However, this is inconsistent with previous reports (e.g. Grossmann et al., 2006) that have assessed multisensory perception at a later processing stage. Further examining the modality effects, the latency for the comparison between auditory-only and incongruent conditions ranged from 100 until around 500 ms, whereas the comparison between auditory-only and congruent conditions was shorter (~350 ms). This may be in accordance with concepts advanced by Kushnerenko et al. (2002) whereby the P150 and P350 might reflect functionally different mechanisms. Even though the precise function of each component is not clear in terms of the infant's auditory responses, the P350 has been indicated as either an attention shift mechanism for novel and unfamiliar stimuli (Hyde et al., 2011; Purhonen et al., 2004) or as a system for processing higher valence of emotion types (Otte et al., 2015). As such, we assumed that the interactions of the emotion perception from audiovisual information have, most likely, occurred by 230 ms after the presentation of sounds in 6.5-month-old infants. However, emotionally incompatible body expressions to sounds might appear novel to infants, resulting in them taking longer to process following the timing of the P350.

The topographical distribution of the P150 and P350 indicate variability in the mechanisms involved in modal emotion processing. Typically, the two ERP components are reported at frontal-central rather than central-parietal regions (Hyde et al., 2011; Otte et al., 2015). We consequently speculated that the modality effects at central-parietal sites were adjusted by the movement (dynamic body stimuli), and elicited a broad positive waveform over a wider topographical array. This is supported by work investigating visual motion (Hirai & Hiraki, 2005; Marshall & Shipley, 2009; Reid, Hoehl, & Striano, 2006) whereby a larger positive amplitude was elicited 300 ms over parietal and temporal areas in young infants in response to the presentation of point-lights depicting human movement compared to scrambled or inverted stimuli. More recently, Missana and her colleagues (Missana et al., 2015; Missana et al., 2014; Talsma & Woldorff, 2005) observed distinct ERP topographies in terms of timing and regions when viewing an emotional body with motion and posture cues in 8-month-old infants. Their findings showed that the responses were modulated by static body expressions at frontal and central sites whereas the dynamic body appeared to modify the component towards temporal and parietal locations. The results of these studies strongly suggest that biological motion perception develops by the first postnatal year, which may contribute to the integration of emotional information in the current study.

With regard to the congruency effects, the ERP response to emotional congruency across auditory and visual information was elicited prominently at an early stage of processing. This was inconsistent with our expectation that the congruency effects would occur at a later stage. With a similar paradigm, the congruency effect was elicited between 180-330 ms (P2) but this was not observed within the 90-180ms (N1) range in adults (Yeh et al., 2016). It is possible that this difference is connected to the different latency definition was used for the P150 in that it was much broader (100-230 ms) in the current study relative to the latency of the adult P2. As the congruency effects were slightly salient for the P150 (p <*.05*) and marginally achieved significance for the P350 in infants, congruency detection and interpretation may occur between the time windows of the two components. In addition, the P150 is likely to overlap the visual emotion effects preceding the presentation of the angry sounds. Missana et al. (2015) found a greater positive response for happy than for fearful body expressions at 700-1200 ms in infants, which was also the epoch in which we began to present the sounds after the presentation of body expressions. The difference that was found in the current study between the congruent and incongruent conditions is unlikely to be fully driven by the visual emotion effects. As the significant modality effects were also observed, that is, an interaction between visual and auditory perceptions occurs, which promotes perceivers to assess the compatibility of information presented across the two modalities.

In the mid latency stage of the congruency effects, a larger N450 was produced for emotionally incongruent than congruent body-voice pairs when an angry voice was presented. The direction of the congruency effect is consistent with Grossmann et al. (2006), which showed that the Nc was more negative for emotionally incongruent compared to congruent pairs. The Nc for visual processing is thought to reflect attention toward salient stimuli in terms of familiarity, novelty or other factors (de Hann, 2007). Based on this explanation, we considered that the emotionally incongruent pair might be non-logical or a socially novel match for infants, causing a more negative response for incongruent compared to congruent pairs. Specifically, this processing stage is likely to be influenced by how the emotion content coheres across the auditory and visual modalities. Grossmann et al. (2006) found the response to be more negative when a happy face was presented with an angry voice than when an angry voice was presented with a happy sound. In addition, Yeh et al. (2016) found different directions of the congruency effects for anger and fear at a similar latency in adults. As a result, the auditory processing appears to be modulated by the preceding visual stimuli either strengthening or weakening the following information. However, the study of early development limits researchers to shorten studies in order to prevent the infant participants from becoming fatigued. In order to maintain their attention for the duration of the ERP study, it is difficult to present more conditions for the visual stimulus (i.e., angry body-only) and reversed pairs (i.e., angry/fearful body with fearful sounds) to further understand this stage of processing. At this stage, it can only be inferred that infants at this age are able to use body expressions beyond those examined in this study to aid them in effectively processing emotionally matched sounds.

The significant congruency effects implies that infants at 6.5-month have the ability to discern the difference between angry and fearful body expressions. However, it is not known whether infants at this age fully understand the emotions of anger and fear. Also it is an open question whether they can precisely match the body information to emotionally congruent vocalizations, rather than mapping general information across the two modalities. It has been indicated that infants by 10 months of age possibly discriminate emotions from others on the basis of emotional valence or perceptual features (Widen & Russell, 2008). It is not until 10 to 12 month of age that the ability to recognize emotions from faces and voices develops, that is, they can understand the underlying meaning of emotional expressions (Walker-Andrews, 1997). Infants acquire the meanings by interactions with caregivers connecting to events that contain punishing or rewarding elements. Following this explanation, the strategy that 6.5-month-old infants use for discriminating emotions from body expressions might be different from adults. As adults can extract the emotional meanings from displays of emotion via bodies, perceptual features might be crucial cues for young populations to discern emotions. In our studies, the angry body expressions were presented with forward movement and clenched fists, while the fearful body showed defensive movements of hands with backward steps. These visual materials might be prototypes of emotional expressions that are interpretable for infants. Regarding other factors, the frequency of movement, illumination and emotional intensity are all controlled in our stimuli. Given that this is the case, these factors are unlikely to be cues with which to discriminate the two emotions in our studies. However, extending the results, it would be of interest to know when and how infants acquire the knowledge of the emotions of anger and fear from body expressions. Further work is required to clarify the current results.

There are some limitations associated with our study that could be addressed in future studies on emotional perceptual integration in infants. For instance, we defined the latency of each component largely based on the visual inspection of the grand average of the mean amplitude. Inspection of individual averages indicates that there is a large variation in the peak latency and amplitude across individual infants. Since the P350 was followed by the N450, the timing of N450 might be a negative polarity for some infants but be a positive polarity for other infants (Trainor, 2007). It is for this reason that we cannot further clarify the differential functions of the P350 from the N450, as both components display similar modality and congruency effects. As the N450 is hypothesized to have a role in assessing and unifying emotional content, it could be a way of contrasting with another group at the same age with reversed incongruent pairs, such as an angry body with a fearful sound, to extend the current results. Furthermore, the present observations involve the emotion of anger only, and thus we cannot generalize results to the bimodal processing for other emotion categories. It is already known that the developmental trajectory of processing differs across emotion types and modalities (Chronaki et al., 2015). It would also be valuable to understand multisensory perception for other types of emotions, like happiness, during early development. Addressing these issues will be useful to further the field of the ontogeny of multisensory processing of emotion information.

4.5. Conclusion

The present study is a preliminary investigation on the processing of emotion that is perceived from the body and sounds during early development. By comparing the auditory responses in auditory-only and audiovisual conditions, the study have shown emotionally perceptual interaction across the two modalities at perceptual and attentional ERP components in 6.5-month-old infants. Just like adults, the two separated processes, one for the interaction of the audiovisual perception and other for the assessment the combined emotion content, have been found before 400ms in infants. This is earlier than observations of congruency effects in other audiovisual studies with infants. Young infants are also likely to be capable of discriminating anger and fear from body expressions, guiding them to effectively process the following angry sounds. The modulation of motion cues possibly plays a key role in recognizing body expressions during the perceptual integration of information.

Prelude to Chapter 5

The observation of the neural mechanism underlying the integration of audiovisual perception from body expression and sounds that develop in young children

In our prior studies, we measured ERP responses in adults and 6.5-month-old infants to understand the change of neural responses in emotional audiovisual perception from early in development to adulthood. However, it still leaves questions open about the maturational effects for emotional multisensory processing. For example, at what age do the neural ERP polarities change to adult-like responses? In behavioural findings, higher accuracy for emotion recognition is usually seen in adults compared to children (e.g. Chronaki et al., 2015). A number of factors could be underlying the differences in behavioural responses, including the maturation of brain cortices on processing speed, or the development of advanced strategies of goal planning and executive function (Brandwein et al., 2011). Consequently, it is worth extending the exploration of the emotional multisensory processing after infancy, aiding us to understand how maturation sensitizes our perceptual integration of emotion. With an ERP measurement, we could observe the maturational trajectories in different components reflecting different stages of processing. This is also the reason why we want to further explore auditory responses in children to preliminarily bridge the developmental trajectory of audiovisual perception between the two age groups.

As discussed in the last few chapters, the typical cortical auditory evoked potentials (CAEP) in adults are sequentially comprised of a P1 (peaking at ~ 50ms), N1 (~100 ms) and P2 (~180 ms) (see Coch & Gullick, 2011, for a review). In contrast, the P1 and N1 are less evoked in infants, as they often showed a broad positive peak (referred to as the P2, ~ 200ms) followed by a broad negative wave (referred to as the N2, ~350 ms). Spanning the periods from four until 15 years, studies have consistently reported that the earlier components, P1 and N1, become more prominent in amplitude than P2 and N2 (Ceponiene, Rinne, & Naatanen, 2002; Shafer et al., 2015; Sussman et al., 2008). As the following study has aimed to observe these auditory components in emotional perceptual integration in young children, we reviewed the literature surrounding the developmental change in each auditory component, P1, N1, P2 and N2, across childhood to adolescence in detail.

The P1 is usually reported peaking at 100 ms after the onset of a sound stimulus, and it mostly recognisable between the ages of 3-5 years. The amplitude of the P1 slightly increases from ages 4 to 10, and then abruptly decreases in the teenage years (Ceponiene et al., 2002; Ponton et al., 2000; Sussman et al., 2008). In addition, the P1 peak latency is reduced with increasing age. The age-related changes for the P1 peak and amplitude are similar to the N1, a negative component often emerges after the P1. The N1 is predominantly elicited in adults at around 100 ms, but is typically smaller or absent in newborns and children younger than 6 years of age. The N1 amplitude increases until 10-12 years of age, and gradually attenuates to adult-like level at the age of 16. As such, the disappearance of the P1 has been thought to be caused by the overlap with the emerging N1 component (Ceponiene et al., 2002).

A relatively large positivity from 100 until 400 ms is often reported in newborns and young infants, which is referred to as the P2. The P2 amplitude decreases with age, whereas the latency does not. Following the P2 is another negative deflection, N2 or N250, which is thought to be a classic characteristic of the child auditory ERP waveform (Sussman et al., 2008). It might be the reason that the N2 amplitude becomes increasingly stable during childhood (~10-year-old) and thereafter decreases to an adult-like value during adolescence (Ceponiene et al., 2002; Ponton et al., 2000; Sussman et al., 2008). However, there are inconsistent results for the N2 latency across childhood, with some showing a reduction (e.g. Shafer et al., 2015; Sussman et al., 2008), or an increase with age (e.g. Ponton et al., 2000), or observing no change (e.g. Ceponiene et al., 2002). Due to similar scalp distributions between 9-year-old children and adults, Ceponiene et al. (2002) proposed that the childhood N2 might be a precursor of the adult N2. The attenuation of N2 amplitude with age might reflect an increasing function of inhibitory control as the adult N2 is greatly activated during sleep (e.g. Nielsenbohlman, Knight, Woods, & Woodward, 1991). This perspective also supports an hypothesis proposed by Wunderlich and Cone-Wesson (2006) which states that the N2 is associated with greater efficiency of higher level processes (Cunningham, Nicol, Zecker, & Kraus, 2000) and is sensitive to task demands (Ceponiene et al., 2002) and attention (Naatanen & Picton, 1986)

Currently, only two electrophysiological studies (Brandwein et al., 2011; Knowland et al., 2014) have examined the typical developmental courses of multisensory perception from childhood until adulthood. Although these studies did not focus solely on researching the emotion domain, the results showed that the effects of multisensory interactions emerge at an early sensory stage of processing (~ 100 ms after sounds) across 6-year-olds to adulthood. The peak latency was also reduced with increasing age. Nonetheless, some key questions remain to be addressed. For example, it is common to determine time windows for infants or children responses from visual inspection of the grand average, or the latency regions of the adult data. However, the analysis might overlook a large variance in peak latency and amplitude across individuals, or different transitions of polarity waveforms among different children at the same epoch time. Consequently it is then difficult to determine statically significant regions for these effects (Trainor, 2007). As such, it may not be adequate to analyze peak amplitudes within a latency of the grand average for the infant and child data. Moreover, the effects of distribution are required to be considered across maturation as the timing and topography of the components will vary relevant to age (Brandwein et al., 2011).

Another issue about emotion perception is that different emotional expressions could lead to inconsistent results. Recent literature has indicated that developmental trajectories are discordant across different emotions associated with modalities. A behavioural study by Chronaki et al. (2015) provided evidence that young children recognized anger and happiness expressions from faces and voices more accurately then sadness. In addition, children are more sensitive to emotions that display happiness compared to other emotions at these early ages (e.g. Montirosso et al., 2010). Considering the above reasons, we therefore measured congruency effects by comparing angry/happy body expression with angry sounds in our fourth study. This is different from our previous studies (study1, study2 and study3) that conducted angry/fearful body expressions with angry sounds. As we assumed that the greater level of emotionally incongruent comparisons are presented across visual and auditory modalities, the more easily processed congruency effects are observed. We also hypothesized that the perceptual processes for the markedly incongruent effects may result in more automatic processing and rely less on later, cognitive stages of multisensory processing in children.

The aim of the next study was to investigate the neural activities underlying the emotion perceptions we drew from body expressions combined with sounds in typically developing 5-6 year-children. The methodology was similar to previous studies that have been applied to infants, presenting three conditions (auditory-only, emotionally congruent and incongruent audiovisual conditions). Children participants were also required to complete non-emotional responses so their attention could be maintained. The two major components, P2 (~ 100 ms) and N2 (~ 250 ms), with the potential emerging components (P1 and N1), were observed for the effects of multisensory interaction and emotional congruency across auditory and visual modalities. Since the morphology still varies in early childhood, the mean amplitude was calculated for each component, with latencies determine from the previous literature (e.g. Ponton et al., 2000) and visual inspection. Based on the lateralization findings in our infant studies, we also considered regions (frontal-central, central-parietal) and hemisphere (left, middle, right) effects in children. To confirm that the children were typically developing, we administrated verbal and nonverbal standardized tasks. These behavioural results were also calculated to examine the relationship with individual neural responses. Based on the current study, we expect to provide a starting point for future research on emotion perception in typically developing as well as clinical populations.

Chapter5 Study4

The integration of emotional perception from body expressions and the voice in early childhood

Abstract

Although body expressions have been indicated to be powerful visual cues of conveying emotional signals, developmental research has not extensively examined multisensory perception of body expressions when combined with other modal information. To observe the maturational changes in the neural correlates underlying audiovisual emotional perception, we measured EEG from 5-6 year-old children by presenting emotion stimuli in auditory-only, emotionally congruent and incongruent audiovisual conditions. Based on the children's responses in the auditory-only condition, double positive peaks (P1, P2) followed by a negative response (N2) were indexed for the effects of perceptual integration. Results showed significant comparisons between the auditory-only and the audiovisual responses within the P1(100-160 ms after the sounds) and P2 (160-260 ms). The P1 and the N2 (~250 ms) responses were also elicited by the emotional congruency across the auditory and visual modalities. Both modality and congruent effects were maximally distributed in the right frontal-central regions. These findings suggest that the integration of emotion processing for body expressions and sounds occurs at a sensory stage in early childhood. For angry expressions, the information from emotionally congruent body expressions may facilitate the specialized processing of sounds.

5.1. Introduction

Multisensory information usually sensitizes our perception, allowing us to make effective responses compared to unisensory information. The capacity to associate relevant features of events or objects across multiple modalities is developed in the first postnatal year, which plays important roles in our perceptual learning related to language, attention, emotion and other social cognition in the environment (see Bahrick & Lickliter, 2012, for a review). As for emotion perception, previous multisensory research has mainly focused on facial expressions combined with sounds in adults (e.g. Pourtois et al., 2002) and infants (e.g. Grossmann et al., 2006). In the past ten years, a growing numbers of studies have demonstrated that body expressions are also important visual cues in conveying emotional information (de Gelder et al., 2010). The ability to extract emotional information from body expressions has already developed by 8-months of age (Missana et al., 2015; Missana & Grossmann, 2015). However, we usually receive emotional information from various modalities rather than via a modal source in the social world. Consequently, it is an important issue to address that how cues, body expressions and other modal information, are perceived as a unified percept rather than separate precepts.

A few studies with adults have discussed the integration of emotion perception on body expressions combined with affective sounds. Behaviourally, the consistency of emotional content across body expressions and sounds improves the accuracy of emotion recognition in contrast to uni-sensory presentation (Van den Stock et al., 2007). Event-related potential (ERP) studies (Jessen & Kotz, 2011; Yeh et al., 2016) further showed the differentiation in negative responses between auditory-only and audiovisual conditions at 100 ms (N100, N1) after the onset of sounds, suggesting that the visual and auditory perception of emotional information interact at an early processing stage (Giard & Peronnet, 1999; Stekelenburg & Vroomen, 2007). Another effect influenced by consistency of emotional content across auditory and visual modalities was also observed at 200 ms (P200, P2). The P2 is modulated by the preceding visual contexts to the following auditory processing, either suppressing or facilitating the auditory responses in the audiovisual conditions (Ho et al., 2014; Yeh et al., 2016). Thus, the P2 may reflect a process for assessing the combined emotional content (Paulmann et al., 2009) or a competition between the two forms of modal information (Stekelenburg & Vroomen, 2007). As the modality and congruency effects are found within the N1 and the P2 respectively, these results imply at least two functionally separate processes for the integration of emotion perception on body expressions and sounds.

At the developmental level, studies investigating the integration of emotion perception from body expressions and sounds are still scare. Existing evidence has shown that the emotional concept of body expressions associated with affective sounds has already developed by 6.5-month-old (e.g. Zieber et al., 2014a). However, no studies have extended the issue of audiovisual emotional perception after infancy. Despite this, several pieces of research have provided relevant evidence for developmental changes in multisensory processing of emotion during childhood. For example, Gil et al. (2016) examined the ability to categorize facial expressions paired with prosody in 5 to 9-year-old children and adults. Based on the mean proportion of behavioural response to sadness along the facial emotion continuum (30%, 60%, 90% of happiness or. sadness), the patterns in adults were similar to those in 9-year-old children but were dissimilar to children aged below 7 years. The authors considered that some processes develop nonlinearly between infancy and adulthood, with changeable weights of emotional cues during this period of time. This finding also suggests the changes in the use of emotional cues during audiovisual emotion recognition between infancy and adulthood.

To broaden our understanding of audiovisual emotional perception across development, EEG/ERPs are the optimal way of measuring infants and children's neural processing of multisensory perception. Because of superior temporal recordings to the order of milliseconds, ERPs can effectively reflect rapid changes in neural activities that correspond to the presentation of stimuli. ERP studies also do not require behavioural responses, so it is suitable for studies with developing and clinical populations. As each component may reflect a specific process, ERPs allow us to explore differences in the cognitive and perceptual systems between children and adults that underlie their behaviour. Previous work focussing on multisensory perception in infants and children, typically measures feature congruency across modalities for the effect of perceptual integration (Bristow et al., 2009; Grossmann et al., 2006; Kushnerenko, Teinonen, Volein, & Csibra, 2008). The Nc, a frontal negative component peaking around 400 ms, is often indexed for effects of multisensory perceptual integration. However, the Nc reflects attention allocation to salient or familiar visual stimuli, which belongs to a higher level of cognitive processing (see Csibra et al., 2008, for a review). As such, it is plausible that the integration of multisensory perception takes place at an earlier processing stage in infants and children.

In order to identify the assumption that an integration of multisensory perception occurs at an early processing stage, it could be a more practical way of comparing auditory responses in unisensory and multisensory contexts. This method has been widely demonstrated in adult studies showing the effects of audiovisual perceptual interaction within the N1 and the P2 (e.g. Giard & Peronnet, 1999; Jessen & Kotz, 2011; Stekelenburg & Vroomen, 2007). The advantages of the design are to observe multisensory processing at both perceptual and cognitive levels. To date, only a few studies have examined perceptual integration in children by comparing unisensory and multisensory responses (Brandwein et al., 2013; Knowland et al., 2014). However, there are open questions related to young children's responses in terms of polarities and latencies. It is the reason that these young populations have often shown unclear auditory responses, or reversed patterns relative to adult's auditory responses, with a broad positive wave followed by a broad negative response within 400 ms (see Coch & Gullick, 2011, for a review). It is also uncertain at what age that the transition of the ERP polarities changes towards adult-like responses. On top of this, the responses are tremendously different between individuals in terms of latencies. While some infants, or children, show positive deflections, negative responses are observed in others during the same time period. The grand average waveform could, therefore, become flat and difficult to identify the certainty of the effects (Trainor, 2007).

Despite great maturational changes in the auditory responses, a growing body of research has disclosed maturational progress for each auditory component (e.g. Ceponiene et al., 2002; Ponton et al., 2000; Shafer et al., 2015; Sussman et al., 2008). This allows us to more confidently index several components in multisensory research with children. According to sequential polarities in adults' auditory responses, P2 (~ 180 ms) and N2 (~ 250 ms) are more frequently present in infants and young children when contrasted to P1 (~ 50 ms) and N1 (~ 100 ms). During early infancy, the first emerging component is often a positive peak from 100 until 400 ms (labelled as P2; Kushnerenko et al., 2002; Wunderlich & Cone-Wesson, 2006). The P2 amplitude increases from infancy until late childhood, and then decreases until an adult-like level across childhood to adulthood. In addition to the P2, the N2 often peaks after the

P2 and is another prominent characteristic of a child's auditory ERP waveform. The N2 is a negative deflection that shows relatively stable amplitudes and latencies until early adolescence. However, it gradually decreases during adolescence and is not often seen in mature adult waveforms (Ceponiene et al., 2002; Ponton et al., 2000; Sussman et al., 2008). The N2 amplitude reflects a greater efficiency for higher level processes (Cunningham et al., 2000) as it is sensitive to task demands (Ceponiene et al., 2002) and attention (Naatanen & Picton, 1986). In contrast, the P1 and N1 are typically smaller or absent in newborns and children younger than 6 years of age (e.g. Ponton et al., 2000). The P1 gradually increases in amplitude from 4 to 10 years-old, and then sharply decreases to an adult-like level at adolescence. The age-related changes for the P1 peak are similar to the N1, which becomes a predominant component in adults at around 100 ms. Since the N1 is often found between the P1 and the P2 during late childhood and early adolescence, it might lengthen the P2 peak latency (Sussman et al., 2008). The P1 may also be attenuated by the overlap with the emergence of the N1 (Ceponiene et al., 2002). Although these auditory components change with age, other factors, such as sound type, also contribute to the developmental changes in the morphology of auditory responses (Sussman et al., 2008).

Each emotion serves a unique function for communication and social adaption. It is likely that there are distinct patterns of neural connectivity and maturational trajectories for the perception of different emotions. Providing behavioral evidence, Nelson and Russell (2011) found that 3 to 5 year-old preschoolers can more accurately recognize anger and happiness relative to fear from facial and body expressions. However, previous studies have not paid attention to neural mechanisms related to emotion perception in early childhood, particularly when considering emotion type and multiple modal resources. Unisensory studies in healthy adults and brain-damaged patients have shown brain asymmetry for emotion perception, but in differential patterns (see Demaree, Everhart, Youngstrom, & Harrison, 2005, for a review). The *Right Hemisphere Hypothesis* states that the right hemisphere is dominant in the processing for all emotions (Borod et al., 1998). Comparatively, the *Valence Hypothesis* specifies the lateralization of emotion processing is relevant to emotional valence, with the left hemisphere dominating for positive emotions (e.g. happiness) while the right specializing for processing negative emotions (e.g. fear) (Davidson, 1995). This is similar to the approach-withdrawal perspective (Balconi & Mazza, 2009; Davidson, 1992a) whereby the hemispheric asymmetry for emotion is associated with the fact that it drives the individual toward or away the stimuli or events in the environment. This hypothesis is also supported by EEG studies in infants that show greater relative left-frontal activation for the approach condition (e.g., happy face), and greater relative right-frontal activation for the withdrawal condition (e.g. sad face) (Davidson & Fox, 1982; Fox & Davision, 1987). Recently, Balconi and Vanutelli (2016) further examined the cortical lateralization for emotion perception in cross-modal contexts. The pictures depicting emotionally (un)comfortable interactions between human and animals were represented for positive/negative visual emotion, with presenting affective sounds in an audiovisual condition. The results highlighted the negative lateralization effect as only negative emotions with emotionally incongruent pairs could more elicit the right-side prefrontal cortex activity. The above evidence illustrates how lateralized processing differs across emotions, which develops in the first postnatal year. The lateralized emotion processing across modalities, especially for negative emotions, or withdrawal contexts, may be more salient compared to positive emotions or approach contexts. However, it remains unclear if there is a developmental shift in the brain asymmetry for emotion processing during childhood.

The current study aimed to investigate the neural mechanisms underlying the audiovisual emotional perception of body expressions and affective sounds during early childhood. We therefore compared the auditory obligatory responses, P2 and N2 (or P1 and N1) in auditory-only and audiovisual contexts in typically developing 5- to 6-year-old children. This methodology also enabled us to further understand the maturational changes in the processing of multisensory interactions at the perceptual and cognitive levels. To more easily facilitate the effects of emotional congruency for auditory and visual information, we presented a happy body expression with an angry sound for an incongruent pairing. Although children's morphology may not be adult-like in their responses, we expected to find a difference in the processing of modality effects at an early stage of processing, with the assessment of the emotional congruency across the two forms of modal information occurring at a later stage. Moreover, the laterality was also considered in order to examine whether brain asymmetry is present for audiovisual emotional processing at this age.

5.2. Method

5.2.1. Participants

Sixteen 5- to 6 year-old children (mean = 5.71 years, *SD* = .26; 8 boys, 8 girls) were in the final sample. Data from seven additional children were excluded due to less than 10 trials (n=4), poor quality of the resulting recording (n=2) and behavioral results (n=1). All children in the study did not have any known psychiatric, genetic or medical condition based on parents' reports. The British Picture Vocabulary Scale (BPVS; Dunn & Dunn, 2009) was used as approximations of verbal ability (% percentile rank) whereas the three subtest of Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III; Wechsler, 2002) - block-design, matrix reasoning and picture concepts, was used as measures of nonverbal intelligence. The exclusion criterion in the study was either children's verbal scores (*M* = 61.94 %, *SD* = 18.03), or the total nonverbal scores were below 10% at their ages (*M* = 72.44 %, *SD* = 18.18)

Prior to the study, verbal consent was obtained from each of the child participants. Parents or guardians also provided written informed consent form and the social communicative questionnaire (SCQ; Rutter, Bailey, & Lord, 2003). Children were not included for further analysis if their SCQ scores were greater than 15 (*M* = 5, *SD* = 2.85). The study was approved by the Lancaster University Ethics Committees. All parents were paid (£10) to cover travel costs, and children were given a book for their participation.

5.2.2. Stimuli

All visual stimuli were obtained from the research group of Beatrice de Gelder and have been utilised in their studies (e.g., Kret et al., 2011). We presented angry sounds with angry or happy body expressions for congruent and incongruent pairs, respectively. The body expressions for anger included shaking a clenching fist and raising the arm, while the happy expressions were characterized by arms being raised above the shoulders. The face area was blurred in all conditions involving the visual modality. The characters were dressed in black and they performed the body movements against a gray background. The luminance of each video clip was analyzed by taking into account each pixel within a frame (33 frames/clip, 480 × 854-pixel/frame). Each pixel was measured on a gray-scale using MATLAB, with values ranging from 0 to 255. The values of all pixels within a frame were the averaged to obtain a luminance score for that frame. This allowed us to explore any potential variations in luminance, which may appear with time due to the velocity and frequency of motion. The average luminance of the individual frames in the dynamic stimuli ranged from 64 to 68, with differences of no more than 1 between two consecutive frames. The auditory stimuli were chosen from the Montreal Affective Voices (MAV) database (Belin et al., 2008), which were edited to last 700 ms. The sounds were two types of male interjections spoken with an angry prosody (mean pitch: 240.47 Hz (*SD* = 60.72); mean intensity: 71.66 db (*SD* = 9.60). The emotional intensity of visual and auditory stimuli have been controlled and validated in Yeh et al. (2016)

5.2.3. Procedure

Children sat comfortably in a dimly lit/darkened room, and were asked to make their responses by pressing a button. Each stimulus was presented using the Psychtoolbox 3.0 in Matlab 2012a. The visual stimuli were presented on a monitor at a distance of 90-100 cm from the participants, and the auditory stimuli were bi-aurally played via two speakers at a sound pressure of 70 dB for all participants. In the study, the auditory stimuli with or without the visual stimuli were presented in the following conditions: auditory-only (A), emotionally congruent audio-visual (CAV), and emotionally incongruent audio-visual conditions (IAV). In the A condition, only a sound was played against a black background. The CAV and IAV conditions played affective sounds with either emotionally congruent (CAV) or incongruent (IAV) body expressions. Each trial started with a 800-ms white fixation on a black screen, followed by the presentation of a video clip (CAV and IAV condition) or a black background (A condition) for 1300 ms. An interval randomised between a fixation and a video clip (visual stimulus) from 800 to 1200 ms. The auditory stimuli were shown 600 ms after the onset of the visual stimulus and ended synchronously with the video clips. A picture showing a penguin's head randomly appeared after the sounds, and all participants were instructed to press the keyboard for the presentation of the picture. The picture disappeared when the participants made their response. In order to avoid learning the regularities of presentation in penguin's heads, in each block we randomly showed them after a trial in less than 60% of the cases (ranging from 20 to 33), by using a custom Matlab script. The penguin's head was presented after a trial less than 5 consecutive times. Each block included 48 trials. The testing started after a practice session consisting of 15 trials. The study consisted of 2 blocks for a total of 96 trials. The study lasted approximately 15 minutes, including breaks.

5.2.4. EEG Recording and analysis

The data were recorded by EGI NetStation system (Geodesic Sensor Nets, Inc., Eugene, OR) with a 128-channel electrode net. The EEG signal was sampled at 500 Hz and the impedances were kept to 50 Hz or less during recording. All electrodes were on-line referenced to vertex (Cz). For computing the ERPs, the data was filtered with a 0.3-30 Hz bandpass filter and segmented off-line from 100 ms before to 1300 ms after the onset of the visual stimuli. Baseline correction was applied to 100 ms prior to each segment before artifact rejections. Trials were rejected with EGI software once the eye movement exceeded +/- 140 uV, and eye blinks exceeded +/- 100 uV. Any channels that exceeded over +/- 200 uV for an electrode were marked as bad. If more than 12 electrodes within a trial were marked as bad, the trial was automatically rejected. The Netstation bad channel interpolation algorithm was then applied to the accepted trials. The remaining trials were re-referenced into an average reference before averaged waveforms were created for each participant for each condition. There were two positive peaks followed by a negative response in the auditory-only condition. Based on visual inspection and existing literature (e.g. Ponton et al., 2000; Sussman et al., 2008), three ERP components were observed: P1 (80-160 ms after onset of sounds), P2 (160-260 ms) and N2 (260-400 ms). Since the peak of each component was not entirely precise, we only analyzed the mean amplitude of each condition within certain time windows.

As the distribution between frontal-central and central-parietal sites showed a reversed polarity of the potentials, the statistical analysis were performed individually, taking the average of six electrode clusters for left frontal-central (13, 29, 20, 24, 19), mid frontal-central (6, 12, 5, 11), right frontal-central (112, 111, 118, 4, 124), left central-parietal (30, 37, 36, 42, 54), mid central-parietal (7, 106, 31, REF/Cz, 80, 55), and right central-parietal (105, 104, 93, 87, 79) regions. Each participant with less than 7 trials was removed from the final analysis. The average number of available trials for each infant was 12.6 for the auditory-condition (*SD* = 4.94), 16.1 for the congruent audiovisual condition (*SD* = 5.51), and 14.4 for the incongruent condition (*SD* = 4.88). A 3 x 3 x 2 repeated-measures ANOVA was conducted on the three time windows, with the factors of *condition* (A, CAV, and IAV), *laterality* (left, midline, right) and *region* (frontal-central, central-parietal). Post-hoc analyses (least significant difference) were run where any significant (*p*-value < .05) interaction effects were reported.

5.3. Results

Figure 5.1. shows the topography distributions of the difference in responses between each condition from 100 to 400 ms after the onset of sounds. **Figure 5.2.** shows the grand-averaged waveform of the P1, P2 and the N2 for each condition at FCz, FC3, FC4, CPz, CP3, CP4 sites. The grand mean amplitudes of each component across *condition, hemisphere* and *site* can be found in **Table 5.1**.

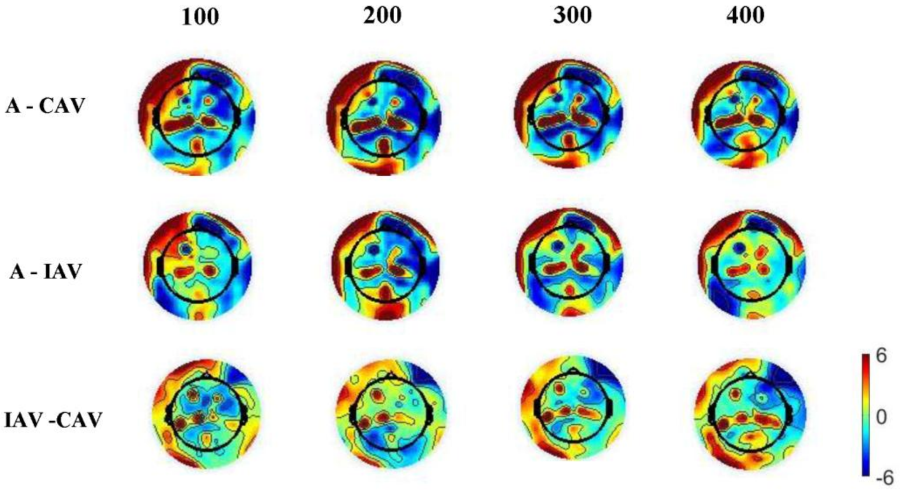


Figure 5.1. The topography distributions for comparisons between each condition from 100 to 400 ms after the onset of sounds in children. (A = auditory-only; CAV = congruent audiovisual condition; IAV = incongruent audiovisual condition)

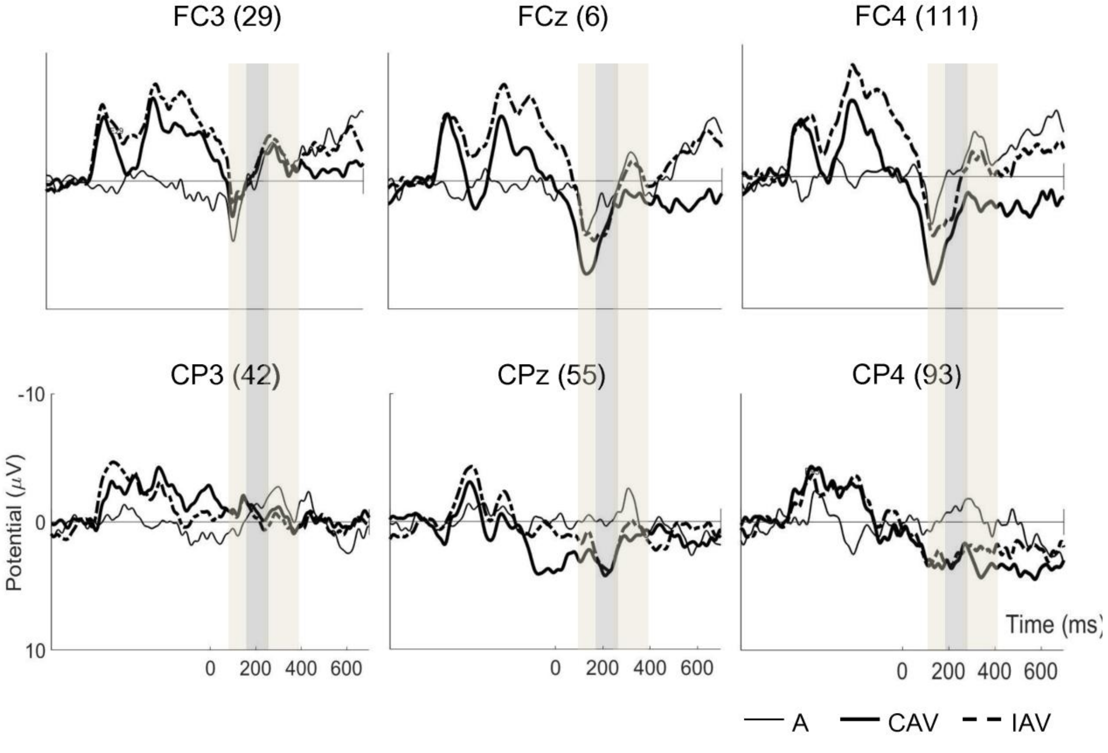


Figure 5.2. The grand average for each condition at left, mid and right frontal-central, and central-parietal electrode sites. The shaded areas were the time windows for P1, P2 and N2, respectively (left to right).

Table 5.1. The mean amplitudes of the P1, P2 and N2 in microvolt across *condition*, *hemisphere* and *site* (SD in parentheses)

		P1			P2			N2		
		L	M	R	L	M	R	L	M	R
FC	A	3.09 (4.91)	2.40 (5.11)	1.94 (3.90)	.60 (3.87)	1.04 (5.14)	.08 (4.97)	-1.62 (5.38)	-.98 (5.52)	-1.94 (4.66)
	CAV	1.91 (4.96)	5.06 (4.38)	6.00 (4.00)	-.20 (5.70)	2.72 (5.02)	3.26 (4.52)	-1.95 (5.76)	.84 (5.63)	1.17 (4.70)
	IAV	.55 (5.69)	2.81 (5.93)	3.08 (5.09)	-.33 (5.37)	2.49 (6.61)	1.89 (6.31)	-2.94 (4.96)	-.79 (5.34)	-1.68 (4.29)
CP	A	1.18 (4.72)	.93 (4.14)	.31 (2.72)	-.58 (3.75)	.68 (3.22)	-.81 (2.86)	-1.57 (4.59)	-.82 (4.14)	-1.71 (3.93)
	CAV	.93 (3.27)	3.28 (3.65)	3.24 (4.32)	.72 (2.90)	3.27 (3.66)	2.73 (4.51)	.09 (3.90)	.89 (3.61)	.85 (3.85)
	IAV	-.35 (2.34)	1.48 (3.47)	2.06 (3.78)	.44 (2.30)	3.00 (4.90)	2.64 (5.48)	-.96 (2.39)	-.42 (3.52)	.61 (3.87)

L= left; M = middle; R= right

5.3.1. P1 amplitude

The main effect of *laterality* was significant ($F(2,30) = 6.28$, $p = .005$, $\eta^2 = .295$), with smaller P1 amplitudes in left ($M = 1.22$ (.47) *u*V) compared to middle ($M = 2.66$ (.40) *u*V, $p = .002$) and right regions ($M = 2.77(.56)$ *u*V, $p = .028$). The *laterality* also significantly interacted with *condition* ($F(4,60) = 4.73$, $p = .002$, $\eta^2 = .240$). Further analysis showed that the responses were more positive in CAV compared to IAV conditions at middle ($d = 2.03$ (.85) *u*V, $p = .030$) and right regions ($d = 2.05(.83)$ *u*V, $p = .026$). The P1 amplitude was marginally larger for CAV compared to A conditions at middle regions ($d = 2.51(1.38)$ *u*V, $p = .089$), but the difference was more pronounced at right regions ($d = 3.49$ (.92) *u*V, $p = .002$). No other main effects or interactions were found.

5.3.2. P2 amplitude

The main effect was found for laterality (F(2,30) = 6.66, p = .004, η^2 = .308), with smaller P2 amplitudes in left (M = .11 (.38) *u*V) compared to middle regions (M = 2.20 (.54) *u*V, p = .002). The laterality nearly achieved significant interactions with condition (F(4,60) = 2.37, p = .063, η^2 = .136). Further analysis for comparisons between conditions was only found at right regions, with a larger P2 amplitude for CAV compared to the A condition (d = 3.36 (1.06) *u*V, p = .006). There was also a trend for more positive responses in IAV than in A condition (d = 2.63 (1.40) uV, p = .081).

5.3.3. N2 amplitude

No any significant main effects were found but a three-way interactions between *condition*, *laterality* and *condition* marginally reached significance (F(2,60) = 2.47, p = .054, η^2 = .142). Further analysis showed that significant effects for *condition* centred on the right regions. The *post-hoc* showed that larger N2 amplitudes were observed for IAV compared to the CAV condition at right frontal-central sites (d = -2.84 (1.07) *u*V, p = .018), as well as for A compared to the CAV condition at both right frontal-central (d = -3.10 (1.60) *u*V, p = .072) and central-parietal sites (d = -2.56 (1.08) *u*V, p = .032). A lager N2 approached significance in the A than in the IAV condition at right central-parietal sites (d = -2.31 (1.19) *u*V, p = .070).

5.4. Discussion

The goal of the study was to investigate neural processing for the integration of emotion perception on body expressions and sounds in early childhood. This is a preliminarily study comparing children's auditory responses in auditory-only and audiovisual conditions in order to explore the integration of emotional information. This method differs from prior studies in developing populations using the comparison between emotional content across modalities. However, the current methodology enabled us to observe the developmental changes in emotional perceptual processing at a sensory and cognitive level. Due to great maturational changes in children's auditory responses, the ERP components we examined were based on visual inspections for the responses in the auditory-only (A) condition from the onset of sounds until 400 ms, that is, double positive peaks (P1 and P2) followed by a negative response (N2).

In the present findings, both the P1 and P2 amplitudes were significantly larger for the emotionally congruent pairs relative to sound isolation in the middle and right regions. Although the polarities in children were different from adults, the latencies of the difference between auditory and audiovisual responses were in parallel with the findings in prior adult studies with similar paradigms (~ 100 ms after onset of sounds) (Jessen & Kotz, 2011; Yeh et al., 2016). Based on the assumption of multisensory perception (Giard & Peronnet, 1999), differentiation in modal responses reflects the timing for the interaction of auditory and visual perception. Therefore, the interaction of emotion processing for body expressions and sounds might also emerge at an early processing stage in 5-6 year-old children. In addition, there was a trend for the difference in the P2 amplitudes for the incongruent audiovisual compared to the auditory-only condition. This suggests that children at this age also benefit from recognizing angry sounds from emotionally congruent body expressions.

The P1 amplitudes were also differentiated between emotionally congruent and incongruent conditions. The differentiation for the congruency effects was attenuated within the P2, but then emerged again during the N2. Given that this is the case, we infer that there are distinct processes for the congruency effects during the timing of the P1 and N2. For the P1, the congruency effects might be caused by the visual processing for the two opposite valence (anger versus happiness) emotional bodies. The responses differed when the visual emotional content was presented before the involvement of the sounds. The differentiation thus extended to the auditory P1 but diminished within the P2. In addition, the more salient unpleasant stimuli are correlated with increased sustained processing relative to pleasant stimuli (Kujawa, Weinberg, Hajcak, & Klein, 2013). As such, it was plausible that the angry body expressions enhanced processing for the emotionally congruent sounds (angry vocalization), and resulted in perceptual integration with the P1. On the other hand, the happy body expressions can hardly improve processing for the angry sounds. This may also explain why the comparison between incongruent and auditory-only responses was not statistically significant. In contrast to the P1, the congruency effects were more stable within the N2, that is, at a later processing stage. In terms of the function and peak latencies, the N2 in the current data is similar to the P2 observed in adults with a similar paradigm (Yeh et al., 2016). Therefore, the N2 might reflect a function of assessing the combined emotional content, or the competition between visual and auditory information. The results also imply that children at age 5 years have the capacity to associate angry expressions from body expressions with sounds.

It is also noted that right-lateralization was found for both the modality and congruency effects across the three components. This is in line with the findings by Balconi and Vanutelli (2016) showing a higher activation for negative stimuli in incongruent audiovisual conditions relative to visual conditions in the right prefrontal regions. Based on the *valence hypothesis*, the current study only focused on the emotion of anger, which was expected to elicit relatively modality effects at the right hemisphere. However, we did not find any difference in the responses in the three regions (left, middle and right hemisphere) when auditory-only was presented. The hemisphere effects were only present in the congruent audiovisual conditions, suggesting that body expressions largely contributed to the right-lateralized processing for the negative emotions. For children, the angry expressions from the auditory modality might not be as salient as the visual modality, so it cannot elicit a specialized processing for emotion. Another reason could be related to immature systems inducing the lateralized processing for the angry vocalization. Alternatively, the materials we used did not evoke lateralization at this age. Although the current results do not dissociate these explanations, it has revealed that the congruent body expressions can enhance processing for angry sounds for children at the age of 5 years.

There are several limitations present, in a number of respects. We determined the latency of children's auditory components based on prior findings (e.g. Ponton et al., 2000) and visual inspection, but the problem of individual variance in terms of peak latencies and polarities is not completely resolved. Particularly, the current data showed a trend for a negative deflection between the P1 and P2 in the auditory-only condition but this was absent in the audiovisual condition. This is also observed by Sussman et al. (2008) whereby a negative inflection slowly emerged between the two positive peaks in children's auditory morphology. The authors considered that it is the positive components that overlapped the emerging N1 and caused a difference in peak latency in individuals. Given that the two polarities occur simultaneously, it should be noted that the current analysis of mean amplitudes could eliminate the real effects. In addition, we have less knowledge about the developmental changes in the responses to the audiovisual emotional condition across childhood and into adolescence. Therefore, additional work is needed to further separate the responses that related to the maturational changes and the modulation of body expressions.

Regarding the lateralization effects, it is possible that general effects from the visual emotion contribute to these effects, to the point where lateralisation reflects this rather than the desired combined emotional content. Although we controlled the emotional intensity of stimuli from the same modalities, the intensity of body expressions was significantly higher than that of sounds. In that case, the visual emotion, that is body expressions, might elicit lateralized processing for emotion, and enhance the lateralized effects by the presentation of sounds. Another extending question is whether the right-lateralized processing is only specialized for the negative emotions or for the emotion-related expressions in children. As we only focused on the emotion of anger, it is unknown how the neural processing of the opposite valence emotion in these children would manifest. Future studies could examine the ERP responses in another group with positive, or approach conditions, to confirm the hypothesis of brain asymmetry for emotion perception in childhood. .

In conclusion, the present observations in children showed the differentiation in responses between auditory-only and audiovisual modalities at approximately 100 ms. This study could argue that the integration of emotional perception on body expression and sounds occurs at an early sensory stage rather than a later stage. For anger, the current findings also suggest that body expressions improve the processing of the angry vocalization, with a specialized neural system for the integration of emotion perception.

Chapter 6 General Discussion

6.1. Introduction to the Discussion: Revisiting the theoretical background

In the natural environment, we usually perceive an event or object with multiple sources of information via our different sensory modalities, such as eyes or ears. Multisensory experience is ubiquitous and benefits our perception in terms of detecting the changing world (Bahrick & Lickliter, 2012). Despite this, in essence, it is challenging to process multisensory information, as it requires the ability to select relevant information from different modalities and then to synthesize these into a unified percept. However, the capacity to integrate multisensory inputs has been constructed by the end of the first postnatal year (Murray et al., 2016). A wealth of behavioural evidence with infants has shown that multisensory experience enhances learning abilities in social and non-social events when contrasted to unisensory experience (e.g. Flom & Bahrick, 2007), which highlights how important multisensory perception is and why we need to understand multisensory process related to our life. As outlined in the introduction section (in Chapter 1), the two prominent hypotheses, *Multisensory perceptual narrowing* and *Intersensory Redundancy Hypothesis* describe how multisensory perception might be shaped in early development. However, both hypotheses were built based on behavioural findings. This is still insufficient to concretely specify how the connections between one perception to others are built and affected by other influences (e.g. attention) during multisensory processing.

Related to emotion perception, a behavioural study by Zieber et al. (2014b) demonstrated that infants at 6.5-months-old can appropriately understand the emotional relationships between body expressions and vocalization. However, the findings did not provide a clear answer regarding how and when emotion perception to visual and auditory information is integrated. This is also a constraint to prior audiovisual studies with behavioural measurement. As such, we conducted ERP measurements to examine the developmental trajectory of audiovisual perception on body and vocal emotional expression. This technique allows us to establish the time course of audiovisual processing for emotional information in infants, young children and adults. ERP results from adults further showed the influence of other factors (e.g. emotional intensity and movement from body expressions) at different periods of processing for audiovisual emotional perception. Results in adults have also showed that attention can be characterized by both bottom-up and top-down influences on multisensory perception (Talsma et al., 2010). Bottom-up attention is an automatic process driven by salient properties of an object or an event. Emotions naturally induce attention when encountered in the environment. Such stimulus-driven attention achieves greater ecological validity when contrasted with paradigms containing specific attentional instructions. By contrast, top-down attention is based on our expectations and motivations. As some emotions can be more easily perceived from one compared to another modality, top-down attention can modulate modality dominance for different emotions (Focker, Gondan, & Roder, 2011; Takagi et al., 2015).

Through ERP evidence, we can understand what and when different processes of audiovisual emotion perception change from infancy to adulthood. We also attempted to fit these findings into the findings of existing behavioural studies and hypotheses to complementarily interpret the developmental courses for audiovisual perception to emotional body expressions and sounds. In the thesis, we starts with two theories, *Perceptual Narrowing* and *Intersensory Redundancy Hypothesis (IRH)*, to elucidate how integration of multisensory perception occur in an early life and the advantage of multisensory compared to unimodal learning. However, the current findings from developing populations cannot explicitly feed back to either of these theories. In both infant and children studies, the responses to body-voice pairs and vocalization responses differed by 400 ms. Despite this, these results cannot be interpreted by what the *IRH* highlights as the beneficial learning from audiovisual relative to auditory-only information.

From a *multisensory perceptual narrowing* perspective, it is unknown whether infants at 6.5-month can specifically match angry body expressions to angry sounds. The significant congruency effect found in this experiment could be due to recognition of angry from fearful body expressions, with the consequence that the differentiation might be unrelated to the involvement of vocal information. However, behavioural studies have indicated children at 5 years can discriminate angry, fearful and happy body expressions (Nelson & Russell, 2011). Given that this is case, the congruency effect in children possibly indicates well-developed specific associations between angry expressions and angry sounds.

Regarding adults studies, a shorter N1 peak latency and reduced N1 amplitude in the audiovisual compared to the auditory condition can reflect the efficiency of multisensory information. This could be accordant with *IRH* whereby the redundant information elicits greater attentional salience compared to the same events present through one modality. In addition, the significant comparisons between emotionally congruent and incongruent body-voice combinations is in congruence with *perceptual narrowing*, revealing the neural tuning to particular emotional pairs across body expressions and sounds. Nevertheless, *perceptual narrowing* and *IRH* hypotheses do not explain other influences. For example, how does modality dominance relate to neural perceptual narrowing? In addition, the two hypotheses are based on bottom-up attention, which has a focus on salient attributes of events or stimulus. In contrast, the modulation of top-down attention on perceptual tuning remains unexplored, although this approach does not logically interact with the first study showing distinct attentional modulation on audiovisual perception to angry and fearful expressions.

Taken together, the studies with developing groups do not have the capacity to provide evidence for either of these two dominant theories. Due to the present setting only focusing on the emotion of anger, the results are not available in order to demonstrate the accuracy of the hypothesis. Also, more work is needed, for example, with further investigations on different ages of children, in order to understand the neural mechanisms underlying the maturational changes related to perceptual narrowing. On the other hand, other factors should also be considered, such as attention, in both frameworks.

6.2. Summary of findings

The present studies exploring audiovisual emotional perception from body expressions and sounds in infants and children are preliminary work. Due to substantial variation in auditory responses during both infancy and early childhood, it is more challenging to determine accurate time windows for modality effects. As such, it may be arbitrary to use an open approach that may reflect emotional perceptual integration, for example, examining the ERP time course for angry stimuli pairs and then looking for the same component or deflection for another emotions. Rather than that, we conducted the traditional method utilized when examining perceptual integration in adults, that is, comparing between auditory-only and audiovisual responses. We firstly applied the paradigm to adults and the finding showed a significant difference between auditory-only and audiovisual responses at approximately 100 ms after onset of an auditory stimulus. This is also consistent with other audiovisual studies (Jessen & Kotz, 2011), suggesting the paradigm could also be practical for exploring the issue in developing populations. On top of this, the same paradigm can directly examine the neural processing of developmental change in the integration of emotion perception. Although the morphology in young populations is distinct from adults, an increasing number of studies have indicated some robust components during infancy (i.e., P150-P350-N450) and young childhood (i.e.,P1-N1) that might reflect specific processing (e.g., attention to salient stimulus, Missana et al., 2017; Otte et al., 2015; Purhonen et al., 2004). In the following sections, the specific findings from each study are also summarized.

6.2.1. An exploration of audiovisual emotional perception on body and sounds in adults

In order to understand developed neural patterns of audiovisual emotional perception on body expressions and sounds, we observed two audiovisual ERP components, the N1 and P2, in auditory and audiovisual conditions in adults. We also examined the influences of attention, emotion types and emotional intensity of body expressions. Consistent with prior audiovisual research (Besle et al., 2004; Jessen & Kotz, 2011; Pourtois et al., 2002; Stekelenburg & Vroomen, 2007), the auditory N1 amplitude, a negative peak at 100 ms after onset of sounds, was reduced for the audiovisual compared to the auditory-only conditions. This suggests that an interaction between visual and auditory perception takes place at an early sensory stage (van Wassenhove et al., 2005). Further, the modality effects were distinct across emotion types and as a function of attention instruction. This was evidenced by two groups of adults that were presented the same stimuli and conditions but with different attentional instructions. When attention was directed to visual non-emotional content, the N1 amplitudes for angry sounds were reduced in both emotionally congruent and incongruent audiovisual conditions in contrast to auditory-only conditions. For fearful sounds, decreased N1 amplitudes were only found for incongruent pairs compared to auditory-only conditions. In contrast, when attention was directed to emotional sounds, the modality effects were decreased for anger, but were absent for fear. This may be accounted for by modality dominance (Spence & Squire, 2003), whereby some characteristics are more easily perceived by one modality than by others.

Regarding the P2, a positive response peaking at 200 ms, showed different directions of responses to the emotional congruency across auditory and visual modalities. For fearful sounds, attenuated P2 amplitudes were found for incongruent compared to congruent responses. On the contrary, the P2 amplitudes for angry sounds were reduced for congruent compared to incongruent conditions. It is likely that the P2 amplitude was reduced when angry body expressions were presented beforehand. This is in line with the assumption that the P2 might be a process indexing a competition across modal information (Stekelenburg & Vroomen, 2007) or an assessment of the emotional content (Paulmann et al., 2009).

Taken together, the first study indicated three main findings. Firstly, the N1 and the P2 components are likely to reflect two functionally different processes during audiovisual perception. While the N1 reflects a process for the integration (or interaction) of audiovisual perception, the P2 is relevant for detecting audiovisual combined content. Secondly, the N1 amplitude can be modulated by attentional instructions but the influence was not evident within P2. A final point is that modality dominance differed for anger and fear; however, attention and emotional intensity can modulate the effects upon both emotional expressions.

6.2.2. The modulation of body type during the integration of emotion perception

Neuroimaging evidence has shown that there are different neural circuits for perceiving dynamic and static bodies displaying the same emotional expressions (Grezes et al., 2007; Pichon et al., 2008). Based on the first study in the thesis, we further examined the influence of body types on audiovisual emotional perception in adults. Similar to the method in Chapter 2, the study in Chapter 3 presented factors of condition and emotion type, in additional to visual types, whereby body expressions were presented with movements (dynamic) and without movements (static).

The results confirmed our hypothesis that the types of body expressions impact audiovisual emotional processing for body expressions combined with sounds. With presentation of dynamic body expressions, the N1 amplitudes were reduced for both angry and fearful sounds in the audiovisual conditions compared to the auditory-only conditions. However, when body expressions were static, the modulation was distinct between the two emotional expressions. The modality effects can still be observed for fearful expressions within the N1 amplitudes, but were absent for angry expressions. Likewise, the factor of body types also differently modulated the congruency effects for angry and fearful expressions. For angry sounds, congruency effects within the P2 amplitudes were observed for dynamic body expressions but not static expressions. By contrast, the congruency effects were observed for fearful sounds presented with dynamic and static body expressions. Although neuroimaging studies provided evidence that that body expressions with movement activate brain areas related to emotional understanding (e.g. superior temporal sulcus; Gallese et al., 2004; Iacoboni, 2005), the study reported in this thesis indicated that dynamic information is not a compulsory cue to recognizing fear from body expressions. Comparatively, motion plays a more important role in perceiving an angry body.

6.2.3. Neural correlates indicate the integration of emotion perception in infancy

In Chapter 4 we extended findings in adults by exploring the integration of emotion perception in early development. Zieber et al. (2014b) showed that the capacity to match emotional relationships between body expressions and sounds might emerge by 6.5-months of age. We therefore measured the neural activities in this age of infants, aiming to understand the neural mechanism underlying audiovisual emotional perception. We applied the same techniques as in the previous studies with adults, comparing responses in the auditory-only conditions to audiovisual conditions in infants, to observe the developmental change in the neural processing of the integration of emotion perception. The method also enables us to explore the processes at a perceptual and a cognitive level in infants. However, it is challenging to clearly define infants' auditory components as they are usually changeable in terms of latencies, amplitudes and polarities. Based on several studies investigating evoked auditory responses during infancy (e.g. Kushnerenko et al., 2002), we indexed three relatively stable components, P150, P350 and N450, for the comparisons between modal responses and emotional congruency across the auditory and visual responses.

The results were similar to the results in adults in some ways. Firstly, the infants' responses also reduced in amplitude for audiovisual compared to auditory-only conditions at 100 ms post-stimulus, particularly when fearful body expressions were presented (emotionally incongruent audiovisual condition). This effect extended to the P350 amplitudes at frontal to central-parietal regions. This may be accounted for because the interaction of emotion perception occurs at an early sensory stage in 6.5-month-old infants, rather than at a later processing stage (e.g. ~ 400 ms; Grossmann et al., 2006). Since the P350 is thought to be related to attentional processing (Kushnerenko et al., 2002; Otte et al., 2015), we inferred that the emotionally incompatible audiovisual combination might appear salient and novel to infants, causing processing to take longer, up until P350. Secondly, congruency effects were also observed, broadly elicited across the three components at left frontal-central regions. However, the effects within P150 and P350 are likely to be confounded with the processing of visual-only emotions (anger vs. fearful body). As the 600-ms body expressions preceded sounds, the responses might be differentiated for the two emotional body expressions before the presentation of sounds.

In contrast, both modality and congruency effects were left-lateralized in the study, whereas no any lateralization was found in adults. Although the current results provides more compelling evidence for left over right lateralization offered in previous studies (e.g. Missana et al., 2015), in either case, our results in adults suggests that there might be a maturational strategy of lateralized processing for emotion-related information during infancy.

6.2.4. The maturation process for the integration of emotion integration in early childhood

Chapter 5 describes the neural processing for the integration of emotion perception in early childhood. Behavioural studies have pointed out that strategies of using cues for perceiving emotion could shift during childhood (e.g. Aguert, Laval, Le Bigot, & Bernicot, 2010; Gil et al., 2016). The neural circuits underlying behavioural responses appear to re-weight strategies of recognizing emotion across modalities between infancy and adulthood. In order to realize the maturational change in the neural patterns of perceptual integration, we measured 5- to 6-year-old children with the same design as our infant study (Chapter 4). Although the auditory responses still greatly vary in terms of latency and polarity during childhood, an increasing number of studies have been indicated the developmental trajectory of auditory responses (e.g.

Ponton et al., 2000; Sussman et al., 2008). Based on this literature and visual inspection of our data, the three components, P1 (80-160 ms), P2 (160-260 ms) and N2 (260-400 ms) were observed for the comparisons between modal responses and emotional congruency responses in the final study.

Results showed that the P1 and P2 amplitudes reduced for emotionally congruent audiovisual conditions compared to auditory-only conditions, with a slightly attenuated P2 amplitude for incongruent audiovisual when contrasted with the auditory-only conditions. Despite a different polarity from adults, the results indicate that the interaction of emotion perception emerges at a sensory level in early childhood. This also suggests that the emotionally congruent body expressions (angry body) contributed to the processing for angry sounds for children at the age of 5. Regarding congruency effects, they were activated within the P1 and the N2 amplitudes. However, the effects might reflect different processes underlying the two components. As the body expressions in the study were presented with two opposite valences of emotion (anger vs. happiness); plausibly the effects on emotional content might be greatly elicited before the sounds were presented. Therefore, the congruency effects within the P1 might be involved with the comparison of visual-only emotional expressions. Alternatively, the congruency effect at a later stage may be associated with the processing of combined emotional content.

It is also noted that both modality and congruency effects were lateralized at the right hemisphere in young children. Similar findings were also seen in a recent study by Balconi and Vanutelli (2016) showing right-lateralization for incongruent negative stimuli when (un)comforting pictures with sounds were displayed to adults. Although it is unknown whether the right-lateralized processing is specific to negative valence or general-category emotion, the results highlight that the right hemisphere plays an important role in processing angry expressions in early childhood.

6.3. Implication of the findings

6.3.1. Resolving the integration of emotion perception at different processing stages in adults

In the thesis, we compared auditory responses among conditions with other factors across three age groups. This enables us to understand the developmental changes in the timing of specific processing, either dissociated or overlapping, during the interaction of emotional perception of body expressions and sounds. Results in adults set out the influence of condition, emotion type, emotional intensity, and type of body exhibition within the N1 and P2 components (Chapter 2 and Chapter 3). In accordance with the study of Jessen and Kotz (2011) and other perceptual domains (e.g., Stekelenburg & Vroomen, 2007), our studies showed differentiation in the amplitudes between auditory-only and audiovisual conditions within the N1 time window. This indicates audiovisual processing for the body and vocal emotional expressions occurring as early as 100 ms after presentation of sounds. Furthermore, the effects on emotional congruency across visual and auditory modalities were observed within the P2 but not within the N1. The P2 component has been indicated in deeper processing related to cross-modal content in emotion and other perceptual domains (Ho et al., 2014; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). This explanation fits with our adult findings on different directions of congruency effects; either increasing or decreasing responses to emotionally congruent pairs in contrast to incongruent pairs. Collectively, our work with adults was in agreement with other multisensory research that the N1 and P2 components reflect different functional processes during audiovisual perception (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). More specifically, the current study suggests that the N1 reflects a process of interaction between multisensory perceptual information, whereas the P2 is associated with competition between auditory and visual information and the assessment of combined emotional content.

In reality it is difficult to divide multisensory integration into multiple sub-progresses as they occur in parallel rather than sequentially. For example, in our studies both N1 and P2 components were modulated by the factors of emotional intensity and types of body that were exhibited. Therefore, underlying the N1 or the P2, processes related to various factors take place. On the other hand, the N1 is likely to be specifically modulated by attention. In Chapter 2, the group attending to non-emotional features showed the modality effects within the N1. However, the modality effects nearly disappeared for both emotions when attention was directed to the sounds. By contrast, the manipulation of attention appears to not affect the P2. This is consistent to the finding by Ho et al. (2014) showing the attentional influence within N1 relative to the P2 time window. The present findings also indicate that attention can influence multisensory integration in a bottom-up (stimulus-driven) and top-down (motivation-driven) fashion at the pre-attentional processing stage (Talsma et al., 2010).

6.3.2. Diverse findings on anger and fear during the integration of emotion processing

The work in the thesis focused on perceptual integration of the emotions of anger and fear. On the emotional dimension, both anger and fear are high arousal and negative valence emotions. However, they convey different social signals. While anger often displays ongoing aggression from the expressers, fear exhibits potential environment threat by the expressers (Adams, Gordon, Baird, Ambady, & Kleck, 2003). In contrast to fear, anger appears to be an more interactive message that the observers are required to modify their behaviour in tune with the forthcoming interaction (Pichon et al., 2009). The difference can be supported by neuroimaging research showing different neural circuits for the two emotional body expressions (Pichon et al., 2009). Two ERP studies (Jessen & Kotz, 2011; Jessen et al., 2012) have also provided evidence of different timing for processing the two emotions, with a shorter N1 peak latency for anger than for fear in audiovisual and auditory-only conditions.

In the studies with adults we replicated Jessen's findings and also found significant differences between auditory-only and audiovisual responses for both emotional expressions. With the involvement of other factors, these modality effects were diverse for the two emotions. In Chapter 2, we investigated attentional influence on two groups with different attentional instruction. When the group was instructed to attend to non-emotional features, the N1 amplitudes to angry sounds differed in auditory-only conditions compared to both emotionally congruent and incongruent audiovisual conditions. For fearful sounds, the modality effects were mostly from auditory-only in comparison with incongruent audiovisual conditions. The other group were instructed to judge emotional sounds and showed attenuated modality effects in both emotional expressions, particularly for angry body expressions. For this group, the modality effects largely disappeared. We assumed that not only neural circuits of processing but also modality dominance (Spence & Squire, 2003) are distinct for anger and fear, causing different attentional modulations on the two emotions. Prior behavioural work has demonstrated the existence of modality dominance in emotions, showing different accuracy across emotion recognition for face-sound pairs (Collignon et al., 2008; Focker et al., 2011; Takagi et al., 2015). Takagi et al. (2015) further examined the influence of attention and modality dominance on multisensory perception. Without attentional instruction to a specific modality (e.g. attend to face or sound), fear is prone to be perceived via the auditory modality whereas there is no difference in performance for recognition of anger from face or sound stimuli. However, modality dominance for anger altered with attentional instructions in our study. Based on these findings, sound might be sufficient information for recognizing fear, whereas anger can be easily perceived by both visual and auditory channels in isolation. We therefore inferred that visual or auditory modality containing angry information can benefit the processing of anger perceived from another modality. In contrast, emotionally congruent information across multisensory channels might not be advantageous for recognizing fear. Furthermore, if the incongruent body expressions paired fearful sounds, were presented via a dominant modality, the processing for fear might be inferred. Since angry expressions can be strongly conveyed by the visual modality, this may explain the present results that only the observation of different responses to fearful sounds in auditory-only compared to incongruent audiovisual conditions. However, the modality dominance for each emotion can be differently modulated depending on how attention is instructed.

We also found that body exhibitions differently modulate emotion processing for anger and fear. In Chapter 3, we explored the influence of body types on audiovisual emotional perception by presenting dynamic and static body expressions. The results indicated that the N1 amplitudes were reduced for sounds paired for both types of body expressions in contrast to sound-only conditions. However, the modality effects were not observed for angry expressions when body expressions are static. Evidence from behavioural (Atkinson et al., 2004; Coulson, 2004) and neuroimaging studies (Grezes et al., 2007; Pichon et al., 2008) have shown that body expressions with movements help to improve emotional recognition. Despite this, it is likely that each emotion has a specifically optimized feature that can be most readily recognized (Coulson, 2004). In the natural environment, an angry body is usually characterized by high-velocity movements, whereas fear shows comparatively low-velocity or even less movements (Roether et al., 2009a). In that case, dynamic information is relatively important to perceiving angry bodies; however, it is plausible to successfully recognize fearful body expressions from dynamic or static exhibitions. The findings of our second study also suggest that the type of body expressions differently modulate processing for each emotion during the perceptual integration of emotion.

6.3.3. The meaning of audiovisual integration of emotion perception in infancy

In Chapter 4 we see that infants' responses differed in the auditory-only when contrasted with both congruent and incongruent audiovisual conditions at around 100 ms. Infants' auditory patterns showed a broad positive peak followed by a negative waveform, which are distinct from those observed in adults. Despite this, a dissociation between unisensory and multisensory responses was observed, which implies that the integration of audiovisual perception occurs (Giard & Peronnet, 1999). To date, little ERP work has focused on the issue of audiovisual emotional perception during infancy; therefore, we cautiously interpreted the findings on the dissociations in the responses to modality as well as emotional congruency in infants. Here, we addressed a question related to construction of the emotion relationship between body expressions and sounds. It is debatable whether infants substantially understand the meaning of the two emotions from body expressions. Infants by 10-months discriminate emotional faces in terms of emotional valence or physical features (e.g. Soken & Pick, 1992). It is not until 10- to 12-months of age that they begin to understand emotional meaning by connecting others' emotions to environmental events, which is called social referencing (Widen & Russell, 2008). Reviewing prior infant research that demonstrated the ability to discriminate emotional body expressions by 8-months (Missana et al., 2015; Missana et al., 2014; Zieber et al., 2014b), we found that body expressions were presented with opposite emotional valence in these studies (e.g. anger/fear versus happiness). This is important as happiness might be a relatively familiar emotion in the infants' environment (e.g. Striano, Brennan, & Vanman, 2002; Walker-Andrews, 2008). Although both emotions (anger and fear) we used have negative valence, we still obtained effects for emotional congruency with the presentation of sounds. This implies that 6.5-month-old infants are likely to be able to discriminate angry from fearful body expressions. As we controlled the amount of body movement and luminance for body expressions it is unlikely that motion accounted for this discrimination. Therefore, it is unknown whether the young populations extracted emotional meaning from body expressions or categorized the visual stimuli with other cues beyond valence.

There is another question about the certainty of emotional perceptual integration in the present work with infants. Results showed that the congruency effects emerge at the P150 time window, and then gradually decreased within the N450 epoch. This is similar to results in children (Chapter 4) but opposite to those in adults, who showed stable congruency effects at a later stage (~ 200- 330 ms, in Chapter 2 &3). The congruency effects were elicited earlier in young populations than those seen in adults. Since body expressions were presented before the sounds for 600ms, we considered that the congruency effects might also be confounded with the response to visual-only emotion discrimination. On top of this, the *perceptual narrowing hypothesis* (Murray et al., 2016) discussed in Chapter 1 suggests the neural tuning for multisensory stimuli at an immature age might be convergent rather than interactive.

Even if multisensory integration happens, the neural process broadly tunes to the shared features across the modalities. It is conjectured that this stage of perceptual tuning is still relatively category-general rather than specific to the multisensory attributes.

This study in this thesis with infant participants compared responses in modalities to examine audiovisual emotional perception. The differences in the responses to modality as well as emotional congruency were found within the P150 and P350 components, suggesting that the integration of emotion perception may occur at a sensory processing stage in 6.5-month-old infants. Despite the questions we mentioned, our work offers evidence that the method of comparing modality responses enables us to observe the process of multisensory emotional information integration in infants at a sensory level. This is at variance with prior work using a congruency paradigm to consider the ability to process specific features across modalities, like speech, synchronized timing, emotion, all of which have reported differences at nearly 450 ms after the presentation of emotional sounds during infancy (Grossmann et al., 2006; Hyde et al., 2011; Reynolds et al., 2014)

6.3.4. The developmental course of the integration of emotion perception

In Chapter 5, we further explored the maturational changes in the neural mechanisms underlying the integration of emotion perception. We examined 5- to 6-year-old typically developing children with the same paradigm outlined in Chapter 4 but with a different emotional contrast (happiness versus anger) across modalities. Since the two emotions are categorized as opposite in valence, greater congruency effects were expected to be observed in these young children.

The results in young children were similar to our findings in adults. The responses to the auditory-only and audiovisual conditions differed from 100 ms. In particular, the difference in conditions was more pronounced when comparing responses in auditory-only and emotionally congruent audiovisual conditions. This differed from the findings in infants (Chapter 4), which showed greater differences between auditory-only and incongruent audiovisual responses. An explanation could be that maturational changes in neural activities cause the discrepant responses to emotion congruency effects in infant and children. In our infant study, the fearful body expressions were presented with angry sounds for emotionally incongruent pairs. Based on the perceptual narrowing perspective (Bahrick & Lickliter, 2012), at an immature stage that neural tuning may still be broad and responses are general-category for multisensory attributes. As both fear and anger are threatening emotions, infants might mistakenly identity the fearful body and angry sounds as emotionally suitable pairs. It could also be the fact that infants were relatively more familiar with fearful than angry body expressions. With an increasing age or more experience to the environment, perceptual tuning becomes narrower and more constrained (Murray et al., 2016). On the other hand, the incongruent visual stimuli were happy body expressions in the study with children. Due to oppositely valenced emotions, it may be easier for children to detect differences between the two emotional expressions. In that case, children can rapidly perceive the difference between happy and angry body expressions. The presentation of angry bodies therefore enhances the processing of the angry emotional sounds at a sensory stage, causing differentiation in the P1 and P2 amplitudes to sounds-only and congruent audiovisual conditions. In contrast, the happy body expressions failed to evoke perceptual interaction to angry sounds.

In the same vein, the comparisons between congruent and incongruent responses were found in the P1 and N2 components. For the P1, visual-only emotion effects may be confounded as the body expressions preceded sounds for 600 ms. Alternatively, the congruency effects robustly presented within the N2 might essentially reflect the similar function as the P2 observed in adults, either assessing the emotional content or assessing the competing information across modalities at a later processing stage. This thesis indicates that young children can discriminate anger and happiness from body expressions as well as connect the concept of angry expressions between body expressions and sounds.

6.3.5. The function of lateralization in early development

In the thesis we investigated the integration of emotion processing in three age groups with a similar paradigm. Results showed that the modality and congruency effects were lateralized in both infants and young children, whereas adults did not show significant lateralization for any effects. When we further examined the findings on infants and children, it is interesting that the two groups showed opposite lateralized processing for congruency effects. While the congruency effects distributed in left hemisphere electrodes in infants, right-lateralized modality and congruency effects were found in children. The distinction might indicate different understandings of emotional expression for the two groups. As mentioned in Section **1.2.2** of the thesis, several hypotheses focus upon the asymmetry of emotion processing. The current findings are unlikely to support *Right Hemisphere Hypothesis* (Borod et al., 1998) stating that emotion-related information is processed by the right hemisphere. In contrast, the results of this thesis can be accounted for by the *Valence-Specific* (Ahern & Schwartz, 1985) or the *Approach-Withdrawal hypothesis* (Davidson, 1992b) whereby two hemispheres are respectively dominant in processing for positive-negative valence of emotion or approach-withdrawal behaviour. However, the two theories have contradictory perspectives about anger; the emotion that we examined in infants and young children. From a *Valence-Specific* perspective, anger is a negative emotion and so more engagement of brain activity was expected to be observed within right hemisphere electrodes. In the *Approach-Withdrawal* theory, anger and happiness are classified as approach behaviours; therefore, it is supposed to elicit higher activation in left brain areas. It appears that our findings cannot completely be explained by either of the hypotheses. Alternatively, the different lateralization may be more relevant to the fact that infants and young children differently read emotions from body expressions. As discussed in Section 6.2.2, it is unclear whether infants understood the meaning of angry expressions or generalized the angry expressions as a positive emotion and an approach behaviour. As for young children, they are capable of recognizing the angry body expressions but might consider it as a withdrawal behaviour. Other factors, like neural maturation, could also interpret the developmental difference, but with unknown timing of the formation of adult-like patterns. Despite insufficient evidence to explain the developmental changes in lateralization, our work with children and infants provides evidence of asymmetric audiovisual emotional processing in early development.

6.4. Limitations and Direction for Future Studies

6.4.1. Potential limitations for the studies with adults

In the present study with adults, the techniques that were employed might cause some questions related to attention between the visual and auditory modalities. In the tasks without any attention instructed to emotion-related properties, the participants were required to judge non-emotional features (i.e. clothes, gender) (in Chapter 2 and Chapter 3). The aim of the design was to ensure there were no attention biases in emotion processing across the two modalities, rather than to draw attention away from the auditory processing. This is because we believe that emotional sounds would still be automatically processed without instruction. As indicated by Talsma et al. (2010), attention within multisensory integration is considered to be from both bottom-up and top-down mechanisms. Whereas bottom-up attention is an automatic process that is driven by salient events or objects in the environment, top-down attention is a selective bias process for the events that are aligned with the observers' expectations. Without instructions related to directing attention, bottom-up attention can still happen within multisensory integration (Talsma & Woldorff, 2005). Illusion (Shams, Kamitani, & Shimojo, 2000) and the McGurk effect (McGurk & MacDonald, 1976) are classical examples where the presentation of sounds can automatically influence the perceived properties of visual information. However, even if the instructions were not directing attention to emotional attributes within a specific modality, we cannot rule out the possibility of attentional biases for visual emotional content.

We are also aware of unbalanced conditions in the work with adults. The blocks presenting high-low visual intensity (in Chapter 2) or dynamic-static body expressions (in Chapter 3) were contrasted to the same emotional sounds. That is, the number of auditory-only conditions was twice as many as the other three conditions, which might enable these participants to become habituated to the repetitive sounds. We therefore inspected responses to the auditory-only conditions, and found no statistical difference between the blocks presented with dynamic or static body expressions. From another perspective, it is a problem that the auditory-only condition was presented less often than the audiovisual conditions (congruent and incongruent conditions). Due to less frequent presentation of auditory-only stimuli and a lack of visual information, this may increase attention to sounds in isolation. To examine the contribution of the number of trials on the ERP, we created an ERP by artificially reducing the audiovisual condition to the same number as in the auditory-only conditions for each participant. The results were similar to those reported in the relevant chapter, with the modality effects clearly still present within the N1 and P2 for both emotions. Despite designing the study to minimize attentional bias, we still cannot entirely exclude potential problems related to attention. Nonetheless, our results in adults show that the significant comparison between the auditory-only and audiovisual conditions are unlikely to be completely generated by some sort of "pop-out" effect from the auditory-only condition.

6.4.2. Potential limitations for the studies with infants and children

In order to explore the maturational changes in audiovisual emotion processing, we modified the adult paradigm in this thesis to better suit the physical states of infants and young children. The simplified designs of these studies allowed us to obtain valid data from the two groups, with a reduced number of trials and conditions when contrasted with the adult studies. Therefore, both infant and children groups were individually presented with angry sounds in three conditions: auditory-only, emotionally congruent and incongruent audiovisual conditions. However, these studies excluded the presentation of reverse incongruent pairs (angry bodies with fear sounds in our infant study, or angry body with happy sounds in our study with young children). This does not allow us to further understand the influence of visual context on auditory processing within later processing stages (~ 250 ms). Without contrasting these emotional expressions, we cannot identify whether the lateralized processing we observed is for specific or general-categorized emotions for a certain hemisphere, and cannot determine whether the *Valence-specific* or *Approach-Withdrawal* hypothesis better explains asymmetric emotion processing during early development. It is also unclear whether developmental changes in neural systems explain why we found the opposite lateralization for the modality effects between infants and children.

Another limitation in the present studies with developing populations relates to the certainty of ERPs responses. This is also a typical issue in ERP developmental studies (e.g. Grossmann et al., 2006; Knowland et al., 2014; see Trainor 2007, for a review). In Chapter 4 and Chapter 5, the timing of ERP components are based on visual inspection, which is standard in previous literature on developmental changes in auditory components. As the components in young populations are not as clearly defined as in adult samples, we calculated the effects with a mean amplitude analysis. This traditional analysis is also less influenced by high-frequency noise compared to other analyses, such as peak amplitude. Despite this, as mentioned in the introduction 1.1.2., the question of substantial individual variation should still be considered in the study. The grand average waveforms could be attenuated in amplitude, with some children showing positive peaks but others showing negative deflections during the same period (Trainor, 2007). Although we tried to find different time windows, it was not precisely clear when to determine infants and young children's ERPs in terms of latency, polarity and distribution. Moreover, the traditional approach we used requires the averaging of all participants' responses together and the extraction within a certain time window, which makes it difficult to separate the responses to presentation of stimuli from those that correspond to irrelevant information (e.g. eye blinks). This particularly happens for developing individuals who do not pay attention to the stimuli for a sustained period of time and make lots of movements. Despite these limitations, the current series of studies constitutes a preliminary exploration of this research area. A key contribution that we make is to provide a clear direction for future research to follow, which we now discuss.

6.5. Future Directions

The focus of the thesis was audiovisual perception of angry and fearful expressions. However, the process of audiovisual perception might be distinct for other emotions (e.g., happiness, sadness) as has been indicated by evidence suggesting there are different neural routes for each emotion related to maturational courses and modalities (e.g., Chronaki et al., 2015; Nelson & Russell, 2011). It would be worth exploring other emotions that serve different purposes in our social life. On top of this, a question concerning the developmental findings in this thesis is that both infants and young children showed lateralized emotion-related processing but in different hemispheres. It is unknown whether processing for emotions changes with increasing ages or whether other factors caused the effect. Furthermore, whether the asymmetry is specific for negative (i.e., anger), or general-category emotions during audiovisual processing in early development is currently unknown but poses an empirical question for future research to resolve. Currently both the *Valence-Specific* and the *Approach-Withdrawal* hypothesis account for the lateralization of emotion processing in adults. Comparatively, fewer studies have confirmed the two hypotheses about asymmetry of emotion processing in young populations, particularly during multisensory processing. In the thesis, both infants and children were not presented with conditions for emotions other than anger. The reason for this was that having additional conditions would make the experiment too long for young participants with limited attention spans, resulting in too few trials per condition. Directly contrasting different pairs of these other emotions could be an additional research stream to explore asymmetric emotion processing in early development.

As for research with developing populations, there can be further improvements in data analysis. In this thesis we analyzed mean amplitudes of infant and children's responses. Compared to other traditional analysis (i.e. peak amplitude), the approach is less influenced by high-frequency noise and less biased for certain time windows for some electrodes sites (Steven, 2005). Despite this, there are still some constraints to calculate effects in developmental studies. Firstly, the infant and children's ERP components we reported have been previously present in the developmental literature (Kushnerenko et al., 2002; Ponton et al., 2000; Sussman et al., 2008); however, auditory evoked responses alter with stimulus type and paradigm in infancy and childhood (Ceponiene et al., 2002). In addition, as indicated in section **6.4.2.,** caution needs to be exercised for the results obtained from the traditional analysis due to great individual variance in the same time window. The analysis also makes it difficult to directly make comparisons between adults and developing groups due to different topographical distributions for modality or congruency effects. The results in the thesis showed that the comparison in auditory-only and audiovisual responses was intensively distributed at frontal-central areas in adults, whereas developing groups showed a broader distribution of the effects from frontal to central-parietal sites. This suggests the engagement of brain process for the modality effects might be distinct in regions among the three groups (infants, children, adults). Moreover, it is debatable whether the same comparisons showing in different brain areas reflect the same processing. For example, our infant data showed that P1 amplitudes differed in congruent and incongruent conditions at left frontal-central and central-parietal electrodes. Apart from more studies with an advanced design to examine the certainty of the current findings, an alternative analysis approach could be another way to overcome the limitations of traditional analysis. Principal component analysis (PCA) is an increasingly popular approach for identifying potential responses to the specific stimuli or events in a certain time window (e.g. Kayser, Tenke, & Bruder, 1998). Unlike traditional analysis using visual inspection for pursuing potential effects, PCA constructs factors on the basis of covariance across individual experimental conditions. It is also a mathematical process that can reduce potential for experimenter bias. To date, PCA has been demonstrated to be a powerful approach in developmental studies of cognitive and language issues (see Molfese, Molfese, & Kelly, 2001, for a review). For example, Rivera-Gaxiola et al. (2007) used PCA to assess infants' responses to contrasts of native and non-native speech. Two principal components (P150-250 and N250-550) were identified but distributed at different brain regions. Comparatively, the use of PCA increases our ability to identify certain effects that do not sequentially occur in the same electrode clusters. Although PCA has rarely been utilized in developmental studies on emotion perception, it seems to be a useful approach for isolating effects in developing populations beyond the domain of language perception.

Other research direction that could be more explored is about timing of information presentation across the visual and auditory modalities. In the present studies, the body expressions were presented earlier than vocalizations for 600 ms. As the latency for body expressions were longer than vocalizations, we presented the former stimulus beforehand and made it synchronously disappear with the auditory stimulus. Also, the auditory responses were mainly observed for effects of perceptual integration. In that case, the preceding visual stimulus could provide a predictive context that could guide the auditory processing. This current paradigm is plausible in social situations. However, there could be real-world situations where we hear other' voices before seeing their facial or body expressions or the two sets of modal information fades out asychronously. Studies have indicated that the synchronization of information presentation across the modalities could also impact perceptual integration (Hyde et al., 2011; Reynolds et al., 2014). As such, the results in the thesis may not generalize to all multisensory emotional situations.

The present work could be a beginning for investigating the development of neural mechanisms underlying audiovisual emotional processing of the body and sounds. The eventual aim would be to identify how neural systems develop for audiovisual emotional perception at different maturational stages. Given the achievements of the studies in this thesis, these results can be a reference point for future work examining individuals with deficits in emotion processing, such as in individuals with autism. Were this to be the case, we could advance our understanding of how strategies of multisensory processing for emotions, for example, modality dominance, are different between healthy and clinical groups. Plausibly, such findings could also provide a neural biomarker that contributes to early detection and intervention for infants at risk for autism.

6.6. Conclusions

The series of studies in the thesis comprised ERP measurements to investigate emotion perception of interactions between bodily derived and auditory emotional expressions across development. Results from adult data show decreased auditory N1 amplitudes for audiovisual compared to auditory-only conditions, implying that interactions between visual and auditory perception occurs at an early sensory stage. In contrast, the effects of emotional congruency across two modal sources were mainly found within the P2 time window. Furthermore, the N1 was affected by attention more than P2 effects. Given that this is the case, there might be two functionally different processes underlying the N1 and P2. However, both the N1 and P2 were modulated by emotion intensity and the type of body expressions. The modulation was also distinct between angry and fearful expressions, which may be accounted for by the different modality dominance for the two emotions. Regarding the developmental studies, the components we indexed for observation are stably identifiable when contrasted with other components during infancy and childhood. Although the pattern of auditory responses in infants and young children was not the same as in adults, both developing groups showed the differences in the responses to modality as well as emotional congruency across the modalities at approximately 100 ms. It is also noted that lateralized processing for emotion-related effects were in both groups but in different hemispheres. This is likely because the process for audiovisual emotional perception occurs at an early sensory stage during infancy; however, the strategy for the neural processing of cross-modal emotional information could possibly become more rapid during childhood.

Reference

Ackles, P. K., & Cook, K. G. (1998). Stimulus probability and event-related potentials of the brain in 6-month-old human infants: a parametric study. *Int J Psychophysiol, 29*(2), 115-143.

Adams, R. B., Jr., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science, 300*(5625), 1536. doi: 10.1126/science.1082244

Adams, R. B., Jr., & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychol Sci, 14*(6), 644-647.

Adolphs, R., Jansari, A., & Tranel, D. (2001). Hemispheric perception of emotional valence from facial expressions. *Neuropsychology, 15*(4), 516-524.

Aguert, M., Laval, V., Le Bigot, L., & Bernicot, J. (2010). Understanding expressive speech acts: the role of prosody and situational context in French-speaking 5- to 9-year-olds. *J Speech Lang Hear Res, 53*(6), 1629-1641. doi: 10.1044/1092-4388(2010/08-0078)

Ahern, G. L., & Schwartz, G. E. (1985). Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. *Neuropsychologia, 23*(6), 745-755.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33*(6), 717-746.

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

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

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Dev Psychol, 36*(2), 190-201.

Bahrick, L. E., Lickliter, R., Castellanos, I., & Todd, J. T. (2015). Intrasensory Redundancy Facilitates Infant Detection of Tempo: Extending Predictions of the Intersensory Redundancy Hypothesis. *Infancy, 20*(4), 377-404. doi: 10.1111/infa.12081

Bair, W. N., Kiemel, T., Jeka, J. J., & Clark, J. E. (2007). Development of multisensory reweighting for posture control in children. *Exp Brain Res, 183*(4), 435-446. doi: 10.1007/s00221-007-1057-2

Balconi, M., & Mazza, G. (2009). Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues. ERS/ERD and coherence measures of alpha band. *Int J Psychophysiol, 74*(2), 158-165. doi: 10.1016/j.ijpsycho.2009.08.006

Balconi, M., & Vanutelli, M. E. (2016). Vocal and visual stimulation, congruence and lateralization affect brain oscillations in interspecies emotional positive and negative interactions. *Soc Neurosci, 11*(3), 297-310. doi: 10.1080/17470919.2015.1081400

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav Res Methods, 40*(2), 531-539.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci, 20*(8), 2225-2234. doi: 10.1111/j.1460-9568.2004.03670.x

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., . . . Whalen, J. R. (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology, 12*(3), 446-458.

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: high-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cereb Cortex, 23*(6), 1329-1341. doi: 10.1093/cercor/bhs109

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cereb Cortex, 21*(5), 1042-1055. doi: 10.1093/cercor/bhq170

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2009). Hearing Faces: How the Infant Brain Matches the Face It Sees with the Speech It Hears. *Journal of Cognitive Neuroscience, 21*(5), 905-921. doi: DOI 10.1162/jocn.2009.21076

Canli, T., Desmond, J. E., Zhao, Z., Glover, G., & Gabrieli, J. D. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. *Neuroreport, 9*(14), 3233-3239.

Ceponiene, R., Rinne, T., & Naatanen, R. (2002). Maturation of cortical sound processing as indexed by event-related potentials. *Clin Neurophysiol, 113*(6), 870-882.

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *Br J Dev Psychol, 33*(2), 218-236. doi: 10.1111/bjdp.12075

Coch, D , & Gullick, M. M. (2011). *The Oxford Handbook of Event-Related Potential Components* (E. Kappenman, S & S. J. Luck Eds.). U.S.A: Oxford University Press.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Res, 1242*, 126-135. doi: 10.1016/j.brainres.2008.04.023

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior, 28*(2), 117-139. doi: Doi 10.1023/B:Jonb.0000023655.25550.Be

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732-744. doi: DOI 10.1016/j.clinph.2003.11.021

Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development. In C. A. Nelson & M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience, Second Edition* (pp. 247-262). Cambridge: MIT Press.

Cunningham, J., Nicol, T., Zecker, S., & Kraus, N. (2000). Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. *Ear Hear, 21*(6), 554-568.

Davidson, R. J. (1992a). Anterior cerebral asymmetry and the nature of emotion. *Brain Cogn, 20*(1), 125-151.

Davidson, R. J. (1992b). Emotion and Affective Style - Hemispheric Substrates. *Psychological Science, 3*(1), 39-43. doi: DOI 10.1111/j.1467-9280.1992.tb00254.x

Davidson, R. J. (1995). Cerebral asymmetry, emotion and affective style. In R. J. Davidson & K. Hughdahl (Eds.), *Brain asymmetry* (pp. 361–387). Cambridge: MA: MIT Press.

Davidson, R. J., & Fox, N. A. (1982). Asymmetrical Brain Activity Discriminates between Positive and Negative Affective Stimuli in Human Infants. *Science, 218*(4578), 1235-1237. doi: DOI 10.1126/science.7146906

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience, 7*(3), 242-249. doi: 10.1038/nrn1872

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philos Trans R Soc Lond B Biol Sci, 364*(1535), 3475-3484. doi: 10.1098/rstb.2009.0190

de Gelder, B., de Borst, A. W., & Watson, R. (2015). The perception of emotion in body expressions. *Wiley Interdiscip Rev Cogn Sci, 6*(2), 149-158. doi: 10.1002/wcs.1335

de Gelder, B., & Van den Stock, J. (2011). The Bodily Expressive Action Stimulus Test (BEAST). Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body Expression of Emotions. *Front Psychol, 2,* 181. doi: 10.3389/fpsyg.2011.00181

de Gelder, B., Van den Stock, J., Meeren, H. K. M., Sinke, C. B. A., Kret, M. E., & Tamietto, M. (2010). Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience and Biobehavioral Reviews, 34*(4), 513-527. doi: DOI 10.1016/j.neubiorev.2009.10.008

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion, 14*(3), 289-311.

de Hann, M. (2007). Visual attention and recognition memory in infancy. In M. de Hann (Ed.), *Infant EEG and event-related potentials* (pp. 101-121). New York: Psychology Press.

de Hann, M., & Nelson, C. A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Dev Psychol, 35*(4), 1113-1121.

DeBoer, T., Scott, L. S., & Nelson, C. A. (2007). Methods for acquiring and analyzing infant event-related potentials. In M. de Hann (Ed.), *Infant EEG and Event-relatd potentials* (pp. 5-37). New York, U.S.A: Psychology Press.

Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn Sci, 3*(5), 172-178.

Demaree, H. A., Everhart, D. E., Youngstrom, E. A., & Harrison, D. W. (2005). Brain lateralization of emotional processing: historical roots and a future incorporating "dominance". *Behav Cogn Neurosci Rev, 4*(1), 3-20. doi:

10.1177/1534582305276837

Dunn, L., & Dunn, D. . (2009). *The British Picture Vocabulary Scale (3rd Ed.)*. UK: GL Assessment.

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

Focker, J., Gondan, M., & Roder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica, 137*(1), 36-47. doi: 10.1016/j.actpsy.2011.02.004

Fox, N. A., & Davision, R. J. (1987). Electroencephalogram Asymmetry in Response to the Approach of a Stranger and Maternal Separation in 10-Month-Old Infants. *Developmental Psychology, 23*(23), 233-240.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci, 8*(9), 396-403. doi: 10.1016/j.tics.2004.07.002

Ganesh, A. C., Berthommier, F., Vilain, C., Sato, M., & Schwartz, J. L. (2014). A possible neurophysiological correlate of audiovisual binding and unbinding in speech perception. *Front Psychol, 5*, 1340. doi: 10.3389/fpsyg.2014.01340

Garcia-Larrea, L., Lukaszewicz, A. C., & Mauguiere, F. (1992). Revisiting the oddball paradigm. Non-target vs neutral stimuli and the evaluation of ERP attentional effects. *Neuropsychologia, 30*(8), 723-741.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci, 11*(5), 473-490.

Gil, S., Hattouti, J., & Laval, V. (2016). How children use emotional prosody: Crossmodal emotional integration? *Developmental Psychology, 52*(7), 10644-11072.

Grezes, J., Pichon, S., & de Gelder, B. (2007). Perceiving fear in dynamic body expressions. *Neuroimage, 35*(2), 959-967. doi: DOI 10.1016/j.neuroimage.2006.11.030

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Dev Sci, 9*(3), 309-315. doi: 10.1111/j.1467-7687.2006.00494.x

Herba, C. M., Landau, S., Russell, T., Ecker, C., & Phillips, M. L. (2006). The development of emotion-processing in children: effects of age, emotion, and intensity. *J Child Psychol Psychiatry, 47*(11), 1098-1106. doi: 10.1111/j.1469-7610.2006.01652.x

Hirai, M., & Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Cognitive Brain Research, 22*(2), 301-304. doi: 10.1016/j.cogbrainres.2004.08.008

Ho, H. T., Schroger, E., & Kotz, S. A. (2014). Selective Attention Modulates Early Human Evoked Potentials during Emotional Face-Voice Processing. *J Cogn Neurosci*, 1-21. doi: 10.1162/jocn_a_00734

Hoehl, S., & Wahl, S. (2012). Recording Infant ERP Data for Cognitive Research. *Developmental Neuropsychology, 37*(3), 187-209. doi: 10.1080/87565641.2011.627958

Hyde, D. C., Jones, B. L., Flom, R., & Porter, C. L. (2011). Neural Signatures of Face-Voice Synchrony in 5-Month-Old Human Infants. *Developmental Psychobiology, 53*(4), 359-370. doi: 10.1002/dev.20525

Hyde, D. C., Jones, B. L., Porter, C. L., & Flom, R. (2010). Visual Stimulation Enhances Auditory Processing in 3-Month-Old Infants and Adults. *Developmental Psychobiology, 52*(2), 181-189. doi: 10.1002/dev.20417

Iacoboni, M. (2005). Neural mechanisms of imitation. *Curr Opin Neurobiol, 15*(6), 632-637. doi: 10.1016/j.conb.2005.10.010

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage, 58*(2), 665-674. doi: DOI 10.1016/j.neuroimage.2011.06.035

Jessen, S., Obleser, J., & Kotz, S. A. (2012). How Bodies and Voices Interact in Early Emotion Perception. *Plos One, 7*(4). doi: 10.1371/journal.pone.0036070

Kayser, J., Tenke, C. E., & Bruder, G. E. (1998). Dissociation of brain ERP topographies for tonal and phonetic oddball tasks. *Psychophysiology, 35*(5), 576-590.

Killgore, W. D., & Yurgelun-Todd, D. A. (2004). Activation of the amygdala and anterior cingulate during nonconscious processing of sad versus happy faces. *Neuroimage, 21*(4), 1215-1223. doi: 10.1016/j.neuroimage.2003.12.033

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *Neuroimage, 18*(1), 156-168.

Knowland, V. C., Mercure, E., Karmiloff-Smith, A., Dick, F., & Thomas, M. S. (2014). Audio-visual speech perception: a developmental ERP investigation. *Dev Sci, 17*(1), 110-124. doi: 10.1111/desc.12098

Kokinous, J., Kotz, S. A., Tavano, A., & Schroger, E. (2014). The role of emotion in dynamic audiovisual integration of faces and voices. *Soc Cogn Affect Neurosci*. doi: 10.1093/scan/nsu105

Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants. *Developmental Cognitive Neuroscience, 9,* 56-67.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang, 86*(3), 366-376.

Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. *Prog Brain Res, 156,* 285-294. doi: 10.1016/S0079-6123(06)56015-7

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage, 37*(4), 1445-1456. doi: 10.1016/j.neuroimage.2007.06.020

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., & Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia, 47*(14), 3059-3066. doi: 10.1016/j.neuropsychologia.2009.07.001

Kret, M. E., Pichon, S., Grezes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage, 54*(2), 1755-1762. doi: 10.1016/j.neuroimage.2010.08.012

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog Neurobiol, 72*(5), 341-372. doi: 10.1016/j.pneurobio.2004.03.006

Kujawa, A., Weinberg, A., Hajcak, G., & Klein, D. N. (2013). Differentiating event-related potential components sensitive to emotion in middle childhood: evidence from temporal-spatial PCA. *Dev Psychobiol, 55*(5), 539-550. doi: 10.1002/dev.21058

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Naatanen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport, 13*(1), 47-51. doi: Doi

10.1097/00001756-200201210-00014

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proc Natl Acad Sci U S A, 105*(32), 11442-11445. doi: 10.1073/pnas.0804275105

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu Rev Neurosci, 23*, 155-184. doi: 10.1146/annurev.neuro.23.1.155

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Dev Psychobiol, 56*(2), 292-315. doi: 10.1002/dev.21197

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proc Natl Acad Sci U S A, 103*(17), 6771-6774. doi: 10.1073/pnas.0602027103

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences, 13*(11), 470-478. doi: 10.1016/j.tics.2009.08.004

Lewkowicz, D. J., Sowinski, R., & Place, S. (2008). The decline of cross-species intersensory perception in human infants: underlying mechanisms and its developmental persistence. *Brain Res, 1242*, 291-302. doi: 10.1016/j.brainres.2008.03.084

Marshall, P. J., & Shipley, T. F. (2009). Event-Related Potentials to Point-Light Displays of Human Actions in 5-month-old Infants. *Developmental Neuropsychology, 34*(3), 368-377. doi: Pii 910997920 10.1080/87565640902801866

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychon Bull Rev, 3*(2), 215-221. doi: 10.3758/BF03212421

Mauss, I. B., Bunge, S. A., & Gross, J. J. (2007). Automatic Emotion Regulation. *Personality and Social Psychology Review,, 8*, 220-247.

Mauss, I. B., Cook, C. L., & Gross, J. J. (2007). Automatic emotion regulation during anger provocation. *Journal of Experimental Social Psychology, 43*(5), 698-711. doi: DOI 10.1016/j.jesp.2006.07.003

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Missana, M., Altvater-Mackensen, N., & Grossmann, T. (2017). Neural correlates of infants' sensitivity to vocal expressions of peers. *Dev Cogn Neurosci, 26*, 39-44. doi: 10.1016/j.dcn.2017.04.003

Missana, M., Atkinson, A. P., & Grossmann, T. (2015). Tuning the developing brain to emotional body expressions. *Developmental Science, 18*(2), 243-253. doi: 10.1111/desc.12209

Missana, M., & Grossmann, T. (2015). Infants' Emerging Sensitivity to Emotional Body Expressions: Insights From Asymmetrical Frontal Brain Activity. *Developmental Psychology, 51*(2), 151-160. doi: 10.1037/a0038469

Missana, M., Rajhans, P., Atkinson, A. P., & Grossmann, T. (2014). Discrimination of fearful and happy body postures in 8-month-old infants: an event-related potential study. *Front Hum Neurosci, 8*, 531. doi: 10.3389/fnhum.2014.00531

Molfese, D. L., Molfese, V. J., & Kelly, S. (2001). The use of brain electrophysiology techniques to study language: A basic guide for the beginning consumer of electrophysiology information. *Learning Disability Quarterly, 24*(3), 177-188. doi: Doi 10.2307/1511242

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res, 14*(1), 115-128.

Montirosso, R., Peverelli, M., Frigerio, E., Crespi, M., & Borgatti, R. (2010). The Development of Dynamic Facial Expression Recognition at Different Intensities in 4-to 18-Year-Olds. *Social Development, 19*(1), 71-92. doi: 10.1111/j.1467-9507.2008.00527.x

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory Processes: A Balancing Act across the Lifespan. *Trends Neurosci, 39*(8), 567-579. doi: 10.1016/j.tins.2016.05.003

Naatanen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound - a Review and an Analysis of the Component Structure. *Psychophysiology, 24*(4), 375-425. doi: DOI 10.1111/j.1469-8986.1987.tb00311.x

Naatanen, R., & Picton, T. W. (1986). N2 and automatic versus controlled processes. *Electroencephalogr Clin Neurophysiol Suppl, 38*, 169-186.

Naatanen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and Location Specificity of the Human Vertex N1-Wave. *Electroencephalography and Clinical Neurophysiology, 69*(6), 523-531. doi: Doi 10.1016/0013-4694(88)90164-2

Nelson, C. A., & de Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology, 29*(7), 577-595.

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology, 110*(1), 52-61. doi: DOI 10.1016/j.jecp.2011.03.014

Nielsenbohlman, L., Knight, R. T., Woods, D. L., & Woodward, K. (1991). Differential Auditory Processing Continues during Sleep. *Electroencephalography and Clinical Neurophysiology, 79*(4), 281-290. doi: Doi 10.1016/0013-4694(91)90124-M

Otte, R. A., Donkers, F. C., Braeken, M. A., & Van den Bergh, B. R. (2015). Multimodal processing of emotional information in 9-month-old infants I: emotional faces and voices. *Brain Cogn, 95*, 99-106. doi: 10.1016/j.bandc.2014.09.007

Paulmann, S., Jessen, S., & Kotz, S. A. (2009). Investigating the Multimodal Nature of Human Communication Insights from ERPs. *Journal of Psychophysiology, 23*(2), 63-76. doi: Doi 10.1027/0269-8803.23.2.63

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion, 35*, 192-201.

Pichon, S., de Gelder, B., & Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Social Neuroscience, 3*(3-4), 199-212. doi: Doi 10.1080/17470910701394368

Pichon, S., de Gelder, B., & Grezes, J. (2009). Two different faces of threat. Comparing the neural systems for recognizing fear and anger in dynamic body expressions. *Neuroimage, 47*(4), 1873-1883. doi: 10.1016/j.neuroimage.2009.03.084

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastian-Galles, N. (2009). Narrowing of intersensory speech perception in infancy. *Proc Natl Acad Sci U S A, 106*(26), 10598-10602. doi: 10.1073/pnas.0904134106

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clinical Neurophysiology, 111*(2), 220-236. doi: Doi 10.1016/S1388-2457(99)00236-9

Pourtois, G., Debatisse, D., Despland, P. A., & de Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Brain Res Cogn Brain Res, 14*(1), 99-105.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci, 18*(6), 2188-2199.

Purhonen, M., Kilpelainen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *International Journal of Psychophysiology, 52*(3), 257-266. doi: 10.1016/j.ijpsycho.2003.11.003

Reid, V. M., Hoehl, S., & Striano, T. (2006). The perception of biological motion by infants: An event-related potential study. *Neuroscience Letters, 395*(3), 211-214. doi: 10.1016/j.neulet.2005.10.080

Reynolds, G. D., Bahrick, L. E., Lickliter, R., & Guy, M. W. (2014). Neural correlates of intersensory processing in 5-month-old infants. *Developmental Psychobiology, 56*(3), 355-372. doi: 10.1002/dev.21104

Rivera-Gaxiola, M., Silva-Pereyra, J., Klarman, L., Garcia-Sierra, A., Lara-Ayala, L., Cadena-Salazar, C., & Kuhl, P. (2007). Principal Component Analyses and scalp distribution of the auditory P150-250 and N250-550 to speech contrasts in Mexican and American infants. *Developmental Neuropsychology, 31*(3), 363-378.

Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and Cognition, 69*(2), 269-278. doi: 10.1016/j.bandc.2008.08.007

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009a). Critical features for the perception of emotion from gait. *J Vis, 9*(6), 15 11-32. doi: 10.1167/9.6.15

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009b). Critical features for the perception of emotion from gait. *J Vis, 9*(6), 1-32. doi: 10.1167/9.6.15

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire (SCQ)*. U.S.A.: Western Psychological Services.

Sauter, D. A., Panattoni, C., & Happe, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31*(1), 97-113. doi: 10.1111/j.2044-835X.2012.02081.x

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *J Cogn Neurosci, 15*(8), 1135-1148. doi: 10.1162/089892903322598102

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Brain Res Cogn Brain Res, 14*(2), 228-233.

Shafer, V. L., Yu, Y. H., & Wagner, M. (2015). Maturation of cortical auditory evoked potentials (CAEPs) to speech recorded from frontocentral and temporal sites: Three months to eight years of age. *International Journal of Psychophysiology, 95*(2), 77-93. doi: 10.1016/j.ijpsycho.2014.08.1390

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions - What you see is what you hear. *Nature, 408*(6814), 788-788. doi: Doi 10.1038/35048669

Soken, N. H., & Pick, A. D. (1992). Intermodal Perception of Happy and Angry Expressive Behaviors by 7-Month-Old Infants. *Child Development, 63*(4), 787-795.

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology, 13*(13), R519-R521. doi: 10.1016/S0960-9822(03)00445-7

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. *Neuroreport, 15*(5), 777-780.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci, 19*(12), 1964-1973. doi: 10.1162/jocn.2007.19.12.1964

Steven, J. L. (2005). An Introduction to the Event-Related Potential and Their Neural Origins *An Introduction to the Event-Related Potential Technique* (pp. 1-50). London, England: The MIT Press.

Striano, T., Brennan, P. A., & Vanman, E. J. (2002). Maternal Depressive Symptoms and 6-Month-Old Infants' Sensitivity to Facial Expressions. *Infancy, 3*(1), 115-126. doi: 10.1207/S15327078in0301_6

Sussman, E., Stemschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research, 236*(1-2), 61-79. doi: 10.1016/j.heares.2007.12.001

Symons, A. E., El-Deredy, W., Schwartze, M., & Kotz, S. A. (2016). The Functional Role of Neural Oscillations in Non-Verbal Emotional Communication. *Front Hum Neurosci, 10*, 239. doi: 10.3389/fnhum.2016.00239

Takagi, S., Hiramatsu, S., Tabei, K., & Tanaka, A. (2015). Multisensory perception of the six basic emotions is modulated by attentional instruction and unattended modality. *Front Integr Neurosci, 9,* 1. doi: 10.3389/fnint.2015.00001

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex, 17*(3), 679-690. doi: 10.1093/cercor/bhk016

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*(9), 400-410. doi: 10.1016/j.tics.2010.06.008

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J Cogn Neurosci, 17*(7), 1098-1114. doi: 10.1162/0898929054475172

Trainor, L. J. (2007). Event-Related Potential (ERP) Measures in Auditory Development Research. In L. A. Schmidt. & S. J. Segalowitz (Eds.), *Developmental Psychophysiology: Theory, Systems, and Methods* (pp. 69-102). New York: Cambridge University Press.

van de Riet, W. A., Grezes, J., & de Gelder, B. (2009). Specific and common brain regions involved in the perception of faces and bodies and the representation of their emotional expressions. *Soc Neurosci, 4*(2), 101-120. doi: 10.1080/17470910701865367

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7*(3), 487-494. doi: 10.1037/1528-3542.7.3.487

van Heijnsbergen, C. C., Meeren, H. K., Grezes, J., & de Gelder, B. (2007). Rapid detection of fear in body expressions, an ERP study. *Brain Res, 1186,* 233-241. doi: 10.1016/j.brainres.2007.09.093

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A, 102*(4), 1181-1186. doi: 10.1073/pnas.0408949102

Volkova, E. P., Mohler, B. J., Dodds, T. J., Tesch, J., & Bulthoff, H. H. (2014). Emotion categorization of body expressions in narrative scenarios. *Front Psychol, 5,* 623. doi: 10.3389/fpsyg.2014.00623

Walker-Andrews, A. S. (1986). Intermodal Perception of Expressive Behaviors - Relation of Eye and Voice. *Developmental Psychology, 22*(3), 373-377. doi: Doi 10.1037/0012-1649.22.3.373

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: differentiation of multimodal information. *Psychol Bull, 121*(3), 437-456.

Walker-Andrews, A. S. (2008). Intermodal emotional processes in infancy. In J. M. Lewis, Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3 ed., pp. 364-375). New York: Guilford Press.

Walker-andrews, A. S., & Lennon, E. (1991). Infants Discrimination of Vocal Expressions - Contributions of Auditory and Visual Information. *Infant Behavior & Development, 14*(2), 131-142. doi: Doi 10.1016/0163-6383(91)90001-9

Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence (3rd ed.)*. London, UK: The Psychological Cooperation

Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Percept Psychophys, 39*(4), 294-300.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol Bull, 88*(3), 638-667.

Widen, S. C., & Russell, J. A. (2008). Young children's understanding of others' emotions. In A. Bremner, D. J. Lewkowicz & C. Spence (Eds.), *Multisensory development* (pp. 348-363). Oxford, England.: Oxford University Press.

Wunderlich, J. L., & Cone-Wesson, B. K. (2006). Maturation of CAEP in infants and children: A review. *Hearing Research, 212*(1-2), 212-223. doi: 10.1016/j.heares.2005.11.008

Yeh, P. W., Geangu, E., & Reid, V. (2016). Coherent emotional perception from body expressions and the voice. *Neuropsychologia, 91*, 99-108. doi: 10.1016/j.neuropsychologia.2016.07.038

Zieber, N., Kangas, A., Hock, A., & Bhatt, R. S. (2014a). The development of intermodal emotion perception from bodies and voices. *Journal of Experimental Child Psychology, 126*, 68-79. doi: 10.1016/j.jecp.2014.03.005

Zieber, N., Kangas, A., Hock, A., & Bhatt, R. S. (2014b). Infants' perception of emotion from body movements. *Child Dev, 85*(2), 675-684. doi: 10.1111/cdev.12134