PhD thesis

Culture and cultural immersion modulate the brain response to human expressions of emotion: electrophysiological evidence

Pan Liu

School of Communication Sciences and Disorders

Faculty of Medicine

McGill University, Montreal

November, 2014

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Acknowledgements

First and foremost, I would like to express my utmost gratitude to my supervisor, Dr. Marc Pell, who has been supporting me through my PhD program with his exceptional supervision and guidance, unfailing patience and encouragement, immense knowledge and inspirational insights. Has his guidance not only helped me during the entire time of conducting my PhD research and writing up this thesis, but also edified my learning and growing as researcher. Words fail me to express how much I am indebted to him. I would also like to acknowledge the financial support for my PhD research and stipend from Social Sciences and Humanities Research Council of Canada operating grants awarded to Dr. Marc Pell.

I am also deeply grateful to the rest of my thesis committee, Drs. Shari Baum and Simon Rigoulot. I would like to thank Dr. Shari Baum for her always prompt responses, constant encouragement, sharp questions, enriching suggestions, and meticulous editorial advice on my English writing. I am also thankful to her for kindly agreeing to act as the internal examiner for my thesis evaluation. A special thank you goes to Dr. Simon Rigoulot, for all the lengthy yet inspiring discussions he had with me since the very first draft of my thesis proposal, through the entire process of EEG testing and data analyses, until the last stages of thesis writing and revision; and, equally importantly, for his unique sense of humour which made me laugh out loud when I was having a hard time.

My gratitude also goes to Dr. Keiko Ishii, who kindly consented to act as the external examiner for this thesis regardless of a very tight working schedule of her own. I sincerely appreciate her efforts and expertise to be spent on this time-consuming work and I am looking forward to reading her comments which undoubtedly will of great value to improve my research work.

While being unable to exhaust the list of people without whom this thesis would not be complete on time, I want to express my thankfulness to all my participants for being tortured by the EEG testing and offering me good data. I also owe my deepest appreciation to my dear colleagues and friends of Pell Lab, both the former and the present ones, for providing an environment of learning that is both stimulating and heartening, for the immeasurable academic assistance and spiritual support they generously offered all the time, also for the countless dinners that we went to together and the joy and friendship we have shared and enjoyed over the past years.

In addition to my fellow Pell Labbers, I would also like to thank my other Beatty friends, Myrto Brandeker, Kristina Kasparian, Aruna Sudarshan, and Katja Haeuser; thank you for all the therapeutic coffee/dinner time we spent together chatting, gossiping, and laughing, which has become an integral part of my life since long ago.

When I was still inadvertently using the term “Beatty friends”, our department had already recently moved out of Beatty Hall, after decades—for me, five years—of residence there. So, I would also like to thank Beatty Hall, for providing a home-like accommodation with a special charm for the most part of my PhD life, for the five-year-long daily mountain climbing from McGill Ghetto to my office every morning. To all of us, the Beatty-related memories will never be replaceable.

Last but definitely not the least, I would like to thank my beloved parents. In spite of the geographical distance their love has to travel from China to reach me, it never fails to offer me the most prompt, solid, and perpetual support in all my pursuits.

Statement of Originality

The two studies in this thesis present a set of novel investigations to better understand the effects of different cultural experiences on the neural (and behavioural) processing of emotional expressions across faces and voices. In Study 1, the role of *cultural origin* during this process was explored by comparing two cultural groups, Chinese and English North Americans, with ERP measurements. A second aim of this thesis was to examine the effect of a different cultural experience, cultural immersion, on multisensory emotion processing, which was addressed in Study 2 by testing a third group, Chinese immigrants to Canada. The two studies have been written in manuscript format for publication elsewhere. At the time of official submission of this thesis to McGill University, Study 1 has been submitted to *Neuropsychologia* for review. Parts of the results of Study 1 have been presented at the *Plenary Meeting of International Society for Research on Emotion,* in August, 2013. Parts of the results of Study 1 and Study 2 have also been presented at the *International Conference on Multilingualism*, in October, 2013, and *Annual Meeting of Society for Neuroscience*, in November, 2013. Study 2 will be submitted for publication before the oral examination.

Contribution of Authors

The current thesis is a manuscript-based one comprised of two studies. Both studies were conceptualized, designed and conducted by the first author, Pan Liu, under the guidance of her supervisor, Dr. Marc Pell (the third author in both manuscripts), and her thesis committee member, Dr. Simon Rigoulot (the second author in both manuscripts). The results were primarily analyzed by the first author, with advice from the second and third authors. Both manuscripts were written in full by the first author, with revision advice and editorial suggestions from the second and third authors.

Table of contents

Abstract…………………………………………………………………………………1

Résumé…………………………………………………………………………………….3

1. **Introduction**…………………………………………………………………………..5

1.1 Emotion……………………………………………………………………….....7

1.1.1. Emotion theories…….…………………………………………………...10

1.1.2. Emotion perception……………………………………………………....18

1.2. Cultural effects on emotion perception………..……………………………….......30

1.2.1. Theories on cross-cultural differences in emotion communication……….......32

1.2.2. Cross-cultural differences in emotion perception………………………......37

1.2.3. Effects of cultural immersion on emotion perception………………………...44

1.3. The present study……………………………………………………………........49

2. **Study 1**……………………………………………………………………………......53

Culture modulates the brain response to human expressions of emotion: electrophysiological evidence

3. **Study 2**……………………………………………………………………………….114

Cultural immersion alters emotion perception: electrophysiological evidence from Chinese immigrants in North America

4. **General discussion**……………………………………………………………………..145

5. **References**………………………………………………………………………......163

Abstract

Communicating feelings and emotions with each other using multiple cues (e.g., facial expressions, prosodic cues, body postures, etc.) is a common and essential part of our daily life and vital to human's survival as social beings. Meanwhile, as the phenomenon of cultural exchanges is multiplying around the world, it is not unusual that emotional communication occurs between people from different cultural backgrounds, where misunderstandings are likely to be encountered due to culturally-varied rules and routines for social communication. A comprehensive investigation of the culturally-specified biases in communicating multi-sensory emotions, which might potentially help eliminate misunderstandings and promote relations between different cultural groups, is still ongoing. To understand the impact of different types of cultural experiences on both off-line and on-line responses to emotional information, this thesis examined emotion processing in three cultural groups by measuring behavior and event-related brain potentials (ERPs). Study 1 compared individuals from two cultures, English-speaking North Americans and Chinese, in how they process multimodal emotional cues. Results of an emotional Stroop task and an Oddball task demonstrated marked group differences at distinct processing stages as reflected in behavioural accuracy, N400, and visual MMN ERP components, with English participants showing higher sensitivity to faces than Chinese, whereas Chinese were influenced to a larger extent by vocal cues. Study 2 further explored how the experiences of cultural immersion and adaptation modulate emotion processing by testing a third group--Chinese immigrants to Canada--in identical tasks to those of Study 1. In comparison with the data of the Chinese and English groups from the first study, the immigrants showed a similar behavioural response pattern to the English group in processing multi-sensory emotions, but analogous patterns in N400 and vMMN to the Chinese group. Interestingly, correlation analyses indicated that longer stays of the immigrants in Canada were associated with ERP patterns resembling the English group. This thesis provides significant evidence on the modulating effects of culture and cultural immersion on human's behavioural and neural responses to emotional expressions at multiple processing stages, and advances our understanding of human's neuro-cognitive plasticity in socio-emotional communication.

Résumé

Communiquer aux autres nos sentiments et nos émotions au travers de supports multiples (tels que nos expressions faciales, notre ton de voix, nos attitudes corporelles ou autres) est une activité courante mais essentielle de notre vie quotidienne, vitale pour notre survie en tant qu'êtres sociaux. Aussi, de par la multiplication des échanges culturels dans le monde, il est facile d'imaginer que de tels échanges puissent se faire entre deux personnes de culture différente, pouvant entrainer des incompréhensions liées à des règles de communication sociale ne correspondant pas. Une étude approfondie des biais culturels dans la communication multimodale des émotions est donc nécessaire. Dans cette thèse, deux études ont ainsi été mises en place pour comprendre l'impact des origines culturelles sur le traitement des informations émotionnelles, au niveau comportemental et cérébral (potentiels évoqués). Dans la première étude, des individus anglophones issus d'Amérique du Nord ont été comparés à des participants chinois dans leur traitement d'informations émotionnelles multimodales (audiovisuelles). En utilisant une tâche de Stroop émotionnel et une tâche de déviance, cette étude a permis d'observer que les participants anglophones étaient plus sensibles aux informations émotionnelles visuelles (visages) alors que les participants chinois étaient largement plus influencés par le contenu émotionnel des informations auditives (voix), suggérant des différences

marquées dans le traitement émotionnel, aussi bien au niveau comportemental qu'au niveau cérébral (N400, MMN), selon l'origine des participants. Le but de la seconde étude, était d'explorer comment l'immersion culturelle et l'adaptation à une nouvelle culture peuvent moduler le traitement des informations émotionnelles, en utilisant les tâches décrites précédemment sur un troisième groupe de participants, des Chinois immigrés récemment au Canada. Ce troisième groupe montre un profil de réponse comportementale similaire à celui du groupe anglophone, mais une activité cérébrale (N400, MMN) analogue à celle du groupe de Chinois. De manière intéressante, des analyses de corrélation mettent aussi en évidence que plus la durée du séjour au Canada est longue, plus le profil de réponses cérébrales des participants immigrés se rapprochent de celui observé chez les participants anglophones. Cette thèse apporte donc des éléments nouveaux dans la compréhension de l'influence de l'origine culturelle et de l'immersion culturelle des individus sur leurs réponses comportementales et cérébrales aux informations émotionnelles et permet donc une meilleure appréhension des effets de plasticité neuro-cognitive humaine dans un cadre de communication sociale.

1. Introduction

Emotion, which is one of the most essential aspects of human nature, has been a topic researched by philosophers and scientists for centuries, dating back to the Ancient Greek period. Theories on emotions were conceived as responses to certain events of concern to an individual, triggering physiological changes and subjective feelings, and motivating corresponding changes in behaviour, including characteristic displays in various sensory channels (Ruberg, 2009). Communicating such responses through sensory channels, i.e., notifying other people how we feel towards certain events and meanwhile, understanding how others feel about it, is of vital importance for human's survival as social beings. For instance, we convey fearful feelings to other people to signal to them the potential danger in the immediate environment; or we successfully interpret the sadness of a friend who is seeking social support and help.

However, as we are living in a rapidly globalizing world nowadays, it is not unusual that this friend is from a different cultural background and follows distinct culturally-learnt rules to express sadness, as well as other emotions. Moreover, emotion is a multidimensional process and can be displayed via multiple sensory channels, such as facial, vocal, lexical, gestural, and postural means (de Gelder, 2006; Grandjean, Banziger, & Scherer, 2006; Kipp & Martin, 2009; Wilson & Wharton, 2006). While each of these sources plays a role in communicating emotional information, the two most commonly involved sources are facial expressions and emotional prosody; e.g., a terrified face is almost always accompanied by a trembling voice in different cultures. Thus, correctly interpreting such multi-sensory emotional displays of other people, especially those who are from a different cultural group, is vital to maintain social rapport and avoid unnecessary misunderstandings between individuals from different cultural backgrounds. However, in spite of the importance of efficient cross-cultural communication, culture-specific cognitive biases and their underlying neural mechanisms in communicating multi-sensory emotions are not fully understood. Therefore, the present work focuses on how various cultural experiences, i.e., cultural origin and cultural immersion, modulate human's neuro-cognitive responses to multi-sensory emotional expressions.

This study aims to contribute to the understanding of how culture shapes multi-sensory emotion perception by examining different cultural groups in processing emotional face-voice dyads. Two major questions were addressed in this thesis. The first topic of concern is whether culture plays a role at multiple temporal processing stages of multi-sensory emotion processing. By using sensitive *event-related brain potentials* (ERPs) measures, the question examined was whether people from two cultural origins, Chinese and English-speaking North American, differed from each other at both early and late on-line processing stages of multi-sensory emotion perception, in addition to behavioural performance. Following the first study, the second question aims at further investigating whether a distinct type of cultural experience, cultural immersion in a foreign environment, also has an impact on the neuro-cognitive mechanisms of multi-sensory emotion perception. To this aim, a third cultural group, Chinese immigrants to Canada, were examined using identical ERP methods as in the first study and compared to the Chinese and North American groups. By exploring these questions, this thesis is one of the first studies that systematically investigates the temporal features of the process for processing and combining emotional cues in the face and voice, and more importantly, will provide novel evidence regarding the extent to which cultural origin and cultural immersion shape the way the neuro-cognitive system operates during multi-modal emotion perception.

1.1 Emotion

Due to its vital role in the human mind, emotion has been a prevailing topic in psychology and relevant research disciplines, where it has been defined as a conscious mental experience characterized by subjective states, physiological responses, behavioural expressions and reactions (Ruberg, 2009). Emotions are believed to play a fundamental role in human's evolution and adaptation to the physical, social and cultural environments and serve a wide variety of significant mental functions (Darwin, 1998; Ekman, 1992b; Izard, 1977, 1992; Plutchik, 1980).

They are considered as the driving force behind motivation, which promotes various thoughts and behaviours (Izard, 1977, 1992; Naqvi, Shiv, & Bechara, 2006). Emotions are also believed to be closely associated with mood, temperament, personality, and disposition (Stemmler & Wacker, 2010). Therefore, it is essential to understand the processes and mechanisms involved in emotion and a tremendous amount of research has been devoted to this topic, among which multiple theoretical frameworks have been proposed.

The conception of emotion is complicated and studies on emotion have been conducted from different perspectives and approaches. In the literature, Darwin's notion, which considers emotion as innate biologically-hardwired aspect of human nature, has been the most influential one in the modern psychology of emotion (Darwin, 1998). In observing the similarities in facial expressions between non-human primates and humans, Darwin argued that emotional expressions originally developed to perform adaptive biological functions to increase the chances of survival (e.g., by rejecting noxious contaminants); meanwhile, traits facilitating the communication of emotions among conspecifics would be passed on to the next generation by natural selection. This evolutionary perspective on emotion has influenced a variety of the following theories, for example, the discrete emotion theory, with the notion of *universality* as a working assumption in emotion research (Ekman & Friesen, 1969; Izard, 1977).

However, the universality of emotion has been questioned by increasing knowledge of cross-cultural differences in some human emotional behaviours which were originally assumed to be instinctual and biologically invariant. For example, anthropologists observed that Japanese (Hearn, 1894) and some African (Gorer, 1949) individuals tend to mask their negative emotions with smiles and laughter, while the Utku Eskimos tend to conceal their negative feelings by neutral faces (Briggs, 1970). In contrast, the Kiowa Tribe of Oklahoma encourages enthusiastic emotional expressions during specific events, even in the absence of an internal feeling (Labarre, 1947). With these observations of culture-specific emotional behaviours, the idea of the universality of emotions continues to be hotly debated, and different notions have been proposed that emotions are socially learned to a certain extent, instead of completely instinctual (Klineberg, 1940; Labarre, 1947; Mead, 1975).

Therefore, with the development of theoretical frameworks on emotion, a variety of theories have been proposed to depict the mechanisms and characteristics of emotion in the literature; however, there has not been a generally accepted consensus among researchers. For instance, from a Darwinian evolutionary perspective, one of the most classical theories has proposed that there is a set of innate basic emotions with adaptive functions experienced by all humans across cultural and linguistic boundaries (Ekman, 1992b). Another prevalent theory defines emotion by the common features shared by different emotions along certain affective dimensions, e.g., valence and arousal (Bradley & Lang, 2000; Russell, 2003). In addition, some models focus on the notion of emotional 'appraisal' mechanisms, according to which emotion is the result of a series of cognitive evaluations of the immediate situation and events (Scherer, 1986, 2003). Therefore, no agreement has been achieved on the theoretical account of emotion; rather different aspects of emotion have been the focus of different approaches. In the next section, an overview of the major psychological theories of emotion will be discussed, followed by a succinct review of existing research on emotion perception.

1.1.1 Emotion theories

1.1.1.1 Discrete emotion theory

In the emotion literature, a dominant and widely-cited theoretical account of emotion is the *discrete (basic) emotion* theory, which was first proposed by Tomkins (Tomkins, 1962, 1963) and further developed by Ekman (Ekman & Friesen, 1969) and Izard (Izard, 1977). This theory defines emotion in reference to a set of discrete emotion categories, e.g., happiness, sadness, and fear, each of which is considered as a unique experiential state that results from distinct causes. While the exact number of categories of basic emotions varies from six to eleven, at least six emotions are widely accepted and adopted in emotion research: happiness, sadness, anger, fear, disgust, and surprise (Ekman, 1992b).

In particular, this core set of discrete emotions are believed to be distinct from each other in the patterns of preceding incentives, appraisal, physiological responses, subjective experiences, behavioral expressions and reactions. Empirical evidence supporting such discrete patterning was first reported in the visual modality for facial expressions (Ekman, 1992b; Izard, 1994) and later in the auditory (vocal) modality (Bryant & Barrett, 2008; Scherer, Banse, & Wallbott, 2001). It is suggested that there is an independent neural system subserving each basic emotion (Ekman, 1992b). For example, a meta-analysis of functional magnetic resonance imaging (fMRI) studies demonstrated that each of the basic emotions (fear, anger, disgust, sadness, and happiness) was characterized by consistent and discriminable neural correlates across different studies, as defined by reliable correlations with regional brain activations (Vytal & Hamann, 2010). These prototypical patterns of the basic emotions are considered as biologically hard-wired, present from birth, and serving distinct adaptive functions to increase the chances of survival (Izard, 1992; Stein & Oatley, 1992). Therefore, it can be said that they share universal expressions and responses across cultural boundaries (Ekman, 1992a, 1992b; Izard, 1994). It is also assumed that all affective states can be derived from this core set of universal and innate basic emotions (Ekman, 1992a, 1992b).

While basic emotion theory has been considered as one of the most classical approaches to emotion in the literature, it has been criticized by some researchers who suggest that its prototypical patterns of basic emotions cannot generalize to more complex emotions, e.g., moral emotions and aesthetic emotions such as guilt and awe; the assumed universality of the innate patterns also fails to account for the differences across cultures and languages in emotion categorization (Scherer & Ellgring, 2007; Schorr, 2001). In addition, some of the reported neuro-imaging and physiological evidence in emotion recognition has been incompatible with the assumption that there exists an independent neural system subserving each of the discrete emotions (Barrett & Wager, 2006). In spite of these shortcomings, this approach has been widely employed in emotion studies in recent decades. The current study also follows the framework of basic emotion theory and examines emotion perception of two basic emotions, sadness and fear.

1.1.1.2 Dimensional theory of emotion

Instead of the prototypical patterns of categorical basic emotions, an alternative approach to define emotions focuses on a number of independent affective *dimensions*, which combine with each other to comprise specific emotions. This notion was first suggested by Wundt (1896), who proposed the distinction among three dimensions of emotions: pleasure-displeasure, arousal-calmness, and tension-relaxation (Wundt, 1896). In the recent literature, two dimensions have been most widely accepted and best understood: emotional *valence* which indicates the hedonic value of a specific emotion (pleasure–displeasure); and emotional *arousal* that refers to the intensity or strength of the emotional state, i.e., activation–deactivation (Bradley & Lang, 2000; Russell, 2003). Both valence and arousal can be defined as subjective experiences (Russell, 1980); meanwhile, they are thought to represent two independent neurophysiological systems and act as a motivational guide to set the priority for a response (Cabanac, 1996, 2002). According to the dimensional theory, all affective states are considered as arising from these two independent systems; in other words, different emotions can be understood as combinations of varying degrees of valence and arousal (Posner, Russell, & Peterson, 2005; Russell, 1980). Both valence and intensity can be measured in the form of parametric variables (Cabanac, 2002) and evidence has also been reported that they exist across cultural boundaries (Russell, 1994) and are present early in life (Russell & Bullock, 1985). In addition to the two dimensions, other researchers suggest a third dimension, *dominance*, which indicates the feeling of being in control versus being controlled (Bradley & Lang, 2000); or an *approach-withdrawal* dimension indicating the tendency to approach or escape from the event that triggers the emotion (Davidson, 1995).

Despite the prevalence of the dimensional theory in the emotion literature, there has been some evidence reported on the model's incompatibility with emotion data (Schimmack & Bockenholt, 2002; Schimmack & Grob, 2000). For instance, this theory has been criticized for its lack of differentiation between emotions that are close neighbours in the valence-arousal space, such as anger and fear which are supposed to be similar according to the dimensional model but are actually very distinct from each other in many aspects, including subjective feelings and physiological-behavioural responses (Tellegen, Watson, & Clark, 1999). It has also been reported that the two-dimensional model is not able to account for all the varieties in some domain-specific emotions, e.g., music-elicited emotions (Bigand, Vieillard, Madurell, Marozeau, & Dacquet, 2005; Collier, 2007; Ilie & Thompson, 2006; Vieillard et al., 2008).

1.1.1.3 Appraisal theory of emotion

Distinct from the discrete and dimensional models of emotion, the appraisal approach has conceptualized emotion from a new perspective, considering emotion as a dynamic process of the individual's evaluation of the event that triggers the emotion. Specifically, the term 'appraisal' refers to the process of evaluating how the event, object, or person affects us and how to cope with the situation. For example, an event is assessed to determine whether it has a positive or a negative consequence, following which an emotional response would be triggered to cope with the situation. Strong claims have been made that appraisal processes are mandatory for identification of emotion and that the emotion category is determined based on the appraisal of the significant event (Averill, 1980; Frijda, 1987; Scherer, 2001).

Within the framework of appraisal theory, Scherer and colleagues (1986, 2001, 2003) have proposed a *component process model* that describes the nature and features of emotion, with specific reference to the processing of vocal emotional stimuli. According to this model, a significant event and its consequences are appraised with a set of *stimulus evaluation checks*, i.e., evaluation criteria, which relate to the individual's needs and objectives. The main components of the *stimulus evaluation checks* are relevance detection, implication assessment, an evaluation of coping potential, and normative significance. Relevance detection refers to the appraisal of the features of the event, e.g., pleasantness and importance to the individual. For example, an event that is threatening to one's survival would lead to the detection of unpleasantness and the significance of the event. Next, the implications and consequences of the event will be assessed in terms of the cause and the outcome probability of the event. The situation triggered by the event is also evaluated to see how urgently a responsive action is required. Meanwhile, the individual's ability to cope with the situation is evaluated to determine whether the person has the power to change the event or adjust to the situation. Finally, an individual evaluates the situation according to social norms to decide whether it is socially acceptable and how it will be interpreted by other group members (Scherer, 1986, 2001, 2003).

This appraisal process will have an impact on the individual's motivational state; this in turn affects the autonomic nervous system yielding emotional physiological responses (e.g., cardiovascular and respiratory changes) and activation of the somatic nervous system, resulting in different forms of emotional expressions in multiple sensory channels (e.g., face, voice, and body postures, etc.). All different component processes are centrally integrated and represented with continuous updating as the appraisal is ongoing. It is also argued that most of these appraisal processes are automatic and independent of consciousness, with the exception that some deeper analyses of emotional expressions may require conscious processing (Scherer, 1986, 2003). While this model has been appealing to many researchers, concerns have also been raised, for example, it was argued that the component processes at different functional levels are more general processes of cognition and not specific to emotion processing (Russell, 2009).

Comparing the different theories reviewed above, whereas the appraisal model of emotion emphasizes the dynamic process of cognitive evaluation of the event that induces an emotion, the discrete and dimensional theories of emotion are more focused on the structural features of specific emotions. However, despite the fundamental differences among these theories, one common aspect shared by all theories is that each emotional state is associated with specific expressive patterns in multiple sensory modalities, including facial and vocal expressions, gestures, and postures. Consequently, it has been assumed that emotions can be communicated by these expressive patterns from different modalities between humans (Ekman, 1992b; Izard, 1992; Scherer, 1986). For example, the appearance of a snake can trigger a specific emotion of fear, for which a set of facial and vocal expressions, physiological responses and voluntary actions that are uniquely characteristic of *fear* will be evoked, either directly or via cognitive appraisal. Communicating these expressions to other people to signal the danger of the snake, therefore, is crucial to the survival and welfare of conspecifics.

In the literature, the perception of emotional expressions has been extensively investigated in single modalities, especially for facial expressions (Keltner, Ekman, Gonzaga, & Beer, 2003) but also for speech prosody (Scherer, 2003), emotional vocalizations (Frühholz & Grandjean, 2013; Scott, Sauter, & McGettigan, 2010), hand gestures (Kipp & Martin, 2009), and body postures (de Gelder, 2006). Curiously, the understanding of multi-sensory emotional signals, which are more common and natural in our daily life, is relatively limited. Therefore, the current thesis will focus on two major sensory channels in emotional processing: facial expression and speech prosody. In the following sections, key studies on emotion perception from facial expressions and speech prosody will be briefly reviewed and discussed.

1.1.2 Emotion perception

The most common and natural way of communicating emotions occurs in face-to-face interactions between two or more people, during which information from multiple sensory sources is involved. The two major sources are visual and auditory modalities: visual information encompasses facial expressions, hand gestures, and body postures; auditory information includes both linguistic information and non-linguistic vocal cues. In this thesis, two non-linguistic channels of sensory information, facial expressions and prosodic cues, will be investigated in emotional communication. While language is the most essential tool of human communication, from an evolutionary point of view, however, non-linguistic communication such as facial expressions and prosody are relatively more ancient and basic (Hauser, 2004). For instance, non-human animals, and pre-linguistic ancient humans, used facial and body movements as well as vocalizations to communicate purposefully to conspecifics; human infants also use various non-verbal cues to communicate their feelings and needs before speech is developed. As noted earlier, the significance of non-linguistic cues is especially germane in communicating emotional messages, and understanding how such messages are processed has also been an important research question in psychology and cognitive sciences.

1.1.2.1 *Uni-sensory emotion perception from facial expressions or emotional prosody*

In the literature of emotion perception/recognition, the most extensive work has been conducted on the facial channel, i.e., facial expression recognition. A facial expression consists of one or more motions or positions of the muscles beneath the skin of the face, which serves the function of conveying various emotional states of an individual. As a primary form of nonverbal communication, facial expressions play a significant role in conveying social/emotional information between humans, but they also occur in most other mammals and some other animal species (Waller & Micheletta, 2013). For human participants, emotions can be readily identified from facial expressions. This ability develops early in infancy, e.g., some basic emotions can be discriminated by 7-month old infants (Saarni, Mumme, & Campos, 1998; Soken & Pick, 1992) , and persists across the life span (Moreno, Borod, Welkowitz, & Alpert, 1993), although evidence of an aging effect on facial expression recognition, i.e., a decrease in recognition accuracy has also been reported (Adolphs, Damasio, Tranel, & Damasio, 1996). Interestingly, recognition of different emotions from facial expressions is associated with different performance accuracies; typically, it has been reported that happy faces show a recognition advantage when compared to many basic emotions (Leppänen & Hietanen, 2004; Leppänen, Tenhunen, & Hietanen, 2003), probably due to their high salience and distinctiveness of the “smile” when compared to other facial expressions.

In addition to behavioural studies on facial expression perception, neural evidence has been reported demonstrating the brain mechanisms underlying this process; these data come from related electrophysiological (Eimer & Holmes, 2007), neuroimaging (Alves, 2013; Phan, Wager, Taylor, & Liberzon, 2002; Posamentier & Abdi, 2003) and neuropsychological studies (Balconi, 2012). In particular, event-related potential (ERP) studies have investigated the time course of facial expression processing, in which face-sensitive P1, N170, and vertex positive potential (VPP) components have been examined. The P1 component is thought to correlate with early rapid categorization of emotional information that occurs before more in-depth, fine-grained perceptual processing (Linkenkaer-Hansen et al., 1998; Liu, Higuchi, Marantz, & Kanwisher, 2000; Luo, Feng, He, Wang, & Luo, 2010; Pizzagalli et al., 2002) and effects of facial emotions on P1 have been reported, e.g., larger amplitude was elicited by fearful faces than by neutral faces (Batty & Taylor 2003; Halgren , Raij , Marinkovic , Jousmäki , & Hari 2000). The N170 is considered an index of the structural processing of faces (Bentin , Allison , Puce , Perez , & McCarthy 1996; Bötzel , Schulze , & Stodieck 1995; Eimer, 2011) and some researchers have found that this component is also modulated by emotional expressions (Batty & Taylor 2003;

Caharel, Courtay, Bernard, Lalonde, & Rebaï, 2005; Miyoshi, Katayama, & Morotomi, 2004), e.g., noting a larger mean amplitude of N170 for fearful relative to neutral faces (Batty & Taylor 2003). The VPP is a positive deflection at the fronto–central area with an early latency similar to that of the N170. According to recent findings, larger mean VPP amplitudes are observed in response to fearful faces relative to happy and neutral faces (Williams, Palmer, Liddell, Song, & Gordon, 2006).

In addition to facial expressions, the display of emotions in the form of vocal cues or speech *prosody* is another integral part of social/emotion communication. Prosody refers to a set of supra-segmental features of speech, representing variations in a number of acoustic parameters, e.g., frequency, amplitude, duration of speech sounds, etc. These acoustic properties serve a variety of communicative functions, including emotional expression, supplying information about speaker identity, etc. From such acoustic features, emotions can readily be recognized in speech (Scherer, 1995) and this ability is argued to be widespread or universal (Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; Sauter, Eisner, Ekman, & Scott, 2010). In the growing literature on emotional prosody, studies have been conducted to explore how prosody *encodes* discrete emotions via various acoustic features as well as how listeners *decode* prosody conveying discrete emotions (Banse & Scherer, 1996; Castro & Lima, 2010; Liu & Pell, 2012; Pell, Paulmann, et al., 2009; Scherer et al., 2001; Thompson & Balkwill, 2006; Trimmer, Meyer-MacLeod, Cuddy, & Balkwill, 2007). For instance, accuracy data suggests that anger and fear tend to be recognized more accurately from vocal speech cues than other emotions across different cultures; this could support an evolutionary view of emotion communication, whereby vocal signals associated with threat must be highly salient over long distances and across language systems to ensure human survival (Castro & Lima, 2010; Liu & Pell, 2012; Pell, Paulmann, et al., 2009). These studies have advanced our knowledge of the particular perceptual-acoustic characteristics of vocal emotions and provide clues about how emotional communication may differ in the voice when compared to the face (Paulmann & Pell, 2011).

Other studies have used neuropsychological (Adolphs, Damasio, & Tranel, 2002; Adolphs, Tranel, & Damasio, 2001; Pell & Leonard, 2003; Ross & Monnot, 2008; Thompson & Balkwill, 2006; Trimmer et al., 2007) and neuroimaging (Grandjean et al., 2005; Mitchell & Ross, 2008; Paulmann & Kotz, 2008; Paulmann, Pell, & Kotz, 2008) methods to investigate the neuro-cognitive mechanisms underlying the processing of emotional prosody. For example, some ERP studies have identified a *voice-specific response* (VSR) elicited by sung voices compared to instrumental sounds peaking around 320ms after stimulus onset with a fronto-central distribution (Levy, Granot, & Bentin, 2001). Similarly, amplitude differences of ERPs were reported as early as 164ms post stimulus onset between voice and non-voice (bird songs and environmental sounds) stimuli which peaked around 200 ms on fronto-temporal (positivity) and occipital (negativity) electrodes; this suggests a rapid brain discrimination of sounds of voice at latencies comparable to the face-specific N170 component (Charest et al., 2009). More importantly, emotional effects were also observed in the neural processing of prosody: for instance, the P200 component was shown to be modulated by emotionality and marginally by arousal (Kotz & Paulmann, 2011; Paulmann, Bleichner, & Kotz, 2013; Paulmann & Kotz, 2008) reflecting an early emotional encoding of the stimulus. It was also observed that P200 was followed by a late positive component (LPC) between 400 and 750 ms, which was modulated by both emotion and arousal of the stimuli (Paulmann et al., 2013). These results suggest that emotional/valence information is robustly decoded during early and late processing stages, while arousal information is only reliably taken into consideration at a later stage of processing (Schirmer, Chen, Ching, Tan, & Hong, 2013).

Models of how emotional prosody is processed and represented in the human brain have been proposed in the literature. Schirmer and Kotz (2006) proposed a neuro-cognitive model to demonstrate the multiple successive processing stages during recognition of emotional prosody. In their model, three stages are involved: the first stage occurs around 100 ms, when the sensory processing of acoustic cues takes place; the second stage occurs around 200ms, corresponding to the extraction of emotional significance from acoustic cues (Paulmann & Kotz, 2008; Paulmann, Seifert, & Kotz, 2010); the third stage (around 400 ms) is associated with higher-order cognitive processes that allow emotional significance to be assigned to the output of sensory processes, including the integration of the emotional significance of acoustic cues with semantic information (Schirmer & Kotz, 2006). Wildgruber and colleagues (2009) proposed a similar three-step model of emotional prosody processing: during the first step, supra-segmental acoustic information is extracted, which is associated with activation of predominantly right hemispheric primary and secondary auditory processing regions; the second step corresponds to the representation of meaningful supra-segmental acoustic sequences, which is linked to posterior aspects of the right superior temporal sulcus; the third step pertains to emotional judgment, linked to the bilateral inferior-frontal cortex (Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009).

As a summary, tremendous work has been done in the past on how humans process facial expressions and emotional prosody from others in communication, with evidence on both behavioural performance and neural activities. Moreover, based on the empirical literature, theoretical models on emotional prosody processing have been proposed, which provides a general framework on how emotional prosody is processed and represented in the human brain.

1.1.2.2 Multi-sensory emotion perception across facial expressions and emotional prosody

While both facial expression and emotional prosody, as reviewed above, play a vital role in emotion communication, a more natural situation that occurs in daily life is that the information from these two channels exists simultaneously to communicate emotions. Numerous studies have investigated how information from the two sensory channels interacts and integrates with each other in processing emotional messages (see review Klasen, Chen, & Mathiak, 2012). In this literature, one paradigm that has been commonly employed is the presentation of congruent relative to incongruent information from the two sensory channels, i.e., emotional information from one channel is paired with (or followed by) the same or a different emotion from the other channel; subsequently, the analysis of the *effect of congruence* on emotion perception provides insights into the process of multi-sensory emotion integration. By using such paradigms, it has been shown that facial expressions and vocal expressions influence each other in a significant and involuntary way in emotion perception (de Gelder & Vroomen, 2000; Pell, 2005; van den Stock, Grèzes, & de Gelder, 2008). For instance, when an ambiguous facial expression "morphed" between sadness and happiness was accompanied by a sad voice, participants were more likely to identify the face as sad, even though they were instructed to actively ignore the voice; in contrast, when the same ambiguous face was accompanied by a happy voice, they tended to recognize the face as happy. This pattern was replicated when a morphed emotional voice was identified while accompanied by a happy or sad face (de Gelder & Vroomen, 2000). This effect was still evident when the participants had to perform an extra cognitive task (e.g., digit-adding) in addition to emotion identification (Vroomen, Driver, & de Gelder, 2001); suggesting that despite the simultaneous cognitive demand, the emotion identified was influenced by the concurrent channel to a similar extent as that found in the study of de Gelder and Vroomen, 2000. This finding strengthens the suggestion that the integration of emotional cues across modalities occurs as a mandatory process which is unconstrained by attentional resources (Collignon et al., 2008; Nygaard & Queen, 2008).

In addition to behavioural findings, electrophysiological studies that employ a high-temporal-resolution measurement further shed light on the temporal mechanisms of multi-sensory emotion integration. As has been shown in EEG studies, this integration appears very early during emotion processing. De Gelder et al. (1999) employed an Oddball-like task with ERP measurements in which emotional face-voice pairs were presented to participants who were instructed to pay attention to the faces but ignore the voices. In one condition, congruent pairs (angry voices and angry faces) served as the standard trials (85%) and incongruent pairs (angry voices and sad faces) served as deviants (15%); in the second condition, the standard and deviant trials were exchanged. In both conditions, an aMMN-like (auditory Mismatch Negativity) component peaking at 178ms was observed in response to deviant trials relative to standards. An aMMN is considered as a typical index of the early, pre-attentive detection of deviant information in the auditory modality (Näätänen, 1992). In this study, the auditory stimuli were identical across trials (angry voices) while facial stimuli changed to serve as deviance; it was therefore inferred that the deviance from the visual modality was integrated into the auditory modality and elicited the aMMN, suggesting that the multi-sensory integration occurs early at the pre-attentive stage (de Gelder, Bocker, Tuomainen, Hensen, & Vroomen, 1999). An early integration effect has also been observed on the P2 component (Pourtois, Debatisse, Despland, & de Gelder, 2002), where a longer latency for incongruent face-voice pairs relative to congruent pairs was observed, suggesting that the early perceptual processing indicated by this component was delayed by incongruent emotional information across sensory modalities.

In addition to early stages where sensory processing and cue integration occurs, more in-depth integration of multi-sensory emotional cues at later processing stages has also been documented. For example, it has been shown that the N400 component, which is sensitive to the semantic/ emotional meaning of events in certain processing environments (Bower, 1981;

Bowers, Bauer, & Heilman, 1993; Brown, Hagoort, & Chwilla, 2000), is modulated by the emotional congruence of cues between sensory channels. In a cross-modal priming study conducted by Paulmann and Pell (2010), emotional pseudo-utterances of 200ms or 400ms were presented as prime stimuli, followed by emotional facial expressions or unemotional facial "grimaces" as the target events. Participants passively listened to the utterances and were instructed to judge whether the face represented an emotion or not (Pell, 2005). The EEG results revealed an effect of prime-target congruence in both prime duration conditions: in the longer duration condition, incongruent prime-target pairs elicited a classical N400 effect with larger amplitudes when compared to congruent pairs, indicating the difficulty of integrating the *incongruent* facial cues into the preceding prosodic channel. In the short prime duration condition, however, congruent pairs elicited a larger N400 relative to the incongruent ones (i.e., reversed priming). It was inferred that short primes failed to provide sufficient activation of the corresponding emotional meaning, thus leading to an inhibition of congruent emotional concepts in the other modality. Nevertheless, despite the difference in effect directions, the modification of N400 in both conditions provides evidence of cross-talk between vocal and facial modalities that can be detected at later processing stages of semantic evaluation. Consistent findings on N400 were also reported in a study of emotion integration between the prosodic and semantic channels; using a similar affective priming task in which semantically neutral sentences spoken in different emotional tones served as primes, followed by emotional words as targets, a larger N400 with longer latency was elicited in response to incongruent prime-target pairs relative to congruent ones (Schirmer, Kotz, & Friederici, 2002). Moreover, interestingly the temporal course of this process seemed to differ as a function of sex: female participants showed the N400 effect with a shorter interval between the prime and target, whereas male participants exhibited this effect with a longer prime-target interval. Although the role of biological sex in how emotional signals are communicated and processed is a topic of considerable interest and debate in the literature (Atkinson, Tipples, Burt, & Young, 2005; Goos & Silverman, 2002; Schirmer et al., 2008; Schirmer & Kotz, 2003; Schirmer et al., 2002; Schirmer, Kotz, & Friederici, 2005; Schirmer et al., 2006; Schirmer, Striano, & Friederici, 2005), this topic is beyond the scope of the current work that focuses on how other social factors, the effects of culture background and cultural immersion, impact on emotional communication.

A similar N400 effect was also observed after administering an emotional Stroop task, in which pleasant and unpleasant words spoken in either congruent or incongruent emotional prosody were presented to a group of Japanese participants who was asked to identify the valence of the word meaning and ignore the prosody while EEG signals were recorded (Ishii, Kobayashi, & Kitayama, 2010). The ERP results revealed a typical Stroop effect, such that a larger N400 component was elicited when word meaning and the accompanying prosody were incongruent versus congruent. This N400 effect was larger for females than for males, suggesting that females are more sensitive to prosodic information than males in multi-sensory emotion integration, replicating the findings of Schirmer and colleagues (2002). Moreover, they found that the N400 effect was predicted by the participants' scores of *social orientation,* which refers to an individual's preparedness to engage in social relations; in particular, higher social orientation scores were associated with larger N400 effect, indicating that individuals who are more ready to engage in social relations tend to integrate the prosodic information into word processing to a larger extent. However, again, this association was observed in female participants only, meaning that the effects of social orientation on emotion processing remain an empirical question. Nonetheless, both the behavioural and ERP studies reviewed above highlight the complex process whereby emotion in facial expressions and in prosody are integrated and assigned meaning during emotion perception: this integration can be measured not only by examining "off-line" responses at the behavioural level, but can be indexed at distinct temporal processing stages in the event-related potential, including the MMN and N400 components.

1.2 Cultural effects on emotion perception

A major motivation to study cultural differences in social/emotional communication is their prevalence and relevance in our daily life. As today's world has been globalizing at a fast pace and on an extensive scale, exchanges and migrations between different cultures have become commonplace throughout the world. As a consequence, communications that occur between people from sometimes vastly different cultural backgrounds are not unusual any more. Under such circumstances, however, difficulties in communicating and frustrations caused by misunderstandings are encountered on a frequent basis by the interlocutors, even when the same language is being used as the communicative tool. Therefore, exploring why and how such misunderstandings happen has been an important and appealing research question in psychology and relevant disciplines, the answers to which will not only advance our knowledge of cultural differences in communication but also potentially help us avoid miscommunications and promote relations and mutual understanding between individuals from different cultural backgrounds.

During the past decades, extensive research has been conducted on this topic, and different theoretical frameworks have been proposed, to explain the extent and nature of cultural differences in emotional communication. One key concept used to explain cultural differences in social communication refers to external social norms, i.e., *display rules*, that regulate the expression (and perception) of emotions to serve the purpose of culture-specific social values

(Ekman, Sorenson, & Friesen, 1969; Matsumoto, 1990; Matsumoto et al., 2002; Matsumoto, Yoo, & Fontaine, 2008). Alternatively, a different theory focuses on the internal features of emotional expressions produced by different cultural groups, which do not purposely serve any social values and are defined as "dialects" of the universal language of emotion (Elfenbein, Beaupré, Levesque, & Hess, 2007; Tomkins & McCarter, 1964). In this section, a brief overview of these two major theoretical accounts of cultural differences in emotion communication will be discussed.

1.2.1 Theories on cultural differences in emotion communication

1.2.1.1 Display rules theory

A dominant view on cultural differences in emotion is the theory of *display rules*. Display rules refer to social norms in a specific culture about when, how, to whom, and to what extent one's feelings should (or should not) be displayed in various social situations; these social norms are aimed to maintain certain social values in this particular culture (Ekman et al., 1969; Engelmann & Pogosyan, 2013; Matsumoto, 1990; Matsumoto et al., 2002; Matsumoto et al., 2008). For individuals from a particular cultural group, display rules are enforced by learning relevant management techniques that regulate their emotional displays (e.g., amplification/deamplification, qualification, masking, neutralization, etc.) during socialization and development (Ekman et al., 1969; Engelmann & Pogosyan, 2013; Matsumoto, 1990; Matsumoto et al., 2002; Matsumoto et al., 2008). This kind of social learning results in sustained culture-specific experiences that govern emotion communication, including routine participation in behavioural practices and exposure to particular patterns of emotional expressions; these routines, in turn, modulate the way that people tend to decode and interpret emotional displays posed by members of the (real or perceived) cultural group (Nisbett, Peng, Choi, & Norenzayan, 2001).

In comparison between the Western *individualist* culture and East Asian *collectivist* culture, the East Asian culture considers social harmony as the most critical social value, whereas the Western culture prioritizes the value of individuality (Hall & Hall, 1990; Scollon & Scollon, 1995). Accordingly, culture-specific display rules are adopted in each cultural group to maintain their respective social values (Gudykunst et al., 1996; Matsumoto, Takeuchi, Andayani, Kouznetsova, & Krupp, 1998). For instance, in a classical study investigating negative facial expressions (Ekman, 1971), American and Japanese participants watched stressful movies while their facial expressions were recorded either when they were alone or in the presence of an examiner. While alone, participants from both groups displayed similar negative facial expressions; however, when there was an experimenter present, Japanese participants tended to mask their negative emotions by putting on a smile, whereas the Americans continued showing the same negative facial expressions in spite of the presence of the experimenter. These observations were interpreted as indications of different display rules in the two cultures: while East Asian collectivistic culture (hypothetically) encourages the concealing of negative affective expressions to maintain group harmony, Western individualistic cultures tend to endorse expressions of individual feelings (Ekman, 1971; Markus & Kitayama, 1991; Matsumoto et al., 2008; Matsumoto, Yoo, Hirayama, & Petrova, 2005). Moreover, in East Asian cultures, indirectness in speech communication (e.g., unfinished sentences or vague linguistic expressions: a little, about, possibly, etc.) is more common and socially acceptable than direct references when compared to Western cultures (Bilbow, 1997). Such indirectness is used to prevent potential conflicts or embarrassment that might threaten the face of the interlocutors to maintain interpersonal relations (McGloin, 1984). East Asians are also reported as more inclined to avoid eye contact during face-to-face interactions than individuals from Western cultures; they perceive faces with direct eye gaze as unapproachable and unpleasant (Akechi et al., 2013; Sue & Sue, 1977) and consider the avoidance of eye contact as a sign of respect and politeness to others, which is beneficial to social harmony (Hawrysh & Zaichkowsky, 1991; McCarthy, Lee, Itakura, & Muir, 2006; McCarthy, Lee, Itakura, & Muir, 2008). In sum, with such examples of different social norms in communication being observed across different cultures, *display rules* has become one of the most influential views in the cross-cultural literature on social communication, providing a theoretical framework to interpret the phenomenal cultural differences that people frequently encounter in their daily communication.

1.2.1.2 Dialect theory

An alternative theoretical account of cultural differences in emotion is the *dialect theory*, which was first proposed by Tomkins and McCarter (1964) who adopted the linguistic concepts of "dialect" to the context of non-verbal emotional communication. In particular, it considers the differences across cultures in communicating emotions as "dialects" of the universal language of emotion—similar to the dialects of a language which differ in accent, grammar, and vocabulary, the language of emotion may also have dialects that differ from each other in the way that a particular emotion is conveyed (and interpreted) across cultures (Tomkins & McCarter, 1964). Built on these ideas, the dialect theory was further developed and elaborated by Elfenbein and colleagues (Elfenbein & Ambady, 2002b; Elfenbein et al., 2007). Specifically, this theory starts with a universal view of emotion, which argues that different cultural groups share certain basic universal rules in communicating emotions. In addition to the universalities, each cultural group has a *specific affect program* which adjusts the universal rules in a culturally-specified manner and leads to culture-specific differences in the appearance of emotional expressions, i.e., dialects of the language of emotion. While these differences are acquired through social learning, they do not necessarily serve any specific social purposes, but rather are an integral part of emotional displays; this is distinct from the *display rules,* which are conscious management techniques for the benefit of certain social values.

The dialect theory is arguably supported by the widely observed *in-group advantage* in emotion recognition, which indicates that participants from a certain cultural group are more accurate in judging emotional displays from their own group than a different one (Elfenbein & Ambady, 2002b). Similar to the case of language communication, where it is more common to misunderstand someone with a different accent, the presence of dialects in emotional displays could potentially impede emotion recognition across cultures. For example, it was found that while participants tended to confuse facial expressions of fear and surprise in their own culture, this confusion occurred even more frequently when they were identifying facial expressions with a "dialect", i.e., from a different culture (Elfenbein, Mandal, Ambady, Harizuka, & Kumar, 2002). Similarly, in vocal emotions, confusions between fear and sadness, and confusions between happiness and neutrality are reported within the same cultural group, but these confusions are observed more commonly in judging out-group vocal expressions (Scherer et al., 2001). On the other hand, as spoken words with greater redundancy and uniqueness from other words can be understood in speech even with a strong foreign accent, facial expressions with more distinct features compared to other facial emotions (e.g., happy faces with the revealing of teeth) are relatively less frequently confused with other facial expressions in both in-group (Kirita & Endo, 1995; Leppänen & Hietanen, 2004), and out-group judgement tasks (Elfenbein & Ambady, 2002a).

In comparison with each other, the display rules theory focuses on the external social norms in communication which serve the purpose of specific social values, whereas the dialect theory focuses on the internal features of affective displays which do not purposely serve social values but are considered as an integral part of emotional expressions. While these two theories have been debated in the literature (Elfenbein, Mandal, Ambady, Harizuka, & Kumar, 2004; Matsumoto, 2002), they have been accepted as the two most leading theories on cross-cultural differences in social/emotion communication, which approach this issue from two different perspectives.

1.2.2 Cross-cultural differences in emotion perception

1.2.2.1 Cultural differences in uni-sensory emotion perception from facial expressions or emotional prosody

As previously mentioned, multiple sensory modalities exist for human beings to communicate emotions, e.g., faces, voices, hand gestures, body postures, etc. Among them, facial expressions and vocal cues (i.e., prosody) are the two sensory channels that are most frequently involved in our daily communication. Research has suggested that culture plays a role in perceiving emotions from each of the two modalities in a way that 1) participants can successfully understand emotions from faces or voices from another culture, suggesting the existence of a set of shared properties that encode emotions across cultures and languages; however, 2) they always perform better in recognizing emotions from their own culture than from a foreign culture, indicating an in-group advantage or out-group disadvantage in emotion perception (Elfenbein & Ambady, 2002b). One classical study on cross-cultural facial expression recognition was conducted by Ekman and colleagues (1969), in which participants from various cultures, including U.S., Brazil, Japan, New Guinea and Borneo, identified facial expressions posed by American actors conveying six basic emotions (i.e., happy, surprise, fear, disgust, anger, and sad). It was found that while all participants could recognize the six emotions at above chance level, the American group outperformed the other groups in identifying facial expressions from their own culture (Ekman, 1971; Ekman & Friesen, 1969; Ekman et al., 1969; Izard, 1977).

Likewise, in recognizing emotional prosody, existing cross-cultural studies have also reported that listeners can correctly identify the emotional meanings of prosody from a foreign language at high accuracy levels (often 3-4 times chance performance) (Pell, Monetta, Paulmann, & Kotz, 2009; Scherer et al., 2001; Thompson & Balkwill, 2006; Trimmer et al., 2007). Studies that have examined acoustic measures have described a number of central tendencies in the acoustic characteristics of vocal emotion expressions that seem to be similar across language and cultural boundaries (Banse & Scherer, 1996; Castro & Lima, 2010; Liu & Pell, 2012; Pell, Paulmann, et al., 2009). However, again, it is noteworthy that many of the perceptual studies also reported an in-group advantage, indicating that listeners identify vocal emotions in their native language more accurately than in a foreign language, implying that exposure to vocal emotions in a certain language plays a role in these abilities (Elfenbein & Ambady, 2002a).

While the studies reviewed above have demonstrated clear cultural effects on emotion perception from faces or prosody on the behavioural level, relatively less work has been done on how culture affects the neural mechanisms and responses associated with uni-sensory emotion processing. An fMRI study compared Japanese and American participants in how they processed facial expressions, and found that fearful faces from the participants' own culture induced greater amygdala activation than fearful faces from the other culture; however, this pattern was not observed for other emotion types, such as anger, happiness, and neutrality (Chiao et al., 2008). This pattern might be due to the evolutionary significance of fear: fearful cues from in-group conspecifics may be more relevant to the individual's survival than those from a foreign group, which therefore evoked stronger brain responses in that study. In an ERP study,the neural temporal features of emotional picture processing were compared in European and Asian individuals; here, they found that similar electrophysiological responses were evoked in the early stage of emotion processing, but later components showed a significant decrease of amplitudes for Japanese participants compared to the European group (at parieto-occipital electodes). The latter effects may reflect a poorer engagement of parietal areas of the Japanese participants, which were known to be involved in emotional arousal processing. Possibly, the cultural differences observed during the later processing stages are related to the more restrained emotional expressions in the Japanese population relative to the Western culture (Hot, Saito, Mandai, Kobayashi, & Sequeira, 2006; see also Murata, Moser, & Kitayama, 2013).

1.2.2.2 Cultural differences in multi-sensory emotion perception

While numerous studies have been conducted to explore cross-cultural differences in emotion perception in uni-sensory modalities as reviewed above, relatively limited work has investigated cultural differences in multi-sensory emotion perception involving different information sources. In a classical study in this area, Kitayama and Ishii (2002) examined potential differences between Japanese and American participants in how they integrate *prosodic* and *semantic* information during emotion processing. Using an emotional Stroop task, participants were presented semantically negative or positive words spoken with a congruent or incongruent emotional prosody, and were instructed to recognize the valence of the word and ignore the prosody (word task), or to recognize the emotional prosody and ignore the word meaning (prosody task). Analysis of decision response times revealed that in comparison with the American group, the Japanese participants were influenced to a larger extent by task-irrelevant prosody in the word task, but to a lesser extent by task-irrelevant semantic meaning in the prosody task. It was therefore concluded that in emotion perception, Japanese people were more sensitive to prosodic information relative to the Western group, even when instructed to ignore it (Kitayama & Ishii, 2002). These results were replicated in another East Asian group, Philippine participants, who showed larger attention allocation to prosody than to semantics in both Tagalog and English (Ishii, Reyes, & Kitayama, 2003).

By using a similar Stroop task, Tanaka and colleagues (2010) investigated emotion integration between two other sensory modalities, *faces* and *voices*, in Japanese versus Dutch participants. In their study, happy or angry faces were paired with utterances spoken in happy or angry prosody; participants were asked to identify the facial expression as happy or angry and ignore the prosody (face task), or to identify the emotional prosody and ignore the face (prosody task). Behavioural accuracy results showed that Japanese participants were influenced more by task-irrelevant prosody during face recognition than Dutch participants, whereas in the prosody recognition task, a reversed pattern was observed: the Dutch group was influenced more by the to-be-ignored faces than the Japanese participants (Tanaka et al., 2010). These results indicate that Japanese people are more attuned to vocal information than Dutch people in multi-sensory emotion perception, replicating and extending Kitayama and colleagues' findings (2002, 2003) that Japanese and Philippine participants showed higher sensitivity to prosody than word meaning.

In these studies, the theory of *display rules* has been adopted to explain these observed cultural differences in cross-channel emotion perception. It has been proposed that East Asian collectivist culture values harmonious social relations higher than individuals in comparison with Western individualist culture (Hall & Hall, 1990; Scollon & Scollon, 1995), which arguably leads to culture-specific rules of displaying emotions in different cultures (Gudykunst et al., 1996; Matsumoto et al., 1998). Specifically, restraints on expressing individual feelings, including using indirectness in speech (Bilbow, 1997), masking of (negative) emotions (Ekman, 1971; Markus & Kitayama, 1991; Matsumoto et al., 1998; Matsumoto et al., 2008; Matsumoto et al., 2005), and avoidance of direct eye contact (Hawrysh & Zaichkowsky, 1991; McCarthy et al., 2006; McCarthy et al., 2008) are more prevalent in Eastern Asian culture than in the West. Such display rules result in sustained communication experiences in East Asia where both semantic and facial information is relatively less salient than in the West. In order to efficiently infer the feelings of the others, East Asians may have developed cognitive biases that allow prioritized sampling of relevant information from other available sources, e.g., the vocal cues.

The cross-cultural studies reviewed so far demonstrate clear differences between the Eastern and Western cultures in perceiving multi-channel emotions. However, they were mostly limited to one subgroup of the East Asian culture, the Japanese population, and it is not clear whether the culture-specific biases attributed to East Asian culture built on these studies can be generalized to other Eastern cultures, e.g., the Chinese culture. In addition, the current literature is limited to the behavioural level and little work has been done to examine whether the observed culture-specific biases in behaviours also exist at distinct on-line processing stages of emotion perception. Given the evidence that cross-channel emotion perception is a dynamic process involving both early and later temporal processing stages indicated by different ERP components (de Gelder et al., 1999; Ishii et al., 2010; Schirmer & Kotz, 2003; Schirmer et al., 2002), new insights can be gained by comparing the neural responses and sensitivity of different cultural groups during multi-channel emotion processing. Exploring this research question will advance our understanding of the extent to which culture plays a role in modulating brain activities in emotion communication and shed light on the neuro-cognitive plasticity in emotion processing in the context of cultural experiences. The first study of this thesis, which compared Chinese and English North Americans, aims to provide insights to this question.

1.2.3. Effects of *cultural immersion* on emotion perception

The cross-cultural literature discussed above focuses on emotion perception by contrasting two groups of different cultural origins, e.g., Japanese and Dutch (Tanaka et al., 2010). However, given the rapid change in human migration patterns in recent decades which have resulted in large immigrant communities in many different countries, another relevant question to address is whether the experiences of *cultural and language immersion* in a new environment shape how individuals perceive the world, and how the neurocognitive apparantus adapts to these changes. In the domain of emotional communication, a small amount of work has explored the effect of cultural immersion on these processes. For example, Damjanovic et al. (2013) explored how cultural immersion affects facial expression processing in a visual search task composed of Caucasian and Japanese faces. Three groups were tested: Japanese participants in Japan, British-Caucasian participants living in the UK, and Japanese participants living in the UK. They found that while Caucasian British participants displayed a search advantage in response time favouring happy faces, Japanese participants in Japan showed comparable response time for happy and angry faces; importantly, a group of Japanese participants living in the UK showed an advantage for happy faces, resembling the British participants. Furthermore, a sorting task was performed by the three groups to examine how they structured their emotional space for the seven basic emotions expressed by Caucasian and Japanese faces. Grouping patterns made by the British-Caucasian group and the Japanese participants in Japan were more clearly defined for own-race relative to other-race faces, whereas Japanese participants in the UK tended to create equally defined emotion clusters for both Caucasian and Japanese faces (Damjanovic, Athanasopoulos, Krajciova, Kasai, & Roberson, 2013). These results imply that for the group that had lived in a foreign culture, their experiences of cultural immersion influenced the way facial expressions were processed and represented in their mind, demonstrating a *transition* pattern from their native culture to their host culture.

Neuroimaging evidence has also been reported on the cultural immersion effect on facial expression processing. Derntl and colleagues (2009) conducted fMRI studies on two cultural groups, 24 Asians (Chinese, Japanese, Pakistanis, and Indians; average age of 29.2 years) and 24 Caucasian Europeans (Austrians; average age of 28.3 years), who performed emotion recognition tasks involving Caucasian facial expressions. The Asian group had lived in Europe for an average of 6.1 months. Their results indicated that Asian participants showed stronger amygdala responses to Caucasian facial expressions than European participants, which might reflect an alien effect, i.e., greater demands on cognitive resources when processing out-group facial expressions. More importantly, in the Asian participants, a reverse correlation was observed between their duration of stay in Europe and amygdala activation: Asians who had lived in Europe for a longer time showed reduced amygdala responses, a tendency resembling the patterns of the Europeans (Derntl et al., 2009). These results were replicated in another study comparing a more narrowly-defined group of East Asians (Chinese and Japanese) who had been in Europe for an average of 5.8 months and Austrian Europeans (Derntl et al., 2012). Taken together with Damjanovic and colleagues' behavioural findings on facial expression processing (2013), these studies suggest that immersion in a new culture, and the amount of exposure to the host culture, are significant factors that can modify emotion perception at both the behavioural and neural processing levels.

While data highlighting a cultural immersion effect on emotion processing are relatively limited, complementary evidence of this effect has documented in other domains. Kitayama and colleagues (2003) compared four different groups of university students: 32 Japanese adults living in Japan; 40 European Americans in the U.S.; 18 Americans living in Japan (duration of stay = maximum 4 months); and 21 Japanese living in the U.S. (duration of stay = 2 months to 4 years). The participants performed a framed-line test (FLT) where they were presented a square frame, within which a vertical line was printed; they were then shown a second frame of a same or different size as the first one, and were instructed to draw a line within the frame that was identical to the first line in the first frame in absolute length (*absolute task*) or proportional to the height of the surrounding frame (*relative task*). It was found that *Japanese in Japan* were more accurate in the relative task whereas *Americans in the U.S.* were more accurate in the absolute task; interestingly, the other two groups, *Americans in Japan* and *Japanese in the U.S.,* each displayed a pattern that resembled their respective *host culture* (Kitayama, Duffy, Kawamura, & Larsen, 2003). These results support the general hypothesis that people from East Asian/collectivist cultures are more sensitive to contextual information than people from North American/individualist cultures (Hall, 1976); even more specifically, they suggest that immersion experiences in a new culture affect object perception in a way that progressively resembles the host culture.

There is also some electrophysiological evidence of the effect of cultural immersion on cognition in the domain of colour processing. Athanasopoulos et al. (2010) employed a visual Oddball task of colour perception to compare ERP responses of three groups of young adults (aged 20-23 years old): 10 native English speakers living in the UK; 10 native Greek speakers living in the UK for a relatively long duration (averaged 42.6 months); and 10 native Greek speakers living in the UK for a relatively short duration (averaged 7.2 months). In the Greek language, two linguistic terms, *ghalazio* and *ble,* are used to distinguish light and dark blue, whereas in English no such distinction in colour terms exists. The authors hypothesized that Greek speakers would therefore be more sensitive to the luminance contrast of the blue colour than English speakers. In the Oddball task, the standard stimuli were light or dark blue (or green) circles (70%); the deviant stimuli were blue (or green) circles with a luminance opposed to that of the standards (10%); and the target stimuli were dark and light blue (or green) squares (20%). The participants were instructed to detect the squares (both dark and light) by pressing a button. Results showed that a visual mismatch negativity (vMMN) in the 150-220 ms time window was observed for all three groups in each condition. Interestingly, the short-stay Greek group showed a larger vMMN for blue luminance contrasts than for green ones; in contrast, both the English group and the long-stay Greek group showed similar vMMN responses to both blue and green stimuli. These results argue that the linguistic terms of the blue colour in Greek leads to a higher early sensitivity of the Greek speakers to the blue luminance contrast; of even greater relevance to this thesis, the data argue that this sensitivity was lost by the Greek speakers who had lived in the UK for a longer period of time, providing evidence that increased exposure to the new cultural and linguistic environment had altered their colour perception at a very early online processing stage in a way that approaches their host culture (Athanasopoulos, Dering, Wiggett, Kuipers, & Thierry, 2010).

Taken together, these studies demonstrate that in addition to the effect of *cultural origin* which has been more widely explored in the current literature, *cultural immersion* is also likely to play a role in shaping cognitive processes, including emotion processing, in both behavioural performance and when sensitive on-line neural processing measures are examined. However, given the fact that previous studies on emotion have focused only on facial expressions (Damjanovic et al., 2013; Derntl et al., 2009; Derntl et al., 2012), it is unclear how cultural immersion will impact on multi-sensory emotion processing, when faces and vocal expressions are presented in tandem, and to what extent an impact of the host culture will be observed in altered behaviourial tendencies and/or in neural sensitivity to emotional stimuli. The second study of this thesis represents a first major undertaking to investigate this question.

1.3 The present study

The broad aim of this thesis is to investigate how different cultural experiences affect the neuro-cognitive mechanisms underlying multi-sensory emotion perception. In order to examine different neural processing stages that underlying emotion perception, the ERP technique was employed in this thesis in the context of two frequently-used tasks, a Stroop task and an Oddball task. Event-related potentials (ERPs) are the evoked brain responses extracted from the recorded raw EEG (Electroencephalography) signals, which can be linked to specific sensory, cognitive, or motor events. The ERP measurement bears high temporal resolution and has been commonly used to investigate the temporal characteristics of various cognitive processes in the literature. In this thesis, two ERP components are of interest: the N400 component elicited in the Stroop task, which has been typically associated with the processing of semantic meaning (Bower, 1981; Bowers et al., 1993; Brown et al., 2000; Schirmer & Kotz, 2003); and the MMN component observed in the Oddball task, which has been considered an index of the pre-attentive or early perceptual processing of the evoking stimulus (de Gelder et al., 1999; Näätänen, 1992). In addition to neural measurements, the behavioural results of the Stroop task were also analyzed as an indication of the response selection process at the behavioural level. Thus, rather than focus on a single processing stage—the behavioural response—as in most previous cross-cultural studies (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010), this thesis is novel in its more comprehensive examination of both on-line and off-line measures of emotion perception that are thought to reflect multiple stimulus processing stages, as inferred from online neural responses (indexed by vMMN and N400) and off-line behavioural judgments. Together, these data will contribute to descriptions of temporal characteristics of cross-sensory emotion processing, at specific points in time of theoretical importance to the literature.

Equally important, this thesis will provide novel insights about how this process is influenced by culture, by comparing three different groups—Chinese, English-speaking North Americans, and Chinese immigrants—to shed light on how two different types of cultural experiences, cultural origin and cultural immersion, affect multi-sensory emotion perception. Two ERP studies were conducted involving different groups of participants: Study 1 was designed to explore the effect of *cultural origin* on multi-sensory emotion perception by comparing two groups, Chinese and English North Americans. This study was comprised of two EEG experiments, which employed a Stroop (Expt 1) and Oddball (Expt 2) task, respectively. Data from Study 1 were meant to identify the temporal processing stage(s) that culture plays a role in when processing emotional cues from facial and vocal expressions. To further explore the effect of a distinct kind of cultural experience, i.e., cultural immersion and adaptation to a foreign environment, Study 2 involved participants from a third group, Chinese immigrants to Canada, employing identical tasks and measures used in Study 1. Data from Study 2 were then compared to those from Study 1 to clarify the effects of cultural immersion on multi-sensory emotion processing.

The results of these two studies will be crucial to inform new studies in the emerging literature on how culture affects emotion and cognition, as there are limited reports that have addressed if and how the neuro-cognitive processing of multi-sensory emotions is modulated by different cultural experiences, especially in the case of cultural immersion. By elaborating how emotions from two sensory channels, facial expressions and emotional speech prosody, are processed both behaviourally and neurally by different cultural groups, these studies will provide new clues about the extent to which cultural experiences shape the neuro-cognitive system for communicating emotions; in turn, this will advance our understanding of the plasticity of the human brain in the context of communication and the processing of socio-emotional stimuli. In the following two chapters, each of the two studies is presented in detail in terms of the motivation, method, results, and interpretation of the findings. This is followed by a General Discussion that ties together the two studies and provides an analysis of overarching issues.

2. Study 1

Culture modulates the brain response to human expressions of emotion: electrophysiological evidence

Pan Liu[1], Simon Rigoulot[2], Marc D. Pell[1]

[1] *School of Communication Sciences and Disorders, Centre for Research on Brain, Language, and Music (CRBLM), McGill University, Montréal, Canada*

[2] *International Laboratory for Brain, Music and Sound Research (BRAMS), Centre for Research on Brain, Language, and Music (CRBLM), McGill University, Montréal, Canada*

Corresponding author
Pan Liu
School of Communication Sciences and Disorders, McGill University
2001 McGill College, 8th floor, Montréal, Québec H3A 1G1, Canada
pan.liu@mail.mcgill.ca
1-514 398 4400 ext. 00010

Abstract: To understand how culture modulates on-line neural responses to social information, this study compared how individuals from two cultures, English-speaking North Americans and Chinese, process multimodal emotional cues. Behavioral and event-related potential data were compared between cultures at distinct processing stages that tap semantic and early perceptual processing of emotions. In Experiment 1 (emotional Stroop task), participants were presented face-voice pairs expressing congruent or incongruent emotions in conditions where they judged the emotion of one modality while ignoring the other (face or voice task); results demonstrated marked differences in accuracy and N400 amplitudes between groups, with English participants showing more interference from irrelevant faces than Chinese. In Experiment 2 (Oddball task), participants passively viewed emotional faces with or without simultaneous vocal expressions to test how culture effects early processing of emotions; here, a significantly larger visual MMN was observed for Chinese versus English participants when faces were accompanied by voices, suggesting that Chinese were influenced to a larger extent by vocal cues. Our data illuminate distinct biases in how adults from East Asian versus Western cultures process socio-emotional cues, supplying new evidence that cultural learning modulates neurocognitive responses to emotional expressions rapidly and at multiple stages of stimulus processing.

Key words: Cross-cultural, emotion, ERP, facial expression, vocal tones

1. *Introduction*

In today's highly globalized world, social interactions between individuals from different cultures are not unusual, despite the fact that acquired rules and routines for interacting socially, and possibly the neurocognitive system that supports social communication, vary in marked ways across cultures. One aspect of communication that is vital for interpersonal rapport is the ability to correctly interpret facial and vocal displays of emotion. This study represents one of the first attempts to investigate whether culture shapes the on-line *neural* response to human emotional expressions across faces and voices at two critical stages of event processing. These data will supply new information on how culture modulates the cortical response to socio-emotional cues that are an integral part of (inter-cultural) communication.

Cultural differences in multi-channel emotion processing

Humans typically utilize multiple information sources to express and understand emotions in daily life (Grandjean et al., 2006; Paulmann & Pell, 2011), e.g., a happy utterance is often accompanied by a smiling face. However, recent psychological studies imply that cultures vary in how they process emotions from different sources, at least when behavioral responses are examined. In a key study (Kitayama & Ishii, 2002), Japanese and American participants performed an emotional Stroop task consisting of semantically negative or positive words spoken in an emotionally congruent or incongruent vocal tone, where they had to identify the emotion of the word meaning while ignoring the tone, or vice versa. Recall that in the Stroop paradigm, a

Stroop or *congruence effect* on behaviour is expected in the form of longer response times and/or lower accuracy for incongruent versus congruent trials, which suggests that incongruent cues from the to-be-ignored source impede performance of the target task; likewise, congruent information from the to-be-ignored channel facilitates the target task. Kitayama and Ishii (2002) reported that Japanese participants demonstrated a significantly larger congruence effect in response times than Americans when judging word meanings; in contrast, they exhibited a smaller congruence effect when identifying the tone. This suggests that Japanese participants were influenced to a larger extent by to-be-ignored vocal tones, whereas Americans were influenced more by irrelevant semantic cues. The authors concluded that Japanese people are more likely to allocate their attention to vocal cues in speech, whereas Americans are more attentive to semantic information, even when instructed to ignore it.

Compatible cultural differences were reported in a study of Japanese and Dutch participants who performed an analogous Stroop task, but one composed of emotionally congruent or incongruent face-voice pairs (Tanaka et al., 2010). These researchers found that Japanese participants' ability to identify facial expressions was reduced by conflicting to-be-ignored vocal cues to a greater degree than for Dutch participants, whereas the opposite pattern occurred between cultures in identifying vocal emotions. In line with the previous findings, these results exemplify that during multichannel emotion perception, East Asians bear higher sensitivity to vocal information whereas Westerners are more oriented towards semantic or facial cues, which is also suggested in a study of Tagalog-English bilinguals tested in the Philippines (Ishii et al., 2003). These data make a compelling argument that cultures vary in how they process emotional expressions, although evidence is so far restricted to behavioral judgments.

To explain these cultural differences, it has been suggested that *display rules*, which refer to culture-specific social norms regulating how emotions are expressed in socially appropriate ways, play a role in emotion perception (Engelmann & Pogosyan, 2013; Ishii et al., 2003; Park & Huang, 2010). In contrast to Western individualist cultures, East Asian collectivist cultures consider harmonious social relations as most important (Hall & Hall, 1990; Scollon & Scollon, 1995). In order to maintain social harmony, certain display rules are adopted (Gudykunst et al., 1996; Matsumoto et al., 1998); for example, in East Asian cultures indirectness in communication (e.g., unfinished sentences or vague linguistic expressions) is more commonly used than in Western cultures to prevent embarrassment between interlocutors (Bilbow, 1997). East Asian cultures also endorse the control of facial expressions, especially negative ones, whereas Western cultures encourage explicit emotional expressions by faces (Ekman, 1971; Markus & Kitayama, 1991; Matsumoto et al., 1998; Matsumoto et al., 2008; Matsumoto et al., 2005). Moreover, East Asians are more inclined to avoid eye contact than Westerners, as a sign of respect and politeness to others that benefits social harmony (Hawrysh & Zaichkowsky, 1991; McCarthy et al., 2006; McCarthy et al., 2008). Due to such culture-specific display rules, East Asians may have to rely more on vocal cues to express their emotions and on the other hand, may have developed a higher sensitivity to vocal cues in interpreting the feelings of others during interpersonal communication. In contrast, Western cultures tend to rely more on facial and semantic information which is typically more salient and accessible during their social interactions (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010).

These findings lead to the first research question of our study: do cultural differences in emotion processing previously attributed to "East Asian" cultures generalize to Chinese? Given that the Chinese culture constitutes a major part of East Asia—one that shares the essential collectivist display rules of previously-studied cultures such as Japanese—it would be extremely valuable to examine how Chinese process cross-channel emotions when compared to a major Western culture (e.g., English-speaking North Americans). From the perspective that display rules could promote differences in how Eastern and Western cultures process facial and vocal cues to emotion, it was expected that Chinese participants would show a similar pattern as the Japanese population when compared to Western participants, i.e., higher sensitivity to vocal cues than to facial cues. Testing this hypothesis in a manner that includes both behavioral and on-line *neural* responses that begin to inform the nature and time course of these cultural effects was another major goal of this study.

Cultural effects on electrophysiological activities

While cultural differences in *behavioral* judgments of emotional displays have been reported, there is still little evidence that culture exerts a notable effect on brain responses that occur as humans process multichannel emotions in real time. Electrophysiological evidence shows that multichannel emotion perception is a dynamic process marked by various temporal stages that can be indexed by different event-related potential (ERP) components (de Gelder et al., 1999; Ishii et al., 2010; Schirmer & Kotz, 2003). In an ERP study using a Stroop task, Schirmer and Kotz (2003) presented positive, negative, and neutral words spoken in congruent or incongruent emotional vocal tones to German participants, who identified the valence of the word meaning and ignore the tone, or vice versa. They observed a larger *N400*—a component considered sensitive to the integration of matching/mismatching semantic meanings from different sources (Bower, 1981; Bowers et al., 1993; Brown et al., 2000; Schirmer & Kotz, 2003)—in response to incongruent word-tone pairings relative to congruent pairs in the word valence judgment in female participants. Here, the N400 response can be examined to investigate the potential impact of *culture* at the stage of on-line semantic processing of multi-sensory emotions—i.e., the point when the emotional meaning of the two expressions is registered and compared around 400 milliseconds (ms) after event onset—allowing deeper specification of current behavioral findings.

To further examine the role of culture in multi-channel emotion processing, Schirmer and colleagues (2006) conducted a similar EEG study in which they presented an emotional Stroop task to a different cultural group, Cantonese speakers from Hong Kong. Two-syllable Cantonese words spoken in congruent or incongruent emotional tones were presented to Cantonese speakers whose brain potentials were recorded while they were identifying the emotion of the word meaning and ignoring the tone. Interestingly, the results failed to exhibit a typical Stroop/N400 effect in the Cantonese as had been reported for Germans (Schirmer & Kotz, 2003), but showed a late positivity peaking around 1500 ms, which showed larger amplitude in the incongruent versus congruent condition. The authors argued that in contrast to the N400 effect found in German participants, the delayed congruence effect they witnessed in Cantonese speakers was related to the use of lexical tones and a higher degree of homophony in the Cantonese language; this could have rendered lexical processing more effortful and postponed the integration of emotional tone with the semantic message (Schirmer et al., 2006).

While this conclusion seems inconsistent with previous behavioural data for Japanese participants who displayed higher sensitivity to emotional prosodic cues than to word meanings, it does not necessarily contradict the hypothesis that East Asians prioritize emotional prosody over other information sources when processing emotions. Schirmer et al.'s (2006) observations may be accounted for by several factors; for example, since the emotional salience of the Cantonese stimuli was not matched between the two channels, word meaning and emotional prosody, it is possible that prosody was emotionally less salient than word meanings and easy to ignore, exerting little or delayed influence on judging the word meaning. This could be inferred by the participants' high accuracy rates (around 95%) in the word judgment task (Schirmer et al., 2006). From a cultural perspective, given the long-term history of Hong Kong as a British colony and the fact that Hong Kong people are mostly educated in the English language, it is possible that differences in cultural exposure may also have affected how Schirmer et al.'s (2006) participants process emotional information (Athanasopoulos et al., 2010; Damjanovic et al., 2013), even in their native language. Processing emotions from two-syllable words versus full utterances could also potentially affect the extent to which biases for vocal cues are recorded in different studies. Therefore, by testing a cultural group from Mainland China using well controlled materials, the present study aims to clarify the role of culture in on-line processing of cross-channel emotions.

In addition to the relatively late, semantic processing, it is possible that culture affects brain functioning at earlier stages of multi-sensory emotion processing that is much sooner than 400ms post-stimulus onset, which can be typically measured by using an Oddball paradigm. In such a paradigm, a stream of repetitive stimuli is presented interspersed by infrequent, task-irrelevant changes, in response to which a Mismatch Negativity (MMN) component is typically elicited. In an ERP study using the Oddball task, de Gelder et al. (1999) presented participants angry voices paired with angry or sad faces. The congruent pairs (*angry face-angry voice*) served as standard trials (85% of total trials) while the incongruent pairs (*sad face-angry voice*) served as deviants (15%). Participants passively viewed the faces while ignoring the voices. An auditory MMN (aMMN)-like component peaking at 178ms was observed in deviants relative to standards, which is typically associated with deviance detection in the *unattended* auditory channel and considered as a neural index of the early pre-attentive processing of auditory input (Näätänen, 1992). Note that in this study, the auditory stimuli were identical across all trials (angry) while facial expressions (angry or sad) were manipulated as deviants; it was thus inferred that the detected discrepancy indexed by aMMN actually referred to the facial channel which bore the deviant information (de Gelder et al., 1999). This suggests that the cross-modal integration of emotion takes place at a very early stage, prior to 200ms post-stimulus onset (Jessen & Kotz, 2011;

Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000), although current data do not reveal whether cultural factors somehow modulate brain responses at this early processing stage.

Thus, while research shows that the processing of cross-channel emotions is a dynamic process marked by distinct neural processing stages as indexed by ERP components, it is still not clear how *culture* plays a role at these on-line stages beyond the behavioral level. This leads to the second research question of our study: do cultural differences in emotion processing exist during different neural processing stages beyond the behavioural level? Our hypothesis that culture modulates neural responses to multi-sensory emotions, while still untested, is built upon recent evidence that certain ERP components, such as N400 and MMN, are sensitive to cultural differences in other cognitive domains. In one study, cultural background modulated the N400 when (otherwise unspecified) East Asian and European American participants were presented an affective background picture, followed by a superimposed facial expression that was emotionally congruent/incongruent with the background scene (e.g., a happy face on a positive scene). When identifying the facial expression, East Asians showed a larger N400 for incongruent picture-face pairs compared to congruent pairs, whereas the other group exhibited no difference across conditions, suggesting that East Asians were influenced to a larger degree by background information than the other group. The authors concluded that it is the process of semantic integration, indexed by N400, that is modulated by culture (Goto, Yee, Lowenberg, & Lewis, 2013).

Likewise, the earlier component MMN was reported to be modified by culture/language background in an Oddball task of colour perception (Thierry et al., 2009). In particular, Greek speakers, in whose language there are two different terms distinguishing *light blue* and *dark blue* (*ghalazio* and *ble*, respectively), showed a larger visual MMN component in response to deviants of *dark blue* circles relative to standards of *light blue* circles than did English speakers, in whose language both colours are called *blue* and no such distinction exists. Similar to auditory MMN, the vMMN is considered as a neural index of the early, automatic detection of discrepancy in the *visual* input, which was found to be modulated by culture/language background in this study wherein the Greek speakers exhibited a higher early sensitivity in discriminating dark and light blues due to the specific characteristics of their language. This small but growing literature provides a foundation for new studies of how culture shapes on-line neural responses as humans process emotional expressions.

The present study

Building on recent work, this study presents two ERP experiments that explored how culture modulates neurophysiological responses to emotional expressions (face-voice dyads) at two distinct cognitive processing stages (semantic, early perceptual integration). Adults from two major cultural groups, *English North Americans* and *Chinese*, were tested. In order to examine how emotions are processed by each group in a context that resembles how these displays are typically encountered in daily life, emotional (pseudo) speech and facial expressions were adopted as experimental stimuli. In our first experiment, an emotional Stroop task was conducted to examine cultural effects on behaviour *and* on neural semantic responses to emotions (indexed by N400), and in a second experiment, participants encountered the stimuli in an Oddball task to determine whether culture shapes very early stages in processing multi-sensory emotions (indexed by the MMN). Based on previous findings suggesting that East Asian display rules promote culture-specific biases in emotion processing for Japanese participants (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010), it was hypothesized that Mandarin-speaking Chinese participants, who share cultural display rules promoting social harmony, will be relatively more attuned to vocal expressions than facial expressions when processing cross-channel emotions, whereas English North Americans will be more oriented towards facial cues.

In particular, given evidence that neural semantic processing (Goto, Ando, Huang, Yee, & Lewis, 2010; Goto et al., 2013) and early perceptual processing (Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009) are each sensitive to cultural variables in certain cognitive domains, we anticipated culture-specific biases to be observed at both neural processing stages, as indexed by the N400 and MMN, as well as in behavioural performance. In the Stroop task (Experiment 1), we predicted a larger Stroop effect in the Chinese participants when they were judging the emotion of the faces and ignoring the voices in both the behavioural and N400 data, and a larger Stroop effect in the English group when they were judging the emotion of the voices and ignoring the faces. In the Oddball task (Experiment 2), we expected that the MMN component would be modulated by culture in a way that the Chinese group would be influenced to a larger extent than the English group by vocal emotion cues (indicated by an enhanced MMN component). However, this latter prediction was less certain given the lack of previous evidence demonstrating culture effects on MMN in the *emotional* domain.

1. Experiment 1: Effects of culture on semantic level processing of emotional expressions

To explore how culture influences behavioural and neural responses when processing emotional meanings, we employed an emotional Stroop paradigm, composed of emotionally congruent and incongruent face-voice pairs, with ERP measures in two cultural groups, English North Americans and Chinese.

2.1. Participants

Two groups of participants were recruited. The first group consisted of 19 *English-speaking North American*s (10 female, 9 male; mean age = 25 ± 3.91 years; years of education = 14.18 ± 2.19 years). Given the diversity of cultural backgrounds in native English speakers in North America, a screening questionnaire was administered to ensure that each participant (1) was born and raised in Canada or the U.S. and spoke English as their native language; (2) had at least one grandparent on each of the maternal and paternal sides of British descent and the other grandparents of descent of other Western European countries. The second group consisted of 20 *Mandarin-speaking Chinese* participants (10 female, 10 male; mean age = 24.55 ± 2.75 years; years of education = 16.45 ± 2.68 years), for whom a questionnaire was administered in advance to ensure that they were all born and raised in Mainland China as native Mandarin speakers and had lived out of China for *less than one year*. No difference in age ($F(1, 37) = 0.253$, $p = 0.618$) or years of education ($F(1, 37) = 1.039$, $p = 0.319$) was found between groups, which will be referred to henceforth as the “English” and “Chinese” groups for ease of description. None of the participants reported any hearing impairment, and all had normal or corrected-to-normal vision, as determined by self-report. All gave informed consent before participation, which was approved by the Institutional Review Board of the Faculty of Medicine at McGill University. All participants were financially compensated for their participation.

2.2. Stimuli

The vocal stimuli were fragments of Chinese and English *pseudo-sentences* uttered in sad and fearful emotional tones adopted from previously validated recording databases (Liu & Pell, 2012; Pell, Paulmann, et al., 2009). They were produced by 4 native speakers of each language (2 female, 2 male) in a sad or a fearful tone; in total, 8 fear and 8 sad pseudo-utterances (2 items × 4 speakers) were included. Pseudo-sentences are composed of pseudo content words conjoined by real function words, rendering them meaningless but resembling the phonetic-segmental and supra-segmental properties of the target language (Banse & Scherer, 1996; Pell & Baum, 1997); such stimuli have been used effectively in studies interested in the perception of voice tone independent of the linguistic-semantic content during speech (Castro & Lima, 2010; Paulman & Pell, 2010; Pell, Monetta, et al., 2009; Rigoulot & Pell, 2012; Rigoulot, Wassiliwizky, & Pell, 2013). In the present study, our interest also lies in the perception of emotions in the context of *speech* processing, which occurs most frequently in daily communication. To meet this end, emotionally inflected pseudo-utterances were adopted as vocal stimuli to maximally imitate human speech processing while the semantic content was well controlled.

The facial stimuli were black and white Chinese and Caucasian faces expressing sadness and fear adopted from the Montreal Set of Facial Displays of Emotions (Beaupré & Hess, 2005). In this database, 6 adult actors (3 female, 3 male) from Chinese or Caucasian North American groups were instructed to produce faces of different emotions according to the Facial Action Coding System (Ekman & Friesen, 1978). Six fear and six sad faces (1 item × 6 actors) of each group were selected for the current study (Figure 1).

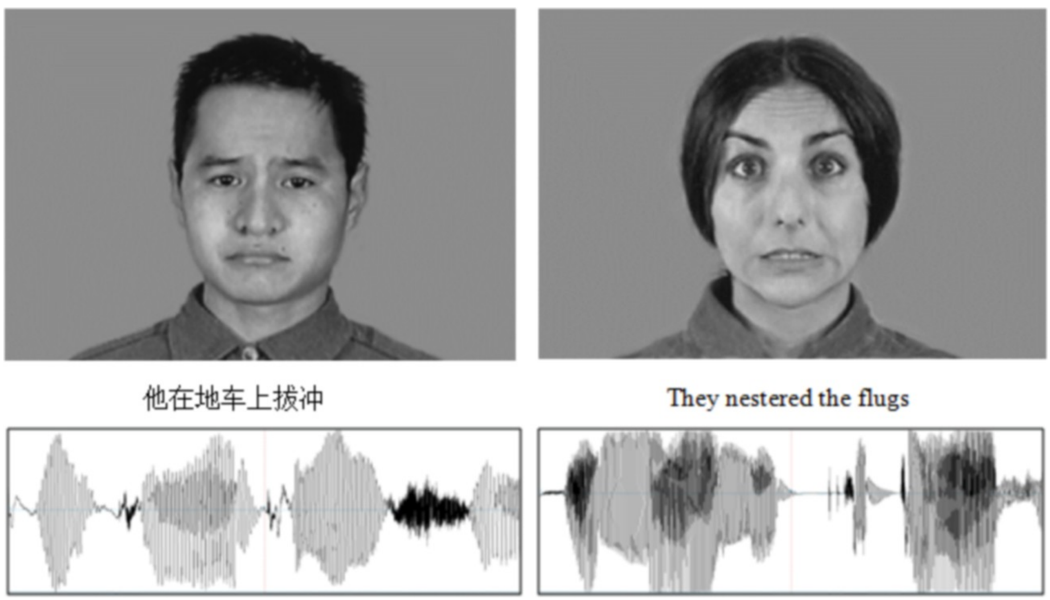


Figure 1. Examples of facial and vocal stimuli. Top left, example of Chinese sad face; bottom left, example of Chinese pseudo-sentence and waveform of this sentence spoken in sadness. Top right, example of Caucasian fear face; bottom right, example of English pseudo-sentence and waveform of this sentence spoken in sadness.

In a pilot study designed to further validate and select the stimuli, 562 English pseudo-utterances and 553 Chinese pseudo-utterances conveying six emotional meanings (fear, sadness, happiness, disgust, anger, and neutrality) were initially selected from the vocal database (Liu & Pell, 2012; Pell, Paulmann, et al., 2009). In order to keep the amount of expressed information consistent across stimuli, the original utterances with varied duration were cut into fragments of 800ms measured from sentence onset using Praat software (Boersma & Weenink, 2001); this duration approximates the average length of utterances presented elsewhere in the immediate literature. For the facial stimuli, 48 Caucasian faces and 48 Chinese faces of the six emotion categories were selected from the facial database. Both the 800ms long vocal stimuli and the facial expressions were then subjected to a validation study, in which each stimulus was validated by 20 in-group native speakers (i.e., 20 English-speaking European Canadians and 20 Mandarin-speaking Chinese, respectively). For each stimulus, the participant made two consecutive judgments: they first identified the emotion being expressed by each item in a six forced-choice emotion recognition task (with happiness, sadness, anger, disgust, fear, neutrality as the 6 options); immediately after, they rated the intensity of the emotion they had selected in the previous recognition task on a 5-point rating scale, where 0 indicated “not intense at all” and 4 indicated “very intense”.

For stimulus selection, recognition rates were calculated for each item across participants; intensity ratings for each item were calculated across participants with correct responses in the emotion recognition task. A three-way ANOVA was run on recognition rates and intensity ratings separately, with sensory channel (voice and face), emotion (fear and sadness), and group (English and Chinese) as three independent variables. To control the emotional salience of the stimuli and ensure that the expected culture-specific bias is not confounded with stimulus salience and clarity (Matsumoto, 2002), stimuli were selected in a way that both the recognition rates and intensity ratings were matched between modalities (vocal and facial), emotions (fear and sadness), and cultural groups (English and Chinese); i.e., there were no significant differences in recognition rates or intensity ratings between vocal and facial modalities, fear and sad emotions, and the two groups. In order to meet all these criteria, 8 pseudo-utterances (2 item × 4 speakers) and 6 facial expressions (1 item × 6 actors) of two emotional categories, *sadness* and *fear*, were selected for each group in the present study. Emotion recognition rates and intensity ratings of these selected stimuli were of medium-to-high level for both groups (Table 1). In the following tasks, only in-group stimuli were presented to each group aiming to investigate emotion processing in the context of their native culture (i.e., Chinese faces and voices to the Chinese group, Caucasian faces and English voices to the English group). Specifically, for each group, the in-group facial and vocal stimuli were paired with each other to construct bi-sensory emotional stimuli, both congruent (fear face & fear voice, sad face & sad voice) and incongruent (fear face & sad voice, sad face & fear voice). In each pair, the static face was synchronized to the dynamic voice, resulting in a length of 800ms for each face-voice pair.

Table 1. Means of recognition rates (in percentage) and intensity ratings (on a 5-point scale with 0 refers to *very weak* and 4 refers to *very intense*) of vocal and facial stimuli for each emotion of each group (standard deviation in parentheses)

	Voices	Faces

	English		Chinese		English		Chinese	
	Fear	Sadness	Fear	Sadness	Fear	Sadness	Fear	Sadness
Recognition rates	90.6(1.8)	92.5(2.7)	91.2(5.2)	91.2(5.2)	90.8(8.6)	91.7(9.3)	91.2(7.6)	90.9(10.0)
Intensity ratings	2.4(0.3)	2.4(0.4)	2.5(0.3)	2.4(0.3)	2.4(0.3)	2.4(0.4)	2.3(0.4)	2.6(0.5)

2.3. Task

An *emotional Stroop* paradigm with ERP measures was used. Two experimental conditions were employed: *face task* condition and *voice task* condition. Each condition consisted of 1 block of 192 trials, in which 96 trials were composed of congruent face-voice pairs (48 fear face-voice pairs, 48 sad face-voice pairs), while the other 96 were incongruent pairs (48 fear face-sad voice pairs, 48 sad face-fear voice pairs). In the face task, the participants were required to identify the facial expression as fear or sad and ignore the voice; in the voice task, they identified the vocal emotion while ignoring the face.

2.4. Procedure

The experimental stimuli were presented and controlled by Presentation® software (Version 16.0, www.neurobs.com). Each trial started with a 1000ms fixation cross at the center of the monitor followed by a 500ms blank screen, followed immediately by the face-voice pair which lasted for 800ms. The face was displayed at the center of a computer monitor (18 × 12.5 cm) and the voice was presented binaurally via insert headphones at a comfortable listening level. The participants were instructed to respond as accurately and as quickly as possible to identify the emotion of the face or the voice as fear or sad, by pressing one of two buttons on a response box, while ignoring the other channel. The assignment of the two response labels, *fear* ("害怕" in

Chinese) and *sad* ("伤心" in Chinese), was counterbalanced to the right and left response buttons. After a response (or after 3500 ms from the offset of the face-voice stimulus in case of no response), a variable inter-trial interval of 500-1000 ms occurred before the next trial began (specifically, a series of random numbers within the range of 500-1000 were generated in Excel and used as the length of the inter-trial interval for each experimental condition). Figure 2 illustrates the sequential structure of each trial. All trials were presented randomly within each block. Each condition started with a practice block of 10 trials (which did not appear in the experiment) to familiarize participants with the experimental procedure. The order of the two conditions (*face task* condition and *voice task* condition) was counter-balanced among participants, and a 10min break was inserted between blocks.

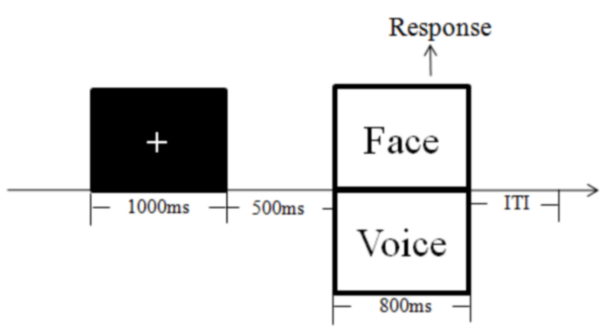


Figure 2. Trial structure in the Stroop task.

2.5. EEG recording and preprocessing

After preparation for EEG recording, the participants were seated at a distance of approximately 65cm in front of a computer monitor in a dimly lit, sound-attenuated and electrically-shielded testing booth. While performing the experiment, EEG signals were recorded from 64 cap-mounted active electrodes (10/10 System) with AFz electrode as the ground, FCz electrode as the on-line reference (Brain products, ActiCap, Munich), and a sampling rate of 1000 Hz. Four additional electrodes were placed for vertical and horizontal electro-oculogram recording: two at the outer canthi of eyes and two above and below the left eye. The impedance for all electrodes was kept below 5 kΩ. The EEG data were resampled off-line to 250 Hz, re-referenced to grand average, and 0.1-30Hz bandpass filtered by EEGLab (Delorme & Makeig, 2004). They were epoched from -200 to 800ms relative to stimulus onset with a pre-stimulus baseline of 200ms (-200 to 0ms). The data were then inspected visually to reject unreliable channels and trials containing large artifacts and drifts, after which EOG artifacts were rejected by means of ICA decomposition. For statistical analyses, only trials with correct behavioural responses were included. On average, approximately 78.4% of the data were retained for analysis after excluding artifacts and incorrect responses (Chinese face task: 80.8%; Chinese voice task: 80.2%; English face task: 76.1%; English voice task: 76.6%). At a final step, data were averaged according to the experimental conditions for each group (congruent/incongruent trials of the face task, congruent/incongruent trials of the voice task).

2.6. Statistical analyses

Based on our predictions of the Stroop effect in this experiment, both behavioural measures (accuracy rates and response time) and N400 data were of interest in the analyses. The N400 literature suggests that this component is typically largest in the centro-parietal region during the 200-600ms post-stimulus time window (see review Kutas & Federmeier, 2011), with more frontal distribution for visual stimuli (Ganis, Kutas, & Sereno, 1996) and auditory language stimuli (see review Kutas & Van Petten, 1994). Therefore, we expected to observe the N400 component in this corresponding area falling within the 200-600ms time span and restricted our ERP data analyses as such (for a discussion of ERP data quantification, see Handy, 2004). Visual inspection of the averaged ERPs for each condition of each group conforms to these expectations, exhibiting a negative deflection after 250ms post-stimulus in the incongruent trials relative to congruent ones, with a main distribution in the middle fronto-centro-parietal area. Accordingly, 15 electrodes from this region (FC1, FC2, Cz, C1, C2, C3, C4, CPz, CP1, CP2, CP3, CP4, Pz, P1, P2) were selected for subsequent analyses. For these electrodes, an exploratory examination of the peak latency of the negativities within the 200-600ms time window yielded an average peak latency of 337ms (range 327-363ms) across conditions and groups. Based on this observation, which conformed to both previous literature and visual inspection, the 250-450ms time window was selected for the N400 component of interest. To quantify this component, mean amplitude values were extracted from this time window for the 15 selected electrodes for statistical analyses.

For both the behavioral and N400 data, a three-step analysis was performed. First, two-way repeated-measures ANOVAs were conducted on accuracy, RT, and the mean amplitude in the 250-450ms time window across selected electrodes, separately in each condition of each group, to examine whether there was an effect of *congruence* elicited in each condition. For each ANOVA, *congruence of the trials* (incongruent and congruent) and *emotion of the attended channel* (i.e., voices in the voice task, faces in face task; fear and sadness) were included as two within-subject factors.

Second, behavioural and neural indices of the congruence effect were calculated. Two *difference scores* were computed for the behavioural data for each participant in each condition: *difference accuracy* as the subtraction of accuracy in the incongruent trials from that in the congruent trials (congruent - incongruent); and *difference RT* as the subtraction of RT in the congruent trials from that in the incongruent trials (incongruent - congruent). For the ERP data, a *difference wave* (i.e., difference N400) was obtained from each selected electrode by subtracting the mean amplitude in the 250-450ms time window in the congruent trials from that in the incongruent trials (incongruent - congruent). Differences between congruent and incongruent trials in behavioural or neural responses reflect the extent to which the information from the task-irrelevant channel is *integrated* into the task-relevant channel (i.e., how much the mismatching information from the task-irrelevant channel interferes with the target task, and how much the matching information from the task-irrelevant channel facilitates the target task). Therefore, higher difference scores and larger difference waves indicate a larger congruence effect.

Third, to specify the impact of culture on the integration effect, a three-way repeated-measures ANOVA was conducted separately on difference accuracy, difference RT, and the mean amplitude difference of the N400 component across selected electrodes *Task* (face task and voice task) was considered as the within-subjects factor, while *Group* (Chinese and English) and *Gender* (female and male) were taken as two between-subjects factors.

In addition to hypothesized changes in the N400, visual inspection showed an earlier positive-going deflection around 200-300ms after stimulus onset in the parieto-occipital region in the voice task of the English group, but not in the other conditions. To quantify this potential earlier effect, 11 parieto-occipital electrodes (Oz, O1, O2, POz, PO3, PO4, Pz, P1, P2, P3, P4) were selected based on visual inspection and the mean amplitude in the time window 200–300ms was extracted for each electrode. The amplitude values were then subjected to a two-way repeated-measures ANOVA in each condition of each group, separately, with *congruence of the trials* (incongruent and congruent) and *emotion of the attended channel* (fear and sadness) as two within-subjects factors.

Finally, to explore potential associations between the participants' behavioural and brain responses to multi-sensory emotional stimuli, Pearson correlation analyses (two-tailed) were conducted between the behavioural difference scores (accuracy; RT) and the amplitude of the N400 difference wave for each experimental condition of each group, respectively.

2.7. Results

Behavioural results

The ANOVAs performed on the behavioural data for each condition for each group showed a significant main effect of *congruence* on accuracy rates (Chinese face task, $F(1,19) = 7.114$, $p < .01$, r[1] = 0.522; Chinese voice task, $F(1,19) = 6.967$, $p < .01$, $r = 0.518$; English face task, $F(1,18) = 5.406$, $p < .01$, $r = 0.481$; English voice task, $F(1, 18) = 8.197$, $p < .01$, $r = 0.559$) and on response times (Chinese face task, $F(1,19) = 9.212$, $p < .01$, , $r = 0.571$; Chinese voice task, $F(1,19) = 8.031$, $p < .01$, $r = 0.545$; English face task, $F(1,18) = 5.037$, $p < .05$, $r = 0.468$; English

[1] Effect size, calculated as $r = \sqrt{F/(F + df_{error})}$ (Rosnow & Rosenthal, 2003)

voice task, F(1,18) = 6.859, p < .01, r = 0.525). Overall, lower accuracy and longer response times were observed when participants encountered emotionally incongruent face-voice pairs than congruent pairs. No significant effect involving the *emotion of the attended channel* was found for these analyses (all ps >.07).

Analysis of the difference scores revealed a significant interaction between *Task* and *Group* on accuracy (F(1, 35) =4.767, p < .05, r = 0.346), but no such effect was found for RT difference scores (p = .17). Simple effects analysis of difference accuracy indicated that the Chinese group showed no difference in performing the face task versus the voice task (p = .26); however, the English group exhibited a significant *Task* effect (F(1, 35) = 20.921, p < .01, r = 0.612) explained by a larger congruence effect in the voice task than in the face task. The effect of culture was significant in the voice task (F(1, 35) = 3.912, p < .05, r = 0.317), as Chinese participants showed a smaller congruence effect than the English group when attending to vocal emotions; no such group difference was noted in the face task (p = .20 ; see Figure 3). No effect involving Gender was found (ps > .21).

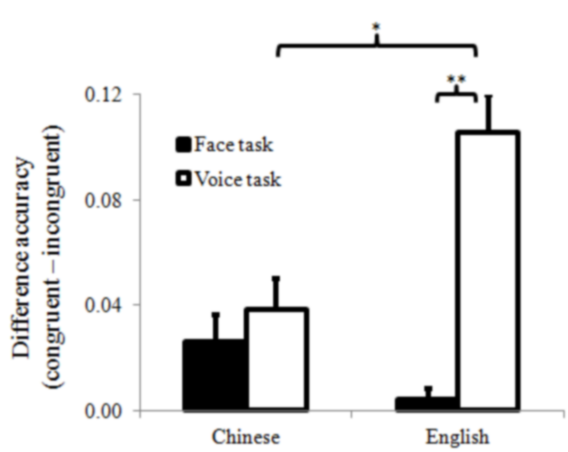
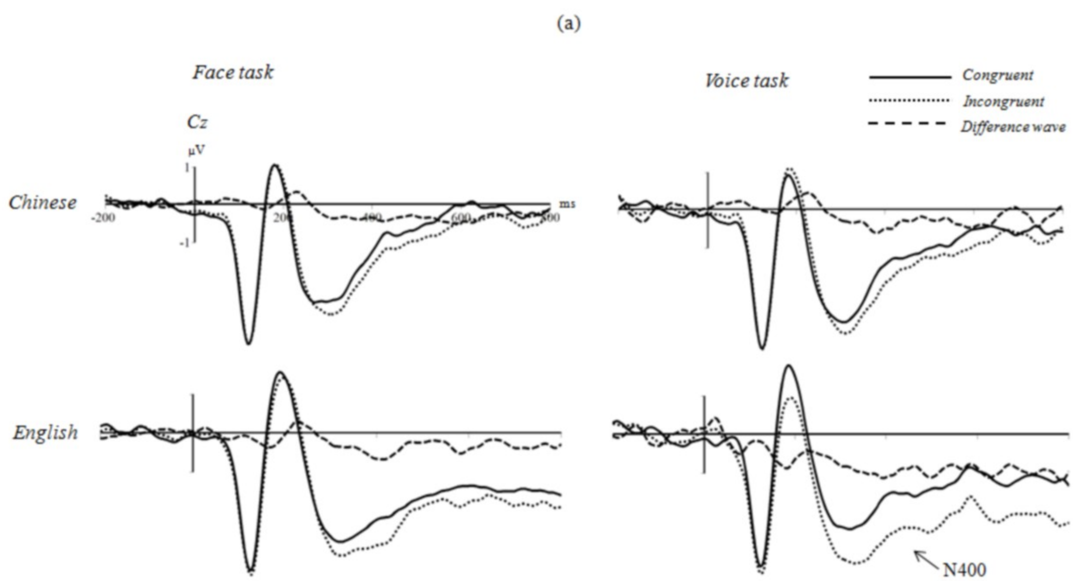


Figure 3. Difference accuracy in the voice and face task for each cultural group.

ERP results

The two-way ANOVAs performed on the mean amplitude in the 250-450ms time window showed a significant main effect of *congruence* across the selected electrodes (i.e., evidence of the N400 component) in each condition for each group (Chinese face task, $F(1, 19) = 6.140$, $p < .01$, $r = 0.494$; Chinese voice task, $F(1, 19) = 6.375$, $p < .01$, $r = 0.501$; English face task, $F(1, 18)= 5.431$, $p < .05$, $r = 0.481$; English voice task, $F(1, 18) = 9.786$, $p < .01$, $r = 0.593$). More negative-going amplitudes were observed in response to emotionally incongruent face-voice pairs relative to congruent pairs, demonstrating a Stroop/congruence effect elicited at the neural processing level. Again, no significant effect involving the *emotion of the attended channel* was found (*ps* >.27; see Figure 4a).



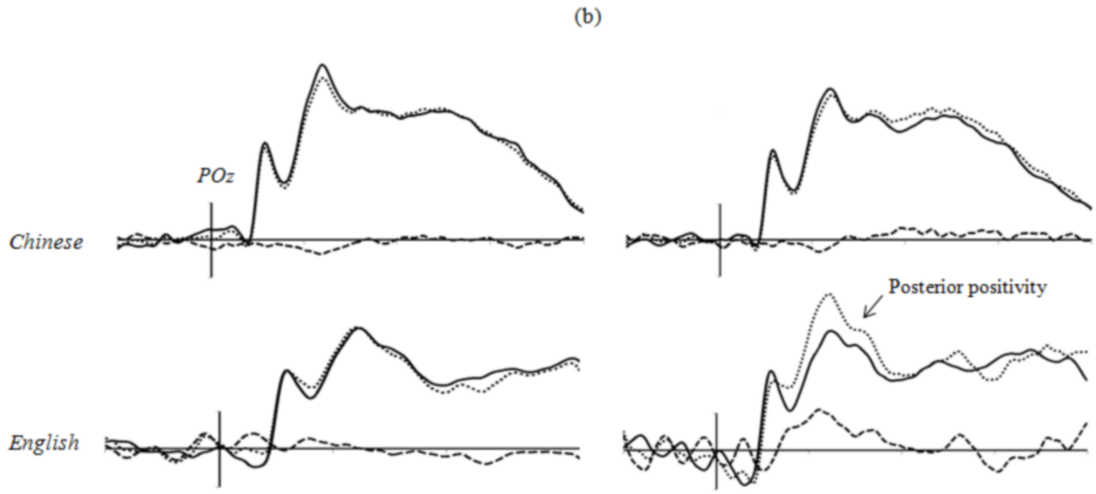
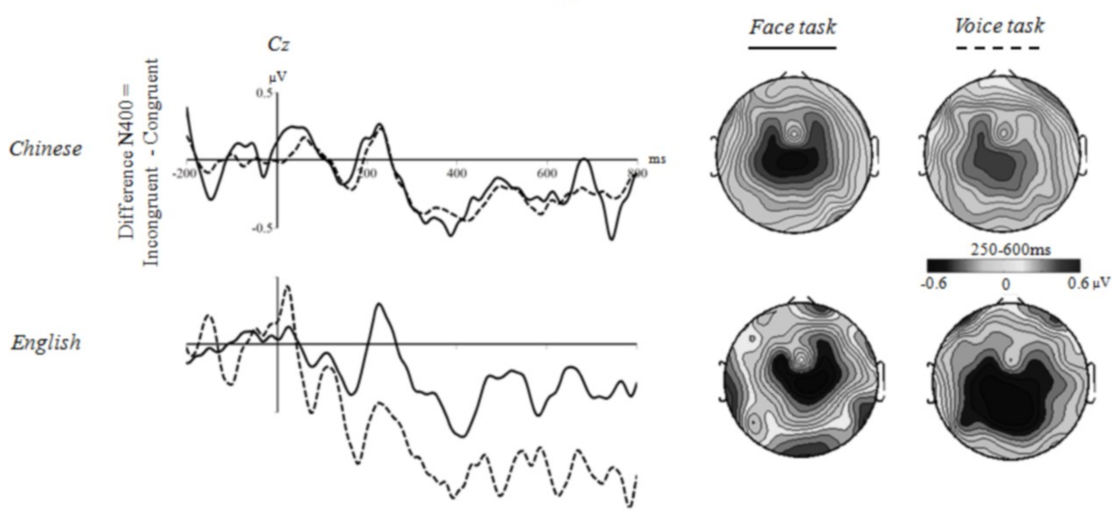


Figure 4. (a) Grand averages elicited by congruent trials (solid lines), incongruent trials (dotted lines), and the difference wave (dashed lines; incongruent - congruent) at Cz electrode for each condition of each group; (b) Grand averages elicited by congruent trials (solid lines), incongruent trials (dotted lines), and the difference wave (dashed lines; incongruent - congruent) at POz electrode for each condition of each group.

The following analysis on the N400 difference waves revealed a significant interaction between *Task* and *Group* ($F(1, 35) = 5.619$, $p < .05$, $r = 0.372$). Step-down analyses indicated that while no effect of *Task* was observed in the Chinese group ($p = .31$), significant task differences characterized the English group ($F(1, 35) = 10.587$, $p < .01$, $r = 0.482$) who showed a larger N400 difference in the voice task than in the face task. A *Cultural* effect was significant in the voice task only ($F(1, 35) = 7.257$, $p < .01$, $r = 0.414$), with the Chinese group showing a smaller N400 difference than the English group (see Figure 5). No effect involving *Gender* was found (*p*s > .11). See Table 2 for a summary of the means and standard deviations for the behavioural and N400 data.

For the earlier positivity observed in the parieto-occipital area, the two-way ANOVAs showed a significant main effect of *congruence* on the mean amplitude in the 200-300ms time window in the voice task of the English group only ($F(1, 18) = 8.906$, $p < .01$, $r = 0.575$), not in any other condition (*ps* > .13). The incongruent trials in the voice task elicited a more positive-going amplitude relative to the congruent ones in the English group, but not in the Chinese group (see Figure 4b). No significant effect involving the *emotion of the attended channel* was found (*p*s > .24).

(a)



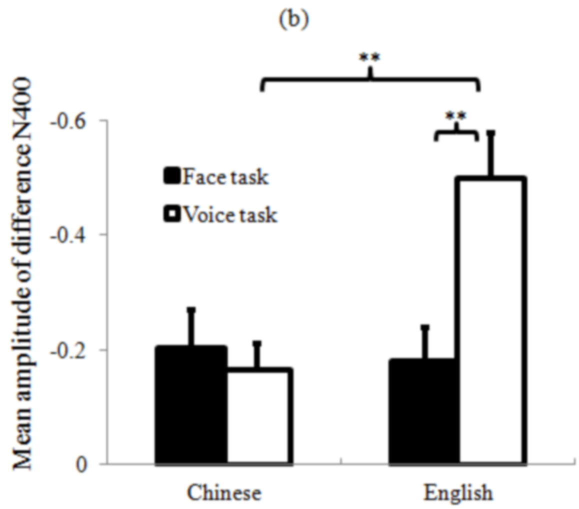


Figure 5. (a) Grand averages at Cz electrode and topographic maps of difference N400 for each condition of each group. (b) Mean amplitude values of difference N400 averaged across selected electrodes.

Table 2. Means and standard deviations (in parentheses) of behavioural accuracy, response time (RT), and mean amplitude of N400 in Experiment 1 for each experimental condition

				Accuracy	RT (ms)	N400 (μV)
Chinese	Face task	Congruent	Sad face-sad voice	0.92(0.05)	745(263)	-1.65(1.50)
			Fear face-fear voice	0.91(0.04)	755(273)	-1.64(1.23)
		Incongruent	Sad face-fear voice	0.89(0.09)	776(313)	-1.86(1.61)
			Fear face-sad voice	0.88(0.06)	770(256)	-1.84(1.89)
	Voice task	Congruent	Sad face-sad voice	0.87(0.08)	784(354)	-1.71(1.72)
			Fear face-fear voice	0.87(0.05)	780(332)	-1.69(1.56)
		Incongruent	Sad face-fear voice	0.84(0.08)	810(183)	-1.87(1.92)
			Fear face-sad voice	0.83(0.07)	802(347)	-1.89(1.58)
English	Face task	Congruent	Sad face-sad voice	0.94(0.05)	699(184)	-1.68(0.98)
			Fear face-fear voice	0.93(0.07)	700(215)	-1.68(1.21)
		Incongruent	Sad face-fear voice	0.91(0.05)	718(165)	-1.88(1.30)
			Fear face-sad voice	0.90(0.06)	715(268)	-1.87(1.29)
	Voice task	Congruent	Sad face-sad voice	0.91(0.08)	760(305)	-1.66(1.47)
			Fear face-fear voice	0.90(0.07)	757(271)	-1.64(1.61)
		Incongruent	Sad face-fear voice	0.80(0.13)	781(318)	-2.15(2.00)
			Fear face-sad voice	0.79(0.09)	782(298)	-2.09(1.97)

Correlation results

Pearson correlations demonstrated significant/marginally significant positive correlations between accuracy (difference score) and N400 amplitudes (difference score) for each condition of each group: Chinese face task, r(18) = .298, $p < .05$; Chinese voice task, r(18) = .207, $p = .061$; English face task, r(17) = .250, $p = .05$; English voice task, r(17) = .301, $p < .05$. No significant correlations between difference RT and N400 difference were found (ps > .07). In general, larger difference accuracy was associated with larger N400 differences for each group, suggesting that the cross-modal emotion integration occurred at both behavioral and neural semantic levels in a consistent and related manner.

2.8. Discussion

As predicted, results of Experiment 1 demonstrate significant cultural differences between Chinese and English groups in the semantic-level integration of multi-sensory emotional stimuli: whereas Chinese participants did not show significant differences between the face and voice tasks, the English group was influenced to a larger extent by facial cues in the voice task than by vocal cues in the face task. These differences were evident in behavioural accuracy, the N400 component, as well as a positive deflection that occurred around 200ms post-stimulus onset in select conditions (as elaborated below).

First of all, it should be noted that regardless of cultural origin, a significant congruence effect was observed at both the behavioural and electrophysiological levels in each experimental condition, as participants performed worse in incongruent versus congruent trials with an increased N400 component. This confirms previous reports that emotion integration across faces and voices can be observed at both the behavioural and neural levels (Collignon et al., 2008; de Gelder & Vroomen, 2000; Föcker, Gondan, & Röder, 2011; Ishii et al., 2010; Massaro & Egan, 1996; Mitchell, 2006; Nygaard & Queen, 2008; Schirmer & Kotz, 2003); moreover, as the congruence effect was not influenced by the emotional meaning of the facial or vocal expressions, it seems that the process of cross-channel semantic integration is relatively independent of the specific emotions being appraised. Importantly, while the congruence effect was observed in each attentional condition for each group, there was a significant distinction between groups in the *size* of the congruence effect as a function of task. The English group exhibited a larger effect in the voice task than the face task, indicated by larger difference accuracy scores and larger N400 differences in the voice task, whereas no between-task difference was observed for Chinese participants. These observations suggest that the English participants were distracted to a larger degree by to-be-ignored faces in the voice task than by to-be-ignored voices in the face task, whereas no such pattern was observed in the Chinese group. These results are compatible with the hypothesis that individuals from Western cultures are relatively more oriented to facial versus vocal cues during communication in comparison with individuals from East Asian cultures (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010).

The N400 effect observed in each experimental condition showed similar temporal and spatial characteristics to those of the semantic-related N400 reported in previous literature (Kutas & Federmeier, 2011). In particular, it has been argued that the N400 effect reflects the processing of mismatching semantic/emotional cues from different information sources, and the larger N400 amplitude is associated with increased cognitive processing resources necessary to integrate mismatching meanings from different sources (Brown & Hagoort, 1993; Brown et al., 2000; Chwilla, Brown, & Hagoort, 1995; Holcomb & Neville, 1990; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008; Paulmann et al., 2008; Schirmer et al., 2002). While this is the typical interpretation of the Stroop-induced N400 effect adopted by previous studies that are directly relevant to this one (Ishii et al., 2010; Schirmer & Kotz, 2003), an alternative interpretation of the N400 effect is that it reflects top-down conflict control functions. When mismatching information from two different sources is presented simultaneously in interference tasks, the participant has to focus on one source and inhibit conflicting information activated by the other source. It has thus been argued that the Stroop-elicited N400-like component in the fronto-central area is associated with conflict monitor and inhibition control (Dong, Zhou, & Zhao, 2011; Liotti, Woldorff, Perez III, & Mayberg, 2000), which may arise from the anterior cingulate cortex and frontal cortex (Etkin, Egner, Peraza, Kandel, & Hirsch, 2006; Haas, Omura, Constable, & Canli, 2006). In the present study, it is possible that this conflict control process is vulnerable to the modulatory effect of acquired cultural biases about the importance of specific social cues in communication which impact on multi-sensory emotion perception, yielding our observation that English participants found it more difficult to inhibit irrelevant facial information than participants in the Chinese group.

In addition to the hypothesized N400 effect, a significant congruence effect on a parietal-occipital positivity around 200-300ms was also observed in the EEG data (i.e., larger amplitude elicited in incongruent trials relative to congruent trials), but exclusively in the voice task for the English group. The temporal and spatial features of this deflection resemble those of a posterior P2 component reported in the literature, which is considered originating from posterior visual association cortex and related to visual attention; more specifically, larger posterior P2 has been observed in response to attended affective stimuli (Carretié, Mercado, Tapia, & Hinojosa, 2001; Eimer, Cockburn, Smedley, & Driver, 2001; Talsma, Slagter, Nieuwenhuis, Hage, & Kok, 2005). In our experiment, this increased positivity found in the voice task may reflect an involuntary attention shift to the irrelevant facial information. Interestingly, this was only observed for English and not Chinese participants, providing potentially new support for the argument that individuals from many Western cultures are more attentive to facial cues when compared to East Asian participants, based on highly sensitive, on-line brain responses to emotional stimuli.

While the English group showed the expected cognitive biases to irrelevant faces in the voice task (e.g., larger congruence effect in accuracy and N400), no difference between the two tasks was observed in the Chinese group. This is somewhat unexpected due to the assumption that Chinese participants, like Japanese individuals (Ishii et al., 2003), would be more oriented to vocal cues and therefore show a larger congruence effect in the face task than the voice task. This pattern could be especially pronounced in the context of processing negative emotions as in the present study, given that Chinese/East Asian display rules tend to mask negative facial expressions (Ekman, 1971), meaning that these individuals may rely more on tonal cues to understand negative feelings (Ishii et al., 2010). However, it should be noted that in spite of the lack of differences between tasks for the Chinese participants, when compared to English participants they showed significantly smaller *congruence effects* in the voice task, although not in the face task. This suggests that in the voice task, Chinese adults had less difficulty ignoring the faces and more easily attuned to the voices when compared to English listeners; this pattern is compatible with results reported in the behavioural literature involving the Japanese population (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010).

It is worth mentioning that, although facial and vocal stimuli were carefully validated and controlled for perceived emotional salience between communication channels, a majority of participants in each group reported that it was more demanding to identify the emotion of vocal expressions in pseudo-sentences than of facial expressions when debriefed after the experiment. Thus, since pseudo-sentences were somewhat unusual and less familiar than facial displays, it is possible that for both groups, vocal expressions were relatively easier to ignore in the face task, whereas the faces are more difficult to ignore in the voice task. If true, the larger congruence effect expected in the face task relative to the voice task for the Chinese group might have been mitigated by differences in stimulus features; likewise, in the English group, the large congruence effect observed in the voice task may have been facilitated for this reason. However, even though the familiarity of stimuli may have varied somewhat in the two communication channels, it merits re-emphasizing that emotional recognition rates and intensity ratings of stimuli in both channels were overall high and carefully matched across channels. Moreover, any influence of presenting vocal cues embedded in pseudo-utterances on our data should be comparable between groups, in spite of which the results still exhibit a clear and robust difference between the two groups in perceiving emotions, i.e., the English participants showed higher susceptibility to the faces than to the voices, whereas the Chinese participants did not.

Another potential factor for the absence of the expected bias in the Chinese group possibly involves the negative valence of the two emotions, sadness and fear, that were adopted in this experiment. Given display rules whereby Chinese/East Asians tend to restrain their facial expressions, especially negative ones such as sadness and fear (Ekman, 1971; Matsumoto et al., 2008; Matsumoto et al., 2005) to maintain social harmony, it is possible that the faces that they were presented during the Stroop task, which explicitly expressed negative emotions, were of higher-than-usual salience to them. Consequently, these faces might have drawn their attention to a greater extent and mitigated the size of the Stroop effect in the voice judgment condition. To clarify the potential effect of emotional valence on culture-specific patterns in cross-channel emotion processing, future studies employing positive emotions would be helpful.

This experiment extends previous findings on emotion perception in East Asians by contributing new ERP evidence demonstrating cultural differences in emotional meaning integration during neural processing stages beyond the behavioural level. While there are previous claims that the N400 is modulated by culture in the domain of visual perception (Goto, Ando, Huang, Yee, & Lewis, 2009; Goto et al., 2013), our ERP findings extend the applicability of this component as a neural index of cultural effect in audio-visual emotion perception. Our results also highlight a posterior positivity occurring around 200-300ms post-stimulus onset that could represent another neural index of acquired cultural differences on attention allocation at earlier stages of emotion perception. This implies that, in addition to behavioural and neural semantic processing, culture may have an impact on much earlier processing stages during multi-sensory emotion perception, as also suggested by recent work on colour perception described by Thierry et al. (2009). To test this hypothesis, a second experiment using an Oddball/MMN paradigm was conducted to explore the early perceptual stage in processing emotions encoded by combined face-voice expressions.

2. *Experiment 2: Effects of culture on early perceptual processing of emotional expressions*

In this experiment, we investigated the role of culture on an earlier neural processing stage of emotion integration by analyzing a visual MMN component elicited in an Oddball paradigm, comparing groups of English and Chinese participants. True to our objectives, data obtained from an Oddball task permit further elaboration of culture effects on emotion processing that were highly expected at the semantic processing level (i.e., Stroop task, Expt. 1), allowing these to be tested at a different stage, the earlier perceptual processing, for the same stimuli and participants.

3.1. Participants

Participants were the same English and Chinese adults who completed Experiment 1.

3.2. Stimuli

The same vocal and facial stimuli used in Experiment 1 were employed in Experiment 2. In addition, two pure tone auditory stimuli lasting 800ms in duration were constructed as non-vocal stimuli to present to each group (four pure tone stimuli total). The frequency of each pure tone stimulus was determined by calculating the mean fundamental frequency values of the fear and sad vocal expressions produced by speakers of each language (Chinese fear voices: 267 Hz, Chinese sad voices: 249 Hz; English fear voices: 266 Hz, English sad voices: 187 Hz).

3.3. Task and design

An Oddball task was adopted to explore the early cross-modal integration in emotion perception. Three experimental conditions were employed. In Condition 1, facial expressions were presented without any auditory stimuli, which served as the control condition to examine the classical visual MMN effect elicited by facial stimuli (*face-only condition)*. In Condition 2, emotional voices were paired with the same facial expressions as in the face-only condition, to test the influence of concurrent vocal information on the early processing of faces (*face-voice condition).* In Condition 3, the *face-tone condition*, pure tone stimuli were paired with the same faces, in order to exclude the possibility that effects observed in the face-voice condition would be simply attributable to the presentation of audio-visual stimuli, regardless of their emotional meanings.

For each experimental condition, 4 blocks of 600 trials each were presented, including 480 standard trials (80%), 60 deviant trials (10%), and 60 target trials (10%). In Condition 1, fear faces served as standard trials and sad faces served as deviant ones in blocks 1 and 2, while this pattern was reversed in blocks 3 and 4 (i.e., sad faces serve as standards and fear faces served as deviants). In addition to faces, 60 pictures of circles were randomly inserted in each block as *target trials* to which the participants had to press a button as response. This was to ensure that the participants were actively attending to the targets and viewing the faces passively.

In Condition 2, each face was paired with an emotional voice to construct a *bimodal face-voice condition*. Specifically, faces were paired with fear voices in block 1 and 3, whereas faces of block 2 and 4 were paired with sad voices. This led to the four bi-modal blocks: in two blocks, congruent pairs (*fear face-voice pairs* in block 1*, sad face-voice pairs* in block 4) served as standards while incongruent pairs (*sad face-fear voice pairs* in block 1*, fear face-sad voice pairs* in block 4) served as deviants; in the other two blocks, incongruent pairs (*fear face-sad voice pairs* in block 2, *sad face-fear voice pairs* in block 3) served as standards and congruent pairs (*sad face-voice pairs* in block 2, *fear face-voice pairs* in block 3) served as deviants. The purpose of exchanging standard and deviant trials as congruent or incongruent pairs was to examine whether the *congruence* of face-voice pairs was relevant in evoking the vMMN. Again, in each block, 60 circles were randomly inserted as *targets*.

Similarly, in Condition 3, each face was paired with a pure tone of 800ms to create a *face-tone condition*. In blocks 1 and 3, the faces were paired with a tone with a frequency matched with the mean f0 of the fear voices; in blocks 2 and 4, the faces were paired with a tone with a frequency matched with the mean f0 of the sad voices. Sixty circles were included as targets. Overall, the visual stimuli, the frequency and proportion of different types of trials (standards, deviants, and targets) were identical in all three conditions; differences lay only in the accompanying auditory stimuli (see Table 3).

Such a visual MMN paradigm (the visual stimulus sequence consists of standards, deviants, and targets) is different from the classical auditory MMN paradigm in which the participants would be focusing on the visual modality and ignoring the auditory channel (Näätänen, 1992). This paradigm has been typically used in the visual MMN literature and it is assumed that while the classical auditory MMN paradigm examines the preattentive processing of auditory stimuli, the vMMN task would tap into the *early perceptual* processing of the deviant versus standard visual stimuli. In particular, the early negativity elicited in such paradigms, i.e., visual MMN, has been considered as an indication of early perceptual detection of deviant information in the visual modality (Athanasopoulos et al., 2010; Stagg, Hindley, Tales, & Butler, 2004; Susac, Ilmoniemi, Pihko, & Supek, 2004; Thierry et al., 2009).

Table 3. Types of standard and deviant trials per block in each condition of the Oddball task

	Condition 1 Face-only	Condition 2 Face-voice	Condition 3 Face-tone
Block 1	Standard: sad face Deviant: fear face	Standard: sad face & sad voice Deviant: fear face& sad voice	Standard: sad face & tone1 Deviant: fear face & tone1
Block 2	Standard: sad face Deviant: fear face	Standard: sad face & fear voice Deviant: fear face& fear voice	Standard: sad face & tone2 Deviant: fear face & tone2
Block 3	Standard: fear face Deviant: sad face	Standard: fear face & sad voice Deviant: sad face & sad voice	Standard: fear face & tone1 Deviant: sad face & tone1
Block 4	Standard: fear face Deviant: sad face	Standard: fear face & fear voice Deviant: sad face & fear voice	Standard: fear face & tone2 Deviant: sad face & tone2

3.4. Procedure

Since participants in Expt. 1 had to focus on either the facial or vocal channel and ignore the other, which could promote conscious attention to the match/mismatch of emotions between the two channels, the Oddball task was always completed by each participant in a single session *preceding* the session when the Stroop task was presented, with an interval of at least a day between sessions. This avoided the possibility that conscious awareness of cross-channel congruence/incongruence in the Stroop task would interfere with responses in the Oddball task, in which participants were instructed to ignore the vocal/auditory channel completely and focus only on the visual channel.

In all three conditions, each block started with a 1000ms fixation cross at the center of the monitor, followed by the trials which were presented pseudo-randomly such that two deviant trials never appeared in immediate succession, and at least three standard trials appeared in a row between two deviant ones. Each trial was presented for 800ms, and the variable inter-trial interval was 500-1000ms. The visual stimuli were presented at the center of the monitor and the auditory stimuli were presented binaurally via headphones at a consistent comfortable listening level. In all conditions, the participants were instructed to detect the circle targets among the faces by pressing the spacebar; in Condition 2 and 3 where a face was paired with a sound, it was emphasized that they should ignore the concurrent auditory stimuli whatever it was (Figure 6). Each condition started with a practice block of 40 trials to familiarize participants with the procedure. The order of the four blocks within each condition, and the order of the three conditions within the experiment, was counter-balanced among participants, and a 10-minute break was inserted between blocks.

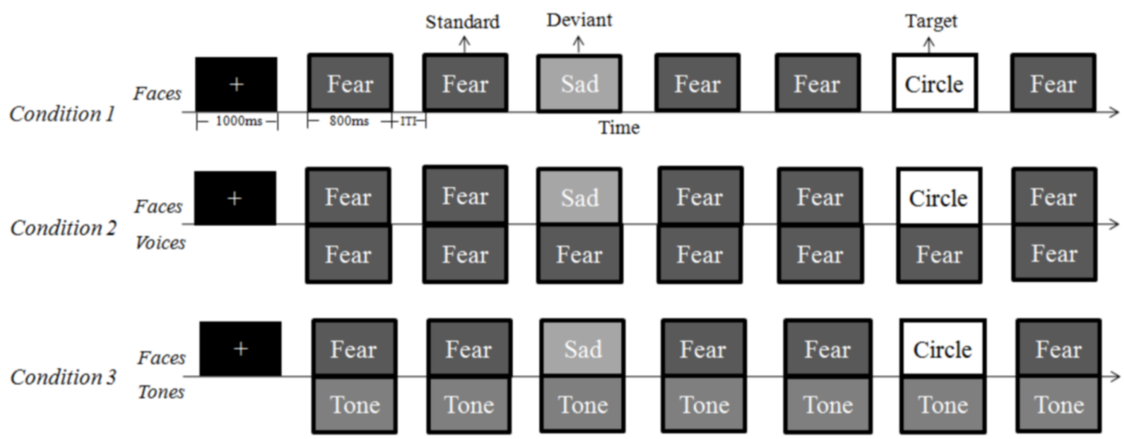


Figure 6. Task procedures for each of the three conditions in the Oddball task. From top to bottom: face-only, face-voice, face-tone.

3.5. EEG recording and preprocessing

The same procedures of EEG recording and preprocessing were employed as those in Experiment 1. For further analysis, all target trials and standard trials that immediately followed deviants were also excluded, which left 480 trials (420 standards, 60 deviants) in each block. After artifact rejection, an average of 81.4% of data were retained for subsequent statistical analyses (Chinese face-only: 79.6%; Chinese face-voice: 81.5%; Chinese face-tone: 81.8%; English face-only: 81.7%; English face-voice: 82.3%; English face-tone: 81.3%).

3.6. Statistical analyses

Built on the hypothesis of the early MMN effect in the Oddball task, the visual MMN component was of interest in the analyses. Research suggests that MMN elicited by visual/facial stimuli is typically maximal in the occipital-parietal area across the 100-300ms time span post-stimulus (Athanasopoulos et al., 2010; Schröger, 1998; Thierry et al., 2009; Wei, Chan, & Luo, 2002; Zhao & Li, 2006). Based on this evidence, we focused our analyses on the same temporal and spatial regions to test our hypotheses. Visual inspection of the waveforms of grand averaged ERPs revealed more negative-going deflections elicited by deviant trials relative to standard ones during the 100-200ms time window in the *parieto-occipital* region in each condition, confirming our expectations. Accordingly, 14 electrodes were selected from the occipital-parietal region (POz, PO3, PO4, PO7, PO8, PO9, PO10, Oz, O1, O2, P5, P6, P7, P8) for further analyses. For these electrodes, an exploratory investigation of the peak latency of the difference wave between the deviant and standard trials (deviant - standard) yielded an average peak latency of 151ms (range 143-158ms) across conditions, consistent with previous literature and our visual inspection of the data; accordingly, the 100-200ms time window was selected for analysis of MMN mean amplitude values for the 14 selected electrodes.

A three-step analysis was performed on the EEG data. First, repeated-measures ANOVAs were conducted on the mean amplitude in this time window across selected electrodes, in each of the three conditions for each group respectively, to verify whether there was a deviance effect in each condition. Specifically, in the face-only condition, *deviance* (standard and deviant) and *facial expression of deviants* (fear and sadness) were adopted as within-subjects factors for a two-way repeated-measures ANOVA; in the face-voice condition, *deviance* (standard and deviant), *facial expression of deviants* (fear and sadness), and *congruence of deviants* (congruent and incongruent) served as within-subjects factors for a three-way repeated-measures ANOVA; in the face-tone condition, *deviance* (standard and deviant), *facial expression of deviants* (fear and sadness), and *tone* (frequency 1 and frequency 2) were included as within-subjects factors for a three-way repeated-measures ANOVA.

Second, *difference waves* were obtained by subtracting the mean amplitude in the standard trials from that in the deviant trials (deviant - standard) for the 100-200ms time window in each condition. In the *face-only* condition where there were only visual-evoked potentials, the difference wave was considered as a visual Mismatch Negativity (vMMN) elicited by deviant faces relative to standard ones (Astikainen & Hietanen, 2009; Susac et al., 2004; Zhao & Li, 2006). In the *face-voice* and *face-tone* conditions, while facial stimuli were varied as deviant and standard trials, the auditory stimuli were identical across all trials (fear or sad voices in the face-voice condition; pure tones in the face-tone condition). Thus, the same auditory brain potentials were evoked in all trials while different visual potentials were elicited in standard versus deviant trials. Therefore, by subtracting the ERPs in the standard trials from those in the deviants, the auditory potentials were eliminated and the obtained difference wave was also considered as a visual-evoked MMN component.

Third, to further clarify how the difference wave was modulated by culture, a second three-way repeated-measures ANOVA was conducted on the amplitude of vMMN across selected electrodes, with *Condition* (face-only, face-voice, and face-tone) as the within-subjects factor, while *Group* (Chinese and English) and *Gender* (female and male) served as two between-subjects factors.

Last, in order to further examine whether the two groups of participants' brain responses in the two experiments were correlated with one another, Pearson correlation two-tailed tests were conducted for each group between the mean amplitude of vMMN in the face-voice condition of the Oddball task and the amplitude of the N400 difference wave in each of the two conditions of the Stroop task, i.e., face judgment and voice judgment condition, respectively.

3.7. Results

First, the main effect of *deviance* was found to be significant on the mean amplitude in the 100-200ms time window across the selected electrodes for each condition performed by each group (Chinese face-only condition, $F(1, 19) = 7.011$, $p < .01$, $r = 0.519$; Chinese face-voice condition, $F(1, 19) = 15.256$, $p < .01$, $r = 0.667$; Chinese face-tone condition, $F(1, 19) = 6.963$, $p < .01$, $r = 0.518$; English face-only condition, $F(1, 18) = 7.709$, $p < .01$, $r = 0.548$; English face-voice condition, $F(1, 18) = 6.910$, $p < .01$, $r = 0.527$; English face-tone condition, $F(1, 18) = 6.884$, $p < .01$, $r = 0.526$). This means that deviant trials elicited a more negative going ERP amplitude than standard trials, implying that a visual MMN effect was evoked in each experimental condition. No significant effect involving *facial expression of deviants*, *congruence of deviants*, and *tone frequency* was found (*ps* > .37; see Figure 7).

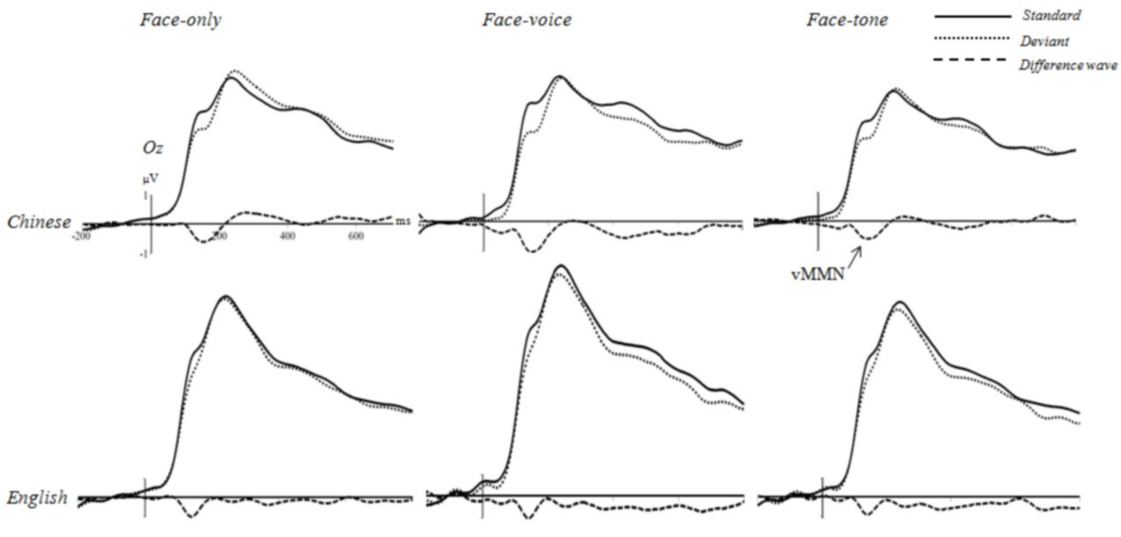
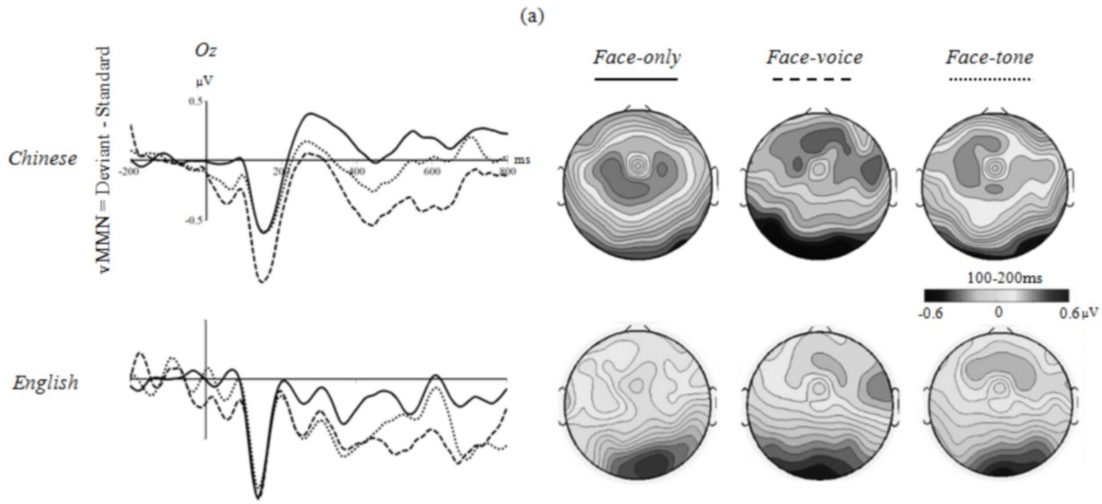


Figure 7. Grand averages elicited by standard trials (solid lines), deviant trials (dotted lines), and the difference wave (dashed lines; deviant - standard) at Oz electrode for each condition of each group.

The following three-way repeated measures ANOVA on vMMN revealed a significant main effect of *Condition* ($F(2, 66) = 32.718$, $p < .01$, $r = 0.576$). Overall, a larger vMMN was observed in the face-voice condition than the other two conditions. More importantly, the interaction of *Condition* and *Group* was significant ($F(2, 66) = 5.263$, $p < .05$, $r = 0.272$). Simple effects analysis specified that the effect of *Condition* was significant in the Chinese group ($F(2, 32) = 5.530$, $p < .05$, $r = 0.384$), which showed a larger vMMN in the face-voice condition than the face-only and face-tone conditions. No such effect was observed in the English group ($p = .32$). The effect of *Group* was significant in face-voice condition, where the Chinese showed a larger vMMN than the English group ($F(2, 32) = 6.297$, $p < .01$, $r = 0.405$; see Figure 8). No effect involving Gender was found (*p*s > .27) [2]. See Table 4 for a summary of the means and standard deviations for the amplitude values of the vMMN component.



[2] We reanalyzed the MMN data with equal number of standards (those preceding the deviants) and deviants and consistent results were found. In the first ANOVA, the main effect of *deviance* was significant for the mean amplitude in the 100-200ms time window across the selected electrodes for each condition of each group (Chinese face-only condition, $F(1, 19) = 6.717$, $p < .01$, $r = 0.511$; Chinese face-voice condition, $F(1, 19) = 13.046$, $p < .01$, $r = 0.638$; Chinese face-tone condition, $F(1, 19) = 6.352$, $p < .01$, $r = 0.501$; English face-only condition, $F(1, 18) = 5.513$, $p < .05$, $r = 0.484$; English face-voice condition, $F(1, 18) = 7.106$, $p < .01$, $r = 0.532$; English face-tone condition, $F(1, 18) = 6.141$, $p < .01$, $r = 0.504$). No significant effects involving *facial expression of deviants*, *congruence of deviants*, or *tone frequency* were found (*ps* > .20). The second ANOVA on vMMN revealed a significant main effect of *Condition* ($F(2, 66) = 28.441$, $p < .01$, $r = 0.549$) and a significant interaction of *Condition* and *Group* ($F(2, 66) = 6.017$, $p < .05$, $r = 0.289$). Simple effects analysis specified that the effect of condition was significant in the Chinese group ($F(2, 32) = 5.976$, $p < .01$, $r = 0.397$), which showed a larger vMMN in the face-voice condition than the face-only and face-tone conditions. No such effect was observed in the English group ($p = .17$). No effect involving Gender was found (*p*s > .10).

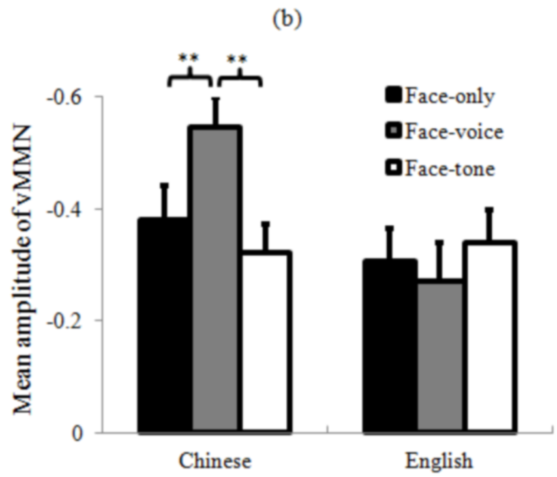


Figure 8. (a) Grand averages at Oz electrode and topographic maps of vMMN for each condition of each group; (b) Mean amplitude values of vMMN averaged across selected electrodes for each condition of each group.

Table 4. Means and standard deviations (in parentheses; μV) of the mean amplitude of vMMN in Experiment 2 for each experimental condition (Std: standard trials; Dvt: deviant trials)

		Chinese	English
Face-only	Std: sad face; Dvt: fear face	-0.37(0.81)	-0.30(0.94)
	Std: fear face; Dvt: sad face	-0.38(0.90)	-0.29(0.88)
Face-voice	Std: sad face & sad voice; Dvt: fear face& sad voice	-0.53(0.97)	-0.29(0.62)
	Std: sad face & fear voice; Dvt: fear face& fear voice	-0.54(1.00)	-0.28(0.64)
	Std: fear face & sad voice; Dvt: sad face & sad voice	-0.53(0.87)	-0.28(0.70)
	Std: fear face & fear voice; Dvt: sad face & fear voice	-0.55(1.02)	-0.28(0.60)
Face-tone	Std: sad face & tone1; Dvt: fear face & tone1	-0.31(0.79)	-0.34(0.78)
	Std: sad face & tone2; Dvt: fear face & tone2	-0.32(0.83)	-0.35(0.91)
	Std: fear face & tone1; Dvt: sad face & tone1	-0.32(0.67)	-0.33(0.86)
	Std: fear face & tone2; Dvt: sad face & tone2	-0.31(0.75)	-0.34(0.81)

Finally, correlation analyses conducted on the ERP data between the two experiments showed that for the Chinese group, vMMN in the face-voice condition was not significantly correlated with the N400 difference wave in either condition of the Stroop task (*p*s > .09). On the other hand, the English group exhibited a trend towards a negative correlation between vMMN in the face-voice condition and the N400 difference in the voice judgement condition of the Stroop task ($r(17) = -.21$, $p = .07$), but not in the face judgement condition ($p = .17$). This result suggests that the English participants who showed larger N400 differences in judging the voice of the Stroop task tended to yield smaller vMMN in the face-voice condition in the Oddball task.

3.8. Discussion

Our ERP results of Experiment 2 provide solid evidence of the effect of culture on an earlier perceptual processing stage of emotion processing for combined facial and vocal expressions. In particular, the Chinese group exhibited a larger vMMN component in the face-voice condition relative to the other two conditions, while no such pattern was witnessed in the English group. This suggests that Chinese participants were more attentive/susceptible than English participants to the influence of concurrent vocal cues during early stages of face processing, consistent with the idea that individuals from East Asian cultures are more sensitive to vocal cues in communication, as early as 100ms after stimulus onset.

As expected, a visual MMN component was observed in each experimental condition for each participant group, indicated by a difference wave between the more negative going potentials in deviant versus standard trials. Given the fact that auditory stimuli in the *face-voice* and *face-tone* conditions were identical across standard and deviant trials, which evoked identical auditory potentials, the MMN in these conditions is considered as a pure visual component (vMMN), where the auditory potentials have been eliminated. The observation of vMMN in all three conditions suggests that participants of both groups detected the changes in facial expressions at an early perceptual stage, even though they were actively looking for the target circles rather than the facial expressions. In addition, this component was not altered by exchanging the standard and deviant trials (i.e., no significant effects involving emotion, congruence, or tone type in the three conditions), suggesting that the MMN effect was not influenced by the specific emotional meanings of the facial and vocal stimuli, or the acoustic properties of the tones (Maekawa et al., 2005; Schirmer, Striano, et al., 2005; Stagg et al., 2004).

Of greater theoretical import, our experiment supplies initial evidence that the cultural background of participants modulates the processing of emotional cues even at early stages of (pre-semantic) stimulus processing. Specifically, individuals in the Chinese group showed larger vMMN amplitudes in the face-voice condition than the face-only and face-tone conditions, whereas no difference was observed between conditions for English participants. The amplitude of the MMN component is thought to reflect the magnitude of discrepancy of the deviant stimulus, inducing a larger MMN when participants detect greater discrepancy (Campbell, Winkler, & Kujala, 2007). Therefore, the larger vMMN of the Chinese group in the face-voice condition suggests that Chinese participants detected larger deviancy in this condition than in the other two conditions. Given the fact that the facial expressions in all conditions were identical, it is inferred that the larger MMN effect in the face-voice condition is due to the presence of concurrent vocal information; in other words, the accompanying voices enhanced the MMN effect for the Chinese group, but not for the English group. These findings support our hypothesis that Chinese participants are more attentive to vocal cues than English participants even at the very early perceptual processing stage indexed by the MMN component. While this expected bias was not observed in the Chinese group in the Stroop task due to limitations of the stimuli, results of Experiment 2 clearly demonstrated the predicted higher sensitivity to vocal cues of the Chinese.

Interestingly, no difference was found in the face-tone condition compared with the face-only condition. This might suggest that this effect is unique to *human* vocal expressions that bear special significance to human communication (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). This is compatible with the species-specific effect on cross-channel cue integration observed in previous literature which, for example, reported that the recognition of emotional body posture was influenced by human vocalizations to a larger extent than by animal sounds (van den Stock et al., 2008). Another possible reason that the face-only and face-tone conditions yielded similar results is that compared to the face-voice condition, the other two conditions are more similar to each other. While the vocal stimuli in the face-voice condition consisted of a variety of different utterances, auditory stimuli in the face-tone condition (i.e., a single pure tone) were identical across trials, which may have allowed participants to easily habituate to these unchanging stimuli in the face-tone condition similar to the face-only condition. Future studies using non-vocal auditory stimuli with similar degrees of variety and complexity to the vocal stimuli (e.g. environmental sound) would help to clarify this question.

In the English group, no difference in vMMN was found between the face-voice condition and the other two conditions, implying that these participants were not influenced by simultaneous vocal cues when they were processing the faces, at least not to the extent that the visual MMN effect was significantly enhanced. In contrast, a previous study reported that an auditory MMN component was induced by discrepant information in concurrent facial cues for Dutch participants (de Gelder et al., 1999). This finding, coupled with our results, implies an asymmetric pattern in Western participants, whereby vocal cues do not automatically influence early facial expression processing indexed by the vMMN in the same way that facial displays automatically modulate the aMMN (i.e., early processing of vocal cues). A similar asymmetry was documented in letter-speech sound processing in Dutch participants, where the aMMN in response to speech sounds was modulated by concurrent visual letters, whereas the vMMN in response to letters was not influenced by concurrent speech sounds (Froyen, van Atteveldt, & Blomert, 2010; Froyen, van Atteveldt, Bonte, & Blomert, 2008). Given the fact that these findings were all observed in participants from the Western culture (English North Americans, Dutch), this asymmetric pattern in the MMN effect, showing that English participants were influenced by faces but not voices, is in keeping with findings of Experiment 1 where the same participants were influenced more by the to-be-ignored faces in the voice task than by the to-be-ignored voices in the face task. Taken together, findings from Experiment 2 underline important distinctions in how individuals from Western and East Asian cultures respond to different elements of human emotional expressions as they are processed in real time, establishing differences in on-line brain responses to these stimuli at both semantic *and* pre-semantic processing stages for the first time.

3. General discussion

The present study investigated cultural differences between a Western group, English-speaking North Americans, and an East Asian group, Mandarin-speaking Chinese, in how they perceive emotions across faces and voices. Our data demonstrate robust differences between the two groups in multi-sensory emotion perception: while processing emotional face-voice pairs, the English group experienced greater interference from facial expressions, whereas Chinese participants were influenced to a relatively greater extent by emotional voices. Importantly, this distinction was observed at both behavioural and neural levels, reflected by behavioural accuracy and ERP components at different processing stages, e.g., the later N400 and early vMMN.

To our knowledge, this is one of the first studies that examined how *Chinese* participants perceive multi-sensory emotions in comparison to individuals from a Western culture. The pattern observed in the Chinese group was consistent with that found when Japanese participants processed multimodal emotional stimuli, where they showed higher sensitivity to vocal cues than both semantic information (Ishii et al., 2003; Kitayama & Ishii, 2002) and facial displays

(Tanaka et al., 2010) in comparison to Westerners. Our results are also in line with a previous observation in non-emotional communication that Japanese speakers use visual cues less than English speakers when interpreting audiovisual speech (Sekiyama & Tohkura, 1991). The consistency between the present study and the relevant literature suggests that the cultural characteristics previously found in the Japanese population (i.e., higher reliance on vocal cues) are shared by the Chinese group, which is another major East Asian collectivist culture that shares certain cultural similarities with the Japanese culture. More importantly, this study for the first time demonstrates the cultural effect on the neural activities underlying emotional processing and communication. The modulatory effect of culture was found not only in behavioural performance, but also in ERP components during both late and earlier processing stages (e.g., N400 and vMMN). In the literature on visual perception, a cultural effect has been reported on the N400 component in visual emotion/object perception (Goto et al., 2010; Goto et al., 2013), and on the MMN component in colour perception (Thierry et al., 2009). The present study provides the first electrophysiological evidence that these two ERP components are also sensitive to culture in the domain of audio-visual emotion perception, and broadens the knowledge of the role of cultural learning on perception and cognition in general.

How culture guides emotion processing and communication

While it is increasingly clear that cultural differences shape the perception and cognitive evaluation of emotional expressions, leading to distinct cultural biases in the uptake of particular social cues, a pertinent question is how to explain these differences and how they shape underlying neurocognitive mechanisms. In the case of emotion communication, as has been mentioned earlier in the introduction section, it has been argued that culture-specific *display rules*, which play a critical role in regulating how emotions are expressed in socially appropriate ways, also play a role in emotion perception (Engelmann & Pogosyan, 2013; Ishii et al., 2003; Park & Huang, 2010). In a specific culture, display rules refer to social norms about when, how, to whom, and to what extent one's feelings should be displayed in various social settings, if at all (Ekman et al., 1969; Engelmann & Pogosyan, 2013; Matsumoto, 1990; Matsumoto et al., 2002; Matsumoto et al., 2008).

In contrast to Western individualist cultures, East Asian collectivist cultures consider the social harmony within a group as more valuable and important than the needs of individuals (Hall & Hall, 1990; Scollon & Scollon, 1995), meaning that culture-specific display rules are adopted to avoid social conflicts and to maintain harmony (Gudykunst et al., 1996; Matsumoto et al., 1998). For instance, research shows that in East Asian cultures, indirectness in communication is more common and socially-acceptable than direct references when compared to Western cultures (Bilbow, 1997). Indirectness is typically expressed by unfinished sentences or vague linguistic expressions (e.g., a little, about, possibly, etc.) to prevent potential conflicts or embarrassment that might threaten the face of an interlocutor, to such an extent that the listener has to make a reasonable inference as to what is intended (McGloin, 1984). It has been reported that Japanese participants tend to mask their negative facial expression (e.g., by putting on a smile) when another person is present, whereas Americans convey similar negative faces irrespective of the presence of another person; this suggests that Western cultures encourage explicit emotional expressions by faces, whereas East Asian cultures endorse the control of facial expressions, especially negative ones, to maintain harmonious relations within a group (Ekman, 1971; Markus & Kitayama, 1991; Matsumoto et al., 1998; Matsumoto et al., 2008; Matsumoto et al., 2005). Moreover, East Asians are more inclined to avoid eye contact during face-to-face interactions than individuals from Western cultures. They perceive faces with direct eye gaze as unapproachable and unpleasant (Akechi et al., 2013; Sue & Sue, 1977) and consider the avoidance of eye contact as a sign of respect and politeness to others, which is beneficial to social harmony (Hawrysh & Zaichkowsky, 1991; McCarthy et al., 2006; McCarthy et al., 2008).

These culture-specific display rules and related procedures for regulating emotion expression in daily life (e.g., amplification/deamplification of emotions, masking, etc.) are learned during critical developmental stages and maintained throughout the lifespan (Ekman et al., 1969; Engelmann & Pogosyan, 2013; Matsumoto, 1990; Matsumoto et al., 2002; Matsumoto et al., 2008). This results in sustained culture-specific learning experiences about emotional expressions, including routine participation in behavioural practices and exposure to particular patterns of emotional expressions of others, which in turn shapes perceptual and attentional biases for extracting and processing emotional information (Nisbett et al., 2001). In East Asian cultures, the more frequent use of incomplete sentences and vague expressions in speech, the constrained facial expressions, and reduced eye contact in comparison with Western cultures, may lead to the fact that East Asians need to rely more on cues from other sources to convey their feelings effectively. Indeed, it has been found that in East Asian communication, the proportion of information conveyed by semantic content is relatively smaller than that in Western cultures, and contextual and nonverbal cues are likely to play a relatively larger role (Ambady, Koo, Lee., & Rosenthal, 1996; Hall, 1976; Markus & Kitayama, 1991). This pattern results in sustained social communication experiences in which both semantic and facial information is relatively less available for East Asians, and the efficiency of sampling emotional information from semantic content or faces is reduced. In order to infer the feelings of the others, East Asian individuals may develop cognitive biases that allow prioritized sampling of relevant information from other available information sources, e.g., the vocal cues, as a strategy for efficient communication.

In contrast, Americans are known to show a bias to infer others' intentions from what is literally being said, while tending to neglect other social constraints on the speaker (Gilbert & Malone, 1995). Thus, cultural specificities in the availability of emotional information from different sources, which to a large extent are determined by display rules, may contribute to people's habitual patterns of perceptual and attentional biases in perceiving emotions. Arguably, our data show that these biases are rooted in specific cultural learning to such a degree that they have an in-depth impact, not only on behaviour but on different levels of brain activity, indexed by N400 and MMN components, that are implicated in the full semantic elaboration of multi-sensory emotion expressions in both the present study and in previous literature (de Gelder et al., 1999; Ishii et al., 2010; Schirmer & Kotz, 2003).

While acquired knowledge about *display rules* may underlie cultural differences in the electrophysiological and behavioural responses of Chinese and English participants tested in this study, an alternative interpretation centres on the linguistic characteristics of the language spoken by each group. Mandarin Chinese is a tone language in which syllables are marked by four different pitch patterns to lexically distinguish words with otherwise similar sound patterns, whereas no such lexical tones exist in the English language. It is therefore possible that Mandarin speakers, who unlike English speakers constantly attend to variations in pitch cues to ascertain word meanings, have developed a higher sensitivity to subtle changes in voice/pitch cues as early as the MMN effect, not only in lexical processing but when vocal cues are encountered in other speech contexts, such as when they convey emotions. In this study, this may be partly responsible for the observed sensitivity to emotional vocal cues when Mandarin speakers processed emotional pseudo-speech. Interestingly, the two other East Asian languages that have been studied previously in this literature, Japanese (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010) and Tagalog (Ishii et al., 2003), are not complex tone languages; rather, they both use "pitch accents"—a simplified use of tonal information to mark certain word meanings (e.g., in Japanese, "ame" with a low-high pitch means "candy", whereas "ame" with a high-low pitch means "rain"). Since the two previously-studied Western languages, English and Dutch, use neither tonal contrasts or 'pitch accents', it is plausible that even a simplified use of tone/pitch accents in East Asian languages promotes a bias to pay greater attention to vocal (emotion) cues, although perhaps to a lesser extent than in Chinese or other languages with complex tone systems.

While both cultural display rules and the tonal structure of a language could each contribute to culture-specific biases in how emotional expressions are processed, our current data cannot disambiguate the contribution of these two factors within the current experimental design. Further data will be needed to disentangle the role of display rules and linguistic competence in the processing of tonal contrasts in a language as possible determinants of the cultural differences we observed. For example, a study employing the same tasks but adopting emotional vocalizations, instead of pseudo-utterances, as vocal stimuli would help determine whether similar cultural differences would be observed when no linguistic information is involved in emotion processing. Another option is to administer our current protocol to a group of participants who come from an East Asian culture that shares many of the display rule features of collectivist cultures, but who speak a language without strong tonal features at the lexical level (e.g., standard Korean); or to a group who come from a Western individualist culture but speak a language with certain features of lexical tones (e.g., Swedish or Norwegian). The extent to which such a group would show a similar culture-specific bias in processing cross-channel emotions in comparison to the Chinese and English North American groups would shed important light on these outstanding issues.

Future directions

The present study supplies significant new evidence of how culture modulates human behaviour and real-time brain responses during emotional communication; in addition, it points to several new directions for future studies. First, due to the strict selection and matching of the facial and vocal stimuli and the limited number of items from the available databases, only two negative emotions, sadness and fear, were included in this study. Future studies adopting other emotional categories would clarify whether the observed cultural differences in this study may be generalized to other discrete emotions, especially positive emotions. Given the different socio-cultural meanings that are often attributed to positive expressions and events (Lu & Gilmour, 2004; Oishi, Graham, Kesebir, & Galinha, 2013; Uchida, Norasakkunkit, & Kitayama, 2004), it would be interesting to explore whether there would be any different culture-specific pattern in processing positive multi-sensory emotions relative to negative ones.

The Chinese participants tested in the present study had just immigrated to Canada or recently arrived in Montreal to attend university. Despite the fact that most of them had fair/good knowledge of English and were being exposed to a Western culture, our results illustrate robust differences between the Chinese and English groups. Nonetheless, future studies that clarify how different types of cultural experiences (e.g., short- vs. long-term cultural immersion, different degrees of acculturation, etc.) shape brain functioning during communication will be greatly beneficial. For instance, it would be valuable to investigate how Chinese individuals who immigrated to North America as young adults, and have been living in the new host culture for an extended time period, perceive multimodal emotional stimuli in comparison to both their native Chinese culture and the host North American culture. A follow-up study involving such a group of Chinese immigrants is currently underway.

In addition to the between-group differences observed in emotion perception, individual differences in certain cultural values (e.g., independence vs. interdependence) have also been associated with cognitive biases in the processing of emotions. For instance, using similar emotional Stroop tasks, participants with higher social orientation scores (i.e., an individual's preparedness to participate in social interactions) displayed a larger N400 effect when identifying the emotion of word meaning while ignoring the spoken tone (Ishii et al., 2010); also, participants with higher interdependent self-construal scores exhibited larger N400 effects when processing focal versus background visual information of an emotional nature (Goto et al., 2010). Future studies examining the relations between individual differences in cultural/social values and the neuro-cognitive patterns in processing facial-vocal emotions would further demonstrate the way the neuro-cognitive system works in accordance with various cultural, social, and individual factors.

Last, while the present study focuses on cultural effects on the *time course* of emotion perception as elucidated by high temporal-resolution ERP measurements, future studies are needed to localize the neural generators involved in these cultural effects with fine spatial resolution. Previous functional neuroimaging evidence suggests that culture affects temporal and parietal lobe functions based on the fact that American participants showed larger activation in these brain areas relative to East Asians when they were viewing focal as opposed to background visual objects (Gutchess, Welsh, Boduroglu, & Park, 2006). Neuroimaging evidence also indicates that culture shapes fronto-parietal functions involved in top-down executive control: compared to culturally-favoured cognitive tasks, culturally-unfavoured tasks, which require greater attentional resources, induced greater activity in this region (Hedden, Ketay, Aron, Markus, & Gabrieli, 2008). It is possible that changes in these brain areas, which are also implicated in multi-sensory emotion processing (Ethofer, Pourtois, & Wildgruber, 2006), contribute to the observed cultural effects on the N400 which possibly involve both semantic processing and conflict control processes. In addition, cultural differences during emotional face processing appear to modulate activation of the amygdala (Adams et al., 2010; Chiao et al., 2008; Derntl et al., 2009; Derntl et al., 2012; Moriguchi et al., 2005), a structure associated with rapid, selective processing of emotion-related information independent of attention and consciousness (Morris et al., 1996; Pessoa, 2010). This suggests that culture-specific patterns affecting *early* stages of emotional information processing, such as the MMN effects we observed in Experiment 2, can also be elucidated by future work that focuses on how functional brain networks are modulated by cultural experiences and related social factors. In particular, a combined EEG/fMRI technique would provide a more comprehensive picture integrating both the temporal and spatial characteristics of neural mechanism underlying cultural differences in emotion perception, and contribute greater insights into the relations between culture, emotion, and the brain.

Funding

This research was supported by an Insight Grant (435-2013-1027) awarded to M.D. Pell from the Social Sciences and Humanities Research Council of Canada, and by a doctoral fellowship (Chinese Government Award for Outstanding Self-financed Students Abroad) awarded to P. Liu.

Preface to Study 2

Study 1 was designed to elaborate on the differences between two cultural groups, Chinese and English North Americans, with respect to how they perceive multi-sensory emotions encoded by faces and voices. To this aim, we investigated how these two groups process emotional face-voice pairs in two ERP experiments. The findings of the first study provided novel insights about cultural differences in multi-sensory emotional communication: while the English North Americans showed higher sensitivity to facial information, the Chinese group was found to be more sensitive to vocal cues. Importantly, these differences were observed at both early and later temporal processing stages as indicated by behavioural accuracy, N400 and vMMN components. Therefore, our results underscore the notion that culture shapes not only how individuals assign significance to emotional stimuli through explicit behavioral judgements, but also the underlying neural responses to multi-sensory emotional stimuli as they are processed and cognitively elaborated in real time.

The observed differences between cultural groups in Study 1 provides further motivation to address the research question posed in Study 2: in addition to cultural origins, does cultural immersion have an impact on how emotional expressions are processed when individuals adopt a new host culture? To investigate this question, a third group, Chinese immigrants to Canada, was tested using identical ERP methods as those in Study 1. The two groups tested in the first study, Chinese and English North Americans, were adopted as control groups in Study 2 to understand the effects of cultural immersion and adaptation on patterns in the data.

3. Study 2

Cultural immersion alters emotion perception:

electrophysiological evidence from Chinese immigrants in North America

Pan Liu[1], *Simon Rigoulot*[1,2], *Marc D. Pell*[1]

[1] *School of Communication Sciences and Disorders, Centre for Research on Brain, Language, and Music, McGill University, Montréal, Canada*

[2] *International Laboratory for Brain, Music and Sound Research (BRAMS), Montréal, Canada*

Corresponding author
Pan Liu
School of Communication Sciences and Disorders, McGill University
2001 McGill College, 8th floor, Montréal, Québec H3A 1G1, Canada
pan.liu@mail.mcgill.ca
1-514 398 4400 ext. 00010

Abstract: To explore how cultural immersion and adaptation modulate human responses to emotional cues, this study tested how a group of Chinese immigrants to Canada process multi-sensory emotional expressions and then compared their performance to existing data from two groups, Chinese and English North Americans. Experiment 1 (Stroop task) presented face-voice pairs expressing congruent or incongruent emotions and participants judged the emotion of one modality while ignoring the other. The behavioural response pattern of Chinese immigrants was similar to that of the English group, with more interference from irrelevant faces than voices, although their neural (N400) response was more similar to the Chinese group. In Experiment 2 (Oddball task), Chinese immigrants passively viewed facial expressions with or without simultaneous vocal emotions; here, they exhibited a larger visual MMN potential for faces accompanied by voices, mirroring patterns observed for Chinese as opposed to English participants. Correlation analyses indicated that the duration of cultural immersion in Canada was associated with neural (N400 *and* vMMN) response patterns resembling those of English participants within the Chinese immigrant group. Our data suggest that in the context of multi-sensory emotion processing, adopting a new host culture first leads to accommodation at the behavioural level followed by alterations in the brain activities of immigrants, supplying new evidence on human's neurocognitive plasticity in (emotional) communication.

Key words: Cultural adaptation, emotion, ERP, facial expression, vocal tone

1. Introduction

As immigration proliferates worldwide and the mobility of migrants increases, the experience of immersing oneself in a new, often distant environment with distinct cultural conventions is increasingly commonplace. In many regions of the world, immigrants constitute a large part of the population in their host culture; for these individuals, there are growing signs that cultural immersion has a significant impact on their psychological and neurocognitive functioning over time, including changes in how socio-emotional cues are processed.

For instance, in a cross-cultural study that used a visual search task composed of facial expressions, Damjanovic et al. (2013) found that Caucasian British participants displayed a search advantage for happy faces, whereas Japanese participants in Japan showed comparable performance for happy versus angry faces. Interestingly, a group of Japanese participants living in the U.K. showed an advantage for happy faces resembling the *British* participants, implying that their behavior had accommodated to the host culture (Damjanovic et al., 2013). In the neuroimaging literature, related studies report that Asian participants show stronger amygdala responses to Caucasian facial expressions than European participants, although these activation patterns seem malleable to the effects of prolonged cultural immersion: Asians who had lived in Europe for a longer period of time showed reduced amygdala responses, a tendency resembling the pattern of the host (European) culture (Derntl et al., 2009; Derntl et al., 2012). Further evidence that individuals who live longer in a foreign culture show more similar cognitive/neurocognitive patterns to their host culture is taken from studies of colour perception

(Athanasopoulos, 2009; Athanasopoulos et al., 2010), object classification (Athanasopoulos, 2007), and length perception (Kitayama et al., 2003). These findings suggest that immersion in a new culture, and the *amount* of cultural exposure, shape and modify certain cognitive processes on both behavioural and neural levels.

While our impressions of the world and other people typically depend on cues from multiple sensory modalities—such as when individuals encounter emotional cues encoded in both a speaker's face and voice—few have investigated the effects of cultural immersion on *multi-sensory emotion processing*, a task critical for effective interpersonal communication in society. Recently, we used event-related potentials (ERPs) to examine neural responses to multi-sensory emotional stimuli at two distinct processing stages, in order to compare these data between two groups of adults with distant cultural origins, Chinese and English North Americans (Liu, Rigoulot, & Pell, under review). Using an emotional Stroop task composed of face-voice pairs, in which participants had to identify the emotional meaning of one modality and ignore the other, we found that English participants showed large interference effects from unattended faces but little interference from unattended vocal cues at the presumed stage of accessing emotional meanings; this pattern characterized both their behavioral judgments *and* N400 response to the stimuli. In sharp contrast, Chinese participants performed in a relatively comparable manner in the two attention conditions for both measures. To examine whether cultural differences would also emerge at an earlier stage of emotional perceptual processing, participants then performed an Oddball task consisting of the same facial-vocal stimuli; here, we found that the Chinese group, but not the English group, exhibited a larger visual mismatch negativity (vMMN) component when processing facial expressions accompanied by emotional voices than when processing faces alone or accompanied by auditory pure-tone stimuli. Together, these results argue that while the English group showed an attentional bias to facial expressions over voices, the Chinese were generally more attuned to vocal information; importantly, these cultural biases were observed not only on the behavioural level (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010) but also in neural activities gauged by N400 and vMMN.

Cross-cultural differences in processing multi-sensory emotions are arguably due to culture-specific display rules (Engelmann & Pogosyan, 2013; Ishii et al., 2003; Liu et al., under review; Park & Huang, 2010) but could also be influenced by tonal features of the languages spoken by particular groups of interest (Liu et al., under review). In East Asian/Chinese collectivist cultures, specific display rules are adopted to avoid conflicts and maintain social harmony, including: indirect speech acts (Bilbow, 1997; McGloin, 1984), less eye contact (Hawrysh & Zaichkowsky, 1991; McCarthy et al., 2006; McCarthy et al., 2008), and restrained facial expressions (Ekman, 1971; Markus & Kitayama, 1991; Matsumoto et al., 1998). These conventions may reduce the availability of linguistic/semantic and visual/facial cues during social perception, yielding a higher reliance on tonal/vocal cues to infer others' intentions in many East Asian cultures such as Chinese (Liu et al., under review) or Japanese (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010). In contrast, in North America, facial and semantic information tends to be more accessible because many Western individualist cultures favour direct speech, more eye contact, and more expressive/less constrained facial expressions, which could promote higher sensitivity to faces than vocal information in Western cultures. The possibility that speakers of a tone language such as Chinese, who constantly attend to fine tonal variations during speech processing, become more sensitive to vocal information in general, must also be considered when explaining these results.

But how does cultural immersion affect the way that immigrants respond to emotional stimuli, and does the neurocognitive apparatus adapt to the host culture in the domain of social communication? Given evidence that English North Americans and Chinese differ in both behavioral and ERP responses to multi-sensory emotional displays, this study tested a group of *Chinese immigrants to Canada* using the same experimental protocols described by Liu et al. (under review), treating data from the Chinese and English groups as anchors for evaluating the effects of cultural immersion on emotion processing. To our knowledge, this is the first study to test how *cultural immersion* shapes operation of the neurocognitive system for perceiving and assigning meaning to complex emotional events that ubiquitously occur in face-to-face social interactions. Based on the literature, we hypothesized that increased exposure to Western display rules and language features (e.g., lack of phonemic tone contrasts in English) would modulate Chinese immigrants' responses to combined face-voice emotional expressions to accommodate conventions of the host culture over time. Specifically, Chinese immigrants should exhibit response patterns that increasingly resemble the English group tested by Liu et al. (under review) when compared to their native culture (Chinese): analyses of behavioral judgements and N400 response in the emotional Stroop task (Experiment 1) should demonstrate an increasing bias for facial information in the immigrant group; and analysis of vMMN in the Oddball task (Experiment 2) should point to reduced interference of vocal cues across conditions when passively viewing emotional faces in the Oddball task. We predicted that patterns reflecting a shift in response from the native (Chinese) to host (English) culture would be associated with the *duration* of exposure to Western culture, although the exact time course of these effects and their relative impact on behavior versus on-line neural activities could not be predicted with any certainty from the literature.

2. Experiment 1: Emotional Stroop task

2.1. Participants

Eighteen Chinese immigrants to Canada (10 female, 8 male) were tested. They were native Mandarin speakers born and raised in China who immigrated to Canada between 10 to 18 years old; they had lived in Canada for *a minimum of five years* by the time of testing and used English and/or French as their second language(s) for at least *20 hours per week*. All participants gave informed consent before entering the study and were financially compensated. In addition, the Chinese (n = 20) and English North American (n = 19) participants tested in our previous study were adopted as control groups for cross-cultural comparisons, with Chinese as the group of *native culture* and English as the group of *host culture*. The immigrants were younger than the other two groups on average ($F(2, 54) = 3.525$, $p < .05$), while the three did not differ in years of education (F(2, 54) = 0.093, *p* = .911; Table 1).

Table 1. Background information of the three cultural groups

	Chinese immigrants	Chinese	English
Years of age (standard deviation)	21.82(2.04)	24.55(2.75)	25(3.91)
Years of Education (standard deviation)	15.29(1.13)	16.45(2.68)	14.18(2.19)
Years of age of arrival in Canada (range)	12.55 (11-18)	N/A	N/A
Years of stay in Canada (range)	9.35 (6-12)	0.82	25
Hours of second language use per week (range)	71 (68-81)	N/A	N/A

2.2. Stimuli

This study employed the same Chinese pseudo-utterances and Chinese facial expressions used by Liu et al. (under review). Eight (2 items × 4 speakers) Chinese pseudo-utterances spoken in sadness or fear were adopted as vocal stimuli (Liu & Pell, 2012), which were cut into 800ms segments from the onset to ensure consistent length across items. Pseudo-utterances were composed of pseudo content words conjoined by real function words, rendering the utterances meaningless but possessing natural segmental/supra-segmental properties of the target language (Banse & Scherer, 1996; Pell, Paulmann, et al., 2009; Rigoulot et al., 2013). Facial stimuli consisted of 6 black-white Chinese faces (1 item × 6 actors) expressing sadness or fear (Beaupré & Hess, 2005). Both vocal and facial stimuli had been validated in our previous study (for details see Liu et al., under review); they all have medium to high recognition rates and emotional intensities, which were matched between modalities (vocal and facial) and emotions (fear and sadness) to control for emotional salience. Faces were synchronized with voices to construct face-voice pairs of 800ms, including both congruent (fearful face & fearful voice, sad face & sad voice) and incongruent (fearful face & sad voice, sad face & fearful voice) trials.

2.3. Task, EEG recording and preprocessing

An *emotional Stroop* task composed of two conditions was administered: *face task* and *voice task* conditions. Each condition consisted of 1 block of 192 trials, with 96 congruent trials (48 fearful face-voice pairs, 48 sad face-voice pairs) and 96 incongruent trials (48 fearful face-sad voice pairs, 48 sad face-fearful voice pairs). Each trial started with a 1000ms fixation cross at the center of the monitor followed by a 500 ms blank screen, then immediately followed the presentation of the face-voice pair. The face was displayed at the center of the monitor (18 × 12.5 cm) and the voice was presented binaurally via headphones. The participants had to respond as quickly and as accurately as possible to identify the emotion of the face (face task condition) or the voice (voice task condition) as conveying fear or sadness while ignoring the other modality. Participants responded by pressing one of two buttons on a response box. The assignment of the labels of response options, *fearful* ("害怕" in Chinese) and *sad* ("伤心" in Chinese), to the right and left buttons was counterbalanced. Following a response (or after 3500ms post-offset of the face-voice stimulus without a response), a variable inter-trial interval of 500-1000ms occurred before the next trial (Figure 1). All trials were presented randomly within each block. The order of the two conditions was counter-balanced among participants, and a 5-10min break was inserted between blocks.

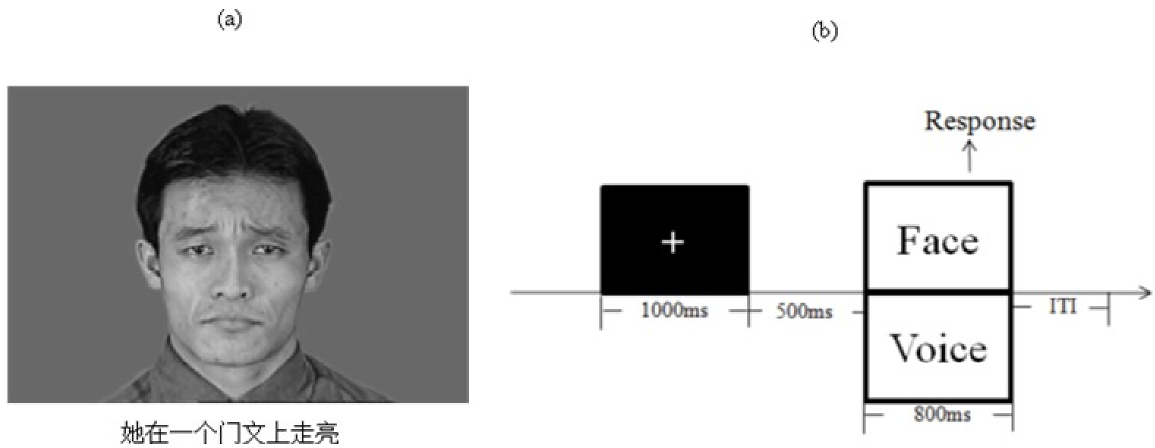


Figure 1. (a) Examples of facial expression and pseudo-sentence in Chinese. (b) Trial procedure of the Stroop task.

EEG signals were recorded during the task from 64 cap-mounted active electrodes (10/10 System) relative to FCz reference (Brain products, ActiCap, Munich) with a sampling rate of 1000 Hz. Bipolar horizontal and vertical electrooculograms (EOG) were recorded for artifact rejection. All electrode impedance was kept below 5 kOhm. The EEG data were off-line resampled to 250 Hz, re-referenced to grand average, and 0.1-30 Hz bandpass filtered. They were then epoched from -200 to 800 ms relative to the facial-vocal stimulus onset with 200ms pre-stimulus baseline, and inspected visually to reject bad channels and trials containing large artifacts. EOG artifacts were rejected by means of ICA decomposition. For further analysis, only trials with correct responses were included. After all, an average of 81% of the data were retained for statistical analysis (face task condition: 79.2%; voice task condition: 82.3%).

2.4. Statistical analyses

Statistical analyses followed an identical protocol to Liu et al. (under review), focusing on both behavioural (accuracy and response time, i.e., RT) and N400 data. Based on our predictions of the Stroop effect and previous N400 literature, we expected to observe N400 differences in the middle fronto-centro-parietal region falling within the 200-600ms time span (see reviews Ganis et al., 1996; Kutas & Federmeier, 2011; Kutas & Van Petten, 1994) and restricted our analyses as such (for a discussion of ERP data quantification, see Handy, 2004). Visual inspection of the averaged ERPs conformed to these expectations. Accordingly, 15 electrodes from this region (FC1, FC2, Cz, C1, C2, C3, C4, CPz, CP1, CP2, CP3, CP4, Pz, P1, P2) were selected, for which an averaged peak latency of the negativities within 200-600ms time window was observed as 335ms across conditions. Based on this observation, which conformed to both previous literature and visual inspection, the 250-450ms time window was selected, from which mean amplitude values were extracted of 15 electrodes for statistical analyses.

Two-way repeated-measures ANOVAs were first conducted on accuracy, RT, and the mean amplitude of 250-450ms across selected electrodes in each condition, with *congruence of the trials* (incongruent and congruent) and *emotion of the attended channel* (fear and sadness) as 2 within-subject factors. Second, behavioural and neural indices of the Stroop/congruence effect were calculated as *difference accuracy*, *difference RT*, and *difference wave of N400* (always incongruent – congruent for each measure). Higher difference accuracy, higher difference RT, and larger difference wave indicate a larger congruence effect.

To specify the impact of *cultural immersion* on the Stroop effect, the data of the Chinese and

English groups from our previous study were included with data for the immigrants and entered into separate two-way repeated measures ANCOVAs conducted on *difference accuracy*, *difference RT*, and mean amplitude of *difference N400*. *Task* (face task and voice task) served as the within-subject factor while *Group* (Chinese immigrants, Chinese, English) served as the between-subject factor, with *Age* as the covariate.

Last, to clarify the relations between the immigrants' cultural immersion experiences and their responses to multi-sensory emotions, correlation analyses were conducted in each condition between each of the two background factors, *length of stay in Canada* and *age of arrival in Canada*, and *difference accuracy*, *difference RT*, and *the opposite number of the amplitude value of difference N400*[3], respectively.

2.7. Results

Behavioural results

The first ANOVAs focusing on Chinese immigrants showed a significant main effect of *congruence* in each task condition for both accuracy (face task, $F(1,17) = 8.796$, $p < .01$, r[4] = 0.584; voice task, $F(1,17) = 6.831$, $p < .01$, $r = 0.535$) and RTs (face task, $F(1,17) = 5.014$, $p < .05$, $r = 0.477$; voice task, $F(1,17) = 6.478$, $p < .01$, $r = 0.525$). Overall, lower accuracy and

[3] The opposite number of the negative amplitude values of difference N400 in Experiment 1(and vMMN in Experiment 2), instead of the original values, was used in the correlation analyses as a more direct indication of the size of the amplitude of the component (i.e., the larger the opposite number, the larger the size of the component).

[4] Effect size, calculated as $r = \sqrt{F/(F + df_{error})}$. (Rosnow & Rosenthal, 2003)

longer RTs were observed in response to emotionally incongruent face-voice pairs than to congruent ones. No significant effect involving the *emotion of the attended channel* was found (*ps* >.075).

The following ANCOVA on data for the three groups revealed a significant interaction between *Task* and *Group* on difference accuracy (F(2, 53)=5.204, $p < .05$, $r = 0.304$), but not on difference RT ($p = .24$). Simple effect analysis of difference accuracy indicated that the Chinese group showed no difference between the two task conditions ($p = .11$); however, both the English group and the Chinese immigrants exhibited a significant *Task* effect (F(2, 53) = 10.218, $p < .01$, $r = 0.402$; F(2, 53) = 6.776, $p < .01$, $r = 0.337$) indicating that they had a larger difference accuracy in the voice task than in the face task.

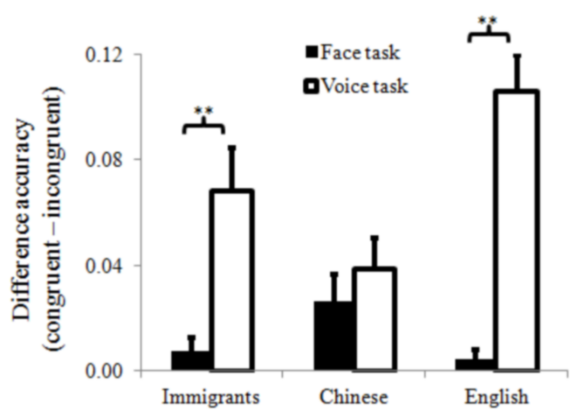


Figure 2. Difference accuracy in the voice and face task for each cultural group.

ERP results

A significant main effect of *congruence* on the mean amplitude of 250-450ms (i.e., N400) across selected electrodes was found in each condition for the Chinese immigrants (face task, F(1,17) = 7.474, $p < .01$, $r = 0.553$; voice task, F(1,17) = 8.913, $p < .01$, $r = 0.586$). More negative going amplitudes were observed in response to emotionally incongruent face-voice pairs relative to congruent pairs, demonstrating a typical Stroop/congruence effect on the neural level. Again, no significant effect involving the *emotion of the attended channel* was found (*ps* >.20; Figure 3).

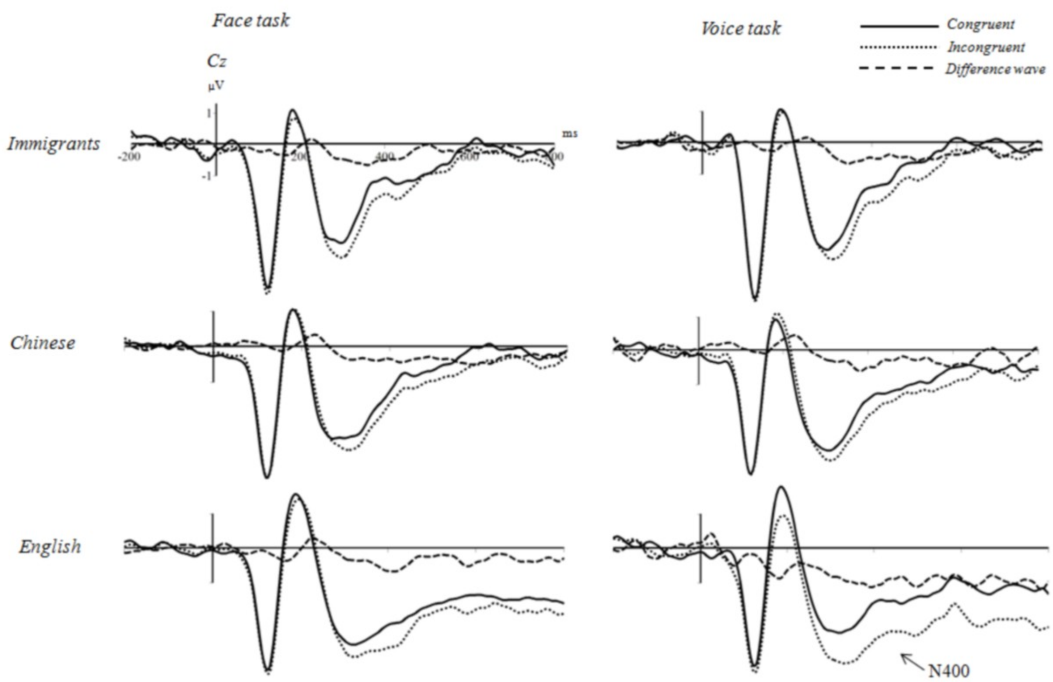


Figure 3. Grand averages elicited by congruent trials (solid lines), incongruent trials (dotted lines), and the difference wave (dashed lines; incongruent - congruent) at Cz electrode for each condition of each group.

The following analysis on difference N400 including the two control groups revealed a significant interaction between *Task* and *Group* (F(2, 53) = 8.340, $p < .01$, $r = 0.369$). Step-down analysis indicated that while no effect of *Task* was observed in the Chinese (p = .13) and Chinese immigrants (p = .20), significant between-task difference was revealed in the English group (F(2,53) = 9.779, $p < .01$, r = 0.395) who showed a larger difference N400 in the voice task than in the face task (Figure 4). See Table 2 for a summary of the means and standard deviations for the behavioural and N400 data for each of the three groups.

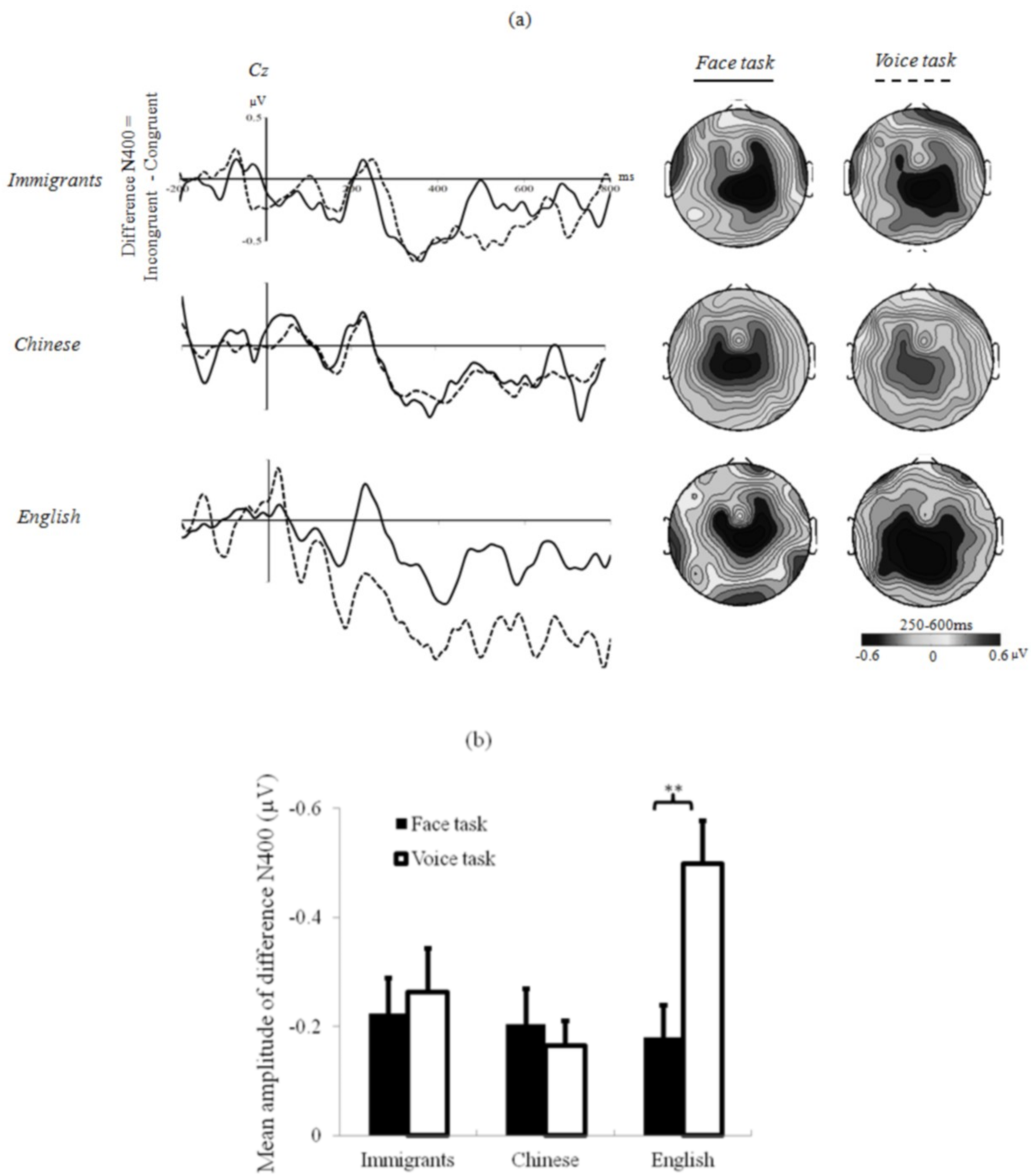


Figure 4. (a) Grand averages at Cz electrode and topographic maps of difference N400 for each condition of each group. (b) Mean amplitude values of difference N400 averaged across selected electrodes.

Table 2. Means and standard deviations (in parentheses) of behavioural accuracy, response time (RT), and mean amplitude of N400 for each experimental condition of each group

Group	Task	Condition	Emotion(face-voice)	Accuracy	RT (ms)	N400 (μV)
Immigrants	Face task	Congruent	Sad-sad	0.93(0.06)	725(180)	-1.68(0.88)
			Fear-fear	0.93(0.05)	728(195)	-1.67(1.01)
		Incongruent	Sad-fear	0.89(0.06)	745(175)	-1.90(1.19)
			Fear-sad	0.90(0.05)	750(189)	-1.88(1.09)
	Voice task	Congruent	Sad-sad	0.89(0.09)	780(285)	-1.67(1.50)
			Fear-fear	0.91(0.10)	777(272)	-1.69(1.71)
		Incongruent	Sad-fear	0.79(0.09)	803(291)	-1.91(1.83)
			Fear-sad	0.79(0.08)	797(289)	-1.89(1.79)
Chinese	Face task	Congruent	Sad-sad	0.92(0.05)	745(263)	-1.65(1.50)
			Fear-fear	0.91(0.04)	755(273)	-1.64(1.23)
		Incongruent	Sad-fear	0.89(0.09)	776(313)	-1.86(1.61)
			Fear-sad	0.88(0.06)	770(256)	-1.84(1.89)
	Voice task	Congruent	Sad-sad	0.87(0.08)	784(354)	-1.71(1.72)
			Fear-fear	0.87(0.05)	780(332)	-1.69(1.56)
		Incongruent	Sad-fear	0.84(0.08)	810(183)	-1.87(1.92)
			Fear-sad	0.83(0.07)	802(347)	-1.89(1.58)
English	Face task	Congruent	Sad-sad	0.94(0.05)	699(184)	-1.68(0.98)
			Fear-fear	0.93(0.07)	700(215)	-1.68(1.21)
		Incongruent	Sad-fear	0.91(0.05)	718(165)	-1.88(1.30)
			Fear-sad	0.90(0.06)	715(268)	-1.87(1.29)
	Voice task	Congruent	Sad-sad	0.91(0.08)	760(305)	-1.66(1.47)
			Fear-fear	0.90(0.07)	757(271)	-1.64(1.61)
		Incongruent	Sad-fear	0.80(0.13)	781(318)	-2.15(2.00)
			Fear-sad	0.79(0.09)	782(298)	-2.09(1.97)

Correlation results

Correlation analyses on data for the immigrants revealed a positive correlation between length of stay in Canada and difference accuracy (r(16) = .356, $p < .05$) and a positive trend between length of stay in Canada and the size of the amplitude of difference N400 (r(16) = .277, $p = .06$) in the voice task only (Figure 5), not the face task (*ps* > .21). No significant correlation involved difference RT (*p*s > .10) or age of arrival in Canada (*p*s > .13) was found.

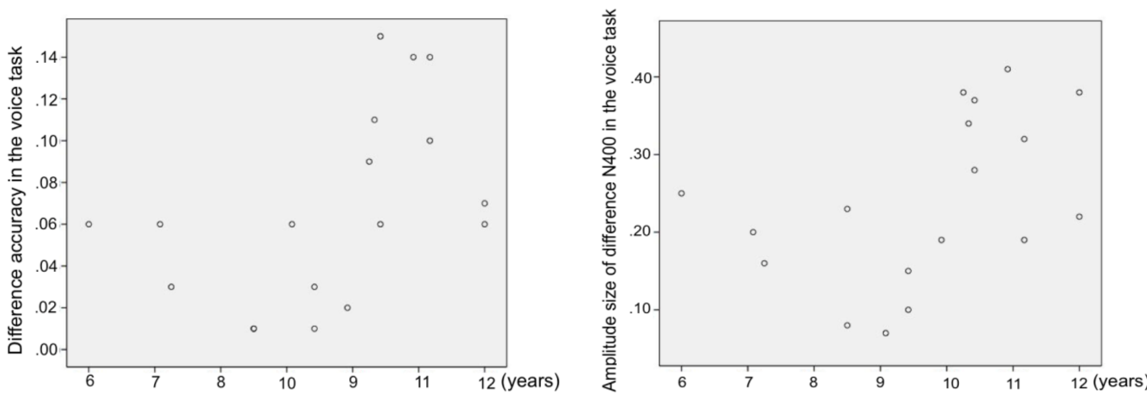


Figure 5. Scatter plots of the difference accuracy (left) and the size of difference N400 (right) of the immigrants in the voice task condition in relation to their years of stay in Canada

3. Experiment 2: Oddball task

3.1. Participants

The same group of Chinese immigrants tested in Experiment 1 participated in Experiment 2.

3.2. Stimuli

The same vocal and facial stimuli as those in Experiment 1 were employed in the Oddball task. Additionally, 2 pure tones lasting 800ms were included as non-vocal stimuli, the frequencies of which were determined by the mean f0 of the fearful (267 Hz) and sad (249 Hz) vocal stimuli.

3.3. Task, EEG recording and preprocessing

An *Oddball* task was employed to investigate the early perceptual processing of multi-sensory emotions. Three conditions were included: Condition 1 was the *face-only condition* without any auditory stimuli where fearful and sad faces served as standard or deviant trials; Condition 2 was the *face-voice condition*, in which the same standard and deviant faces as in Condition 1 were paired with fearful or sad voices; Condition 3 is the *face-tone condition*, where the same faces were paired with one of the two *pure tones*. Each condition consisted of 4 blocks with 600 trials each, including 480 standard trials (80%), 60 deviant trials (10%), and 60 randomly-inserted circle targets (10%; Table 3). Overall, the visual stimuli, the number and proportion of each type of trials were identical across all three conditions, while the differences lied only in the accompanying auditory stimuli.

Table 3. Details of standard and deviant trials in each bock of each condition in Oddball task

	Condition 1 Face-only	Condition 2 Face-voice	Condition 3 Face-tone
Block 1	Standard: sad face Deviant: fearful face	Standard: sad face & sad voice Deviant: fearful face& sad voice	Standard: sad face & tone1 Deviant: fearful face & tone1
Block 2	Standard: sad face Deviant: fearful face	Standard: sad face & fearful voice Deviant: fearful face& fearful voice	Standard: sad face & tone2 Deviant: fearful face & tone2
Block 3	Standard: fearful face Deviant: sad face	Standard: fearful face & sad voice Deviant: sad face & sad voice	Standard: fearful face & tone1 Deviant: sad face & tone1
Block 4	Standard: fearful face Deviant: sad face	Standard: fearful face & fearful voice Deviant: sad face & fearful voice	Standard: fearful face & tone2 Deviant: sad face & tone2

In all three conditions, each block started with a 1000ms fixation cross at the center of the monitor, followed by the trials which were presented pseudo-randomly so that two deviants never appeared in succession, and at least three standards appeared in a row between two deviants. Each trial was presented for 800ms, and the variable inter-trial interval was 500-1000ms. The visual stimuli were presented at the center of the monitor and the auditory stimuli were presented binaurally via headphones. In all conditions, the participants were instructed to detect the circle targets among the faces by pressing the spacebar (to ensure that they were viewing the faces passively); in Condition 2 and 3, it was further emphasized that they ignored the concurrent auditory stimuli (Figure 6). Both the order of the four blocks within each condition and the order of the three conditions were counter-balanced among participants, and a 5-10min break was inserted between blocks. As explained by Liu et al. (under review), the Oddball task was always completed by each participant in a single session *preceding* the administration of the Stroop task (which occurred at least one day later); this was to avoid any potential biasing effects of focusing on emotional meanings of the stimuli, or conscious attention to the match/mismatch of emotions, in the Stroop task on responses in the Oddball task where participants were instructed to ignore the vocal/auditory channel completely and focus only on the facial channel.

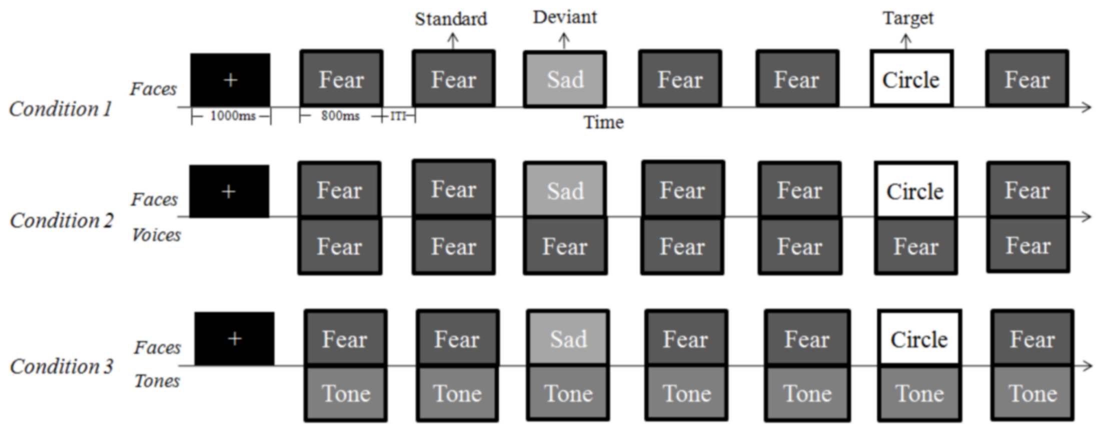


Figure 6. Task procedures for each of the three conditions in the Oddball task. From top to bottom: face-only, face-voice, face-tone.

EEG data were collected during the task. The same procedures of EEG recording and preprocessing were followed as those in Experiment 1. For further analyses, all target trials and standard trials that immediately followed deviants were excluded. After artifact rejection, an average of 79% of the data were retained for statistical analyses (face-only condition: 81.0%; face-voice condition: 77.7%; face-tone condition: 79.6%).

3.4. Statistical analyses

Based on our hypothesis of visual MMN, our analyses focused on this component which was expected to be maximal in the occipital-parietal area across 100-300ms time span post-stimulus according to previous literature (Athanasopoulos et al., 2010; Schröger, 1998; Thierry et al., 2009; Wei et al., 2002; Zhao & Li, 2006). Visual inspection of the averaged ERPs revealed more negative-going deflections elicited by deviant trials relative to standard ones during the 100-200ms time window in the *parieto-occipital* region in each condition, confirming expectations.

Accordingly, 14 electrodes were selected from this region (POz, PO3, PO4, PO7, PO8, PO9, PO10, Oz, O1, O2, P5, P6, P7, P8), for which an averaged peak latency of the difference wave between the deviant and standard trials (deviant - standard) was observed at 152ms across conditions. Therefore, the 100-200ms time window was selected for analysis from which mean amplitude values were extracted for the 14 electrodes.

Repeated-measures ANOVAs were first conducted on the mean amplitude of 100-200ms across selected electrodes in each condition. In the face-only condition, *deviance* (standard and deviant) and *facial expression of deviants* (fear and sadness) served as within-subject factors; in the face-voice condition, *deviance* (standard and deviant), *facial expression of deviants* (fear and sadness), and *congruence of deviants* (congruent and incongruent) served as within-subject factors; in the face-tone condition, *deviance* (standard and deviant), *facial expression of deviants* (fear and sadness), and *tone* (frequency 1 and frequency 2) served as within-subject factors.

Next, *difference waves* were obtained by subtracting the mean amplitude of 100-200ms in the standard trials from that in the deviant trials in each condition. In the *face-only* condition, the difference wave was considered as a visual MMN elicited by deviant relative to standard faces. In the *face-voice* and *face-tone* conditions, while faces varied as deviants and standards, the auditory stimuli were identical across all trials that evoked the same auditory potentials. Therefore, by subtracting the amplitude of the standards from the deviants, the auditory potentials were eliminated and the obtained difference wave can be considered a visually-evoked MMN component.

To examine how MMN was modulated by cultural immersion, the data of Chinese and English groups from the previous study were compared to the immigrants. A two-way repeated-measures ANCOVA was conducted on the mean amplitude of vMMN across selected electrodes, with *Condition* (face-only, face-voice, and face-tone) as the within-subject factor, *Group* (Chinese immigrants, Chinese, and English) as the between-subject factor, and *Age* as the covariate.

Finally, correlation analysis was conducted on the immigrants with factors of *length of stay in Canada* and *age of arrival in Canada,* and the opposite number of the amplitude value of vMMN in each condition, to demonstrate potential relations between extent of cultural experiences and early brain responses to multi-sensory emotions.

3.7. Results

The main effect of *deviance* was significant for the mean amplitude in the 100-200ms time window across selected electrodes for each condition in the immigrant group (*face-only* condition, $F(1, 17)= 6.866$, $p < .01$, $r = 0.536$; *face-voice* condition, $F(1, 17)= 6.269$, $p < .01$, $r = 0.519$; *face-tone* condition, $F(1, 17)= 5.975$, $p < .01$, $r = 0.510$). The deviant trials elicited more negative going amplitudes than the standards, suggesting a visual MMN effect evoked in each condition in Chinese immigrants, consistent with our previous observations in the two control groups. No significant effect involving *facial expression of deviants*, *congruence of deviants*, and *tone* was found (*p*s > .27; Figure 7).

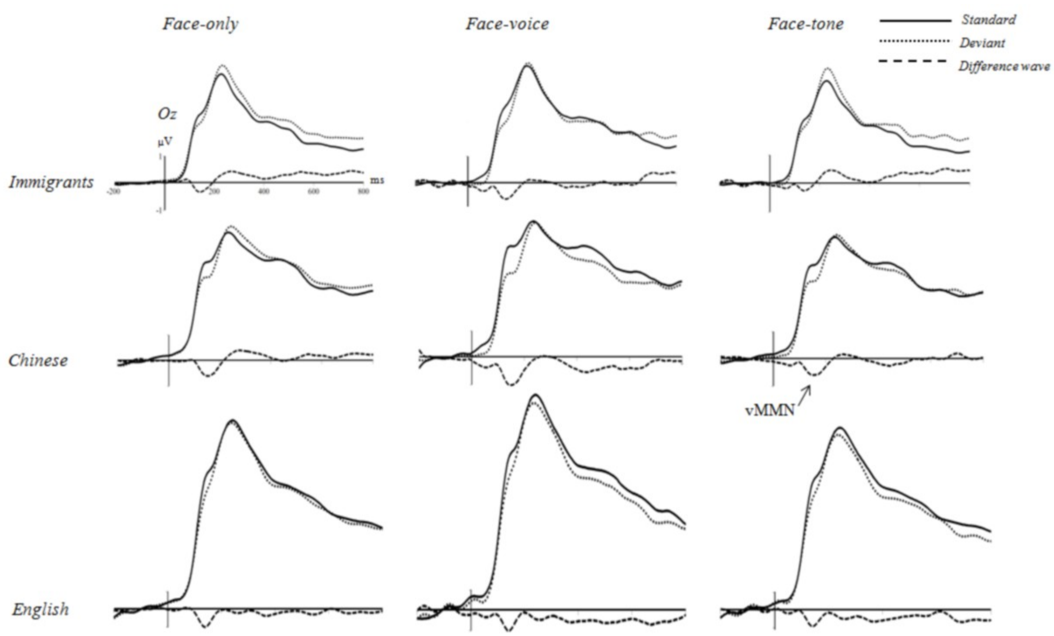


Figure 7. Grand averages elicited by standard trials (solid lines), deviant trials (dotted lines), and the difference wave (dashed lines; deviant - standard) at Oz electrode for each condition of each group.

The following ANCOVA on vMMN for the three groups revealed a significant effect of *Condition* (F(2, 106) = 14.869, p < .01, r = 0.351): overall, a larger vMMN was observed in the face-voice condition than the face-only and face-tone conditions. Importantly, a significant interaction of *Condition* and *Group* (F(4, 106)= 7.226, p < .01, r = 0.253) was found. Simple effect analysis specified that the effect of *Condition* was significant in the Chinese (F(2, 52) = 7.939, p < .01, r = 0.364) and Chinese immigrants (F(2, 52) = 8.017, p < .01, r = 0.365), who showed larger vMMNs in the face-voice condition than the other two conditions; however, no such effect was noted in the English group (p = .17; Figure 8). See Table 4 for a summary of the means and standard deviations for the amplitude values of the vMMN component for the three groups.

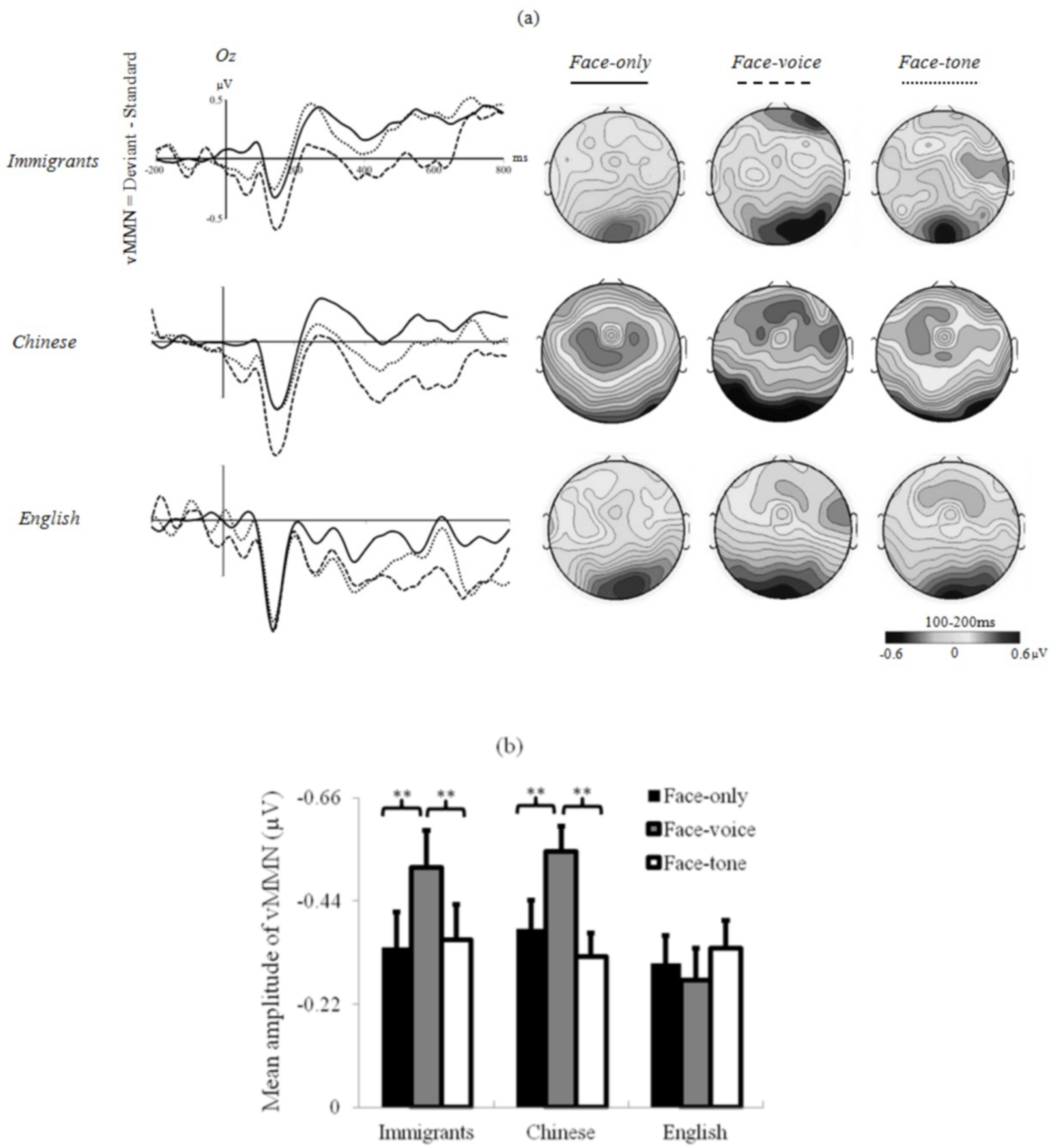


Figure 8. (a) Grand averages at Oz electrode and topographic maps of vMMN for each condition of each group; (b) Mean amplitude values of vMMN averaged across selected electrodes for each condition of each group.

Table 4. Means and standard deviations (in parentheses; µV) of the mean amplitude of vMMN for

each experimental condition of each group (Std: standard trials; Dvt: deviant trials)

		Immigrants	Chinese	English
Face-only	Std: sad face; Dvt: fear face	-0.39(0.74)	-0.37(0.81)	-0.30(0.94)
	Std: fear face; Dvt: sad face	-0.37(0.80)	-0.38(0.90)	-0.29(0.88)
Face-voice	Std: sad face & sad voice; Dvt: fear face& sad voice	-0.55(0.79)	-0.53(0.97)	-0.29(0.62)
	Std: sad face & fear voice; Dvt: fear face& fear voice	-0.52(0.68)	-0.54(1.00)	-0.28(0.64)
	Std: fear face & sad voice; Dvt: sad face & sad voice	-0.55(0.71)	-0.53(0.87)	-0.28(0.70)
	Std: fear face & fear voice; Dvt: sad face & fear voice	-0.54(0.70)	-0.55(1.02)	-0.28(0.60)
Face-tone	Std: sad face & tone1; Dvt: fear face & tone1	-0.36(0.79)	-0.31(0.79)	-0.34(0.78)
	Std: sad face & tone2; Dvt: fear face & tone2	-0.37(0.89)	-0.32(0.83)	-0.35(0.91)
	Std: fear face & tone1; Dvt: sad face & tone1	-0.38(0.76)	-0.32(0.67)	-0.33(0.86)
	Std: fear face & tone2; Dvt: sad face & tone2	-0.37(0.75)	-0.31(0.75)	-0.34(0.81)

Correlation results

Correlation analysis on the immigrants revealed a significant negative correlation between the length of stay in Canada and the size of the amplitude of vMMN in the face-voice condition ($r(16) = -.321$, $p = .05$; Figure 9). No significant correlations were observed in the face-only and face-tone conditions (*p*s >.17). No significant correlation involved age of arrival in Canada was found (*p*s >.15).

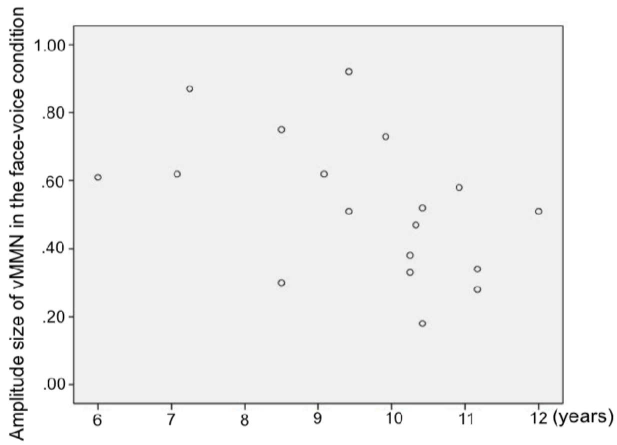


Figure 9. Scatter plot of the size of vMMN of the immigrants in the face-voice condition in relation to their years of stay in Canada

4. Discussion

This study investigated the effect of cultural immersion on multi-sensory emotion processing by testing a group of Chinese immigrants to Canada, in comparison with their native and host cultural groups. Overall, the Chinese immigrants exhibited a behavioural pattern resembling the English group, whereas their neural responses in N400 and vMMN appeared to be more analogous to the Chinese group.

Consistent with our previous observations in the two control groups (Liu et al., under review) the immigrants showed a significant *Stroop/congruence effect* in both voice judging and face judging conditions of the Stroop task, indicated by worse behavioural performance with increased N400 amplitude in incongruent relative to congruent trials. Moreover, a deviance/MMN effect was found in each condition of the Oddball task, indexed by more negative going early potentials in response to deviant relative to standard faces. These results suggest that multi-sensory emotion integration can be detected at different neural processing stages in addition to behavioural performance; moreover, this integration seems independent of the emotional meanings of the stimuli as no effect of *emotion* was found in each task.

Importantly, when comparing Chinese immigrants to the two control groups, our accuracy results for the Stroop task reveal a significant impact of cultural immersion, whereby the immigrants showed a similar behavioural pattern to the English (host) culture but not the Chinese (native) culture. A larger congruence effect in the voice versus face condition was observed in both the immigrants and English participants: their behavioural performance was interfered with more by to-be-ignored faces when identifying vocal emotions than by to-be-ignored voices when judging facial expressions, indicating a higher susceptibility to facial cues during information processing by both groups. In contrast, the Chinese group showed no bias between conditions. These results indicate that the immigrants' behaviour in perceiving multi-sensory emotions was modulated by living in North America and acquired characteristics resembling their host culture, consistent with our expectations based on previous behavioural literature (Athanasopoulos, 2007, 2009; Damjanovic et al., 2013; Kitayama et al., 2003).

However, this effect was not observed in the ERP data as predicted: the immigrants exhibited analogous patterns in N400 and vMMN to the native Chinese group rather than the English. In the Stroop task, whereas a larger difference N400 was observed in the English group in recognizing the voices than the faces, the immigrants and the Chinese did not exhibit distinct

N400 patterns between conditions. In the Oddball task, both immigrants and Chinese participants showed facilitated vMMN by accompanying voices in the face-voice condition than the face-only and face-tone conditions, which was not witnessed for the English group. This suggests that while the immigrants showed similar behavioral tendencies to their host culture, they retained characteristics of their native culture, i.e., higher sensitivity to vocal emotions, during on-line neural processing of emotions at stages of early perceptual processing and retrieval of emotional meanings (vMMN and N400).

It is noteworthy that there was considerable variability among the 18 Chinese immigrants tested in their cultural experiences (e.g., their length of stay in Canada varies from 6 to 12 years), which might be related to the absence of the expected shifting effect in the ERP data. Correlation analyses conducted between the participants' background information and their responses in the two tasks showed that a longer stay in Canada was significantly associated with a larger *difference accuracy* and *difference N400* in the voice judgment of the Stroop task, and smaller vMMN in the face-voice condition of the Oddball task. A larger difference accuracy and difference N400 when judging vocal emotions reflects greater interference by to-be-ignored facial expressions in the Stroop task, while a smaller vMMN in the face-voice condition of the Oddball paradigm implies less sensitivity and enhancement of an accompanying voice on face processing; both of these patterns are characteristic of the English group tested by (Liu et al., under review). Given our evidence that these patterns were associated with the *degree* of cultural immersion (duration of stay), it can be inferred here that: (1) while the immigrants displayed similar behaviours to the English group for accuracy in Experiment 1, those who had been in Canada longer showed an even stronger tendency to resemble their host culture; and (2) while immigrants as a group demonstrated similar neural responses to members of their *native* culture, those who had been in Canada longer might have begun to acquire certain neural response characteristics similar to their host culture (Athanasopoulos et al., 2010; Derntl et al., 2009; Derntl et al., 2012). In other words, while an overall cultural immersion effect was not observed at the neural level in our study, our results allow speculation that this occurs over time as the amount of experience with the host culture increases. Interestingly, no correlation with *age of arrival in Canada* was found, suggesting that for these immigrants who moved to Canada as teenagers, it is the *amount* of exposure to the new culture that plays a role in their adaptation process, not the precise time when this began. When put together, our results imply that the effect of cultural immersion is a dynamic process of transition that occurs over time: behavioural adjustment takes place first, followed by adaptation at the neural level.

As mentioned earlier, culture-specific display rules and/or tonal characteristics of a target language may play important roles in determining the observed differences between Chinese and English groups in processing multi-sensory emotions (Liu et al., under review). For the Chinese group, East Asian display rules which aim to avoid conflicts and maintain harmony (e.g., indirect speech acts, less eye contact, and restrained facial expressions), and possibly linguistic features of the Chinese language which has a complex tone structure, may be responsible for promoting cognitive biases to vocal cues versus facial (and semantic) cues. As native Mandarin speakers, the Chinese immigrants in this study had been exposed since birth to Chinese display rules that seem to favour vocal cues in communication and reduced attention to facial cues (at least when compared to Western participants). After immigrating to Canada as teenagers, however, these immigrants would be exposed to distinct culture-specific forms of emotion expression (Elfenbein & Ambady, 2003) and different attitudes about the implicit value of particular social cues, as encoded in speech or through facial expressions, that often conflict with their early experiences in China and lead to miscommunications in the new host culture (Elfenbein, 2013; Ward, Bochner, & Furnham, 2001; Ward & Kennedy, 1993). To avoid misunderstandings and foster positive social impressions, it would therefore be necessary to adapt to patterns of communication and information processing of the host environment, for example by allocating more attention to the significance of facial versus vocal cues during multi-sensory emotion processing. Since the immigrants tested here all used English and/or French as their dominant language(s) in school, work and social interactions (an average of 71 hours/week), and only used Chinese as a secondary language with family members (Xie, 2011), it is reasonable to assume that levels of exposure to the host culture *and* relative patterns of usage of the native versus host language each play a role in predicting the extent to which individuals adopt behavioural and neural response patterns characteristic of the host culture, pending new data to test these claims.

Our suggestion that the cultural adaptation process begins at the behavioural level, followed by neural reorganization as experience with the host culture increases, begs new studies to test this idea with different cultural groups and in different cognitive domains. Here, in the context of cultural differences in the perceived importance of communicative cues to emotion, it might be easier for immigrants to first accommodate their behaviours by adopting acculturation strategies (Berry, 1990) that mimic the host culture (e.g., intentionally allocating more attention to faces); however, with increased exposure this may lead to a "re-wiring" of underlying neural circuitry to interact effectively with the host environment. Elucidating the time course of cultural adaptation processes on behavioural and neural reorganization awaits further study, with close attention to not only the amount of exposure that drives these processes but the *types* of cultural experiences of different immigrant groups and their relationship to the host culture. Longitudinal studies with several time points would also be useful to determine whether there is a discernible threshold where immigrants begin to shift to the pattern of their host culture, on both behavioural and neural levels.

Funding

This work was supported by an Insight Grant (435-2013-1027) awarded to M.D. Pell from the Social Sciences and Humanities Research Council of Canada, and by a doctoral fellowship (Chinese Government Award for Outstanding Self-financed Students Abroad) awarded to P. Liu.

4. General discussion

The investigation of cultural differences that mark social behavior, cognition, and communication has become an appealing research question in psychology and relevant disciplines over the past decades. Although the number of cross-cultural studies in this area is increasing, the exact role of culture in emotion perception is still not fully understood, especially with respect to how cultural factors affect the neurocognitive system dedicated to emotion processing. The purpose of this thesis was to provide insights into this question; in particular, experiments were designed to shed light on how two different kinds of cultural experiences, cultural origin and cultural immersion, affect the neural temporal mechanisms underlying emotion perception for multimodal stimuli composed of facial expressions and emotional prosody. The first study explored the effect of cultural origin on emotion processing by comparing Chinese and English Canadians, establishing that there are important differences in both behavioural and electrophysiological responses to multimodal emotions in these two groups. The second study used identical methods to investigate the effect of cultural immersion on emotion processing by testing a group of Chinese immigrants to Canada, comparing these data to those gathered for Chinese and Canadian participants in Study 1; in this study, there were new indications that cultural exposure leads to progressive changes in how Chinese immigrants process emotional stimuli, yielding changes that more closely resemble the host (English North American) culture, affecting behaviorial, and over time, neural response tendencies. The broader implications of these findings are considered below.

Effect of cultural origin: Study 1

By examining two cultural groups, Chinese and English North American, Study 1 demonstrated the effects of *cultural origin* on different temporal stages of multi-sensory emotion processing in the event-related brain potential. Cultural differences at the behavioural level have been previously reported between Japanese and Western cultural groups in perceiving emotions from multiple information channels; these data indicated that while Japanese participants were more sensitive to emotional prosodic cues, Western participants appeared to be more attuned to semantic and facial information (Ishii et al., 2003; Kitayama & Ishii, 2002; Tanaka et al., 2010). Building constructively on this literature, the first study of this thesis tested whether the previous findings on the Japanese population could be replicated in another major East Asian culture, the Chinese group. More importantly, this study also aimed to investigate whether cultural differences exist at different on-line neural processing stages beyond the behavioural level by employing EEG measurements.

The results of the English group in Study 1 first demonstrated a larger Stroop (congruence) effect in the voice task than the face task in Experiment 1, in both behavioural judgements and N400 responses evoked by the stimuli. This suggests that the English participants were distracted to a larger extent by the to-be-ignored faces in the voice task, than by the to-be-ignored voices in the face task. Meanwhile, in the Oddball task of Experiment 2, data suggest that the English participants were not influenced by accompanying vocal cues while they were passively processing facial expressions. This pattern is consistent with the hypothesis that Westerners, who represent an individualistic culture that encourages direct, explicit emotional expressivity, are likely more attuned to visual/facial cues in communicating emotions than paralinguistic prosodic cues. In contrast, the Chinese group responded in comparable ways in the face and voice conditions of the Stroop task. This is somewhat inconsistent with our expectation that the Chinese participants are expected to show higher susceptibility to paralinguistic vocal cues than facial information as the Japanese population reported in the literature (Tanaka et al., 2010), which might be due to the lower familiarity of the pseudo-utterances as vocal materials than facial expressions. Nonetheless, when compared with the English group, the Chinese exhibited lower sensitivity to facial information in the voice task than the English in both behavioural and N400 data, suggesting that they are less sensitive to facial cues than Westerners. In addition, the results of the Oddball task in Experiment 2 are more in keeping with the idea that Chinese participants are more sensitive to vocal cues: they were influenced to a significantly greater extent than Westerners by simultaneous to-be-ignored vocal information when they were passively processing facial expressions.

These results support our hypothesis of cultural differences characterize emotion processing: Western cultures favour facial displays as a source for communicating emotions to a larger extent than East Asian cultures, while East Asians favour emotional prosodic cues more. More importantly, these differences were not only observed in the behavioural data but also in participants' neurophysiological responses: the behavioural results serve to replicate and extend previous data on Japanese in the literature, suggesting that some cultural/social characteristics are shared by these two groups which may be attributed to a general East Asian/collectivist cultural bias in emotion processing; moreover, the ERP data further demonstrated some of the direct effects of cultural origin on distinct neural processing stages during multi-sensory emotion processing, indexed by N400 and MMN components.

The observed cross-cultural differences between the East Asian and Western groups support the *display rules* theory in emotional communication. The differential sensitivity of the Chinese group to emotional vocal/prosodic cues when compared to facial information is compatible with the assumed culture-specific display rules of the East Asian/collectivist culture: to maintain social harmony, people from this culture are more constrained in their facial expressions, use more indirect/non-literal speech, and tend to avoid direct eye contact with others. Such display rules possibly render facial cues less salient and lead to the fact that they have to rely more on prosody/vocal cues as a compensatory mechanism for effective communication. On the other hand, the higher sensitivity of the English group to facial information is consistent with the Western display rules that the Western/individualistic culture encourages eye contacts and expressive displays of emotions via facial expressions or direct literal/semantic information in speech. An alternative interpretation of the observed cultural differences involves the tonal features of the spoken language of each cultural group: the Chinese participants as native speakers of Mandarin, a strongly tonal language, might have developed a perceptual or attentional bias towards tonal variations in general, including emotional prosody. In contrast, the English North Americans as native English speakers did not develop such a bias to tonal variations, resulting in their lower sensitivity to prosodic cues in the context of multi-sensory emotional cues. While both display rules and the tonal structure of the spoken language could be potential determinants of the observed cultural differences, the current data cannot disentangle the roles of the two factors.

While the findings of Study 1 significantly advance our knowledge of the effects of *cultural origin* on the neuro-cognitive processes underlying multi-sensory emotion processing by comparing two groups, Study 2 provided some insights on the effect of another type of cultural experiences, *cultural immersion*, on emotion processing by examining a third group, Chinese immigrants to Canada, in comparison with the two groups tested in Study 1.

Effect of cultural immersion: Study 2

Based on the findings of Study 1 on the effects of cultural origin on emotion perception, Study 2 of this thesis sought to understand how a different kind of cultural experience, immersion in a new host culture, affects multi-sensory emotion processing. A third group, Chinese immigrants to Canada, was tested; the data of this group were directly compared to those of Chinese and English North American groups of Study 1. The results of the Stroop task first indicated that the Chinese immigrants showed a larger behavioural Stroop effect in the voice task than in the face task; however, in the N400 data, no difference was observed between the two task conditions. In the Oddball task, the immigrants were significantly influenced by concurrent emotional voices while processing facial expressions. These results suggest that the

Chinese immigrants showed a higher sensitivity to faces on the behavioural level, but a higher susceptibility to vocal cues in the ERP data (i.e., the vMMN component). When compared to the two control groups from Study 1, it appears that the immigrants showed a behavioural pattern that resembled their host culture, the English Canadian group; nonetheless, such a pattern was not observed in their ERP data, i.e., N400 and vMMN components, where the immigrants exhibited similar neural response patterns to the Chinese group, their native culture. In other words, a significant effect of cultural immersion in the immigrants was observed only at a relatively late stage of rendering off-line decisions about the meaning of the stimuli, but not in their on-line neural responses while processing the stimuli in real time.

Interestingly, in spite of the fact that Chinese immigrants did not resemble members of their host culture when ERP results were analyzed, the immigrants' length of stay in Canada was found to be significantly correlated with their neural (and behavioural) responses: those immigrants who stayed longer in Canada tended to exhibit behavioural and neural patterns (including both N400 and vMMN components) that were more similar to the English group. These findings strongly suggest that the effect of cultural immersion on multi-sensory emotion processing, and possibly other aspects of communication and cognition, occurs first with adaptation on the behavioural level, which is followed at a later time by actual modulation of brain functions. The amount of exposure to the new cultural environment plays a role in modulating the extent to which neural responses to multi-sensory emotional displays resemble those of the host cultural group. In particular, it is inferred that the reduced exposure to East

Asian display rules and the increased exposure to Western/individualist display rules contribute to the modulation of the way the immigrants process multi-sensory emotional expressions. Alternatively, the changed language environment—reduced exposure to the lexical tonal features of the Chinese language and increased exposure to English without lexical tones—might be another potential factor motivating the observed changes of the immigrants in processing multi-sensory emotions.

Study 2 provides the first evidence on how cultural immersion experiences shape the cognitive/neuro-cognitive mechanisms underlying multi-sensory emotional processing. It was found that the immigrants' behavioural adaptation to the host culture occurred earlier than the modulation on the neural level. This implies that it might be important for them to first adjust their behaviours to fit into the new cultural/social environment for effective social interactions and positive social impressions. Meanwhile, it might also be easier for them to change their behaviours, which are possibly under greater conscious control, by intentionally adopting certain strategies to imitate the host culture. In the clinical literature, training and modifying behaviours has proved a valid approach to improve social interactions in populations with social impairments. For instance, behavioural intervention for population with Autism Spectrum Disorder commonly include teaching of eye contacts with other people (Carbone, O'Brien, Sweeney-Kerwin, & Albert, 2013), training of responding to pointing and gaze shifting of others, and training of pointing with the purpose of sharing, etc. (Dallaire & Schreibman, 2003). Likewise, behavioural treatment employed for patients with Parkinson's Disease involve practices of various social skills, e.g., ordering and paying in a restaurant, asking for a seat on a bus, etc., which could help them to better adapt to social life (Ellgring et al., 1990; Keus, Munneke, Nijkrake, Kwakkel, & Bloem, 2009; Mohr et al., 1996). In addition, behaviourally oriented psychotherapies, which focus on re-shaping the patients' routine behaviours, have been commonly and effectively used to treat clinical populations with a variety of psychological/social disorders (Butler, Chapman, Forman, & Beck, 2006; Lynch, Laws, & McKenna, 2010; Morley, Eccleston, & Williams, 1999). In both the clinical and healthy population, consciously modifying behaviours might be an effective way to improve social communication, including the case of the immigrants who are exposed to a new and unfamiliar cultural/social environment.

Furthermore, with increased exposure to the new cultural and linguistic environment, these behavioural changes may lead to changes on the neural level, i.e., a reorganization of the neural circuitry underlying these communication behaviours. Indeed, immigrants who had been living in Canada for a longer period of time and had experienced more social input from the foreign environment exhibited neural response patterns more similar to their host culture, the English group. Our suggestion that the cultural adaptation process begins at the behavioural level followed by neural modification echoes previous findings in other domains which imply that behavioural alterations could lead to neural reorganization. For instance, it has been found that following five consecutive sessions of 45-minute behavioural training of recognizing other-race faces at the subordinate level of the individual (i.e., *Joe, Bob*), Caucasian participants showed improved performance in recognizing individual African-American and Hispanic faces;

interestingly, they also exhibited an increased expert N250 component in the posterior area. Here, N250 is considered as a biological marker of category specificity, which always yields increased amplitude in response to faces (or objects) that are successfully differentiated at the subordinate/individual level. This finding provides an example of behavioural training-induced neuroplasticity of the adult human brain in processing social information (Tanaka & Pierce, 2009). Another example was reported in which adults who had been trained to juggle for three months showed a transient and selective structural change in brain areas that are associated with the processing and storage of complex visual motion (Draganski et al., 2004). Similar evidence is also documented in the clinical literature: following a certain period of behavioural interventions, individuals with Autism Spectral Disorder showed improved socially-adaptive behaviours; more critically, their neural activities in processing social cues (e.g., faces) appeared to be "normalized", i.e., became more similar to those of typically developing controls in both EEG (Dawson et al., 2012) and fMRI data (Voos et al., 2013). In another clinical study, individuals with Chronic Fatigue Syndrome, whose grey matter volume was significantly lower than healthy controls, were treated with Cognitive Behavioral Therapy. It was observed that following the treatment, their grey matter volume was significantly increased in the prefrontal cortex in addition to improvement in physical activity and health status (de Lange et al., 2008).

Therefore, consistent with the previous evidence in the literature, the results of Study 2 strengthened the idea that there is a dynamic relation exists between behaviours and neural function and anatomy: behavioural changes could potentially lead to "rewiring" of the human brain. More importantly, this study for the first time demonstrated this notion in the domain of multisensory emotion processing in the context of cultural immersion and adaptation. Still, new studies are awaited which would further elaborate the time course of cultural adaptation that occurs on both behavioural and neural levels by examining different cultural groups in different cognitive domains.

Culture as a multi-faceted concept: cultural origin and cultural immersion

This thesis is one of the first to investigate the multi-faceted concept of *culture* in emotional communication by examining two different types of cultural learning experiences, cultural origin and cultural immersion. When results of Study 1 and Study 2 are combined, they suggest that the way people process emotions from different cues in multimodal stimuli is determined in part by their cultural origin and cultural learning; it can be argued that the learning of culture-specific values and corresponding display rules, and possibly differential exposure to different linguistic features (e.g., complex tones) during development, contribute to the observed differences between the two groups studied here, Chinese and English North American. However, the present findings demonstrate that such culture-specific biases are not unalterable over the lifespan; when people leave their native culture and live in a foreign culture for a certain duration of time, their responses to multi-sensory emotions progressively adapt or accommodate to their host culture to a certain degree. This provides evidence of psychological plasticity at both the behavioural and neural processing levels, at least for immigrants who leave their native culture as adolescents or young adults (as was true for participants in Study 2). Indeed, it appears that a dynamic adaptation process occurred to the immigrants during which their cognitive/ neuro-cognitive patterns in processing multi-sensory emotions transitioned from their native culture to the host culture; the transition happens earlier on the behavioural level than on the neural level, and the extent of the transition is modulated by their amount of exposure to the new cultural/linguistic environment.

Such a dynamic transition process that occurs on the neural level reflects the so-called "experience-dependent neuroplasticity", which refers to the neural re-organization that transpires due to various aspects of environment input, cognitive demand, or behavioural experiences (Holtmaat & Svoboda, 2009; Kleim & Jones, 2008; May, 2011). In the literature, ample evidence has been reported on this phenomenon and its potential mechanisms in a variety of domains. For instance, both functional and structural changes in the brain have been found as a result of one's experiences of learning a second language, including increased gray matter density and white matter integrity. These changes can occur with both long-term and short-term language learning/exposure experiences, have been observed in different age groups including adults and the elderly, and are sensitive to age of acquisition, proficiency level, language-specific characteristics, and individual differences (Li, Legault, & Litcofsky, 2014). Consistent observations of similar neuroplasticity have been documented in other domains as a function of non-linguistic experiences as well, including attention, memory, musical expertise, mathematical learning, and visuo-motor learning (Lövdén, Wenger, Mårtensson, Lindenberger, & Bäckman, 2013; May, 2011; Thomas & Baker, 2013; Zatorre, Fields, & Johansen-Berg, 2012). Here, this literature is further extended and enriched by the current thesis which provides insights into the experience-dependent neuroplasticity of human brain in the domain of multisensory emotion processing, which occurs as a function of the immigrants' immersion experiences in a new cultural/linguistic environment.

In sum, by exploring three cultural groups with different types of cultural experiences, this thesis provides a more comprehensive illustration of the role of culture in emotional processing. Future studies examining more refined immigrant groups with distinct cultural experiences will be helpful to further elaborate the dynamic transition between their native and host cultures that occurs during cultural immersion. In particular, given the evidence from the domain of second language acquisition that neuroplasticity in language processing varies as a function of a wide range of factors, including the learners' age, age of acquisition, the amount of language learning experiences, individual differences, etc. (Li et al., 2014), it could be inferred that these factors, too, could possibly influence the neuroplasticity in multisensory emotion processing in the context of different cultural experiences. Future studies that take these factors into consideration, e.g., examining immigrants who moved to a new culture as mid-age adults, will further complement and refine the picture of the relations between culture, brain, and emotion.

Cultural effects on an overall process of emotion perception

By employing EEG measurements with high temporal resolution in two different tasks, this thesis was able to explore various processing stages during multi-sensory emotion perception.

Instead of focusing on one single stage as in previous cross-cultural literature, it for the first time examined key time points during the process of emotion perception, including stages of both online neural processing (early perceptual processing indexed by vMMN and later semantic processing indexed by N400) and off-line behavioural decisions about emotional stimuli. In Study 1, the effects of *cultural origin* on emotion perception (i.e., the differences between the Chinese and English North American groups) are significant at all three stages, demonstrating the robust influence of the origin of culture on the processes for appraising and interpreting emotional expressions during communication. In particular, this cultural effect appears at a very early stage shortly after the onset of the emotional stimuli (100-200ms), continues to the semantic processing stage (around 400ms), and finally affects the explicit behavioural performance in perceiving emotions.

It is especially worth mentioning that the findings of Study 1 demonstrate that the effect of cultural origin on emotion processing occurs particularly early. This is consistent with previous findings that some other social/cultural factors also exert their impact on emotion processing at a very early stage: for instance, racial effect on facial expression processing has been observed as early as the N170 component: compared to inverted other-race faces, inverted same-race faces lead to greater recognition impairment and elicited larger and later N170 amplitudes (Gajewski, Schlegel, & Stoerig, 2008; Vizioli, Foreman, Rousselet, & Caldara, 2010). It was also reported that social/emotional background influenced early processing of facial expressions indicated by N170: facial expressions embedded in backgrounds of fearful social scenes (e.g., a car accident) elicited larger N170 than faces in happy and neutral scenes, suggesting that the early structural processing of emotional faces is influenced by the social/emotional contextual information (Righart & de Gelder, 2008). Together with the results of Study 1, these findings imply that various cultural and social factors, which have been learned by individuals during development and socialization, play an important and prompt role in processing emotional stimuli.

On the other hand, for the effect of *cultural immersion* in Study 2, an interesting dissociation has been observed between the behavioural and neural levels, suggesting that a dynamic process of cultural adaptation occurs during immersion, which begins in the behaviours and is probably followed by neural modulation at later stages of cultural immersion. By testing different processing stages at the same time, this thesis illustrates an overall picture along the temporal dimension of multi-sensory emotion processing, and, more importantly, of how this integral process is influenced by the factors of culture.

Concluding remarks and future directions

Investigating the characteristics of cultural differences in emotion communication is critical to facilitate understanding and relations between different cultural groups. This thesis addresses this important problem by demonstrating the degree to which emotion processing was modulated by various cultural experiences. Specifically, it investigates the effects of cultural origin and cultural immersion on the behavioural and neural temporal mechanisms underlying emotion perception across facial expressions and prosody. It significantly advances our knowledge about the behavioural and neural plasticity of emotion processing in the context of culture. Meanwhile, it also points to several new directions for future research. First, due to the limitation of the available experimental materials, only two negative emotions, sadness and fear, were involved in this thesis. Following the framework of discrete emotions, future studies adopting other emotional categories would clarify whether the observed cultural differences here could be generalized to other basic emotions, especially positive ones, which arguably bear different meanings from both evolutionary and socio-cultural perspectives as opposed to negative emotions (Lu & Gilmour, 2004; Oishi et al., 2013; Uchida et al., 2004).

While this thesis examined the effects of two kinds of cultural experiences, *cultural origin* and *cultural immersion*, on emotion perception, it is true that there exist various other types of cultural experiences that await examination in the future. Given our findings of Study 2 that the length of stay in the new cultural environment correlates with the observed degree of acculturation in emotion processing on both behavioural and neural levels, testing more refined cultural groups with different amount of exposure to the new culture and different types of immigration experiences will be helpful to further disentangle how different cultural experiences shape behavioural and brain functions during emotion communication. Similarly, longitudinal studies with several time points on a certain immigrants group would also be greatly benefical to determine whether there is a temporal threshold when the immigrants begin to adjust their neuro-cognitive patterns in processing multi-sensory emotions to adapt to their host culture.

As discussed in Study 1, while both cultural display rules and the tonal features of a language are potential factors contributing to the observed cultural differences in processing emotional expressions, it is true that this thesis cannot disentangle these two factors by its current data. Further studies are needed to disambiguate these two possible contributors to the reported cultural differences. One promising option is to conduct the same experimental protocol to a group of participants who are from an East Asian culture that shares the collectivist display rules, but speak a language without strong tonal features at the lexical level, such as standard Korean. The extent to which such a group would show analogous culture-specific biases in processing multi-sensory emotions relative to the Chinese group would provide insights to this important issue regarding what motivates the observed cross-cultural differences in emotion communication.

While this thesis focuses on the effects of culture on the *temporal course* of emotion processing as illuminated by ERP measurements with high temporal-resolution, future studies are needed to localize the neural generators involved in these cultural effects with fine spatial resolution, e.g., the fMRI technique. In particular, a combined EEG/fMRI approach would provide a more comprehensive picture integrating both the temporal and spatial characteristics of the brain mechanisms underlying the cross-cultural differences in emotion perception. In addition, while this thesis with ERP data (and future studies employing fMRI technique) contributes important evidence on neuroplasticity in emotion processing on the *functional* level, another future direction is to explore whether the cultural experiences-induced neuroplasticity also occurs on the *structural/anatomical* level, by employing structural imaging approaches (e.g., structural MRI). Given previous findings that structural neuroplasticity has been observed in language acquisition and various other cognitive domains (Draganski et al., 2004; Li et al., 2014; Lövdén et al., 2013; May et al., 2007), it would be interesting to examine whether anatomical changes in human brain would be induced by cultural experiences in emotion processing as well. With both functional and structural neural data, it would then be possible to compare the functional neural changes to the potential structural neural modifications to examine the consistencies (and inconsistencies) between the two levels. Moreover, investigating how the neural data, functional or structural, are correlated with relevant behavioural performance, would be significant to systematically demonstrate the complex behaviour-function-structure relations in processing multisensory emotions in the context of culture and cultural immersion.

Finally, a more biological approach for future research involves identifying the cellular and molecular mechanisms underlying the cultural experiences-induced neuroplasticity in social cognition. In particular, such efforts could potentially help clarify the relations between culture and genes in this domain. For instance, a robust cross-national correlation has been found between the relative frequency of variants in social sensitivity genes (e.g., serotonin and opioid, the frequency of which is found to be positively associated with the degree of emotional responsivity to social events and experiences) and the relative degree of individualism–collectivism in different cultural groups, suggesting that collectivism may have developed in groups with a high frequency of putative social sensitivity alleles (Way & Lieberman, 2010). While the studies of psychological genetics are still developing, future research in this direction would undoubtedly help disentangle the intricate culture-gene relationship and provide significant insights into the long-lasting debate of "nature versus nurture" on the human mind.

In conclusion, this thesis contributes significant findings to the understanding of the role of culture in multisensory emotion processing, which lead to several new research directions that could be usefully addressed in future studies. Specifically, future research on the brain networks in emotion processing in the context of culture (and different cultural experiences) will provide important avenues to a better knowledge of the neuroplasticity as a function of culture. Such studies will have significant implications for our globalized, multicultural world, and providing a window into the adaptive nature of the human behaviour, mind and brain.

5. References

Adams, R. B., Franklin, R. G., Rule, N. O., Freeman, J. B., Kveraga, K., Hadjikhani, N., . . . Ambady, N. (2010). Culture, gaze, and the neural processing of fear expressions: an fMRI investigation. *Social Cognitive and Affective Neuroscience, 5*, 340-348.

Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural Systems for Recognition of Emotional Prosody. A 3-D Lesion Study. *Emotion, 2*(1), 23-51.

Adolphs, R., Damasio, H., Tranel, D., & Damasio, A. R. (1996). Cortical Systems for the Recognition of Emotion in Facial Expressions. *Journal of Neuroscience, 16*, 7678-7687.

Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology, 15*(3), 396-404.

Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T., & Hietanen, J. K. (2013). Attention to eye contact in the west and east: Autonomic responses and evaluative ratings. *PLoS ONE, 8*(3).

Alves, N. T. (2013). Recognition of static and dynamic facial expressions: a study review. *Estud. psicol. (Natal), 18*(1), 125-130.

Ambady, N., Koo, J., Lee., F., & Rosenthal, R. (1996). More than words: Linguistic and nonlinguistic politeness in two cultures. *Journal of Personality and Social Psychology, 70*(5), 996-1011.

Astikainen, P., & Hietanen, J. K. (2009). Event-related potentials to task-irrelevant changes in facial expressions. *Behavioral and Brain Functions, 5*(30).

Athanasopoulos, P. (2007). Interaction between grammatical categories and cognition in bilinguals: The role of proficiency, cultural immersion, and language of instruction. *Language and Cognitive Processes, 22*(5), 689-699.

Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism, 12*(1), 83-95.

Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J. R., & Thierry, G. (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition, 116*(3), 437-443.

Atkinson, A. P., Tipples, J., Burt, D. M., & Young, A. W. (2005). Asymmetric interference between sex and emotion in face perception. *Perception and Psychophysics, 67*(7), 1199-1213.

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Theories of Emotion* (pp. 247-264). New York: Academic Press.

Balconi, M. (2012). Neuropsychology of facial expressions. The role of consciousness in processing emotional faces. *Neuropsychological Trends, 11*(1), 19-40.

Banse, R., & Scherer, K. R. (1996). Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology, 70*(3), 614-636.

Barrett, L. F., & Wager, T. D. (2006). The structure of emotion evidence from neuroimaging studies. *Current Directions in Psychological Science, 15*(2), 79-83.

Batty, M., & Taylor , M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research, 17*, 613 –620.

Beaupré, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology, 36*(3), 355-370.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans *Journal of Cognitive Neuroscience, 8*, 551 –565.

Berry, J. W. (1990). Psychology of acculturation In J. Berman (Ed.), *Cross-cultural perspectives: Nebraska Symposium on Motivation* (pp. 201-234). Lincoln: University of Nebraska Press.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion, 19*(8), 1113-1139. doi: 10.1080/02699930500204250

Bilbow, G. T. (1997). Spoken discourse in the multicultural workplace in Hong Kong: applying a model of discourse as 'impression management'. In F. Bargiela-Chiappini & S. Harris (Eds.), *The languages of business: an international perspective* (pp. 21–48). Edinburgh: Edinburgh University Press.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Bötzel, K., Schulze, S., & Stodieck, S. R. G. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials *Experimental Brain Research, 104*, 135 –

143.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*(2), 129-148.

Bowers, D., Bauer, R. M., & Heilman, K. M. (1993). The nonverbal affect lexicon: Theoretical perspectives from neuropsychological studies of affect perception. *Neuropsychology, 7*(4), 433-444.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology In L. R. D. & N. L. (Eds.), *Cognitive neuroscience of emotion* (pp. 242–276). New York: Oxford University Press;.

Briggs, J. (1970). *Never in anger: Portrait of an Eskimo family*. Cambridge, MA: Harvard University Press.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*(1), 34-44.

Brown, C., Hagoort, P., & Chwilla, D. (2000). An event-related brain potential analysis of visual word-priming effects. *Brain and Language, 72*(2), 158–190.

Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture, 8*, 135-148.

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review, 26*(1), 17-31. doi: http://dx.doi.org/10.1016/j.cpr.2005.07.003

Cabanac, M. (1996). On the origin of consciousness, a postulate and its corollary. *Neuroscience and Biobehavioral Reviews, 20*(1), 33-40.

Cabanac, M. (2002). What is emotion? . *Behavioural Processes, 60*(2), 69-83.

Caharel, S., Courtay, N., Bernard, C., Lalonde, R., & Rebaï, M. (2005). Familiarity and emotional expression influence an early stage of face processing: An electrophysiological study. *Brain and Cognition, 59*(1), 96-100. doi: http://dx.doi.org/10.1016/j.bandc.2005.05.005

Campbell, T., Winkler, I., & Kujala, T. (2007). N1 and the mismatch negativity are spatiotemporally distinct ERP components: distruption of immediate memory by auditory distraction can be related to N1. *Psychophysiology, 44*(4), 530-540.

Carbone, V. J., O'Brien, L., Sweeney-Kerwin, E. J., & Albert, K. M. (2013). Teaching eye contact to children with autism: A conceptual analysis and single case study. *Education and Treatment of Children, 36*(2), 135-159.

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention, and the 'negativity bias', studied through event-related potentials. *International Journal of Psychophysiology, 41*(1), 75-85.

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods, 42*(1), 74-81.

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., . . . Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience, 10*, 127.

Chiao, J. Y., Iidaka, T., Gordon, H. L., Nogawa, J., Bar, M., Aminoff, E., . . . Ambady, N. (2008).

Cultural specificity in amygdala response to fear faces. *J Cogn Neurosci, 20*(12), 2167–2174.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology, 32*(3), 274–285.

Collier, G. L. (2007). Beyond valence and activity in the emotional connotations of music. *Psychology of Music, 35*(1), 110-131.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research, 1242*(C), 126-135.

Dallaire, C. W., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44*(3), 456-468.

Damjanovic, L., Athanasopoulos, P., Krajciova, A., Kasai, C., & Roberson, D. (2013). *Searching for happiness and defining emotional space: Evidence from biculturalism*. Paper presented at the International Society for Research on Emotion, Berkeley, CA.

Darwin, C. (1998). *The expression of the emotions in man and animals (3rd ed.)*. New York: Oxford University press

Davidson, R. J. (1995). Cerebral asymmetry, emotion, and affective style. In R. J. Davidson & K. Hugdahl (Eds.), *Brain asymmetry* (pp. 361-388). Cambridge, MA: MIT Press.

Dawson, G., Jones, E. J. H., Merkle, K., Venema, K., Lowy, R., Faja, S., . . . Webb, S. J. (2012). Early behavioral intervention is associated with normalized brain activity in young children with autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(11), 1150-1159.

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

de Gelder, B., Bocker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neuroscience Letters, 260*(2), 133-136.

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

de Lange, F. P., Koers, A., Kalkman, J. S., Bleijenberg, G., Hagoort, P., Van Der Meer, J. W. M., & Toni, I. (2008). Increase in prefrontal cortical volume following cognitive behavioural therapy in patients with chronic fatigue syndrome. *Brain, 131*(8), 2172-2180.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134*(1), 9-21.

Derntl, B., Habel, U., Robinson, S., Windischberger, C., Kryspin-Exner, I., Gur, R., & Moser, E. (2009). Amygdala activation during recognition of emotions in a foreign ethnic group is associated with duration of stay. *Social Neuroscience, 4*(4), 294–307.

Derntl, B., Habel, U., Robinson, S., Windischberger, C., Kryspin-Exner, I., Gur, R. C., & Moser, E. (2012). Culture but not gender modulates amygdala activation during explicit emotion recognition. *BMC Neuroscience, 13*(1), 13-54.

Dong, G., Zhou, H., & Zhao, X. (2011). Male Internet addicts show impaired executive control ability: evidence from a color-word Stroop task. *Neuroscience Letters, 499*(2), 114–118.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature, 427*(6972), 311-312.

Eimer, M. (2011). The Face-Sensitive N170 Component of the Event-Related Brain Potential. In A. J. Calder (Ed.), *The Oxford Handbook of Face Perception*. London: Oxford University Press.

Eimer, M., Cockburn, D., Smedley, B., & Driver, J. (2001). Cross-modal links in endogenous spatial attention are mediated by common external locations: Evidence from event-related brain potentials. *Experimental Brain Research, 139*(4), 398-411.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*(1), 15-31.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation* (pp. 207–282). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1992a). Are there basic emotions? *Psychological Review, 99*(3), 550-553.

Ekman, P. (1992b). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169-200.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49-98.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A technique for the measurement of facial action*. Palo Alto, CA: Consulting Psychologist Press.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*(3875), 86-88.

Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review, 5*(1), 90-96.

Elfenbein, H. A., & Ambady, N. (2002a). Is there an In-Group Advantage in Emotion Recognition? *Psychological Bulletin, 128*(2), 243-249.

Elfenbein, H. A., & Ambady, N. (2002b). On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis. *Psychological Bulletin, 128*(2), 203-235.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality and Social Psychology, 85*(2), 276-290. doi: 10.1037/0022-3514.85.2.276

Elfenbein, H. A., Beaupré, M., Levesque, M., & Hess, U. (2007). Toward a Dialect Theory: Cultural Differences in the Expression and Recognition of Posed Facial Expressions. *Emotion, 7*(1), 131-146.

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-Cultural Patterns in Emotion Recognition. Highlighting Design and Analytical Techniques. *Emotion, 2*(1), 75-84.

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2004). Hemifacial differences in the in-group advantage in emotion recognition. *Cognition and Emotion, 18*(5), 613-629.

Ellgring, H., Seiler, S., Nagel, U., Perleth, B., Gasser, T., & Oertel, W. H. (1990). Psychosocial problems of Parkinson patients: approaches to assessment and treatment. *Advances in* *neurology, 53*, 349-353.

Engelmann, J. B., & Pogosyan, M. (2013). Emotion perception across cultures: the role of cognitive mechanisms. *Frontiers in Psychology, 4*, 118.

Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in Brain Research, 156*, 345-361.

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron, 51*(6), 871–882.

Föcker, J., Gondan, M., & Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica, 137*(1), 36-47.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion, 1*, 115-143.

Froyen, D., van Atteveldt, N., & Blomert, L. (2010). Exploring the role of low level visual processing in letter–speech sound integration: a visual MMN study. *Frontiers in Integrative Neuroscience, 4*, 9.

Froyen, D., van Atteveldt, N., Bonte, M., & Blomert, L. (2008). Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neuroscience Letters, 430*(1), 23-28.

Frühholz, S., & Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews, 37*(1), 24-35.

Gajewski, P. D., Schlegel, K., & Stoerig, P. (2008). Effects of human race and face inversion on the N170: A cross-race study. *Journal of Psychophysiology, 22*(4), 157-165.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The Search for "Common Sense": An Electrophysiological Study of the Comprehension of Words and Pictures in Reading. *Journal of Cognitive Neuroscience, 8*(2), 89-106.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*(1), 21-38.

Goos, L. M., & Silverman, I. (2002). Sex related factors in the perception of threatening facial expressions. *Journal of Nonverbal Behavior, 26*(1), 27-41.

Gorer, G. (1949). *Africa dances: a book about West African negroes*. New York: J. Lehmann (London).

Goto, S. G., Ando, Y., Huang, C., Yee, A., & Lewis, R. S. (2010). Cultural differences in the visual processing of meaning: Detecting incongruities between background and foreground objects using the N400. *Social Cognitive and Affective Neuroscience, 5*(2-3), 242-253.

Goto, S. G., Yee, A., Lowenberg, K., & Lewis, R. S. (2013). Cultural differences in sensitivity to social context: Detecting affective incongruity using the N400. *Social Neuroscience, 8*(1), 63-74.

Grandjean, D., Banziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in Brain Research, 156*(235-247).

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., &

Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*(2), 145-146.

Gudykunst, W. B., Matsumoto, Y., Ting-Toomey, S., Nishida, T., Kim, K., & Heyman, S. (1996). The Influence of Cultural Individualism-Collectivism, Self Construals, and Individual Values on Communication Styles across Cultures. *Human Communication Research, 22*(4), 510-543.

Gutchess, A. H., Welsh, R. C., Boduroglu, A., & Park, D. C. (2006). Cultural differences in neural function associated with object processing. *Cognitve, Affective, and Behavioral Neuroscience, 6*(2), 102-109.

Haas, B. W., Omura, K., Constable, R. T., & Canli, T. (2006). Interference produced by emotional conflict associated with anterior cingulate activation. *Cognitve, Affective, and Behavioral Neuroscience, 6*(2), 152–156.

Halgren , E., Raij , T., Marinkovic , K., Jousmäki , V., & Hari , R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG *Cerebral Cortex, 10*(69-81).

Hall, E. T. (1976). *Beyond Culture*. New York: Doubleday.

Hall, E. T. & Hall, M. R. (1990). *Understanding Cultural Differences - Germans, French and Americans*. Yarmouth, ME: Intercultural Press Inc.

Handy, T. C. (2004). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press.

Hauser, M. (2004). *The evolution of communication.* . Cambridge, MA: MIT Press.

Hawrysh, B. M., & Zaichkowsky, J. L. (1991). Cultural Approaches to Negotiations:

Understanding the Japanese. *Asia Pacific Journal of Marketing and Logistics, 3*(1), 40 – 54.

Hearn, L. (1894). *Glimpses of unfamiliar Japan*. London: Osgood, McIlvaine and Co.

Hedden, T., Ketay, S., Aron, A., Markus, H. R., & Gabrieli, J. D. E. (2008). Cultural influences on neural substrates of attentional control. *Psychological Science, 19*(1), 12-17.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: a comparison using event-related potentials. *Language and Cognitive Processes, 5*(4), 281–312.

Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat Rev Neurosci, 10*(9), 647-658. doi: http://www.nature.com/nrn/journal/v10/n9/suppinfo/nrn2699_S1.html

Hot, P., Saito, Y., Mandai, O., Kobayashi, T., & Sequeira, H. (2006). An ERP investigation of emotional processing in European and Japanese individuals. *Brain Research, 1122*(1), 171-178.

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception: An Interdisciplinary Journal, 23*(4), 319-330. doi: 10.1525/mp.2006.23.4.319

Ishii, K., Kobayashi, Y., & Kitayama, S. (2010). Interdependence modulates the brain response to word-voice incongruity. *Social Cognitive and Affective Neuroscience, 5*(2-3), 307-317.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous Attention to Word Content Versus Emotional Tone: Differences Among Three Cultures. *Psychological Science, 14*(1), 39-46.

Izard, C. E. (1977). *Human Emotions*. New York: Plenum.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99*(3), 561 - 565.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*(2), 288-299.

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage, 58*(2), 665-674.

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. S. (2003). Facial expression of emotion. In R. Davidson, K. Scherer & H. H. Goldsmith (Eds.), *Handbook of Affective Science* (pp. 415-432). London: Oxford University Press.

Keus, S. H. J., Munneke, M., Nijkrake, M. J., Kwakkel, G., & Bloem, B. R. (2009). Physical therapy in Parkinson's disease: Evolution and future challenges. *Movement Disorders, 24*(1), 1-14.

Kipp, M., & Martin, J. (2009). *Gesture and Emotion: Can basic gestural form features discriminate emotions?* Paper presented at the Proceedings of the International Conference on Affective Computing and Intelligent Interaction, Amsterdam, Netherland.

Kirita, T., & Endo, M. (1995). Happy face advantage in recognizing facial expressions. *Acta Psychologica, 89*, 149-163.

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: A Cultural Look at New Look *Psychological Science, 14*(3), 201-206.

Kitayama, S., & Ishii, K. (2002). Word and Voice: Spontaneous Attention to Emotional Utterances in Two Languages. *Cognitiion and Emotion, 16*(1), 29-59.

Klasen, M., Chen, Y.-H., Mathiak, K. (2012). Multisensory emotions: Perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4), 381-392

Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research, 51*(1), S225-S239.

Klineberg, O. (1940). *Social psychology.* . New York, NY: Holt.

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research, 1151*, 107-118.

Kotz, S. A., & Paulmann, S. (2011). Emotion, language and the brain. *language and Linguistics Compass, 5*, 108–125.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83-143). San Diego, CA: Academic Press.

Labarre, W. (1947). The cultural basis of emotions and gestures *Journal of Personality, 16*(1), 49–68.

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? . *Psychological Research, 69*, 22-29.

Leppänen, J. M., Tenhunen, M., & Hietanen, J. K. (2003). Faster choice-reaction times to positive than to negative facial expressions. Journal ofPsychophysiology. *Journal of Psychophysiology, 17*, 113-123.

Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: Electrophysiological evidence. *Neuroreport, 12*(12), 2653-2657.

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*.

Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magnetoand electroencephalography. *Neuroscience Letters, 253*, 147–150.

Liotti, M., Woldorff, M. G., Perez III, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia, 38*(5), 701-711.

Liu, P., Rigoulot, S., & Pell, M. D. (under review). Culture modulates the brain response to human expressions of emotion: electrophysiological evidence.

Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. *Neuroreport, 11*, 337–341.

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods, 44*(4), 1042-

1051.

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., & Bäckman, L. (2013). Structural brain plasticity in adult learning and development. *Neuroscience & Biobehavioral Reviews, 37*(9, Part B), 2296-2310. doi: http://dx.doi.org/10.1016/j.neubiorev.2013.02.014

Lu, L., & Gilmour, R. (2004). Culture and conceptions of happiness: individual oriented and social oriented swb. *Journal of Happiness Studies, 5*(3), 269-291.

Luo, W., Feng, W., He, W., Wang, N. Y., & Luo, Y. J. (2010). Three stages of facial expression processing: ERP study with rapid serial visual presentation. *NeuroImage, 49*(2), 1857-1867.

Lynch, D., Laws, K. R., & McKenna, P. J. (2010). Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials. *Psychological Medicine, 40*(01), 9-24. doi: doi:10.1017/S003329170900590X

Maekawa, T., Goto, Y., Kinukawa, N., Taniwaki, T., Kanba, S., & Tobimatsu, S. (2005). Functional characterization of mismatch negativity to a visual stimulus. *Clinical Neurophysiology, 116*(10), 2392-2402.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review, 3*(2), 215-221.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and*

Emotion, 14(3), 195-214.

Matsumoto, D., Consolacion, T., Yamada, H., Suzuki, R., Franklin, B., & Paul, S., et al. (2002). American-Japanese cultural differences in judgements of emotional expressions of different intensities. *Cognition and Emotion, 16*(6), 721–747.

Matsumoto, D. (2002). Methodological Requirements to Test a Possible In-Group Advantage in Judging Emotions Across Cultures:Comment on Elfenbein and Ambady (2002) and Evidence. *Psychological Bulletin, 128*(2), 236-242.

Matsumoto, D., Takeuchi, S., Andayani, S., Kouznetsova, N., & Krupp, D. (1998). The contribution of individualism vs. collectivism to cross-national differences in display rules. *Asian Journal of Social Psychology, 1*(2), 147–165.

Matsumoto, D., Yoo, S. H., & Fontaine, J. (2008). Mapping expressive differences around the world: the relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology, 39*(1), 55–74.

Matsumoto, D., Yoo, S. H., Hirayama, S., & Petrova, G. (2005). Development and validation of a measure of display rule knowledge: the display rule assessment inventory. *Emotion, 5*(1), 23–40.

May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences, 15*(10), 475-482.

May, A., Hajak, G., Gänßbauer, S., Steffens, T., Langguth, B., Kleinjung, T., & Eichhammer, P. (2007). Structural brain alterations following 5 days of intervention: Dynamic aspects of neuroplasticity. *Cerebral Cortex, 17*(1), 205-210.

McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2006). Cultural display rules drive eye gaze during thinking. *Journal of Cross-Cultural Psychology, 37*(6), 717-722.

McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2008). Gaze display when thinking depends on culture and context. *Journal of Cross-Cultural Psychology, 39*(6), 716-729.

McGloin, N. H. (1984). Some politeness strategies in Japanese. . In S. Miyagawa & C. Kitagawa (Eds.), *Studies in Japanese language use* (pp. 127-145). Edmonton, Canada: Linguistic Research Inc.

Mead, M. (1975). Review of "Darwin and facial expression". *Journal of Communication, 25*, 209–213.

Mitchell, R. L. C. (2006). Does incongruence of lexicosemantic and prosodic information cause discernible cognitive conflict? . *Cognitive, Affective, & Behavioral Neuroscience, 6*(4), 298-305.

Mitchell, R. L. C., & Ross, E. D. (2008). fMRI evidence for the effect of verbal complexity on lateralisation of the neural response associated with decoding prosodic emotion. *Neuropsychologia, 46*(12), 2880-2887.

Miyoshi, M., Katayama, J., & Morotomi, T. (2004). Face-specific N170 component is modulated by facial expressional change. *Neuroreport, 15*(5), 911-914.

Mohr, B., Müller, V., Mattes, R., Rosin, R., Federmann, B., Strehl, U., . . . Birbaumer, N. (1996). Behavioral treatment of Parkinson's disease leads to improvement of motor skills and to tremor reduction. *Behavior Therapy, 27*(2), 235-255.

Moreno, C., Borod, J., Welkowitz, J., & Alpert, M. (1993). The perception of facial emotion across the adult life span. *Developmental Neuropsychology, 9*, 305-314

Moriguchi, Y., Ohnishi, T., Kawachi, T., Mori, T., Hirakata, M., Yamada, M., & al., e. (2005). Specific brain activation in Japanese and Caucasian people to fearful faces. *Neuroreport, 16*(2), 133–136.

Morley, S., Eccleston, C., & Williams, A. (1999). Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. *Pain, 80*(1–2), 1-13. doi: http://dx.doi.org/10.1016/S0304-3959(98)00255-3

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*(6603), 812-815.

Murata, A., Moser, J.S., & Kitayama, S. (2013). Culture shapes electrocortical responses during emotion suppression. *Social Cognitive and Affective Neuroscience, 8*(5), 595-601.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: L. Erlbaum.

Naqvi, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision making: A cognitive neuroscience perspective. *Current Directions in Psychological Science, 15*(5), 260-264.

Nisbett, R. E., Peng, K. P., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*(2), 291–310.

Nygaard, L. C., & Queen, J. S. (2008). Communicating Emotion: Linking Affective Prosody and Word Meaning. *Journal of Experimental Psychology, 34*(4), 1017-1030.

Nygaard, L. C., & Queen, J. S. (2008). Communicating Emotion: Linking Affective Prosody and

Word Meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 1017-1030.

Oishi, S., Graham, J., Kesebir, S., & Galinha, I. C. (2013). Concepts of happiness across time and cultures. *Personality and Social Psychology Bulletin, 39*(5), 559-577.

Park, D. C., & Huang, C. M. (2010). Culture wires the brain: a cognitive neuroscience perspective. *Perspectives on Psychological Science, 5*(4), 391–400.

Paulman, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: How much is enough? . *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 230-241.

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology, 4*, 345.

Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language, 105*(1), 59-69.

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion, 35*(2), 192-201.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research, 1217*(C), 171-178.

Paulmann, S., Seifert, S., & Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience, 5*(1), 59-75.

Pell, M. D. (2005). Nonverbal emotion priming: Evidence from the 'Facial Affect Decision Task'.

Journal of Nonverbal Behavior, 29(1), 45-73.

Pell, M. D., & Baum, S. R. (1997). Unilateral brain damage, prosodic comprehension deficits, and the acoustic cues to prosody. *Brain and Language, 57*(2), 195–214.

Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective and Behavioral Neuroscience, 3*(4), 275-288.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing Emotions in a Foreign Language. *Journal of Nonverbal Behavior, 33*(2), 107-120.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37*(4), 417-435.

Pessoa, L. (2010). Emotion and cognition and the amygdala: from "what is it?" to "what's to be done?". *Neuropsychologia, 48*(12), 3416–3429.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. *NeuroImage, 16*(2), 331-348. doi: http://dx.doi.org/10.1006/nimg.2002.1087

Pizzagalli, D. A., Lehmann, D., Hendrick, A. M., Regard, M., Pascual-Marqui, R. D., & Davidson, R. J. (2002). Affective Judgments of Faces Modulate Early Activity (~160 ms) within the Fusiform Gyri. *NeuroImage, 16*(3, Part A), 663-677. doi: http://dx.doi.org/10.1006/nimg.2002.1126

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. (pp. 3-33). New York:

Academic press. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 3-33). New York.

Posamentier, M. T., & Abdi, H. (2003). Processing Faces and Facial Expressions. *Neuropsychology Review, 13*(3), 113-143.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715-734.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport, 11*(6), 1329-1333.

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

Righart, R., & de Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: An ERP study. *Social Cognitive and Affective Neuroscience, 3*(3), 270-278.

Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: emotional prosody implicitly guides visual attention to faces. *PLoS ONE, 7*, e30740.

Rigoulot, S., Wassiliwizky, E., & Pell, M. D. (2013). Feeling backwards? How temporal order in speech affects the time course of vocal emotion recognition *Frontiers in Psychology, 4*, 367.

Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian*

Jouranl of Experimental Psychology, 57(3), 221-237.

Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain and Language, 104*(1), 51-74.

Ruberg, W. (2009). Interdisciplinarity and the history of emotions. *Cultural and Social History, 6*(4), 507-516.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161 - 1178.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115*(1), 102-141.

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion *Psychological Review, 110*(1), 145-172.

Russell, J. A. (2009). Emotion, core affect, and psycholgical construction. *Cognition and Emotion, 23*(7), 1259-1283.

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology, 48* 1290-1298.

Saarni, C., Mumme, D., & Campos, J. (1998). Emotional development: Action, communication, and understanding. In N. Eisenberg (Ed.), *Handbook of Child Psychology, Volume 3: Social, Emotional and Personality Development* (pp. 237–310). New York: Wiley.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National*

Academy of Sciences of the United States of America, 107(6), 2408-2412.

Scherer, K. R. (1986). Vocal Affect Expression. A Review and a Model for Future Research. *Psychological Bulletin, 99*(2), 143-165.

Scherer, K. R. (1995). Expression of Emotion in Voice and Music. *Journal of Voice, 9*(3), 235–248.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York: Oxford University Press.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1-2), 227-256.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76-92.

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by apprraisal? , . *Emotion, 7*, 113-130.

Schimmack, U., & Bockenholt, U. (2002). Response styles in affect ratings: Making a mountain out of a molehill. *Journal of Personality Assessment, 78*, 461-483.

Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. European Journal of Personality. *European Journal of PersonalitY, 14*, 325-345.

Schirmer, A., Chen, C. B., Ching, A., Tan, L., & Hong, R. Y. (2013). Vocal emotions influence verbal memory: Neural correlates and inter-individual differences. *Cognitive, Affective and Behavioral Neuroscience,, 13*(80-93).

Schirmer, A., Escoffier, N., Li, Q. Y., Li, H., Strafford-Wilson, J., & Li, W. I. (2008). What grabs his attention but not hers? Estrogen correlates with neurophysiological measures of vocal change detection. *Psychoneuroendocrinology, 33*(6), 718-727.

Schirmer, A., & Kotz, S. A. (2003). ERP Evidence for a Sex-Specific Stroop Effect in Emotional Speech. *Journal of Cognitive Neuroscience, 15*(8), 1135-1148.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24-30.

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

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research, 24*(3), 442-452.

Schirmer, A., Lui, M., Maess, B., Escoffier, N., Chan, M., & Penney, T. B. (2006). Task and sex modulate the brain response to emotional incongruity in asian listeners. *Emotion, 6*(3), 406-417.

Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport, 16*(6), 635-639.

Schorr, A. (2001). Appraisal-The evolution of an idea. In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 20-34). New York:

Oxford University press.

Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, and Computers, 30*(1), 131-145.

Scollon, R., & Scollon, S. W. (1995). *Intercultural Communication: A Discourse Approach*. Oxford: Blackwell.

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Chapter 5.5 - Brain mechanisms for processing perceived emotional vocalizations in humans. In M. B. Stefan (Ed.), *Handbook of Behavioral Neuroscience* (Vol. Volume 19, pp. 187-197): Elsevier.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America, 90*(4 pt 1), 1797–1805.

Soken, N. H., & Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child development, 63*(4), 787-795.

Stagg, C., Hindley, P., Tales, A., & Butler, S. (2004). Visual mismatch negativity: The detection of stimulus change. *Neuroreport, 15*(4), 659-663.

Stein, N. L., & Oatley, K. (1992). Basic emotions: Theory and measurement. *Cognition and Emotion, 6*(3-4).

Stemmler, G., & Wacker, J. (2010). Personality, emotion, and individual differences in physiological responses. *Biological Psychology, 84*(3), 541-551.

Sue, D. W., & Sue, D. (1977). Barriers to effective cross-cultural counseling. *Journal of Counseling Psychology, 24*(5), 420-429.

Susac, A., Ilmoniemi, R. J., Pihko, E., & Supek, S. (2004). Neurodynamic studies on emotional and inverted faces in an oddball paradigm. *Brain Topography, 16*(4), 265-268.

Talsma, D., Slagter, H. A., Nieuwenhuis, S., Hage, J., & Kok, A. (2005). The orienting of visuospatial attention: An event-related brain potential study. *Cognitive Brain Research, 1*(25), 117-129.

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I Feel Your Voice: Cultural Differences in the Multisensory Perception of Emotion. *Psychological Science, 21*(9), 1259-1262.

Tanaka, J. W., & Pierce, L. J. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective and Behavioral Neuroscience, 9*(1), 122-131.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*(4), 297-303.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America, 106*(11), 4567-4570.

Thomas, C., & Baker, C. I. (2013). Teaching an adult brain new tricks: A critical review of evidence for training-dependent structural plasticity in humans. *NeuroImage, 73*(0), 225-236. doi: http://dx.doi.org/10.1016/j.neuroimage.2012.03.069

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica, 158*, 407-424.

Tomkins, S. (1962). *Affect Imagery Consciousness: Volume I, The Positive Affects*. London:

Tavistock.

Tomkins, S. (1963). *Affect Imagery Consciousness: Volume II, The Negative Affects*. London: Tavistock.

Tomkins, S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills, 18*(1), 119-158.

Trimmer, C. G., Meyer-MacLeod, R., Cuddy, L. L., & Balkwill, L. L. (2007). Acoustic cues and recognition accuracy in cross-cultural vocal expression of emotion. *Canadian Acoustics - Acoustique Canadienne, 35*(3), 76-77.

Uchida, Y., Norasakkunkit, V., & Kitayama, S. (2004). Cultural constructions of happiness: theory and emprical evidence. *Journal of Happiness Studies, 5*(3), 223-239.

van den Stock, J. B., Grèzes, J., & de Gelder, B. (2008). Human and animal sounds influence recognition of body language. *Brain Research, 1242*(11), 185-190.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition and Emotion, 22*(4), 720-752.

Vizioli, L., Foreman, K., Rousselet, G. A., & Caldara, R. (2010). Inverting faces elicits sensitivity to race on the N170 component: A cross-cultural study. *Journal of Vision, 10*(1), 1-23.

Voos, A. C., Pelphrey, K. A., Tirrell, J., Bolling, D. Z., Wyk, B. V., Kaiser, M. D., . . . Ventola, P. (2013). Neural mechanisms of improvements in social motivation after pivotal response treatment: Two case studies. *Journal of Autism and Developmental Disorders, 43*(1), 1-10.

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective and Behavioral Neuroscience, 1*(4), 382-387.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience, 22*(12), 2864-2885.

Waller, B. M., & Micheletta, J. (2013). Facial Expression in Nonhuman Animals. *Emotion Review, 5*(1), 54-59

Ward, C., Bochner, S., & Furnham, A. (2001). *The Psychology of Culture Shock*. East Sussex, UK.: Routledge.

Ward, C., & Kennedy, A. (1993). Where's the culture' in cross-cultural transition? Comparative study of sojourner adjustment. *Journal of Cross-Cultural Psychology, 24*(2), 221-249.

Way, B. M., & Lieberman, M. D. (2010). Is there a genetic contribution to cultural differences? Collectivism, individualism and genetic markers of social sensitivity. *Social Cognitive and Affective Neuroscience, 5*(2-3), 203-211. doi: 10.1093/scan/nsq059

Wei, J., Chan, T., & Luo, Y. (2002). A modified oddball paradigm "cross-modal delayed response" and the research on mismatch negativity. *Brain Research Bulletin, 57*(2), 221-230.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology, 11*(4), 277-281.

Williams, L. M., Palmer, D., Liddell, B. J., Song, L., & Gordon, E. (2006). The "when" and "where" of perceiving signals of threat versus non-threat. *NeuroImage, 31*(1), 458–467.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics, 38*(10), 1559-1579.

Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Engelmann.

Xie, M. (2011). *What Their Stories Tell Us: First Language (L1) Maintenance and Attrition among Young Chinese Adult Immigrants: A Multi-Case Study* Saarbrücken, Germany: VDM Verlag

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat Neurosci, 15*(4), 528-536.

Zhao, L., & Li, J. (2006). Visual mismatch negativity elicited by facial expressions under non-attentional condition. *Neuroscience Letters, 410*(2), 126-131.