Neuropsychologia 67 (2015) 1–13

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

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

Pan Liu [a,*], Simon Rigoulot [b], Marc D. Pell [a,b]

[a] School of Communication Sciences and Disorders, McGill University, Montréal, Québec, Canada
[b] International Laboratory for Brain, Music and Sound Research (BRAMS), Centre for Research on Brain, Language, and Music (CRBLM), McGill University, Montréal, Québec, Canada

ARTICLE INFO

Article history:
Received 27 July 2014
Received in revised form
24 November 2014
Accepted 30 November 2014
Available online 2 December 2014

Keywords:
Cross-cultural studies
ERP
Stroop
Facial expression
Emotional prosody

ABSTRACT

To understand how culture modulates on-line neural responses to social information, this study compared how individuals from two distinct cultural groups, English-speaking North Americans and Chinese, process emotional meanings of multi-sensory stimuli as indexed by both behaviour (accuracy) and event-related potential (N400) measures. In an emotional Stroop-like 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 focus task). Results indicated that while both groups were sensitive to emotional differences between channels (with lower accuracy and higher N400 amplitudes for incongruent face–voice pairs), there were marked group differences in how intruding facial or vocal cues affected accuracy and N400 amplitudes, with English participants showing greater interference from irrelevant faces than Chinese. 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 not only behaviour, but the neurocognitive response to different features of multi-channel emotion expressions.

1. Introduction

In today's highly globalized world, social interactions involving people from different cultural backgrounds 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 multi-sensory emotional expressions (combined faces and voices) beyond the level of behavioural performance. 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.

1.1. Cultural differences in cross-channel emotion processing

Humans typically utilize multiple information channels to express and understand emotions in daily life (Grandjean et al., 2006; Paulmann and Pell, 2011); for example, 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 information sources, at least when behavioural responses are examined. In a key study conducted by Tanaka et al. (2010), Japanese and Dutch participants performed an emotional Stroop-like task consisting of emotionally congruent or incongruent face–voice pairs; participants had to identify the emotion of the face while ignoring the voice, or vice versa. 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. Tanaka et al. found that Japanese participants demonstrated a significantly larger congruence effect in accuracy than Dutch participants when judging facial expressions, whereas the Dutch exhibited a smaller congruence effect when identifying the tone. This suggests that Japanese participants were influenced to a larger extent by to-be-

* Correspondence to: School of Communication Sciences and Disorders, McGill University, 2001 McGill College, 8th floor, Montréal, Québec, Canada H3G 1A8. Fax: +1 514 398 8123.
E-mail address: pan.liu@mail.mcgill.ca (P. Liu).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.11.034
0028-3932/

ignored vocal tones, whereas Dutch participants were influenced more by irrelevant faces. The authors concluded that Japanese people are more likely to allocate their attention to vocal cues in speech, whereas Westerners are more attentive to facial information even when instructed to ignore it, highlighting important differences in how each cultural group processes emotional cues.

These results are in line with previous findings pointing to cultural differences in how emotions are perceived from combined vocal cues ("speech prosody") and semantic word meanings. Kitayama and Ishii (2002) presented a Stroop-like task consisting of semantically negative or positive words, spoken in an emotionally congruent or incongruent vocal tone, to groups of Japanese and American participants who had to identify the emotion of the word meaning while ignoring the tone, or vice versa. They reported that for Japanese participants, response times when judging word meaning were influenced to a larger extent by to-be-ignored vocal tones than for American participants, whereas the opposite group pattern was observed when identifying the tone. This pattern again suggests that Japanese participants were more attuned to vocal cues when performing the task, whereas Americans were more sensitive to semantic cues. Coupled with Tanaka et al.'s (2010) findings, these results imply that during multichannel emotion perception, East Asians are relatively more sensitive to vocal information whereas Westerners are more oriented towards semantic or facial cues. This conclusion is also suggested in a study of Tagalog–English bilinguals tested in the Philippines (Ishii et al., 2003). Collectively, these data argue that cultures vary in how they process emotional expressions, although evidence is so far restricted to off-line behavioural 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 central role in emotion perception (Engelmann and Pogosyan, 2013; Ishii et al., 2003; Park and Huang, 2010). In contrast to Western "individualist" cultures, it is believed that East Asian "collectivist" cultures consider harmonious social relations as most important (Hall and Hall, 1990; Scollon and Scollon, 1995). To maintain social harmony, certain display rules are therefore adopted by these groups (Gudykunst et al., 1996; Matsumoto et al., 1998); for example, East Asian cultures tend to favour indirectness in communication (e.g., unfinished sentences or vague linguistic expressions) to a greater extent than 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 facial expressions of emotion (Ekman, 1971; Markus and Kitayama, 1991; Matsumoto et al., 1998, 2008, 2005). East Asians are also more inclined to avoid eye contact than Westerners, as a sign of respect and politeness to others that is thought to benefit social harmony (Hawrysh and Zaichkowsky, 1991; McCarthy et al., 2006, 2008). Due to these culture-specific display rules, it is possible that East Asians rely more on vocal cues to express their emotions and to understand the feelings of others during interpersonal communication; in contrast, Western cultures tend to rely more on facial and semantic information which tends to be more salient and accessible during their social interactions (Ishii et al., 2003; Kitayama and Ishii, 2002; Tanaka et al., 2010).

These ideas lead to the question: 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 valuable to advance this literature by examining 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 promote differences in how Eastern and Western cultures process facial and vocal cues about emotion, one might expect 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 behavioural and on-line *neural* responses—data that begin to inform the nature as well as the time course of cultural effects on multi-channel emotion processing—was another major goal of this study.

1.2. Cultural effects on neurophysiological responses to emotional stimuli

While cultural differences in *behavioural* judgments of emotional displays have been reported, there is still little evidence that culture affects brain responses that occur as humans process cross-channel emotions in real time. Electrophysiological data show that cross-channel emotion perception is a dynamic process marked by various temporal stages that can be indexed by different event-related brain potentials (ERPs) (de Gelder et al., 1999; Ishii et al., 2010; Jessen and Kotz, 2011; Schirmer and Kotz, 2003). For instance, using an Oddball-like task consisting of emotional face–voice pairs that were presented to Dutch participants, de Gelder et al. (1999) observed a mismatch negativity (MMN) component suggesting that early cross-sensory integration of emotional cues occurs very quickly, approximately 100–200 milliseconds (ms) after stimulus onset. Similarly, in an ERP study that presented a Stroop-like task, presented positive, negative, and neutral words spoken in a congruent or incongruent emotional voice to German participants, who identified the valence of the word meaning and ignored the tone, or vice versa. In this study they observed a larger *N400*—a component sensitive to the integration of matching/mismatching semantic meanings from different contextual sources (Bower, 1981; Bowers et al., 1993; Brown et al., 2000; Schirmer and Kotz, 2003)—in response to incongruent versus congruent word–tone pairings in the word valence judgment (although only in female participants). Differences in N400 amplitude were also reported by Ishii et al. (2010) using an emotional Stroop-like task that contrasted emotional tone and emotional word meanings in Japanese, with greater amplitudes when cross-channel emotional cues were incongruent in meaning. Although these studies did not directly compare participants from different cultural backgrounds, they suggest that the N400 response can be meaningfully examined to shed light on whether cultural differences exist at the neural processing stage where the emotional meaning of a face and a vocal expression is registered and compared, around 400 ms after event onset, allowing deeper specification of behavioural findings on face–voice processing (e.g., Tanaka et al., 2010).

In another noteworthy study, Schirmer et al. (2006) conducted an EEG experiment in which they presented an emotional Stroop-like task to Cantonese speakers from Hong Kong. Two-syllable Cantonese words spoken in a congruent or incongruent emotional voice were presented to Cantonese listeners, who identified the emotion of the word meaning while ignoring the vocal tone as brain potentials were recorded. Results of this study failed to uncover a typical Stroop/N400 effect in the Cantonese as that reported for German participants (Schirmer and Kotz, 2003); rather, a late positivity peaking around 1500 ms was observed with 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 listeners was related to the existence of lexical tones and a higher degree of homophony in the Cantonese language; this may have rendered lexical processing more effortful and postponed the integration of emotional voice tone with the semantic message in this study (Schirmer et al., 2006).