Brain and Cognition 95 (2015) 99–106

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Multimodal processing of emotional information in 9-month-old infants I: Emotional faces and voices

R.A. Otte [a], F.C.L. Donkers [a,b,c], M.A.K.A. Braeken [a,d], B.R.H. Van den Bergh [a,e,f,*]

[a] Department of Psychology, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands
[b] Department of Psychiatry, University of North Carolina at Chapel Hill, 101 Manning Drive, Chapel Hill, NC 27514, USA
[c] Department of Cognitive Neuroscience, Maastricht University, Oxfordlaan 55, 6229 EV Maastricht, The Netherlands
[d] Faculty of Medicine and Life Sciences, Hasselt University, Agoralaan, Building C, 3590 Diepenbeek, Belgium
[e] KU Leuven, Department of Psychology, Tiensestraat 102, PO Box 3726, B-3000 Leuven, Belgium
[f] Department of Wellbeing, Health and Family, Flemish Government, PO Box 30, 1030 Brussels, Belgium

ARTICLE INFO

Article history:
Received 31 July 2013
Revised 15 September 2014
Accepted 22 September 2014

Keywords:
Infant
Multimodal processing
Audiovisual processing
Event-related potential (ERP)
Emotion perception
Affective priming

ABSTRACT

Making sense of emotions manifesting in human voice is an important social skill which is influenced by emotions in other modalities, such as that of the corresponding face. Although processing emotional information from voices and faces simultaneously has been studied in adults, little is known about the neural mechanisms underlying the development of this ability in infancy. Here we investigated multimodal processing of fearful and happy face/voice pairs using event-related potential (ERP) measures in a group of 84 9-month-olds. Infants were presented with emotional vocalisations (fearful/happy) preceded by the same or a different facial expression (fearful/happy). The ERP data revealed that the processing of emotional information appearing in human voice was modulated by the emotional expression appearing on the corresponding face: Infants responded with larger auditory ERPs after fearful compared to happy facial primes. This finding suggests that infants dedicate more processing capacities to potentially threatening than to non-threatening stimuli.

1. Introduction

Humans are social beings, often living together in close quarters and as such communication with others is a significant part of our day-to-day life. In this social context human voices are one of the most important stimuli in our auditory environment. They not only convey semantic information, but also information about one's identity and emotional state (Belin, Fecteau, & Bédard, 2004; Grossmann, Oberecker, Koch, & Friederici, 2010; Latinus & Belin, 2011). That voices are special is emphasised by the existence in the adult human brain of voice-selective regions along the upper bank of the superior temporal sulcus dedicated to the processing of human vocal sounds – both speech and non-speech vocalisations (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009). The amygdala, inferior prefrontal cortex, and insula have been found to be involved in the processing of affective information in voices (Belin et al., 2004; Blasi et al., 2011). How emotional information from vocalisations is processed partly depends on information from other modalities, such as visual input from facial expressions. For example, deGelder and Vroomen (2000) and deGelder, Pourtois, and Weiskrantz (2002) found that in adults both recognition and judgement of emotion in voices is modulated by consciously as well as unconsciously recognised emotion in faces, providing evidence that the brain uses information from both modalities to interpret emotions. The goal of the current study was to examine whether this is also true for 9-month-old infants: Do they process emotional vocalisations differently when they have been primed with a visual stimulus conveying the same, versus a different emotion?

From a developmental perspective, studying how emotional information from faces may modulate the processing of emotional content from voices is important, for example for understanding how interpersonal skills develop, such as interaction with others by reading their emotions (Grossmann, 2010; Walker-Andrews, 1997). The ability to process emotional information from different modalities simultaneously appears to develop quite early in human life. Behavioural experiments, for instance, have found that by 3–5 months of age recognition of affect emerges in bimodal

Abbreviations: GA, gestational age; PELS, prenatal early life stress study; Nc, negative component; Pc, positive component.

* Corresponding author at: Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. Fax: +31 (0)13 466 2067.

E-mail addresses: r.a.otte@tilburguniversity.edu (R.A. Otte), franc.donkers@maastrichtuniversity.nl (F.C.L. Donkers), m.braeken@tilburguniversity.edu (M.A.K. A. Braeken), bea.vdnbergh@tilburguniversity.edu (B.R.H. Van den Bergh).

http://dx.doi.org/10.1016/j.bandc.2014.09.007
0278-2626/

stimulation, first in familiar and then in unfamiliar contexts and persons (Kahana-Kalman & Walker-Andrews, 2001; Walker-Andrews, 1997), as evidenced by discrimination between happiness, anger and sadness (Flom & Bahrick, 2007; Leppänen & Nelson, 2008). For successful differentiation at this age, though, it is necessary that there is temporal synchrony between face and voice, i.e. speech should be played in synchrony with lip movement (Flom & Bahrick, 2007). From 7 months of age infants are able to detect happiness, anger and sadness across audio–visual modalities without needing temporal synchrony between faces and voices (Flom & Bahrick, 2007; Soken & Pick, 1992; Walker-Andrews, 1997).

To date, electrophysiological measures such as event-related potentials (ERPs) have been seldom used to study multimodal processing of emotional information in infancy. However, as ERPs can be recorded in the absence of a behavioural response (Nelson & Bloom, 1997), even for unattended stimuli (Sussman, 2007), they are quite suitable for studying emotion processing in infants. Indeed, in previous infant research on emotional faces (e.g. deHaan, Johnson, & Halit, 2003; Leppänen, Moulson, Vogel-Farley, & Nelson, 2007), emotional voices (Grossmann, Striano, & Friederici, 2005) and emotional face/voice pairs (Grossmann, Striano, & Friederici, 2006), the use of ERPs helped to gain insight into the development of the underlying mechanisms. For example, in the study by Grossmann et al. (2006), the authors presented their 7-month-old subjects with a happy or angry static facial expression (a prime). After a 400 ms-delay, a word was spoken in an emotionally congruent or incongruent tone of voice. Faces remained visible until the end of the presentation of the word. The authors found that the emotionally incongruent condition elicited a larger auditory Negative component (Nc) around 500 ms post-stimulus. In contrast, the emotionally congruent condition elicited a larger auditory Positive component (Pc) approximately 800 ms after stimulus onset. Grossmann et al. (2006) concluded that the attenuation of the Nc and enhancement of the later Pc reflected recognition of the familiar/expected face/voice pairs, and that the infants had thus recognised and processed emotions from both modalities.

In comparison with studies on processing of facial emotional stimuli in infancy (e.g. deHaan, Belsky, Reid, Volein, & Johnson, 2004; Nelson & de Haan, 1996; Striano, Brennan, & Vanman, 2002), few studies have addressed the processing of vocal emotional stimuli and even fewer the processing of emotional voices in the context of emotional facial expressions. However, as the auditory system develops earlier than the visual system (Anderson & Thomason, 2013; Anderson et al., 2001), from a developmental perspective, emotional vocalisations may be just as relevant as facial expressions in the first months of life. This is supported by findings that 5-month-olds do respond to emotional vocalisations in the absence of facial emotional expressions, but not vice versa (Fernald, 1993). Also, Caron, Caron, and MacLean (1988) found that 5- to 7-month-olds rely more on auditory than visual input when discriminating emotional expressions. In addition, results from a study by Mumme, Fernald, and Herrera (1996) suggested that information from the mother's voice alone, but not from her face only, can be sufficient in guiding 12-month-olds' behaviour in ambiguous situations. Therefore, in the current study we examined the processing of emotional vocalisations (fearful and happy) in a large group ($N = 84$) of 9-month-old infants after priming them by a visual stimulus conveying the same, versus a different emotion.

We hypothesised that (1) the emotional quality of the Visual Prime will modulate the response to the following voice; and (2) that emotional (in)congruency between the Visual Prime and the following voice will modulate the ERP response to the latter. Based on Grossmann et al. (2006), we expected both the Nc and Pc to be modulated. Research from Kushnerenko et al. (2002) showed that, already from birth, infants respond to auditory stimuli with a P150-N250-P350-N450 ERP complex (where the N450 is approximately equivalent to the Nc described by Grossmann et al. (2006)). These infant components are speculated to be precursors of child and adult components, i.e. P1, N250 (N2), P3a, N450 (N4) (Kushnerenko et al., 2002). The infants in our study were slightly older than those in the study of Grossmann et al. (2006) and might therefore show a more adult-like component structure. Thus, we also took the earlier components (P150, N250, P350) described by Kushnerenko et al. (2002) into account. Because the Nc and Pc are usually used to describe visual instead of auditory ERPs, in the current study, we refer to the five components of interest as P150, N250, P350, N450 (Nc) and P650 (Pc), respectively.

2. Methods

2.1. Subjects

Subjects were 84 infants (one pair of twins) and their mothers from a normal (i.e. non-clinical) population who have been taking part in a longitudinal study on prenatal early life stress (PELS project). The study was approved by the Medical Ethical Committee of St. Elizabeth Hospital in Tilburg, The Netherlands. Informed consent was obtained from all mothers and fathers in accordance with the Declaration of Helsinki. Detailed information on the cohort and its recruitment has been described previously in Otte et al. (2013).

In short, the cohort consists of 190 women – and their partner and child – who have been recruited during pregnancy, either before 15 weeks gestational age (GA; $N = 178$) or between week 16 and 22 ($N = 12$) of gestation, from a general hospital and four midwives' practices in Tilburg, The Netherlands. Women were followed up three times during their pregnancies (measurement waves T1, T2 and T3, respectively) and were invited to the lab for postpartum observations both 2 to 4 months (T4) and 9 to 11 months (T5) after giving birth. Here, we report results from infants measured at T5; data collected at T4 have been discussed elsewhere (Otte et al., 2013; van den Heuvel et al., in preparation).

At T5 147 of the original 190 women came in for testing with their infant (one pair of twins). Forty-three women did not participate in this measurement wave, because of drop out before T5 ($N = 32$), because they could not be reached in time (6), they were ($N = 1$) or their infant was ($N = 2$) too ill, they had miscarried around T2 ($N = 1$), or their infant had passed away ($N = 1$). Three of the 147 mothers had delivered prematurely, and 1 mother had delivered a baby small for gestational age (GA; e.g. birth weight < 2500 g at term delivery). Data for infants of these mothers ($N = 4$) were excluded from analysis beforehand. Data for an additional 60 of the remaining 144 infants were later excluded because of too little remaining data after removing invalid trials (e.g. with movement artefacts and where the infant had not looked at the stimulus; $N = 33$), fussiness ($N = 13$), and technical problems (e.g. severe problems with mastoids; $N = 14$). This attrition rate (41.2%) is similar to other infant ERP studies (DeBoer, Scott, & Nelson, 2007). All infants were healthy and had passed a screening test for hearing impairments (evoked otoacoustic emission), performed by a nurse from the infant health care clinic, between the 4th and 7th day after birth. The mean age at testing of the 84 infants (45 girls) included in the sample was 303 days ($SD = 14$ days). Mean GA and mean birth weight were 39.9 weeks ($SD = 8.7$ days) and 3477 g ($SD = 464$ g), respectively.

2.2. Stimuli

2.2.1. Visual stimuli

Visual stimuli were 18 colour photos of 9 Caucasian women in frontal view, each expressing both happiness and fear. In contrast