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

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

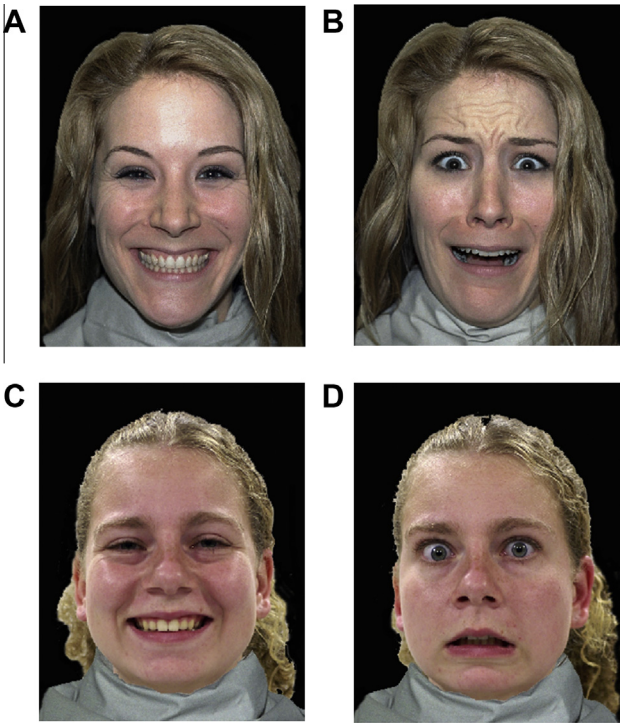

Fig. 1. Examples of a typical happy (A) and fearful (B) visual expression from the NimStim dataset, and of a typical happy (C) and fearful (B) visual expression from the Tilburg dataset.

to the study by Grossmann et al. (2006), we opted to use fear instead of anger, to explore whether we would find similar results with a different emotion. Also, for our ongoing study on prenatal early life stress we were interested in effects of exposure to fearful stimuli (see our twin-paper Otte, Donkers, Braeken, & Van den Bergh., submitted for publication).

Only female and Caucasian identities had been chosen so as to avoid any sex or ethnicity differences from influencing the infant ERPs (e.g. see Ramsey, Langlois, & Marti, 2005; Vogel, Monesson, & Scott, 2012). The emotional faces had been cut out from their original background and pasted onto a black background. Four identities were taken from the validated NimStim Face Stimulus Set (http://www.macbrain.org/resources.htm[1]). Their fearful and happy expression had been recognised at least 75% of the time in the validation study. In Fig. 1A and B examples of a happy and fearful face from the NimStim set can be found. An additional 5 identities were included from a database with emotional facial stimuli from the Cognitive and Affective Neurosciences Laboratory at Tilburg University, The Netherlands. This was done because we wanted to include as many identities as possible to minimise the possibility that potential effects would be due to the identity itself, and the NimStim set did not offer more than 4 female Caucasian identities whose fearful and happy expression survived the validation process. The stimuli from the Tilburg database had been validated in a pilot study in which they were rated for emotion (fear, happiness, anger, neutral, surprise, sadness and disgust), intensity (scale from 1 to 5) and positive/negative affect (scale from1 to 5) by at least 8 participants. For the 5 identities used for this study both the emotion fear and happiness were correctly recognised by the raters 80% of the time or more. In Fig. 1C and D examples of a happy and fearful face from the Tilburg set can be found.

2.2.2. Auditory stimuli

Auditory stimuli were voice recordings of six women expressing fear and happiness with non-verbal vocalisations. We chose to use non-verbal vocalisations as to minimise automatic semantic or verbal processing (Van den Stock, Grèze, & de Gelder, 2008). Because not all vocalisations had been recognised well enough in the validation study (see below), for two of the identities we only used the fearful vocalisation, for two of them we only used the happy vocalisation and for two of them we used vocalisations of both emotions, resulting in 4 fearful and 4 happy auditory stimuli. These stimuli were provided by the Cognitive and Affective Neurosciences Laboratory at Tilburg University also, and were recorded, processed and validated as described by Van den Stock et al. (2008). In short, semi-professional actors were asked to make a frightened or happy sound, based on a specific script describing situations such as an attack by a robber. Audio recordings were made at a 44.1 kHz sampling rate and an intensity level of 75 dB. They were originally of 800 ms duration, and were shortened to 500 ms for the experiment described here. The sounds were validated in a pilot session by 15 participants, who were instructed to categorise as accurately and as fast as possible the emotion expressed by the voices (fear or happiness). All stimuli used here had been correctly recognised 80% of the time or more.

2.3. Procedure

Each of the 18 visual stimuli was paired with both the fearful and happy auditory stimuli. This created a relatively large set of face/voice compounds, minimising the chance that potential effects would be caused by specific compounds. The 144 compounds comprised four experimental conditions: happy face-happy voice (HH), fearful face-fearful voice (FF), happy face-fearful voice (HF), and fearful face-happy voice (FH). They were presented twice during the experiment (288 trials), divided in four blocks of 72 stimuli, each. The presentation order within each of the blocks was randomised and the order between the blocks was counterbalanced. The blocks were presented with small breaks in between as needed.

During stimulus presentation, the infant was seated on its parent's lap in a dimly lit and sound-attenuated room. The parent-infant dyad were seated behind a desk with a computer screen (CRT VGA, 21 inch, 1280 × 1024, 100 Hz) at a distance of approximately 70 cm from the eyes. The visual stimuli measured 18, 5 × 22, 5 cm and the horizontal and vertical visual angles were 7.53° and 9.13°, respectively. Auditory stimuli were presented through speakers positioned on either side of the screen, and at a distance of approximately 90 cm from the infant's head. To prevent the parent from influencing the infant's ERP responses by unconsciously reacting to the auditory stimuli, he or she was wearing headphones through which classical music was playing.[2] Two cameras filmed the experimental session and these data were later used to code whether the infant had looked at a specific trial or not (see also Section 2.4).

Each stimulus block started with the sound of a laughing baby and the presentation of a red dot growing bigger and smaller in the centre of the computer screen to attract the infant's attention. When the infant was looking at the screen, the experimenter started the first trial. Each trial lasted 1400 ms and started with the presentation of a visual stimulus that lasted 900 ms, after which an auditory stimulus was presented. A larger interval

[1] Development of the MacBrain Face Stimulus Set was overseen by Nim Tottenham and supported by the John D. and Catherine T. MacArthur Foundation Research Network on Early Experience and Brain Development. Please contact Nim Tottenham attott0006@tc.umn.edu for more information concerning the stimulus set.

[2] Since for this experiment we were interested in the infants' responses to emotional vocalisations, we did not prevent the mothers from seeing (and potentially responding to) the visual stimuli.

between presentation of faces and voices compared to that in the study by Grossmann et al. (2006) was chosen for decreasing the chance that differences in the late ERP responses to the visual stimuli would overlap and wash out potential response differences in response to the auditory stimuli. The visual stimulus remained in place until the auditory stimulus was played out (after 500 ms). Each trial was followed by an inter-trial interval with variable duration (between 600 and 1000 ms) to reduce temporal predictability. During this interval, the screen was black.

When an infant looked away from the screen, the experimenter tried to recapture the infant's attention by presenting an attractive moving figure in the centre of the screen. As soon as the infant was looking at the screen again, the experiment continued. The experiment was concluded either after all 288 stimuli had been presented or when the infant became too fussy to continue.

2.4. Data acquisition and analysis

EEG was recorded with BioSemi ActiveTwo amplifiers (BioSemi, Amsterdam, The Netherlands) with a sampling rate of 512 Hz.[3] Infants wore head caps with 64 electrode locations positioned according to the revised version of the International 10–20 system. The standard BioSemi reference (CMS-DRL) was used (see www.biosemi.com/faq/cms&drl.htm for details) and two additional electrodes were placed both on the left and right mastoid. Off-line, these were mathematically combined to produce an average mastoids reference derivation (Luck, 2005).

Before the data were processed further, independent raters inspected the data and scored per infant per trial whether or not he or she had indeed looked at the visual stimulus. All trials were scored by 2 different raters (there were 4 raters in total) and these scorings were afterwards compared. Agreement between raters lay between 81% and 99% and was 95% on average. Whenever scorings differed, the trials concerned were re-inspected and scored again. If there was still doubt about whether the infant had actually seen the stimulus, the trial was excluded. Trials during which an infant was crying were also excluded. Only EEG signals from trials during which the infant was looking at the stimulus were used for data analysis.

The EEG data were analysed using BrainVisionAnalyzer software (Brain Products, Munich, Germany). The continuous EEG signals were filtered off-line, with a 50 Hz notch filter (to make sure all line noise would be removed) and a 0.1 to 20 Hzband-pass filter (slope 24 dB). The signals were then segmented into 1000 ms-long epochs, time-locked to the onset of the auditory stimulus. The 200 ms before auditory stimulus onset were used as the baseline. To make sure that this auditory baseline did not differ between happy and fearful facial expressions, we tested it. We did this by first calculating -per subject, per condition – the average amplitude in the auditory baseline period (−200 to 0 ms before voice onset), using the 200 ms before face onset (so −1100 to −900 ms before voice onset) as baseline. Then, we ran a 4-way ANOVA with the factor "Condition" (4 levels: FH, FF, HF and HH) to test whether there were differences between the 4 conditions in the auditory baseline period. The result ($F(3,267) = 2, 192$, $p > .05$) showed no differences between the four conditions, and we concluded the 200 ms before onset of the auditory stimulus were suitable for use as the baseline.[4]

The epochs were averaged separately for each of the four conditions (FF, HH, HF, FH). Epochs with sample-to-sample voltage steps exceeding 80 μV were excluded, as were epochs with amplitudes exceeding 150 μV in any 200 ms-long window within the whole epoch, and those where the amplitude range in any 100-ms window, was below 0.5 μV (i.e., flat lines). Data from infants with less than 14 acceptable responses for any one of the four condition were removed from further analysis (33 infants, see also Section 2.1). The average number of available trials per infant was 36.4 for FF ($SD = 12.6$), 37.5 for HH ($SD = 12.0$), 36.6 for HF ($SD = 11.8$), and 37.1 for FH ($SD = 12.8$).

Time windows for analysis were selected based on visual inspection of the grand average waveform for all responses combined at electrode sites F3, Fz, F4, C3, Cz, C4 (see Fig. 1), where responses were largest. The following windows, each centred around a peak in the grand average waveform, were chosen: 120–200 ms, 200–260 ms, 290–430 ms, 380–540 ms, and 620–680 ms post-stimulus. The fourth (380–540 ms) and fifth (620–680 ms) windows are approximately equivalent to the Nc and Pc described by Grossmann et al. (2006). Together with the first three windows, the 380–540 ms window corresponds to the infant P150-N250-P350-N450 ERP complex in response to auditory stimuli described by Kushnerenko et al. (2002).

Amplitudes measured from the above mentioned windows and from electrode sites F3, Fz, F4, C3, Cz, C4 were analysed by means of a $2 \times 3 \times 2 \times 2$ repeated measures ANOVAs design (run in IBM SPSS 19.0), separately for each window. Within-subjects factors were "Frontal-central" (frontal, central), "Laterality" (left, medial, right), "Visual Prime" (fearful, happy), and "Auditory Emotion" (fearful, happy). As we were especially interested in effects of priming by the visual stimulus on the auditory stimulus, post hoc simple interactions and simple effects including the factor Visual Prime were run to further analyse any significant main/interaction effect. Greenhouse–Geisser correction was used where necessary and the ε correction factor is given, together with the partial η^2 effect size.

3. Results

The grand-averaged waveform for all responses combined can be found in Fig. 2. Fig. 3 shows the grand-averaged responses per condition. For investigating our hypotheses only effects involving Visual Prime will be interpreted here, because this factor represents effects of the visually presented emotion on the processing of the emotional sounds (see Table 1 for statistical information). However, significant results involving other factors can also be found in Table 1.

3.1. Results for P150 (120–200 ms post-stimulus)

The 4-way ANOVA yielded an interaction between the factors Visual Prime and Frontal-Central. Post hoc tests showed that if the visual stimulus had been fearful, the auditory responses had larger positive amplitudes on central electrode sites than when the visual stimulus had been happy ($t(83) = 2.032$; $p < .001$). On frontal electrodes sites response amplitudes for the fearful and happy visual stimuli did not significantly differ from each other ($t(83) = .935$; $p < .05$).

3.2. Results for N250 (200–260 ms post-stimulus)

The analysis yielded a main effect for the factor Visual Prime. Post hoc tests showed that when the visual stimulus had been fearful, responses to the auditory stimulus had been of smaller negative-going amplitudes than when the visual stimulus had been happy ($t(83) = 1.855$; $p < .05$).

[3] On-line filtering options in the BioSemi software are set automatically and depend on the chosen sampling rate. See http://www.biosemi.com/faq/adjustfilter.htm and http://www.biosemi.com/faq/adjustsamplerate.htm.

[4] An alternative would have been to use a baseline before face onset (e.g. −1100 to −900 before voice onset). However, the relatively large amplitude (and variance) of the visual ERPs compared to the auditory ones could lead to the auditory ERP being 'washed out'. In other words, using the visual baseline for estimating the auditory ERP amplitudes could lead to loss of the true possible modulations of the auditory ERP components.

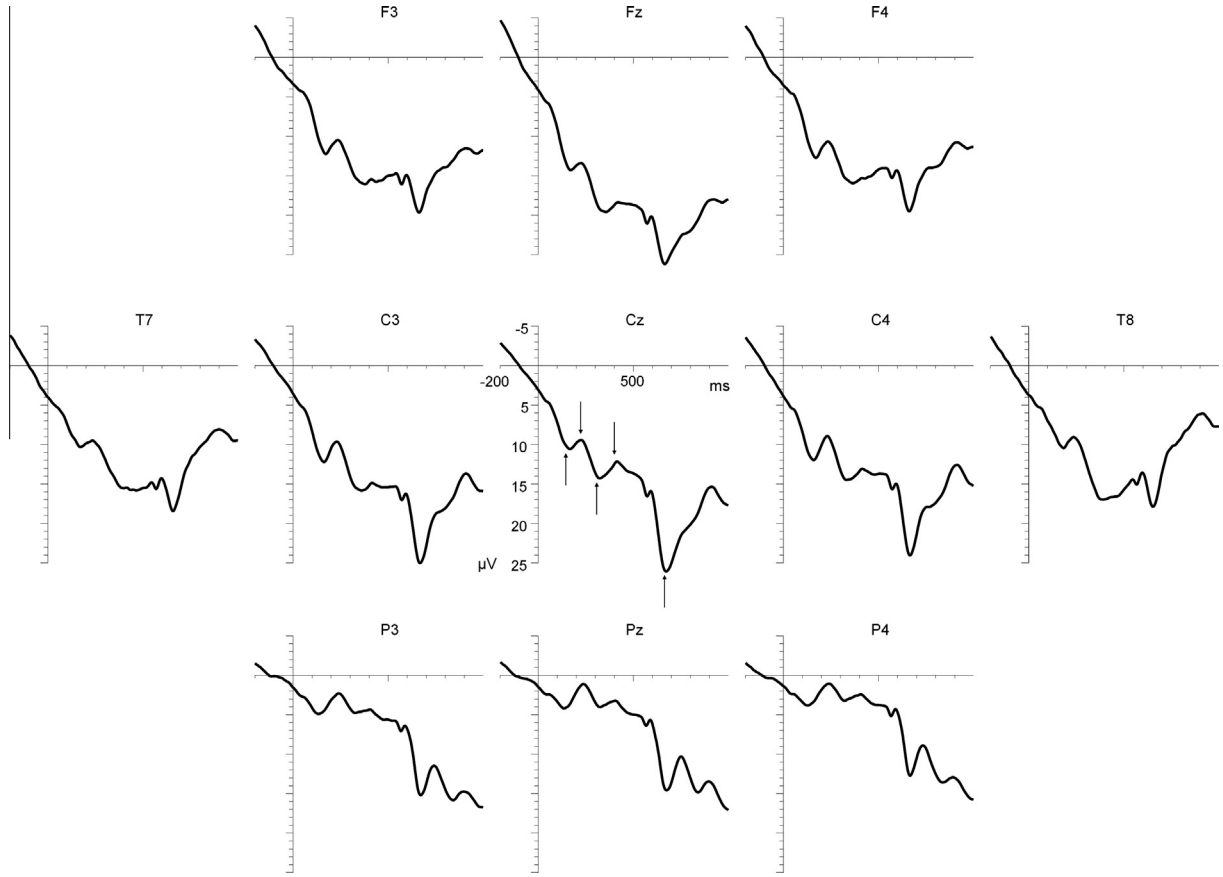

Fig. 2. Grand-averaged waveform for all 4 conditions (FF, HH, FH, HF) combined. Arrows indicate the P150, N250, P350, N450 and P650 area, respectively. Stimulus onset is at 0 ms. Amplitude calibration is at Cz. Negativity is plotted upwards.

3.3. Results for P350 (290–430 ms post-stimulus)

There was a main effect for Visual Prime, and post hoc tests showed that when the visual stimulus had been fearful, responses to the auditory stimulus had been of higher positive amplitudes than when the visual stimulus had been happy ($t(83) = 1.783$; $p < .05$).

3.4. Results for N450 (380–520 ms post-stimulus)

For the N450 results similar to those for the N250 were found. The analysis yielded a main effect for Visual Prime, which was explained by smaller negative-going responses to the auditory stimuli when the preceding facial expression had been fearful ($t(83) = 1.974$; $p < .05$).

3.5. Results for P650 (620–680 ms post-stimulus)

No significant main or interaction effects for Visual Prime were found.

4. Discussion

The current study investigated in 9-month-old infants the processing of emotional (fearful and happy) auditory vocalisations after priming them with a facial expression conveying the same, versus a different emotion. We found that for the P150 and P350 the infants responded with larger positive amplitudes, and for the N250 and N450 with smaller negative amplitudes to emotional auditory stimuli when they had been primed with a fearful visual expression compared to a happy visual expression. The findings confirm our first hypothesis that the presentation of an emotional facial expression can modulate the processing of a vocal expression in 9-month-olds. Contrary to our expectation, we did not find evidence for our second hypothesis. Emotional (in)congruency between the Visual Prime and the following voice did not differentially modulate the ERP response to the latter: When the visual expression had been fearful, responses to both the fearful (congruent) and the happy (incongruent) auditory vocalisations were modulated in the same way (e.g. larger positivity on P150 and P350 and smaller negativity in N250 and N450) compared to when the visual expression had been happy.

Our findings are in accordance with previous studies on audio–visual processing of emotional information in both adults and infants suggesting that how the brain processes emotional information from voices can be modulated by an emotional expression in faces (deGelder & Vroomen, 2000; deGelder et al., 2002; Grossmann et al., 2006). Further, the responses to the auditory stimuli were affected by the presentation of a fearful prime only, and they appear to be most pronounced for the fearful auditory stimuli. The results appear to be driven primarily by the increased positivity at P150 and P350 as opposed to the decreased negativity observed for the N250 and N450. Kushnerenko et al. (2002) have suggested that the P150 component in infants is a precursor of the adult P1 response. This component is thought to indicate preferential attention to auditory input and suppression of unattended information. The P350 component, in turn, has been proposed to develop from its infant form to the P3a in children and adults (Kushnerenko et al., 2002). The P3a has been associated with the orienting response and involuntary attention shifts to change in the environment (Polich, 2003; Polich, 2007). Since in our study

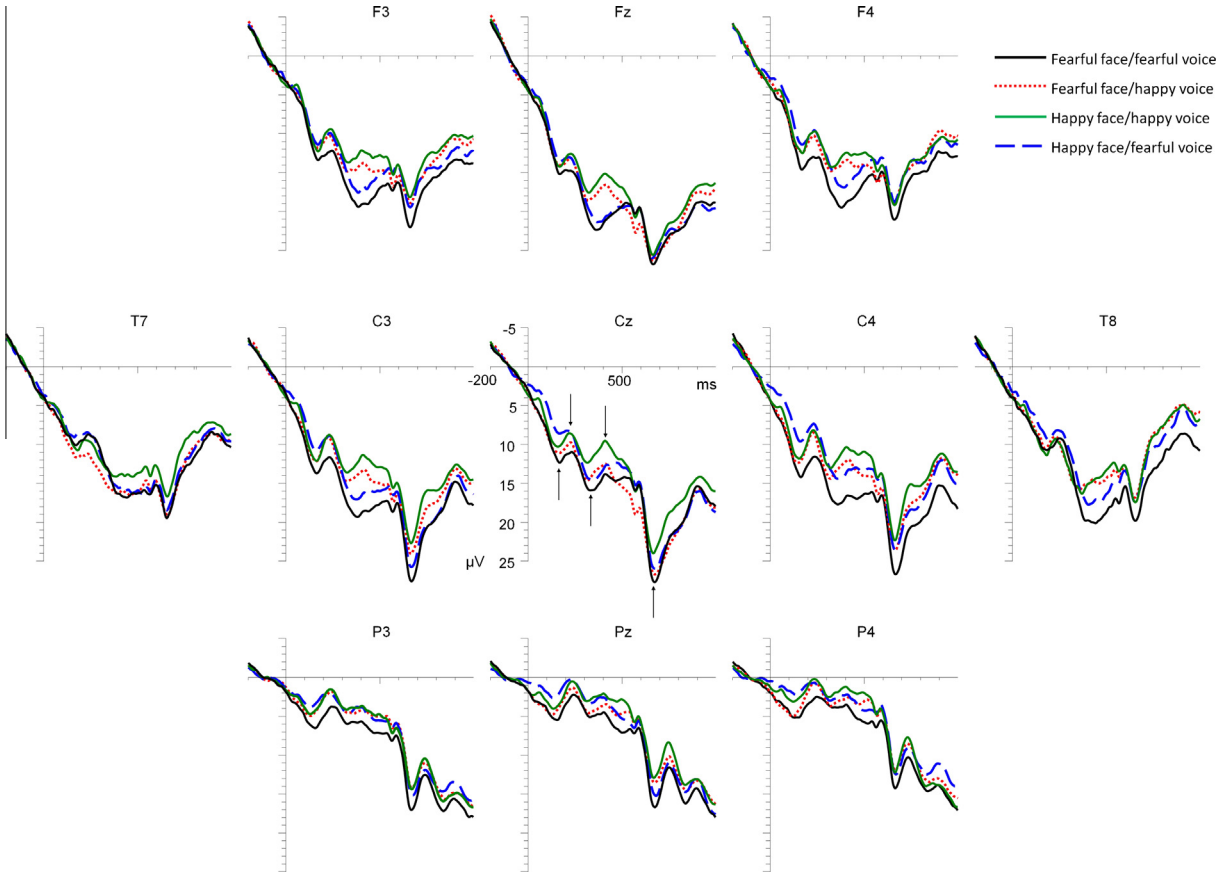

Fig. 3. Grand averages in response to condition FF (solid line), HH (short-striped line), FH (dotted line), and HF (long-striped line), respectively. Arrows indicate the P150, N250, P350, N450 and P650 area, respectively. Stimulus onset is at 0 ms. Amplitude calibration is at Cz. Negativity is plotted upwards.Grand averages in response to condition FF (black), HH (green), FH (red), and HF (blue), respectively. Arrows indicate the P150, N250, P350, N450 and P650 area, respectively. Stimulus onset is at 0 ms. Amplitude calibration is at Cz. Negativity is plotted upwards.

Table 1
Results for the repeated measures ANOVAs.

Latency range (ms)	Factor(s)	*df*	*F*	*p*	ε	η^2
120–200	Frontal-Central	1.83	5204	.025		.059
	Visual Prime	1.83	8762	.004		.095
	Frontal-Central × Laterality	2.166	23,078	.000	.931	.218
	Frontal-Central × Visual Prime	1.83	6295	.014		.071
200–260	Frontal-Central	1.83	10,325	.002		.111
	Laterality	2.166	4511	.015	.910	.052
	Visual Prime	1.83	8494	.005		.093
	Frontal-Central × Laterality	2.166	10,966	.000	.928	.117
290–430	Frontal-Central	1.83	7780	.007		.086
	Visual Prime	1.83	6500	.013		.073
	Auditory Stimulus	1.83	12,625	.001		.132
	Frontal-Central × Laterality	2.166	18,759	.000		.184
380–520	Frontal-Central	1.83	5937	.017		.067
	Visual Prime	1.83	6570	.012		.073
	Auditory Stimulus	1.83	10,747	.002		.115
	Frontal-Central × Laterality	2.166	13,741	.000		.142
	Frontal-Central × Laterality × Auditory Stimulus	2.16	3717	.026		.043
620–680	Laterality	2.166	4957	.008	.810	.056
	Frontal-Central × Laterality	2.166	9004	.000		.098

both the P150 and P350 were larger for auditory stimuli preceded by fearful than by happy stimuli, this may suggest elevated attentional levels to fearful stimuli. This pattern of results is compatible with studies on a negativity bias in infants. These studies showed that "infants attend more to, are more influenced by, and use to a greater degree negative rather than positive facets of their environment" (Vaish, Grossmann, & Woodward, 2008, p. 383).

In the current study, data from a large group of infants ($N = 84$) were included in the analysis. In addition, the infants attended the visual stimuli quite well (73.2% on average), and we were able to retain a relatively large average number of stimuli per condition (37.3). Therefore, the finding that fearful facial expressions modulate the processing of emotional vocalisations appears robust.

In contrast to previous findings, infants in the current study did not respond differentially to emotional auditory vocalisations after priming with a visual stimulus conveying the same, versus a different emotion. There are several possible reasons for this negative finding: (1) The fearful facial stimuli may have elicited higher arousal levels than the happy facial stimuli, such that response amplitudes to the following auditory stimuli were stronger regardless of the type of emotion. This corresponds to previously mentioned findings that infants display a negativity bias to fearful stimuli (Vaish et al., 2008). In the context of the current experiment, however, it does not explain why no (in)congruency effects were found for auditory stimuli following a happy expression; (2) The infants may not have recognised the fearful or the happy expression in the voices and therefore did not match voices to faces expressing the same emotion. Indeed, although there are findings that infants can discriminate between fearful and happy facial expressions from the age of 7 months (Kotsoni, de Haan, & Johnson, 2001; Nelson & de Haan, 1996), studies reporting recognition of fearful vocalisations are very scarce and they were conducted in older infants (Mumme et al., 1996). However, since the largest responses were found for the fearful congruent condition (e.g. fearful face-fearful voice), this explanation does not seem plausible. Also, it does not account for the lack of effects for happy/happy face/voice compounds; (3) The interval between presentation of faces and voices might have been too long. As mentioned in Section 2.3, we had chosen a 900 ms interval to avoid interference from responses to faces in responses to voices. Because of the duration of the interval, perhaps the infants did not interpret faces and voice as a pair. Thus, although processing of the auditory stimulus was influenced by the preceding face, possibly no processing of faces and voice as a compound took place; (4) As the same voices and faces appeared in different combinations in the current experiment, the 9-month-olds may not have been able to associate voices with a particular face. So, on hearing a certain voice, they may have experienced this voice as not belonging to the face that had been (and was being) presented, obviating the need for the face and voice to emotionally match each other. Future studies in which the interval between the facial and vocal expression is smaller (not too small so as to avoid interference from responses to the visual stimuli), and in which a facial identity is always accompanied by the same vocal identity could help answer the question why infants did not respond differentially to emotionally similar versus dissimilar face/voice pairs.

To conclude, the current study investigated multimodal processing of fearful and happy face/voice pairs in 9-month-old infants by means of ERPs. Analysis revealed that the processing of fearful and happy vocalisations was modulated by fearful facial expressions: Infants responded with larger auditory ERPs after having been primed by a fearful visual stimulus. The fact that especially the auditory P150 and P350 were augmented by the fearful visual stimuli suggests that attentional levels are elevated in response to fearful stimuli. This provides evidence for a negativity bias in infants, confirming previous results that infants, just like adults, dedicate more processing capacities to potentially threatening that unthreatening stimuli in their surroundings.

Acknowledgments

The authors are grateful to the parents and infants for their continued participation in our longitudinal study. In addition, they thank dr. Tobias Grossmann for his advice on the multimodal paradigm, and prof. dr. I. Winkler for reading and commenting on previous drafts of the paper. Also, the authors are obliged to Marion van den Heuvel, Malou Menting, and Jaimy Sengers for their help in scoring the videos, and to the students who helped with data collection and processing.

The PELS study is supported by the national funding agencies of the European Science Foundation participating in the Eurocores Program EuroSTRESS programme, i.e. the Brain and Cognition Programme of the Netherlands Organisation for Scientific Research (NWO) for the Netherlands. The PELS study was initiated by BVdB and conducted in collaboration with Vivette Glover (Imperial College London), Stephan Claes (KU Leuven) and Alina Rodriguez (Uppsala University Sweden). BVdB is supported by European Commission Seventh Framework Programme (FP7–HEALTH. 2011.2.2.2-2 BRAINAGE, Grant agreement no: 279281).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandc.2014.09.007.

References

Anderson, A. L., Marois, R., Colson, E. R., Peterson, B. S., Duncan, C. C., Ehrenkranz, R. A., et al. (2001). Neonatal auditory activation detected by functional magnetic resonance imaging. *Magnetic Resonance Imaging, 19*(1), 1–5.

Anderson, A. L., & Thomason, M. E. (2013). Functional plasticity before the cradle: A review of neural functional imaging in the human fetus. *Neuroscience and Biobehavioral Reviews*.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking voices: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*, 129–135.

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

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., et al. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology, 21*, 1220–1224.

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. *Child Development, 59*, 604–616.

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

deGelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in voices is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences, 99*, 4121–4126.

deGelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion, 14*, 289–311.

deHaan, M., Belsky, J., Reid, V., Volein, A., & Johnson, M. H. (2004). Maternal personality and infants' neural and visual responsivity to facial expressions of emotion. *Journal of Child Psychology and Psychiatry, 45*, 1209–1218.

deHaan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: A review. *International Journal of Psychophysiology, 51*, 45–58.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*, 657674.

Flom, R., & Bahrick, L. (2007). The development of infant discrimination in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology, 43*, 238252.

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience, 28*, 219–236.

Grossmann, T., Oberecker, R., Koch, S., & Friederici, A. (2010). The developmental origins of voice processing in the human brain. *Neuron, 65*, 852–858.

Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *NeuroReport, 16*, 1825–1828.

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science, 9*, 309–315.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development, 72*, 352–369.

Kotsoni, E., de Haan, M., & Johnson, M. H. (2001). Categorical perception of facial expressions by 7-month-old infants. *Perception, 30*, 1115–1126.

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

Kushnerenko, E., Čeponienė, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport, 13*, 47–51.

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology, 21*, R143–R145.

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP study of emotional face processing in the adult and infant brain. *Child Development, 78*, 232–245.

Leppänen, J. M., & Nelson, C. A. (2008). Tuning the developing brain to social signals of emotions. *Nature Reviews Neuroscience, 10*, 37–47.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.

Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development, 67*, 3219–3237.

Nelson, C. A., & Bloom, F. E. (1997). Child development and neuroscience. *Child Development, 68*, 970–987.

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

Otte, R., Donkers, F., Braeken, M. & Van den Bergh. (submitted for publication). Multimodal processing of emotional information in 9-month-old infants II: Prenatal exposure to maternal anxiety.

Otte, R., Winkler, I., Braeken, M., Stekelenburg, J., van der Stelt, O., & Van den Bergh, B. (2013). Detecting violations of temporal regularities in waking and sleeping two-month-old infants. *Biological Psychology, 92*, 315–322.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*, 2128–2148.

Polich, J. (2003). Overview of P3a and P3b. In J. Polich (Ed.), *Detection of change: event-related potential and fMRI findings* (pp. 83–98). Boston: Kluwer Academic Press.

Ramsey, J. L., Langlois, J. H., & Marti, N. C. (2005). Infant categorization of faces: Ladies first. *Developmental Review, 25*, 212–246.

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

Striano, T., Brennan, P. A., & Vanman, E. (2002). Maternal depressive symptoms and 6-month-old infants' sensitivity to facial expressions. *Infancy, 3*, 115–126.

Sussman, E. (2007). A new view on the MMN and attention debate: Auditory context effects. *Journal of Psychophysiology, 21*, 164175.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383–403.

van den Heuvel, M., Otte, R., Braeken, M., Winkler, I., Kushnerenko, E., & Van den Bergh, B. (under review). Detecting violations of temporal regularities and novelty in waking two and four month old infants: An auditory ERP study (in preparation).

Van den Stock, J., Grèze, J., & de Gelder, B. (2008). Human and animal sounds influence recognition of body language. *Brain Research, 1242*, 185–190.

Vogel, M., Monesson, A., & Scott, L. S. (2012). Building biases in infancy: The influence of race on face and voice emotion matching. *Developmental Science, 15*(3), 359–372.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*, 437–456.