

# Vocal threat enhances visual perception as a function of attention and sex

Annett Schirmer,<sup>1,2,3</sup> Maria Wijaya,<sup>1</sup> Esther Wu,<sup>4</sup> and Trevor B. Penney<sup>1,2,3</sup>

<sup>1</sup>Department of Psychology, The Chinese University of Hong Kong, Shatin, Hong Kong, <sup>2</sup>Brain and Mind Institute, The Chinese University of Hong Kong, Shatin, Hong Kong, <sup>3</sup>Center for Cognition and Brain Studies, The Chinese University of Hong Kong, Shatin, Hong Kong, and <sup>4</sup>Yale-National University of Singapore College, Singapore, 138527 Clementi, Singapore

Correspondence should be addressed to Annett Schirmer, Department of Psychology, The Chinese University of Hong Kong, 3rd Floor, Sino Building, Shatin, N.T., Hong Kong. E-mail: schirmer@cuhk.edu.hk.

## Abstract

This pre-registered event-related potential study explored how vocal emotions shape visual perception as a function of attention and listener sex. Visual task displays occurred in silence or with a neutral or an angry voice. Voices were task-irrelevant in a single-task block, but had to be categorized by speaker sex in a dual-task block. In the single task, angry voices increased the occipital N2 component relative to neutral voices in women, but not men. In the dual task, angry voices relative to neutral voices increased occipital N1 and N2 components, as well as accuracy, in women and marginally decreased accuracy in men. Thus, in women, vocal anger produced a strong, multifaceted visual enhancement comprising attention-dependent and attention-independent processes, whereas in men, it produced a small, behavior-focused visual processing impairment that was strictly attention-dependent. In sum, these data indicate that attention and listener sex critically modulate whether and how vocal emotions shape visual perception.

**Key words:** ERP; visual attention; emotion; vocal affect; sex differences

## Introduction

We often find ourselves influenced by the stimuli we intend to ignore—especially if these stimuli are affectively relevant and change the way we feel (Lui *et al.*, 2011; Min and Schirmer, 2011; for reviews, see Inzlicht *et al.* 2015; Pessoa 2015). For example, driving a car in heavy traffic is harder when passengers talk as compared to when they are silent (Gaspar *et al.*, 2014), and this difference may be greater for verbal exchanges that are confrontational as compared to benign. Here, we addressed the role of vocal emotions for visual perception. Moreover, we explored the extent to which expected emotion effects require attentive voice processing, how exactly they shape visual perception and whether they differ between sexes.

There is much evidence that the affective value of task-irrelevant stimuli modulates ongoing mental processes (Öhman and Soares, 1994; Globisch *et al.*, 1999; for a meta-analysis, see Schirmer, 2018). For example, this was shown by Vuilleumier and colleagues who presented visual arrays that paired two houses and two faces (Vuilleumier *et al.*, 2001). Color frames around the stimuli indicated which pair was task-relevant. On face trials, participants judged whether the faces were identical, and on house trials, they judged whether the houses were identical. Despite being task-irrelevant, facial expressions modulated brain activity during both face and house trials. Specifically, the amygdala and fusiform gyrus, two structures implicated in face perception, were more strongly activated by fearful expressions as compared with neutral expressions.

**Received:** 12 January 2019; **Revised:** 1 June 2019; **Accepted:** 10 June 2019

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

These and similar findings prompted some to conclude that emotion processing is automatic (for a review, see [Pourtois et al., 2013](#)). However, others have challenged this conclusion ([Pessoa, 2008](#)) and argued instead that an influence of task-irrelevant emotions depends on the availability of processing resources and an undemanding primary task. A study supporting this latter position presented an emotional or neutral face flanked by two bars ([Pessoa et al., 2002](#)). In different experimental blocks, study participants indicated bar orientation (same/different) or face sex (male/female). Differences in brain activity between emotional and neutral faces were observed during the face task, but not during the bar task, suggesting that the processing of emotional information is 'under top-down control'.

The conflict ignited by this original work inspired much subsequent research. Some of this research leveraged on the spatial sensitivity of functional magnetic resonance imaging (fMRI), which allows for a fairly detailed differentiation of emotion systems ([Straube et al., 2011](#); [Schindler et al., 2018b](#)). Additionally, research was accumulated using the electroencephalogram (EEG) and its event-related potential (ERP) technique. Compared to fMRI, EEG/ERP has a better temporal resolution and is thus better suited for dissociating potentially fast-changing effects (e.g. [Schirmer et al., 2005a](#); [Pourtois et al., 2006](#); [Schupp et al., 2006](#); [Kissler et al., 2009](#); [Pourtois et al., 2010](#); [Brosch and Wieser, 2011](#); [Schindler et al., 2018a](#)). Accumulating findings from both fMRI and EEG/ERP paint a complex picture suggesting that automaticity is a differentiated construct that is continuous rather than discreet and that depends on paradigm and methodological choices (for an example of the role of low-level visual features, see [Schindler et al., 2018a](#)). Moreover, extant work highlights the need to consider automaticity more specifically within a particular task context.

The context that was of interest here was modeled on everyday experiences in which individuals perform a demanding visual task (e.g. driving a car) on an auditory backdrop (e.g. passenger conversations). Specifically, we asked how emotional aspects of the auditory backdrop become relevant for modulating visual processes and responses. To the best of our knowledge, this question has been tackled by only a couple of fMRI studies ([Mothes-Lasch et al., 2011, 2012](#)). Moreover, these studies relied on small samples and failed to consider important individual variables such as sex. In fact, there is now much evidence suggesting that, compared with men, women are more sensitive to emotional information if that information is task-irrelevant ([Proverbio et al., 2009](#); [van den Brink et al., 2010](#); [Schirmer et al., 2013a](#); [Proverbio and Galli, 2016](#); [Schirmer and Gunter, 2017](#)). Thus, sex is likely to be important for ongoing efforts at understanding emotion processing automaticity.

With these points in mind, we designed an ERP study in which visual task displays occurred in silence or were accompanied by an angry or neutral voice. In the single-task block, voices were irrelevant, whereas in the dual-task block, speaker sex had to be categorized. Of interest were behavioral responses as well as two negative components in the visual ERP. The first component, the N1, typically peaks at ~160 ms following stimulus onset with larger amplitude to physically salient, emotional or cued stimuli ([Mangun, 1995](#); [Herrmann and Knight, 2001](#); [Schneider et al., 2012](#)). As such, N1 is sensitive to bottom-up mechanisms that recruit attentional resources for stimulus perception. The second component, N2, peaks between 200 and 350 ms and comprises several subcomponents that have been linked to top-down mechanisms and stimulus expectation ([Folstein and Van Petten, 2008](#); [Schneider et al., 2012](#)). In their original form, N1 and N2 reflect responses to the visual field as a whole, but

when one subtracts voltages recorded from electrode sites ipsilateral to the visual target from those recorded contralaterally, they reveal information about hemifield processing and spatial orienting ([Mangun, 1995](#); [Woodman and Luck, 1999](#)). The behavioral measures examined here comprised reaction times and the sensitivity index  $d'$  and were expected to help specify whether and how emotional sounds modulate visual task performance.

We tested the following three predictions. First, if vocal affect modulates visual processing automatically, then differences in the N1, N2 and behavioral responses to angry voices as compared with neutral voices should be present in both the dual and the single task. If, however, vocal affect modulates visual processing as a function of voice-directed attention, then effects on N1, N2 and behavioral responses should be present in the dual task only. Second, vocal background should impair target processing when compared with silence, and this impairment should be largest for angry voices. This prediction was derived from evidence that (i) irrelevant context distracts unless it provides cues (e.g. spatial location) to target processing and that (ii) emotions augment context effects ([Brosch et al., 2009](#); [Gaspar et al., 2014](#)). Third, based on established sex differences (reviewed in [Schirmer, 2013](#)), we expected effects of vocal affect to be more pronounced and more automatic in women than in men.

## Methods

**Participants.** We determined our sample size a priori based on previous research reporting sex differences and counterbalancing needs associated with the paradigm ([Schirmer et al., 2013a, 2018](#); [Schirmer and McGlone, 2018](#)). Sample size, basic methodology and study hypotheses were pre-registered at the Open Science Framework and can be inspected here: [https://osf.io/ur6fk/?view\\_only=bea9c744d0824b6aaea0ad08588a6b04](https://osf.io/ur6fk/?view_only=bea9c744d0824b6aaea0ad08588a6b04). Please note that when visiting this link, you must click on the 'View Registration Form' button to see our registered document. Apart from the hypotheses described here, the registered document mentions stimulus difficulty as an additional variable. This variable was conceived with the intention that slightly and starkly asymmetrical crosses would be presented in separate blocks. However, due to a miscommunication with the programmer, they were presented within the same block. Therefore, the difficulty variable was non-orthogonal, and we excluded it from our analyses as we did not know how to conceptualize the difficulty associated with symmetrical crosses.

We invited 53 participants to this study. The data from five participants had to be discarded due to a data recording issue ( $N=1$ ) or because of their failure to maintain fixation ( $N=4$ ). Twenty-four of the remaining participants were female with a mean age of 22.7 years (s.d. 4.4) and 24 were male with a mean age of 21.6 years (s.d. 2.7). Participants reported normal or corrected-to-normal vision and normal hearing. They received HKD 70 for 1 h of their time.

**Stimuli.** This study used visual and auditory stimuli. The visual stimuli were search displays comprising eight equally-spaced crosses presented along a circle (radius =  $4.5^\circ$  of visual angle) around a central fixation point. Each cross measured  $\sim 1^\circ$  of visual angle in width and height. On a given trial, half the crosses were symmetrical (+) and half were asymmetrical (†). For the asymmetrical crosses, the difference in length between the top and bottom segments of the cross was either 30 or 60% of its total height. Crosses in the search display occurred on a

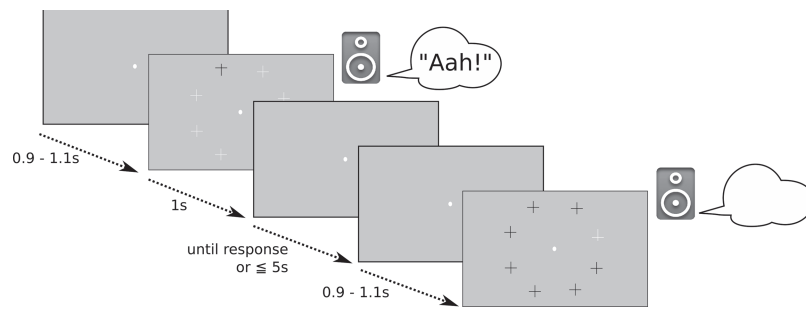


Fig. 1. Research paradigm.

gray background and could be blue or pink—if one cross was blue, the others were pink, and vice versa. Singleton color was balanced across trials and conditions, and identified the target for which participants decided whether it was symmetrical or asymmetrical. Targets occurred on the left and right half of the circle on 50% of the trials, respectively. To reduce neuronal adaptation over time (Luck and Hillyard, 1994), the colors of the search display alternated across trials, i.e. a pink singleton on one trial was followed by a blue singleton on the next trial. Further, we added a small, random jitter to the exact position of each cross ( $\leq 10^\circ$  subtended angle from the screen center) such that cross positions varied slightly from trial to trial.

Auditory stimuli were selected from the Montreal Affective Voices (MAV) (Belin et al., 2008) and from stimuli previously recorded in our laboratory. They consisted of 24 angry and 24 neutral exclamations of the syllable ‘ah’, half of which were spoken by a female and half by a male speaker. An independent group of listeners normed the selected voices. Specifically, 22 listeners (11 female) rated the MAV stimuli, while 20 listeners (10 female) rated the in-house stimuli by indicating their emotion (What is the emotion expressed?) and scoring emotion intensity and arousal on scales ranging from 1 (very weak) to 4 (very strong). The rating results informed stimulus selection. The selected angry and neutral voices differed for intensity ( $t(46) = 6.39$ ,  $P < 0.001$ ,  $M_{\text{angry}} = 3.42$ ,  $s.d._{\text{angry}} = 0.65$ ,  $M_{\text{neutral}} = 2.49$ ,  $s.d._{\text{angry}} = 0.28$ ) and arousal ( $t(46) = 8.73$ ,  $P < 0.001$ ,  $M_{\text{angry}} = 3.56$ ,  $s.d._{\text{angry}} = 0.63$ ,  $M_{\text{neutral}} = 2.33$ ,  $s.d._{\text{neutral}} = 0.30$ ), but not for identification accuracy ( $P > 0.25$ ). Their sound intensity was normalized and sound durations ranged between 228 and 981 ms. A Bayesian  $t$ -test using a joint conjugate prior indicated that mean durations did not meaningfully differ between the angry and neutral conditions ( $t(49) = 0.521$ ,  $P = 0.604$ ,  $CI: -0.103$  to  $0.176$ ).

**Paradigm.** Task displays were presented on a 24 inch LCD monitor and with a viewing distance of 90 cm. The monitor’s refresh rate was 60 Hz. During both a single- and a dual-task block, trials started with a fixation dot (radius  $0.3^\circ$ ) that lasted for 0.9–1.1 s (randomly selected from a uniform distribution) and was followed by a 1 s search display. Participants indicated whether or not the target was symmetrical by pressing one of two buttons on the computer keyboard. The next trial started immediately after a response was made or after 5 s, whichever came first. For two thirds of the trials, search displays were accompanied by an angry or a neutral voice played over speakers positioned to the left and right of the screen. Voices were always presented binaurally. For the remaining trials, search displays were presented in silence (Figure 1).

During the single-task block, participants focused on the visual task only. During the dual-task block, they performed the visual task together with an auditory task that required

them to indicate whether or not the voice, if present, was male. Participants responded to the visual task with the index and middle finger of one hand and to the auditory task with the index and middle finger of the other hand. For a given participant, hand and finger assignment for the visual task was constant across blocks. Across participants, block order and hand and finger assignment were counterbalanced.

At the start of the experiment, participants first gave informed consent and were then prepared for the EEG recording. Subsequently, they were briefed about both tasks. Participants who performed the visual-only block first were briefed about the visual task followed by the auditory task, while participants who performed the visual–auditory task block first were briefed about the auditory task before the visual task. Participants were asked to fixate on the central fixation dot and to not shift their eyes to the crosses. Moreover, they were told to respond as quickly as possible, but without sacrificing response accuracy. Each task block comprised 576 trials distributed equally among the three sound conditions (i.e. angry, neutral and silent) and the two visual hemi-fields. We thus had 96 left and 96 right target trials for each cell in the design. An experimental session lasted about 1 h and 20 min.

**EEG recording and analysis.** The EEG was recorded using a 64-channel EEGO system from ANT. Electrodes were embedded in a cap according to the modified 10–20 system. Five additional electrodes were placed at the two outer canthi, above and below the left eye and the nose. The data were sampled at 500 Hz with a hardware defined non-linear anti-aliasing filter that attenuated frequencies below 183 Hz by  $-6\text{dB}$  and with CPz as an online reference.

Data processing was done in MATLAB R2016B (The MathWorks, Inc., Natick, MA, USA) and EEGLAB 14.1.2.B (Delorme and Makeig, 2004). The data were re-referenced to the average of all electrodes, high-pass filtered at 0.1 Hz (0.1 Hz transition band;  $-6\text{ dB/octave}$ ), low-pass filtered at 30 Hz (7.5 Hz transition band;  $-6\text{ dB/octave}$ ) and epoched by centering a 2 s window around stimulus onset. The resulting epochs were visually scanned for non-typical artifacts caused by drifts or muscle movements, and epochs containing such artifacts were removed. The data were then subjected to an automatic rejection procedure that removed additional epochs in which the HEOG exceeded  $100\text{ }\mu\text{V}$  or the VEOG exceeded  $32\text{ }\mu\text{V}$  within the first 300 ms following stimulus onset. The HEOG cut-off translated to  $2^\circ$  of visual angle and thus less than half the radius of the circle ( $4.5^\circ$ ) on which visual targets appeared. Trials with early HEOG and VEOG movements were excluded in this manner because during these trials, the visual displays would not have been properly processed. Moreover, visual processing would have been suppressed by

the eye movement (Bristow et al., 2005), thus confounding early visual ERPs like N1 and N2.

To prepare our data for an independent component analysis (ICA), we applied a 1 Hz high-pass filter that removed slow drifts and improved component decomposition. The component structure resulting from the ICA was then applied to the original epoched data set with the 0.1 to 30 Hz filter setting (Winkler et al., 2015). Components reflecting the remaining horizontal and vertical eye movements were removed and the data back-projected from component space into EEG channel space. Another automatic rejection procedure was applied that removed epochs in which scalp channels exceeded 100  $\mu$ V. Subsequently, the data were submitted to a current source density (CSD) transformation using the CSD tool box (Kayser and Tenke, 2003) with its default settings. This was followed by a trial number matching procedure whereby the condition with the lowest trial number was identified, and the same number of trials was randomly drawn from the other conditions. Final trial numbers ranged from 33 to 182 per condition and participant due to the fact that we lost many trials for some participants who had difficulties maintaining central fixation. Across participants, each condition averaged to 117 trials (s.d. 40.6). For four participants, one of the channels analyzed here required interpolation.

Following this preprocessing protocol, we re-epoched the data using a -100 to 500 ms time window and applied baseline correction using the 100 ms period before stimulus onset. Early ERP components previously linked to visual attention, including N1 and N2, were of primary interest. We quantified these components in two ways: (1) by averaging over left hemisphere electrodes PO7, PO5 and O1 and right hemisphere electrodes PO8, PO6 and O2 and (2) by subtracting channels ipsi-lateral from those contralateral to the target (i.e. PO7-PO8, PO5-PO6 and O1-O2 for right targets; PO8-PO7, PO6-PO5 and O2-O1 for left targets). Based on visual inspection of component peaks and guided by previous work (Heinze et al., 1990; Woodman and Luck, 1999; Zani et al., 2015), we computed mean voltages from resulting traces in two time windows centered around the N1 (140-190 ms) and N2 (230-270 ms) peaks, respectively. Please note that the N2 window overlapped with the shortest sound offsets (~228 ms). However, given that the offset duration did not differ between conditions, we considered a possible influence of sound offsets on the present N2 modulations to be negligible.

## Results

### Behavioral data

Behavioral results are illustrated in Figure 2.  $d'$  scores were calculated for visual response accuracy by subtracting the normalized

probability of falsely categorizing a cross as symmetrical from the normalized probability of correctly categorizing a cross as symmetrical. The resultant values served as the dependent variable in an ANOVA with task and sound as repeated measure factors and sex as a between-subjects factor. This ANOVA revealed a marginal interaction of sound and sex ( $F(2,92) = 2.61$ ,  $P = 0.079$ ,  $\eta_p^2 = 0.054$ ) and a significant interaction of task, sound and sex ( $F(2,92) = 3.44$ ,  $P = 0.036$ ,  $\eta_p^2 = 0.069$ ). All other effects were non-significant ( $P_s > 0.25$ ). We pursued the three-way interaction by analyzing each task separately. In the dual task, the sound by sex interaction was significant ( $F(2,92) = 5.1$ ,  $P = 0.008$ ,  $\eta_p^2 = 0.099$ ); vocal expressions affected the sensitivity of visual categorizations significantly in women ( $F(2,46) = 3.51$ ,  $P = 0.038$ ,  $\eta_p^2 = 0.132$ ) and marginally in men ( $F(2,46) = 2.81$ ,  $P = 0.07$ ,  $\eta_p^2 = 0.109$ ). Women performed better on angry trials compared with neutral trials ( $F(1,23) = 4.91$ ,  $P = 0.037$ ,  $\eta_p^2 = 0.176$ ) without performance differences between angry and silent or neutral and silent trials (both  $P_s > 0.142$ ). In contrast, men performed better on both neutral ( $F(1,23) = 4.91$ ,  $P = 0.037$ ,  $\eta_p^2 = 0.218$ ) and silent ( $F(1,23) = 4.24$ ,  $P = 0.051$ ,  $\eta_p^2 = 0.156$ ) trials compared with angry trials without performance differences between neutral and silent trials ( $P > 0.25$ ). In the single task, both the sound effect and the sound by sex interaction were non-significant ( $P_s > 0.25$ ).

RTs for correctly categorized targets were submitted to an ANOVA with task and sound as repeated measure factors and sex as a between-subjects factor. This revealed the main effects of task ( $F(1,46) = 126.43$ ,  $P < 0.0001$ ,  $\eta_p^2 = 0.733$ ) and sound ( $F(2,92) = 8.51$ ,  $P < 0.001$ ,  $\eta_p^2 = 0.156$ ), as well as a task by sound interaction ( $F(2,92) = 13.75$ ,  $P < 0.0001$ ,  $\eta_p^2 = 0.23$ ). Follow-up analyses indicated that voices affected performance in the dual task ( $F(2,92) = 12.38$ ,  $P < 0.0001$ ,  $\eta_p^2 = 0.212$ ), but not the single task ( $P > 0.25$ ). In the dual task, angry ( $F(1,46) = 13.48$ ,  $P < 0.001$ ,  $\eta_p^2 = 0.227$ ) and neutral ( $F(1,46) = 16.31$ ,  $P < 0.001$ ,  $\eta_p^2 = 0.262$ ) expressions slowed down RTs relative to silence. However, angry and neutral trials were performed with similar speeds ( $P > 0.25$ ). A marginal effect of sex ( $F(1,46) = 3.08$ ,  $P = 0.086$ ,  $\eta_p^2 = 0.063$ ) suggested that women tended to respond more slowly than men. All other effects were non-significant (all  $P_s > 0.197$ ).

### Event-related potentials

Electrophysiological results are illustrated in Figures 3 and 4. Visual ERPs were explored in two ways. First, we examined components of interest for the entire visual field to determine general effects of voices on visual attention. In a second step, we analyzed the ERP difference between target and non-target hemifields to determine whether voices modulate spatial orienting to targets.

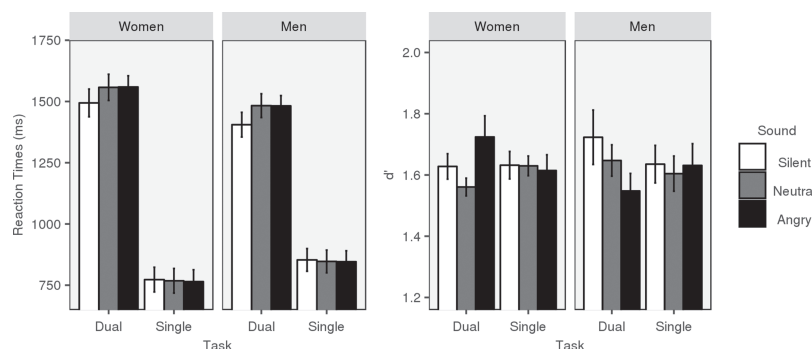


Fig. 2. Behavioral results. Mean  $d'$  scores and reaction times are shown as a function of task, sound and sex. Error bars reflect the within-subject standard error.

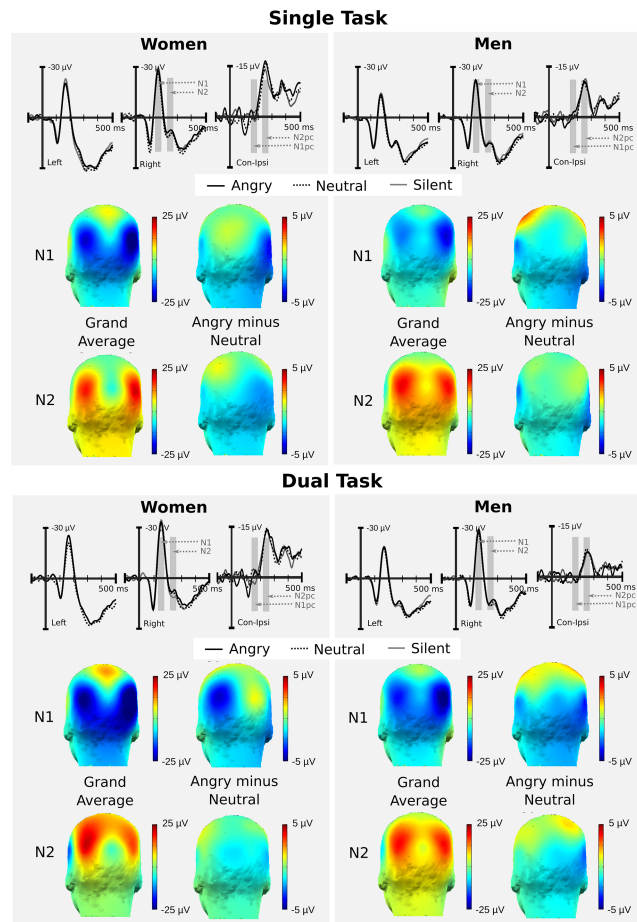


Fig. 3. ERP traces and maps. Mean ERP voltages were derived by separately averaging signals for left occipital electrodes (PO7, PO5 and O1), right occipital electrodes (PO8, PO6 and O2) and the voltage difference between contra- and ipsi-lateral occipital electrodes. Time windows for statistical analysis are marked by the shaded areas. Maps illustrate the mean voltages and condition differences for the statistical analysis windows.

Our first set of analyses revealed effects for both N1 and N2. The N1 was modulated by main effects of task ( $F(1,46)=22.5$ ,  $P < 0.0001$ ,  $\eta_p^2 = 0.328$ ) and sound ( $F(2,92)=6.73$ ,  $P=0.002$ ,  $\eta_p^2 = 0.128$ ) as well as interactions of task, sound and laterality ( $F(2,92)=3.11$ ,  $P=0.049$ ,  $\eta_p^2 = 0.063$ ) and task, sound, laterality and sex ( $F(2,92)=6.12$ ,  $P=0.003$ ,  $\eta_p^2 = 0.117$ ). We pursued the latter interaction by examining data from men and women separately. For men, we found that the interaction of task, sound and laterality was non-significant. Men showed a task effect only ( $F(1,23)=11.87$ ,  $P=0.002$ ,  $\eta_p^2 = 0.34$ ), indicating that N1 was larger in the dual task compared with the single task. No other effects reached the traditional significance threshold (all  $P$ s  $> 0.135$ ). For women, we observed task ( $F(1,23)=10.69$ ,  $P=0.003$ ,  $\eta_p^2 = 0.317$ ) and sound main effects ( $F(2,46)=4.99$ ,  $P=0.011$ ,  $\eta_p^2 = 0.178$ ) as well as an interaction of task, sound and laterality ( $F(2,46)=7$ ,  $P=0.002$ ,  $\eta_p^2 = 0.233$ ). Over the left occipital region, the sound effect differed between tasks ( $F(2,46)=6.65$ ,  $P=0.003$ ,  $\eta_p^2 = 0.224$ ). In the single task, it was non-significant ( $P > 0.25$ ). However, in the dual task ( $F(2,46)=9.05$ ,  $P < 0.001$ ,  $\eta_p^2 = 0.282$ ), N1 amplitudes were larger for silent ( $F(1,23)=12.63$ ,  $P=0.002$ ,  $\eta_p^2 = 0.354$ ) and angry ( $F(1,23)=11.52$ ,  $P=0.002$ ,  $\eta_p^2 = 0.113$ ) trials compared with neutral trials. Silent and angry trials did not differ ( $P=0.101$ ). Over the right occipital region, the sound effect and the sound by task interaction were non-significant ( $P$ s  $> 0.128$ ). There was only a task effect indicating that, similar to men, N1 amplitudes were larger in the dual task compared with the single task.

Analysis of N2 revealed a sound main effect ( $F(2,92)=8.15$ ,  $P < 0.001$ ,  $\eta_p^2 = 0.15$ ) and a sound by sex interaction ( $F(2,92)=3.95$ ,  $P=0.022$ ,  $\eta_p^2 = 0.079$ ; all other  $P$ s  $> 0.109$ ). The sound effect was significant in women ( $F(2,92)=8.77$ ,  $P < 0.001$ ,  $\eta_p^2 = 0.276$ ), but not in men ( $P=0.474$ ). In women, N2 amplitudes were larger for angry trials compared with neutral ( $F(1,23)=4.45$ ,  $P=0.046$ ,  $\eta_p^2 = 0.162$ ) trials and for neutral trials compared with silent trials ( $F(1,23)=5.87$ ,  $P=0.024$ ,  $\eta_p^2 = 0.203$ ).

We explored the target-directed attention effects for both N1 and N2 by computing their posterior-contralateral (pc) indices. For the N1pc, all effects were non-significant. For the N2pc, there was a marginal effect of sex ( $F(1,46)=3.24$ ,  $P=0.078$ ,  $\eta_p^2 = 0.066$ ) suggesting that the N2pc tended to be larger in women than in men. Additionally, a significant effect of task ( $F(1,46)=7.68$ ,  $P=0.008$ ,  $\eta_p^2 = 0.143$ ) indicated that spatial attention toward the target was greater or more effectively allocated in the single task compared with the dual task. All other effects were non-significant ( $P$ s  $> 0.207$ ).

## Discussion

Here we explored the role of attention in enabling effects of vocal threat on visual perception. Additionally, we characterized the nature of these effects and how they unfold in women compared with men. In the following, we will discuss these three points in turn.

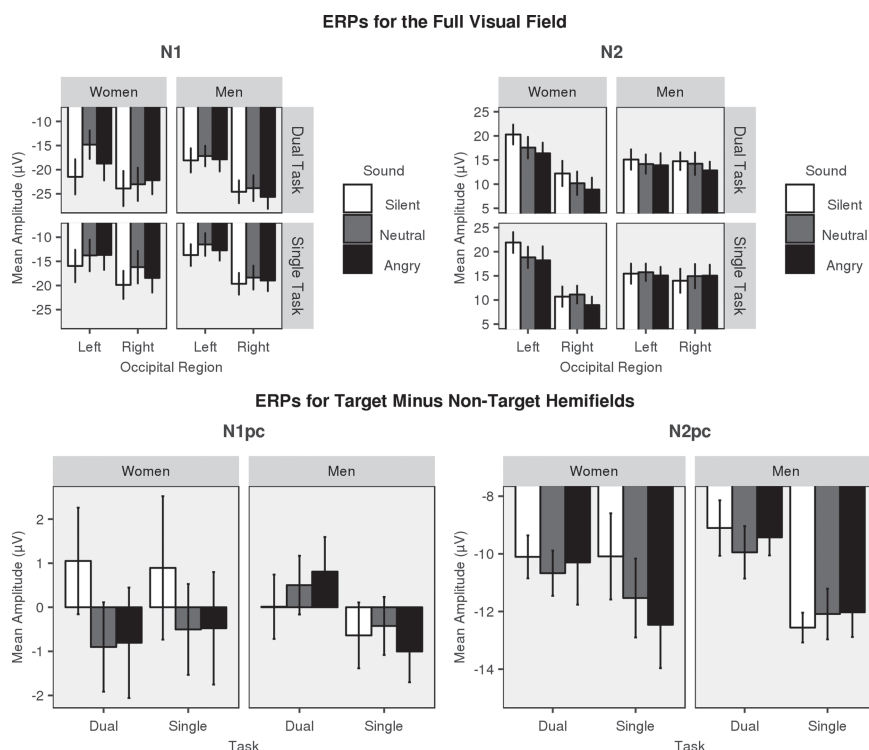


Fig. 4. ERP mean amplitudes. Mean voltages in the N1 and N2 analysis windows are shown as a function of task, sound, sex and region. Error bars reflect the within-subject standard error.

### Are auditory emotion effects on visual processing automatic or controlled?

Much previous work has pursued the interaction between emotion and attention. Of particular interest here are studies examining the cross-modal effect of auditory emotions on visual processing using fMRI (Mothes-Lasch et al., 2011, 2012). Similar to the present paradigm, they presented a visual categorization task against the backdrop of angry and neutral voices (Mothes-Lasch et al., 2011, 2012). Differences in brain activity between these conditions depended on voices being task-relevant, suggesting that auditory affective processing or more specifically, the influence auditory affect on visual processing requires attention. Due to fMRI's sluggish nature, however, these claims have been challenged and studies using a temporally more sensitive approach have been called for (Brosch and Wieser, 2011).

Here we adopted such an approach and, to increase methodological convergence, used the paradigm implemented in earlier fMRI work (Mothes-Lasch et al., 2011, 2012). This, however, meant that volume conduction could compromise the interpretation of occipital scalp effects. Specifically, the concurrent presentation of visual and auditory stimuli could be expected to produce effects extending to auditory and visual recording sites, respectively. We addressed this problem by applying a CSD transformation, thus making recorded voltages reference-free and reducing global effects while enhancing local effects linked to underlying cortical tissue (Kayser and Tenke, 2015).

Our CSD results revealed both task-dependent (N1, behavior) and independent (N2) effects of the vocal affect. As such, they partially disagree with previous fMRI results (Mothes-Lasch et al., 2011, 2012). Moreover, they highlight that both more and less automatic processes may be observed concurrently from the same paradigm with a technique that better captures how these

processes unfold in time. In light of our effects, we conclude that affective influences under automatic processing conditions occurred later than affective influences under controlled processing conditions. In other words, paying attention to affective voices temporally facilitated their integration with attended visual input. Possibly, in the absence of focused attention, the affective processing of auditory signals is too slow or not salient enough to impact on early bottom-up representations in other modalities.

Notably, our conclusion contrasts with previous electrophysiological evidence on unimodal perception. Specifically, intracranial recordings from the amygdala (Pourtois et al., 2010) using a paradigm similar to that of Vuilleumier and colleagues (Vuilleumier et al., 2001) revealed a task-independent early emotion effect starting around 140 ms and a task-dependent later effect starting around 750 ms. Likewise, two studies adopting the paradigm of Pessoa and colleagues (Pessoa et al., 2002) using magnetoencephalography (Luo et al., 2010) or combining EEG and fMRI (Müller-Bardorff et al., 2018) found task-independent effects after 40 ms and task-dependent effects after 280 ms. Thus, future discussions of the relation between emotion and attention must carefully consider the type of processes (e.g. univ vs cross-modal representations) for which task-dependent and independent effects are being assessed.

### How do vocal expressions shape visual perception?

How auditory background shapes visual perception has been of great interest to applied psychologists (Gaspar et al., 2014). Moreover, their work revealed impairment effects which could be replicated here. Specifically, reaction times were longer for voice trials relative to silence when voices were task-relevant. Although part of the latter task effect may arise from the dual

vs single hand motor demands, one may reasonably venture that bi vs unimodal cognitive demands also played a role. In line with this, (neutral) task-relevant but not task-irrelevant voices reduced N1 to visual targets, indicating that the additional demand of attending to a speaker hampered bottom-up visual representations.

Importantly, the present voice effects differed as a function of expression and, unexpectedly, were not generally debilitating. In the single task, N2 was larger for angry trials compared with neutral trials, suggesting that vocal threat benefited associated top-down mechanisms of visual attention. In the dual task, this N2 effect was complemented by a larger N1 and more accurate visual categorizations for angry trials compared with neutral trials. Notably, there were neither impairment effects nor enhancement effects on target-hemifield ERPs, suggesting that spatial orienting was unaffected.

Taken together, we show that unrelated sounds impair aspects of visual performance, but that impairments may be compensated and accompanied by processing benefits as a function of sound affect. Moreover, vocal anger appears to boost a fairly automatic mechanism reflected by N2. Its bilateral topography further points to a modulation of left-lateralized local and right-lateralized global perceptual processes supporting item-specific and display-general representations, respectively (Fink et al., 1996). Additionally, a left-lateralized attention-dependent mechanism reflected by N1 may promote local-over-global processing. Together, both mechanisms seem to enhance resource allocation across an individual's visual field in a catch-all fashion, thus benefiting the integrity of both target and non-target representations.

### Do the sexes differ?

As expected, we found effects of vocal affect to be more pronounced in women than in men. Specifically, the ERP and behavioral results described above were significant in women only. Men, in contrast, showed only marginal accuracy differences between the voice conditions, and angry voices tended to impair rather than enhance visual performance.

These findings fit well with previous evidence that women are more likely than men to process social signals (Proverbio et al., 2009; van den Brink et al., 2010) and emotional expressions that are task-irrelevant. For example, in women, but not men, emotional faces prime lexical decisions (Schirmer et al., 2013a), and vocal emotional oddballs enhance the change detection response in the ERP (Schirmer et al., 2005a). Additionally, women are more likely than men to show enhanced orienting toward an angry voice compared with a neutral voice (Burra et al., 2018). Because these sex differences typically disappear when social signals and their emotions have to be processed in order to perform an experimental task (Schirmer et al., 2005b, 2006), they likely reflect differences in how automatically emotions are accessed. Compared with men, women may require less effort or mental resources for establishing affective or emotion relevance.

### Conclusions

Many visual tasks, such as driving, often occur against an auditory backdrop like the voices of other people. Exploring their influence on primary task performance, we found that although voices in general impaired visual performance, threat, compared to neutral affect, compensated and partially reversed these effects as a function of attention and sex. When the task

focused on visual information only, angry voices relative to neutral voices enhanced the neural correlates of visual attention without consequences for behavior in women, but not in men. When both visual and auditory information were in focus, anger elicited more substantial neural benefits as well as more accurate visual categorization in women only. In sum, we found that automaticity in emotion is not an either/or issue. Instead, emotions emerge in multifaceted ways that may be more or less resource-dependent and that vary as a function of situational and individual factors.

### Conflict of interest

None declared.

### Acknowledgements

Professor and emotion expert Mick Power contributed to the ideas of this project as well as its initial set-up prior to his early death in 2017. He also participated as a collaborator on the Singapore Ministry of Education grant that funded this work and that was awarded to A.S. in 2016 (R-581-000-199-646). We are grateful to Mick for his generous support and friendship. M.P., A.S., T.P. and E.W. designed the study, E.W. and M.W. implemented the study, A.S. and M.W. analyzed the data, A.S. drafted the manuscript, E.W. and M.W. and T.P. commented on/edited the manuscript.

### References

- Belin, P., Fillion-Bilodeau, S., Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–9.
- van den Brink, D., Van Berkum, J.J., Bastiaansen, M., et al. (2010). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, *7*(2), 173–83.
- Bristow, D., Haynes, J.-D., Sylvester, R., Frith, C.D., Rees, G. (2005). Blinking suppresses the neural response to unchanging retinal stimulation. *Current Biology*, *15*(14), 1296–300.
- Brosch, T., Wieser, M.J. (2011). The (non) automaticity of amygdala responses to threat: on the issue of fast signals and slow measures. *Journal of Neuroscience*, *31*(41), 14451–2. <https://doi.org/10.1016/j.cub.2005.06.025>
- Brosch, T., Grandjean, D., Sander, D., Scherer, K.R. (2009). Cross-modal emotional attention: emotional voices modulate early stages of visual processing. *Journal of Cognitive Neuroscience*, *21*(9), 1670–9.
- Burra, N., Kerzel, D., Munoz Tord, D., Grandjean, D., Ceravolo, L. (2018). Early spatial attention deployment toward and away from aggressive voices. *Social Cognitive and Affective Neuroscience*, *14*(1), 73–80.
- Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Fink, G.R., Halligan, P.W., Marshall, J.C., et al. (1996). Where in the brain does visual attention select the forest and the trees? *Nature*, *382*(6592), 626–8.
- Folstein, J.R., Van, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, *45*(1), 152–70.

- Gaspar, J.G., Street, W.N., Windsor, M.B., et al. (2014). Providing views of the driving scene to drivers' conversation partners mitigates cell-phone-related distraction. *Psychological Science*, *25*(12), 2136–46.
- Globisch, J., Hamm, A.O., Esteves, F., Öhman, A. (1999). Fear appears fast: temporal course of startle reflex potentiation in animal fearful subjects. *Psychophysiology*, *36*(1), 66–75.
- Heinze, H.J., Luck, S.J., Mangun, G.R., Hillyard, S.A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection. *Electroencephalography and Clinical Neurophysiology*, *75*(6), 511–27.
- Herrmann, C.S., Knight, R.T. (2001). Mechanisms of human attention: event-related potentials and oscillations. *Neuroscience & Biobehavioral Reviews*, *25*(6), 465–76.
- Inzlicht, M., Bartholow, B.D., Hirsh, J.B. (2015). Emotional foundations of cognitive control. *Trends in Cognitive Sciences*, *19*(3), 126–32.
- Kayser, J., Tenke, C.E. (2003). Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Clinical Neurophysiology*, *114*(12), 2307–25.
- Kayser, J., Tenke, C.E. (2015). On the benefits of using surface Laplacian (current source density) methodology in electrophysiology. *International Journal of Psychophysiology*, *97*(3), 171–3.
- Kissler, J., Herbert, C., Winkler, I., Junghofer, M. (2009). Emotion and attention in visual word processing: an ERP study. *Biological Psychology*, *80*(1), 75–83.
- Luck, S.J., Hillyard, S.A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*(3), 291–308.
- Lui, M.A., Penney, T.B., Schirmer, A. (2011). Emotion effects on timing: attention versus pacemaker accounts. *PLoS One*, *6*(7), e21829.
- Luo, Q., Holroyd, T., Majestic, C., Cheng, X., Schechter, J., Blair, R.J. (2010). Emotional automaticity is a matter of timing. *Journal of Neuroscience*, *30*(17), 5825–9.
- Mangun, G.R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18.
- Min, C.S., Schirmer, A. (2011). Perceiving verbal and vocal emotions in a second language. *Cognition & Emotion*, *25*(8), 1376–92.
- Mothes-Lasch, M., Mentzel, H.-J., Miltner, W.H.R., Straube, T. (2011). Visual attention modulates brain activation to angry voices. *The Journal of Neuroscience*, *31*(26), 9594–8.
- Mothes-Lasch, M., Miltner, W.H.R., Straube, T. (2012). Processing of angry voices is modulated by visual load. *NeuroImage*, *63*(1), 485–90.
- Müller-Bardorff, M., Bruchmann, M., Mothes-Lasch, M., et al. (2018). Early brain responses to affective faces: a simultaneous EEG-fMRI study. *NeuroImage*, *178*, 660–7.
- Öhman, A., Soares, J.J. (1994). Unconscious anxiety: phobic responses to masked stimuli. *Journal of Abnormal Psychology*, *103*(2), 231–40.
- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, *9*(2), 148–58.
- Pessoa, L. (2015). Précis on the cognitive-emotional brain. *The Behavioral and Brain Sciences*, *38*, e71.
- Pessoa, L., McKenna, M., Gutierrez, E., Ungerleider, L.G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11458–63.
- Pourtois, G., Schwartz, S., Seghier, M.L., Lazeyras, F., Vuilleumier, P. (2006). Neural systems for orienting attention to the location of threat signals: an event-related fMRI study. *NeuroImage*, *31*(2), 920–33.
- Pourtois, G., Spinelli, L., Seeck, M., Vuilleumier, P. (2010). Temporal precedence of emotion over attention modulations in the lateral amygdala: intracranial ERP evidence from a patient with temporal lobe epilepsy. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(1), 83–93.
- Pourtois, G., Schettino, A., Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: what is magic and what is not. *Biological Psychology*, *92*(3), 492–512.
- Proverbio, A.M., Galli, J. (2016). Women are better at seeing faces where there are none: an ERP study of face pareidolia. *Social Cognitive and Affective Neuroscience*, *11*(9), 1501–1512. <https://doi.org/10.1093/scan/nsw064>
- Proverbio, A.M., Adorni, R., Zani, A., Trestianu, L. (2009). Sex differences in the brain response to affective scenes with or without humans. *Neuropsychologia*, *47*(12), 2374–88.
- Schindler, S., Kruse, O., Stark, R., Kissler, J. (2019). Attributed social context and emotional content recruit frontal and limbic brain regions during virtual feedback processing. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(2), 239–252. <https://doi.org/10.3758/s13415-018-00660-5>
- Schindler, S., Schettino, A., Pourtois, G. (2018a). Electrophysiological correlates of the interplay between low-level visual features and emotional content during word reading. *Scientific Reports*, *8*(1), 12228. <https://doi.org/10.1038/s41598-018-30701-5>
- Schirmer, A. (2013). Sex differences in emotion. In: Armony, J.L., Vuilleumier, P., editors. *The Cambridge Handbook of Human Affective Neuroscience*, Cambridge, UK: Cambridge University Press, pp. 591–610.
- Schirmer, A. (2018). Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. *Social Cognitive and Affective Neuroscience*, *13*(1), 1–13. <https://doi.org/10.1093/scan/nsx142>
- Schirmer, A., Gunter, T.C. (2017). Temporal signatures of processing voiceness and emotion in sound. *Social Cognitive and Affective Neuroscience*, *12*(6), 902–9.
- Schirmer, A., McGlone, F. (2018). A touching sight: EEG/ERP correlates for the vicarious processing of affectionate touch. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *111*, 1–15
- Schirmer, A., Kotz, S.A., Friederici, A.D. (2005b). On the role of attention for the processing of emotions in speech: sex differences revisited. *Brain Research. Cognitive Brain Research*, *24*(3), 442–52.
- Schirmer, A., Striano, T., Friederici, A.D. (2005a). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, *16*(6), 635–9.
- 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), 406–17.
- Schirmer, A., Chen, C.-B., Ching, A., Tan, L., Hong, R.Y. (2013b). Vocal emotions influence verbal memory: neural correlates and interindividual differences. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(1), 80–93.
- Schirmer, A., Seow, C.S., Penney, T.B. (2013a). Humans process dog and human facial affect in similar ways. *PLoS One*, *8*(9), e74591.
- Schirmer, A., Ng, T., Ebstein, R.P. (2018). Vicarious social touch biases gazing at faces and facial emotions. *Emotion*.
- Schneider, D., Beste, C., Wascher, E. (2012). On the time course of bottom-up and top-down processes in selective visual attention: an EEG study. *Psychophysiology*, *49*(11), 1660–71.



- Schupp, H.T., Flaisch, T., Stockburger, J., Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. In: Anders, G.E.S., editor. *Progress in Brain Research*, Vol. 156, pp. 31–51 Available from: <http://www.sciencedirect.com/science/article/pii/S0079612306560029>.
- Straube, T., Sauer, A., Miltner, W.H.R. (2011). Brain activation during direct and indirect processing of positive and negative words. *Behavioural Brain Research*, 222(1), 66–72.
- Vuilleumier, P., Armony, J.L., Driver, J., Dolan, R.J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*, 30(3), 829–41.
- Winkler, I., Debener, S., Müller, K.-R., Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In: *Conference proceedings: annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society*, pp. 4101–5.
- Woodman, G.F., Luck, S.J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, 400(6747), 867–9.
- Zani, A., Marsili, G., Senerchia, A., et al. (2015). ERP signs of categorical and supra-categorical processing of visual information. *Biological Psychology*, 104, 90–107.