Electro-cortical correlates of multisensory integration using ecologically valid emotional stimuli

Stefanou, M.E.; Dundon, Neil; Bestelmeyer, Patricia; Koldewyn, Kami; Saville, Christopher; Fleischhaker, Christian; Feige, B.; Biscaldi, Monica; Smyrnis, Nikolaos ; Klein, Christoph

Biological Psychology

DOI:
10.1016/j.biopsycho.2019.01.011

Published: 01/03/2019

Peer reviewed version

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Stefanou, M. E., Dundon, N., Bestelmeyer, P., Koldewyn, K., Saville, C., Fleischhaker, C., Feige, B., Biscaldi, M., Smyrnis, N., & Klein, C. (2019). Electro-cortical correlates of multisensory integration using ecologically valid emotional stimuli: Differential effects for fear and disgust. *Biological Psychology*, *142*, 132-139. https://doi.org/10.1016/j.biopsycho.2019.01.011

Electro-cortical correlates of multisensory integration using ecologically valid emotional stimuli: Differential effects for fear and disgust

Stefanou, M.E.[a], Dundon, N.[a,b], Bestelmeyer P.E.G.[c], Koldewyn K.[c] , Saville C.W.N.[c], Fleischhaker, C.[a], Feige, B.[d], Biscaldi M.[a], Smyrnis, N.[e], Klein, C.[a,e,f]

[a] Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, Faculty of Medicine,University of Freiburg, Germany
[b]Brain Imaging Center, Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA, USA
[c] School of Psychology, Bangor University, Bangor, United Kingdom,
[d] Department of Psychiatry and Psychotherapy, Faculty of Medicine,, University of Freiburg, Germany
[e] Department of Psychiatry, National and Kapodistrian University of Athens, Medical School, Athens, Greece
[f] Department of Child and Adolescent Psychiatry, Medical Faculty, University of Cologne, Germany

This research was supported by a stipend to M.E Stefanou by the State Law on Graduate Funding (LGFG), Albert-Ludwigs-Universität Freiburg

Declaration of interest: none

Please address correspondence to:

Prof. Dr. Christoph Klein

Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics,

Medical Faculty,
University of Freiburg
Hauptstrasse 8,
D-79104 Freiburg,
Germany

Email: christoph.klein.kjp@uniklinik-freiburg.de;

Phone: +49 761 270-68804; Fax: +49 761 270-68590

Abstract

Multisensory integration (MSI) is crucial for human communication and social interaction and has been investigated in healthy populations and neurodevelopmental disorders. However, the use of stimuli with high ecological validity is sparse, especially in event-related potential (ERP) studies. The present study examined the ERP correlates of MSI in healthy adults using short (500ms) ecologically valid professional actor-produced emotions of fear or disgust as vocal exclamation or facial expression (unimodal conditions) or both (bimodal condition). Behaviourally, our results show a general visual dominance effect (similarly fast responses following bimodal and visual stimuli) and an MSI-related speedup of responses only for fear. Electrophysiologically, both P100 and N170 showed MSI-related amplitude increases only following fear, but not disgust stimuli. Our results show for the first time that the known differential neural processing of fear and disgust also holds for the integration of dynamic auditory and visual information.

Keywords: Multisensory Integration (MSI); Event Related Potentials (ERPs); Miller's Race Model Inequality; emotional dynamic stimuli

1. Introduction

A prerequisite for a successful interaction with our environment is the ability to adequately integrate the information we receive from different sensory channels (Collignon et al., 2008; Magnée et al., 2011). Multisensory integration (MSI) is accomplished through the redundancy of information across sensory channels, such as in the case of a facial expression of joy and a joyful exclamation complementing each other. Such bimodal presentations of congruent visual and auditory stimuli generate faster reaction times (RT; Brandwein et al., 2011, 2013; Collignon et al., 2013) as well as more accurate responses (Giard & Peronnet, 1999) than the unimodal presentations alone.

Several studies have employed high temporal resolution techniques such as electroencephalography (EEG) in order to trace the unfolding of MSI through time. Studies using non-social stimuli have shown MSI-specific activations starting as early as 40-90 ms and lasting up to 275 ms after stimulus onset at different topographical locations (Giard & Perronet, 1999; Brandwein et al., 2011). Such findings suggest that MSI effects start early on, during pre-attentive and early perceptual stages, before stimuli are fully processed. At the same time, several brain structures such as the sensory cortices, the superior colliculus, middle temporal gyrus and the posterior superior temporal sulcus have been implicated in MSI (Campanella & Belin, 2007; Hornix et al., 2018; Meredith & Stein, 1983; Murray et al., 2016; Stein et al., 2014).

Notably, MSI has been also shown for emotion perception since the two key sensory inputs that provide us with (redundant or complementary) emotion signals are the human face and voice (Gervais et al., 2004). The integration of these two inputs allows us to efficiently recognize the emotions of other individuals (Harms, et al., 2010) and therefore to infer their intentions in social situations (Magnée et al., 2011). So far, healthy adults have been found to show a significant speed-up of responses as a result of MSI with emotional stimuli (e.g. Collignon et al., 2008) and it has been suggested that the integration of face and voice information takes place at the latest around 178 ms after stimulus onset (de Gelder et al., 1999).

Recent evidence shows that the dynamic nature of stimuli is an important factor of MSI. That is, static images would not be representative of our complex environment since in real life we are constantly confronted with changing sceneries and information. Specifically, neuroimaging studies have found that the amygdala and fusiform gyrus respond differently to dynamic facial expressions compared to static faces (LaBar et al., 2003) and MSI effects are stronger when participants are being presented with dynamic compared to static stimuli (for a review see, Campanella & Belin, 2007). Additionally, it has been suggested that the simultaneous presentation of a dynamic stimulus in one modality (e.g., audio) with a static one in another (e.g., still pictures; de Gelder et al., 1999; Magnée et al., 2008) could lead to an incongruity between channels (Jessen & Kotz, 2011) due to the dynamic versus non-dynamic nature of the stimuli in the two channels.

MSI manifestations have been observed in terms of super-additivity elicited by the bimodal condition at temporal-occipital areas at 60–148ms as well as in the form of under-additivity of the N170 component (decreased activity during the bimodal condition; Brefczynski-Lewis et al., 2009). Intriguingly, the audio-visual interaction that Brefczynski-Lewis et al. (2009) found with the use of dynamic emotional stimuli showed a left hemisphere laterality, an effect opposite to the one reported from studies that used non-social stimuli (e.g., Giard & Perronet, 1999). This discrepancy may point to stimulus-specificity of some of the results obtained with the ERP technology which, in turn, questions the generalizability of findings, thus emphasizing the demand for ecologically valid stimuli. Studies accounting for the multimodality of emotions have mainly used stimuli with limited ecological validity such as static images in combination with verbal cues (e.g., de Gelder et al., 1999). The use of different tasks and stimuli between studies also limits the replication of previous findings while studies combining the use of emotional stimuli with high ecological validity and EEG techniques are scarce (e.g., Stekelenburg & Vroomen, 2007).

Based on these considerations, the goal of the present study was to investigate MSI effects on visual components by the addition of auditory signals, with the use of ecologically valid emotional dynamic stimuli. We aimed to replicate in healthy adults the behavioural results of Charbonneau et al. (2013) and additionally trace the temporal structure and neurophysiological basis of MSI through EEG. The replication and validation of an emotional MSI effect using more ecologically valid dynamic stimuli could serve as a solid base for future studies on MSI deficits in ASD. In the absence of clear auditory ERP components and hence focusing on the visual components, we assume to demonstrate the presence of neural MSI, if the latencies and/or amplitudes of these components are modulated in bimodal visual plus auditory emotion presentations when compared to visual only presentations.

According to the literature we formed the following hypotheses. Firstly, we expected to find increased accuracy and speeded RTs in the bimodal compared to the unimodal conditions as well as the presence of a significant redundancy gain (RG, see methods). Secondly, we expected an increase of the visual P100 amplitude and a decrease of the visual N170 in the bimodal compared to the visual condition.

2. Methods

2.1. Participants

A total of eighteen healthy volunteers between the ages of 21 and 28 years were recruited for this study (9 female; age: 24.99±2.86; all right-handed). Participants were excluded in case of unsuitability of the EEG data (N=1). The experimental protocol was approved by the Ethics Committee of the Albert Ludwigs-University Freiburg, and according to the Declaration of Helsinki all data were treated with full confidentiality. In addition to the written information, participants were verbally informed about the purpose of the study and provided signed informed consents.

2.2. Stimuli and procedure

Participants were seated inside a dimly-lit sound-attenuated Faraday cage. They completed a forced-choice discrimination task where they were presented with the emotions of fear and disgust; participants were instructed to press the key “A” with the left-hand index finger if the presented emotion was fear and key “L” with the right-hand index finger for disgust on a Dell keyboard. The two emotions were presented by a female and a male actor in a visual (video with no sound), a bimodal (video) and an auditory (sound only) condition. Each of these three conditions was presented in two blocks of 200 trials each, resulting in a total of 1,200 trials and six blocks that were presented in a counter-balanced order across participants and over a single session. The duration of each stimulus was 500 ms, followed by an inter-stimulus interval (ISI) of 2,000 ms. A white fixation cross on a black background was present during the ISI. Participants were instructed to respond as quickly and as accurately as possible.

Original stimuli were adopted from Simon et al. (2008) and were processed with Adobe Premiere Elements (Adobe Systems, Inc.). The selected videos were segmented to sequences of 500 ms (15 frames) and the audio clips were exported based on these sequences. All stimuli started with a neutral expression for 1 frame evolving into full expression thereafter. The task was presented through a 24” TV monitor (with built-in speakers) connected to a stimulation computer running Presentation V.17.2 Software (Neurobehavioral Systems, USA). Participants were seated at an approximate viewing distance of 80 cm.

EEG was acquired with two BrainAmps DC-amplifiers and using the BrainVision Recorder (Brain Products, Gilching). We used a 64-channel actiCap (Brain Products, Gilching), with electrodes placed according to the International 10-10 System (American

Electroencephalographic Society, 1991). Data were recorded at a sampling rate of 500Hz and impedances were kept below 5kΩ. FCz and AFz electrodes served as reference and ground, respectively. Additionally, two infraorbital channels were placed vertically under each eye, as well as an electrode positioned at Nasion.

2.3 Data Processing and Analysis

2.3.1 Behavioural data

Correct responses were defined as the first response that positively identified the presented emotion within the given time window. Accuracy was then calculated as their relative probability. Median RT was determined as the latency of correct responses ranging from 150-1800ms.

In order to behaviourally assay MSI, we computed redundancy gain (RG) per emotion, i.e., the percentage decrease of the mean RT during the bimodal condition compared to the mean of the fastest unimodal condition for each participant. Furthermore, we applied Miller's Race Model Inequality (RMI; Miller, 1982; as described in Ulrich, Miller & Schröter, 2007), for each emotion separately, in order to ensure that the faster bimodal RT were caused by MSI and not by a probabilistic "race" between two separate signals. MSI was calculated at every 5th percentile and for each individual and was subsequently submitted to a one-sample t-test analysis.

2.3.2. EEG Data Processing

Offline processing of the EEG data was performed in Brain Vision Analyzer (Version 2.0, Brain Products, Gilching). Data were first filtered with a 0.1-45Hz band-pass filter and then down-sampled to 100Hz. Secondly, data sections with a voltage of ≤0.5µV or ≥1,500µV and duration of ≥200ms were considered as artefact-contaminated and were excluded from further analysis (including data ±200ms relative to the artefact). Thirdly, remaining data were segmented into epochs beginning 200ms prior stimulus onset and ending 1,800ms after stimulus onset. All components representing artefacts such as eye blinks, saccades, muscle activity and other movements where identified based on their topographies and time courses through an Infomax Independent Component Analysis (ICA) and removed through a semi-automatic ICA Inverse. After an additional data inspection segments with activity ≤0.5μV or ≥200μV for a period ≥200ms were again excluded, in semi-automatic mode in order to also visually inspect the selected trials. Data were then re-referenced to the average reference and segments were averaged according to condition and emotion. Individual averaged ERPs were created after baseline correction was normalized to the period of 200ms prior stimulus onset.

Based on the grand averages and topographical maps we selected the visual P100 and N170 components. For the visual P100, peak picking was performed separately at electrodes PO7, PO8, PO9 and PO10 at 80-190ms; for the visual N170 at electrodes PO7, PO8, PO9 and PO10 at 140-240ms. A positive peak in the bimodal and visual conditions at approximately 300ms was identified as a P3b and peak-picking was performed at electrodes PO7, POz and PO8 at 200-400ms. The peak for each component was determined separately for each subject and condition, as the maximum peak of each of the aforementioned electrodes (positive or negative according to the component). The time windows for the peak picking process were decided according to the relevant literature and through visual inspection of each component's timing and topography. Peak information was then exported as the mean amplitude over 3 data points (100 Hz sampling rate/30 ms; the peak ± one data-point) around the peak of each component. In order to investigate MSI interactions, individualized sums of the unimodal conditions and difference waves (bimodal-sum) were also calculated.

2.3.3. Statistical Analysis

Median reaction times (RT) and the percentage of correct responses were submitted to two 2*2 repeated measures ANOVA with CONDITION (auditory/visual, bimodal) and EMOTION (fear, disgust) as within-subjects factors. Only the bimodal condition on the one side and the two unimodal conditions on the other were compared since the focus of the study is MSI, and not differences between auditory and visual processing. In order to account for possible time-on-task effects (habituation effects practice etc.) we performed additional analyses with the factor BLOCK. Although participants' responses to the auditory stimuli were faster during the second compared to the first block (*p*=.008), all results from the main ANOVAs remained the same and will therefore be presented without the factor BLOCK.

EEG analysis, too, was limited to comparisons between the bimodal condition and the unimodal conditions. EEG data were submitted to the following analyses. The visual P100 was submitted to a 2*2*4 RM ANOVA with CONDITION (bimodal, visual), EMOTION (fear, disgust) and ELECTRODE SITE (PO7, PO8, PO9, PO10) as factors. The visual N170 was submitted to 2*2*4 RM ANOVA with the factors CONDITION (bimodal, visual), EMOTION (fear, disgust) and ELECTRODE SITE (PO7, PO8, PO9, PO10). Finally, amplitudes and latencies of the P3b were submitted to a 2*2*3 RM ANOVA with CONDITION (auditory, bimodal), EMOTION (fear, disgust) and ELECTRODE SITE (PO7, POz, PO8,) as factors. The amplitude values of the difference waves (time-locked at the P100 and N170 peak latencies of the bimodal condition) were submitted to t-tests against zero. Greenhouse-Geisser corrected values are reported when appropriate. Further analysis of the EEG data with the additional factor of BLOCK, excluded the possibility of habituation since no main effects or interactions involving the factor BLOCK were found to be significant.

In order to further investigate the spatio-temporal evolution of MSI the bimodal and sum conditions were submitted in a cluster-based permutation test implemented in Fieldtrip toolbox (Maris & Oostenveld, 2007; Maris, 2012). Amplitude changes between the two conditions, for each emotion separately, were analysed for every data point (i.e., 10ms), from stimulus onset and until 400ms post stimulus-onset. This analysis was chosen as it elegantly controls for multiple comparisons and accounts for the dependency of the EEG data. This analysis assumes a null hypothesis of no differences and that the data from the two conditions are exchangeable. Therefore, the two conditions are combined and then randomly partitioned in two pseudo-conditions. The randomization was repeated for 10,000 times in order to create a reference distribution for the comparison of the two conditions at each data point. Clusters are defined by adjacent electrodes that show the same effect.

Since MSI effects were expected both for behavioural and EEG data, we sought to investigate possible relations between RTs and ERPs. Individual P100 and N170 amplitudes of the difference waves [bimodal-(visual+auditory)] were submitted to a correlation analysis with the individual RG scores as well as with the individual MSI scores, as defined by Miller's RMI, at the percentile with the maximum MSI for each emotion (40th percentile for fear; 10th percentile for disgust). An additional correlation of the residuals obtained by regression analysis of the P100/N170 amplitude and median RTs (regressing out the visual from the bimodal condition) was performed. Since all the above analyses did not reveal any significant correlations between the MSI effects in behavioural and EEG data, results will not be presented in the following sections. However, a lack of a significant correlation does not explicitly imply different processing effects. Given that RTs include not only stimulus detection and motor response, but additional mental processes (Jaskowski, 1996) such as decision as to the presented emotion in the given study, a correlation of behavioural and visual ERP components that are elicited at detection of stimulus would not always be expected. Finally, one should keep in mind the reliability of differences scores is always an issue (Thomas & Zumbo, 2012).

3. Results

3.1 Behavioural Results.

Bimodal versus Visual. The bimodal condition produced somewhat faster RT and this speed-up was greater for trials of fear ($t_{(16)}$=2.485 p=.024, d=1.24)(-30ms) than disgust ($t_{(16)}$=.566, p=.580, d=.283) (-7ms; CONDITION: $F_{(1,16)}$=2.423, p=.139, η_p^2=.13; CONDITION*

EMOTION: $F_{(1,16)}$=31.454, p<.001, η_p^2=.66; see **Table 1**). Accuracy of responses was higher for the bimodal compared to the unimodal condition and this effect was significantly greater for trials presenting fear than disgust (CONDITION: $F_{(1,16)}$=1.231, p=.284, η_p^2=.07; CONDITION*EMOTION: $F_{(1,16)}$=4.589, p=.048, η_p^2=.22; see **Table 1**).

Bimodal versus Auditory. Participants were responding faster for trials of fear than disgust (EMOTION: $F_{(1,16)}$=8.605, p=.010, η_p^2=.35). There was an additional speedup of RT during the bimodal compared to the auditory condition and this difference was further amplified for trials presenting disgust ($t_{(16)}$=17.370 p<.001, d=8.68) than to fear ($t_{(16)}$=12.341 p<.001, d=6.17; CONDITION*EMOTION: $F_{(1,16)}$=23.247, p<.001, η_p^2=.59; CONDITION: $F_{(1,16)}$=215.992, p<.001, η_p^2=.93; see **Table 1**). There were no significant effects on the accuracy of responses.

MSI. The RG during trials of fear was significantly increased compared to the RG during disgust ($t_{(16)}$=5.76, p<.001; see **Figure 1/A**). Most importantly, as defined by Miller's RMI there was a significant MSI from the 5th until the 55th percentile for fear (see **Figure 2/A**). With regards to disgust, there was no significant MSI; the bimodal and visual conditions were nearly identical with the calculated bound (see **Figure 2/B**).

3.2 EEG Results

3.2.1 Early Perceptual Components

Visual P100. Overall, there was a significant increase of P100 amplitude during the bimodal compared to the visual condition (CONDITION: $F_{(1,16)}$=7.99, p=.012, η_p^2=.33; see **Figure 3/A**). When differentiating the emotions, however, this effect was significant only for the emotion of fear and not for disgust (CONDITION*EMOTION: $F_{(1,16)}$=6.96, p=.018, η_p^2=.30; see **Figure 3/B**; CONDITION effect for fear: $F_{(1,16)}$=11.36, p=.004, η_p^2=.42; disgust: $F_{(1,16)}$=2.559, p=.129, η_p^2=.14). Independent of condition, P100 amplitude was greater for trials of disgust compared to fear (EMOTION: $F_{(1,16)}$=30.9, p<.001, η_p^2=.66). No effect of electrode site was found, and there were no effects on P100 latency.

Visual N170. Overall, the N170 amplitude was greater for the bimodal compared to the visual condition but failed to reached statistical significance (CONDITION: $F_{(1,16)}$=2.325, p=.147, η_p^2=.13; see **Figure 4**). Differentiated according to emotions, however, this increase was significantly greater only for trials of fear but not disgust (CONDITION*EMOTION: $F_{(1,16)}$=5.604, p=.031, η_p^2=.26; CONDITION effect for: fear $F_{(1,16)}$=5.239, p=.036, η_p^2=.25; disgust: $F_{(1,16)}$=0.183, p=.675, η_p^2=.01). In addition, the N170 augmenting effect of fear compared to disgust was significantly greater for electrodes PO7 ($t_{(16)}$=-3.495, p=.003) and PO9 ($t_{(16)}$=-3.150, p=.006), and absent for electrodes PO8 and PO10 (ps>.05) (ELECTRODE*EMOTION: $F_{(1.799,28.791)}$=5.931, p=.009, η_p^2=.27; see **Figure 4**). No effects on N170 latency were found.

3.2.2. Late Cognitive Components

No effects of CONDITION or EMOTION on amplitude and latency of the P3b were found. P3b amplitude was increased at electrode PO8 compared to POz, with this difference being significantly greater for fear than disgust (EMOTION*ELECTRODE: $F_{(1,16)}$=10.08, p=.006, η_p^2=.39). Electrodes PO7 and PO8 showed a shorter P3b latencies than POz (ELECTRODE: $F_{(1.69,23.06)}$=5.24, p=.016, η_p^2=.25).

3.2.3. Bimodal Interactions

Pre-planned cluster-based permutation test. We contrasted the bimodal condition with the sum of the unimodal conditions, separately for fear and disgust. There were no significant differences between these conditions, neither for fear nor disgust.

4. Discussion

The present study set out to investigate the electro-cortical correlates of MSI in the processing of two different, negative valence-emotions using ecologically valid dynamic emotional stimuli. We intended to demonstrate behavioural MSI as speeded RT during the bimodal compared to the unimodal conditions and in the sense of Miller's RMI; furthermore, we intended to investigate MSI effects on the visual components, contrasting the bimodal with the unimodal visual condition. We obtained the following main results. Firstly, behaviourally we observed visual dominance as suggested by the similar visual and bimodal RTs alongside an MSI-related speed up of RTs that was significant only for clips presenting the emotion of fear. Secondly, we found MSI-related increases in the visual P100 and the N170 that were both greater for videos showing fear compared to disgust.

Visual dominance and fear-specific MSI effects. The stimuli and/or participants of our study produced visual dominance, as shown by the faster responses in the visual than the auditory condition despite auditory neural transduction being faster than visual (Pöppel et al., 1990; Pöppel, 1997; King, 2005). Indeed, it has been long established that during bimodal presentations, behavioural responses are often driven by the visual stimulus (Colavita, 1974). When one of the signals of the bimodal stimulus is reliable enough and produces a strong response by itself, the MSI effects are weak; furthermore, MSI effects are stronger in cases where one of the two signals is weak (Kayser & Logothetis, 2007). Visual dominance reduced the behavioural MSI effect as it rendered the additional auditory information in the bimodal condition "non-beneficiary". This, in turn, reduced the need for sensory integration and the strength of MSI as defined by Miller's RMI. Effects of visual dominance leading to a greater speedup of the bimodal RT compared to the visual RT have been previously reported (e.g., Collignon et al., 2008; Diakoniscu et al., 2013). Collignon et al. (2008) reduced the visual dominance in their task by degrading the reliability of the visual signal with Gaussian noise in the bimodal and visual conditions, verifying the hypothesis that MSI is stronger for "weak" signals and vice versa. Thus, our results suggest, that at least in complex stimuli, audio-visual MSI effects may be reduced by visual dominance. If replicated, this result would suggest that somewhat degraded visual stimuli should be used in MSI research employing fairly complex and ecologically valid stimuli like the ones used here. Furthermore, visual dominance has been shown to increase with age (Diakoniscu et al., 2013), explaining therefore the smaller MSI effects in our adult sample compared to the MSI effects reported in Charbonneau whose sample included younger ages as well.

Due to the small sample size, the MSI-effect (main effect of CONDITION) on RTs was at trend level only (two-sided testing of a directional hypothesis), corresponding nevertheless to a Cohen's d of .78. More importantly, however, the significant CONDITION x EMOTION interaction indicated that the RT reduction under the bimodal compared to the visual condition was exclusively driven by a large MSI effect for fear stimuli (p=.024) and a comparatively negligible and non-significant MSI effect for disgust stimuli (p=.580). Also, and as hypothesized, the multimodality of the stimuli produced significantly faster RT and significant MSI as indicated by Miller's RMI. Notably, these MSI effects were facilitated by fear and eliminated by disgust. Although both fear and disgust are negative emotions, they represent distinct emotions both in terms of surface reflectance features and from an evolutionary perspective (Susskind et al., 2008). While fear enhances attention, disgust diverts attention (to reduce exposure; Susskind et al., 2008). Therefore, each emotion's function serves to optimize our actions in specific situations, and is seen in behavioural responses with fear producing faster RT than disgust (Krusemark & Li, 2011; Zhang et al., 2017). Thus, our results show that the opposing effects of fear and disgust do not only alter uni-sensory processing; in addition, the behavioural responses to these two emotions with "opposing" attentional effects are shown here to be differentially sensitive to the early processing stages of MSI as well.

Electrophysiological correlates of MSI.

With regards to the electrophysiological correlates of MSI, we observed the first MSI effects in a larger bimodal than unimodal visual P100. This finding is in line with findings of super-additivity, that is increased bimodal compared to unimodal activity, both in EEG (e.g., Brandwein et al., 2011) and neuroimaging studies (e.g., Pourtois et al., 2005; for a review see, Camplanella & Belin, 2007) caused by the visual-auditory interactions. Importantly, this effect was larger for stimuli showing fear (*p*=.004) rather than disgust (*p*=.129). Since the visual P100 has been linked to facilitation of sensory processing for stimuli at attended locations (for a review, see Mangun, 1995; Hillyard, Vogel & Luck, 1998; Luck et al., 1990) the MSI-related increase of amplitude signifies a stronger facilitation of sensory processing due to the bimodal presentation. That this effect was larger for stimuli presenting fear rather than disgust, would thus reflect increased facilitation of sensory processing and orienting attention due to the emotional charge of the visual-auditory fear stimuli in the bimodal condition. Accordingly, differences in the neural processing of these two emotions are not restricted to the unimodal processing of emotional stimuli, but extend to the modulating influence of the additional auditory emotion exclamation in the bimodal condition. The EEG / ERP technology with its high temporal resolution, furthermore, reveal that these effects start already about 160ms after stimulus, that is, in early perceptual stages of information processing.

The amplitude of the visual N170, peaking around 210ms after stimulus onset, also showed an overall MSI-related increase which, in this case, was significant *only* for videos showing fear (*p*=.036) but *not* for stimuli presenting disgust (*p*=.675). N170 is a component elicited by the presentation and encoding of faces (Luck, 2005; Bentin et al., 1996). The N170 follows the P100 and is more specifically related to the processing of faces than the P100 (but see Thierry et al., 2007).

Emotional stimuli are thought to enhance visual processing (Padmala & Pessoa, 2008) and it has been suggested that, amongst other structures, the pulvinar enhances evoked responses via paths that include the amygdala and orbitofrontal cortex (for a review see Pessoa & Adolphs, 2010) based on the biological significance of the stimuli (for example fear-conditioned objects; Padmala et al., 2010). Despite some studies reporting no overall modulation of N170 amplitude by emotional faces (for example, Eimer et al., 2003 reported similar N170 amplitudes across six basic emotions), the differentiation of fear and disgust has been strongly supported both in terms of their biological "purpose" (Susskind et al., 2008) and from the fact that they activate distinct neural networks – fear produced greater activity than disgust in the amygdala[1] whereas disgust produced greater activity than fear in the anterior insula (Phillips et al., 1998). Accordingly, fear as compared to disgust (amongst other emotions), has been shown to produce a larger deflection of the N170 (Almeida et al., 2016; Batty & Taylor, 2003). Notably, all emotions (anger, disgust, fear, happiness and neutral/calm) Almeida et al. (2016) presented were matched for arousal suggesting that fear elicits a unique response and results in greater facilitation of sensory processing compared to other emotions. Our results add to these findings by showing that the facilitation effects of MSI that were *stronger* for fear than disgust in the early perceptual P100 become *exclusive* for fear and *absent* for disgust, when it comes to processing the proper face information. That fear produces a larger bimodal P100 response in combination with the absence of MSI effects for disgust (see Results) suggests that the divergent effects of these two emotions (engaging sympathetic versus parasympathetic system) enhance (or diminish) MSI effects as well in a manner similar to how they enhance/suppress visual attention – enhancement of visual attention by fear and suppression by disgust (Krusemark & Li, 2011). Furthermore, the *absence* of differential effects on P100 and N170 between neutral, angry and happy bimodal

---

[1] Although amygdala processes would be too slow to account for P100 and N170 emotion differentiation, the comparison is given purely in terms of comparison between fear and disgust effects on neural circuits; fear effects in amygdala have been reported from 200-800 ms (see Krolak-Salmon et al., 2004).

stimuli as for example in the study of Liu et al. (2012) indirectly corroborates the specific and opposing effects of fear and disgust reported in the present study.

Overall, our results suggest that for ecologically valid (videos) stimuli, MSI-effects can be to some degree emotion specific in that these effects are stronger for fear compared to disgust stimuli in the early perceptual processing stages indexed by the visual P100 and exclusive for the encoding of faces, indexed by the N170, and the later decision process as indirectly indexed by the reaction times. This pattern of findings likely reflects that enhanced sensory processing of fear compared to disgust is not only present with unisensory face processing but also when simultaneous auditory stimulation (vocal exclamation) modulate visual processing towards multi-sensory integration.

References

Almeida, P. R., Ferreira-santos, F., Chaves, P. L., Paiva, T. O., Barbosa, F., & Marques-teixeira, J. (2016). Perceived arousal of facial expressions of emotion modulates the N170 , regardless of emotional category : Time domain and time – frequency dynamics. *International Journal of Psychophysiology, 99*, 48–56. doi: 10.1016/j.ijpsycho.2015.11.017

Almeida, P. R., Ferreira-santos, F., Chaves, P. L., Paiva, T. O., Barbosa, F., & Marques-teixeira, J. (2016). Perceived arousal of facial expressions of emotion modulates the N170 , regardless of emotional category : Time domain and time – frequency dynamics. *International Journal of Psychophysiology, 99*, 48–56. doi: 10.1016/j.ijpsycho.2015.11.017

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience, 8*(6), 551–565. doi: 10.1162/jocn.1996.8.6.551

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral Cortex, 21*(5), 1042–55. doi:10.1093/cercor/bhq170

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex, 23*, 1329–1341. doi:10.1093/cercor/bhs109

Brefczynski-Lewis, J., Lowitszch, S., Parsons, M. Lemieux, S., & Puce A. (2009). Audiovisual Non-Verbal Dynamic Faces Elicit Converging fMRI and ERP Responses. *Brain Topogr, 21*: 193. doi:10.1007/s10548-009-0093-6

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. Trends in *Cognitive Sciences, 11*(12), 535–43. doi: 10.1016/j.tics.2007.10.001

Charbonneau, G., Bertone, A., Lepore, F., Nassim, M., Lassonde, M., Mottron, L., & Collignon, O. (2013). Multilevel alterations in the processing of audio-visual emotion expressions in autism spectrum disorders. *Neuropsychologia, 51*(5), 1002–10. doi:10.1016/j.neuropsychologia.2013.02.009

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*(2), 409–412. doi: 10.3758/BF03203962

Collignon, O., Charbonneau, G., Peters, F., Nassim, M., Lassonde, M., Lepore, F., … Bertone, A. (2013). Reduced multisensory facilitation in persons with autism. *Cortex, 49(*6), 1704–10. doi:10.1016/j.cortex.2012.06.001

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research, 1242*, 126–35. doi:10.1016/j.brainres.2008.04.023

de Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neuroscience Letters, 260*(2), 133–136. doi:10.1016/S0304-3940(98)00963-X

Diaconescu, A. O., Hasher, L., & McIntosh, A. R. (2013). Visual dominance and multisensory integration changes with age. *NeuroImage, 65*, 152–166. doi: 10.1016/j.neuroimage.2012.09.057

Eimer, M., & Holmes, A. (2003). The role of spatial attention in the processing of facial expression : An ERP study of rapid brain responses to six basic emotions. *Cognitive, Affective, & Behavioral Neuroscience, 3*(2), 97–110.

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., … Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience, 7*(8), 801–2. doi:10.1038/nn1291

Giard, M., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473–490. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/089892999563544

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*(3), 290–322. doi:10.1007/s11065-010-9138-6

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philosophical Transactions of the Royal Society: Biological Sciences, 353, 1257-1270.

Hornix, B. E., Havekes, R., & Kas, M. J.H. (2018). Multisensory cortical processing and dysfunction across the neuropsychiatric spectrum. *Neuroscience and Biobehavioral Reviews*. doi: 10.1016/j.neubiorev.2018.02.010

Jaskowski, P. (1996). Simple reaction time and perception of temporal order: Dissociations and Hypotheses. *Perceptual and Motor Skills, 82*, 707-730.

Jessen, S., & Kotz, S.A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions, *NeuroImage, 58*(2), 665-674. doi:10.1016/j. neuroimage.2011.2011.06.035.

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure & Function, 212*(2), 121–32. doi: 10.1007/s00429-007-0154-0

King, A.J. (2005). Multisensory Integration: Strategies for Synchronization. *Current Biology, 15*(9), R339-R341. doi: 10.1016/j.cub.2005.04.022

Krolak-Salmon, P., He, M., Vighetto, A., Bertrand, O., Thomas, A., & Lyon, C. B. (2004). Early Amygdala Reaction to Fear Spreading in Occipital, Temporal, and Frontal Cortex: A Depth Electrode ERP Study in Human. *Neuron, 42*, 665–676.

Krusemark, E. A., & Li, W. (2011). Do all threats work the same way? Divergent effects of fear and disgust on sensory perception and attention. *J Neurosci., 31*(9), 3429–3434. doi: 10.1523/JNEUROSCI.4394-10.2011

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex, 13*(10), 1023–1033.

Liu, T., Pinheiro, A., Zhao, Z., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Emotional cues during simultaneous face and voice processing: Electrophysiological insights. PLoS ONE, 7(2). doi: 10.1371/journal.pone.0031001

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

Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalography and clinical Neurophysiology, 75*, 528-542.

Magnée, M. J. C. M., de Gelder, B., van Engeland, H., & Kemner, C. (2011). Multisensory integration and attention in Autism Spectrum disorder: Evidence from Event-related potentials. *PLoS One*, *6*(8), 6–10. doi:10.1371/journal.pone.0024196

Magnée, M. J. C. M., de Gelder, B., van Engeland, H., & Kemner, C. (2008). Atypical processing of fearful face-voice pairs in Pervasive Developmental Disorder: an ERP study. *Clinical Neurophysiology, 119*(9), 2004-2010. doi: 10.1016/j.clinph.2008.05.005

Mangun, G. R. (1995). Neural mechanisms of visual selection attention. *Psychophysiology, 32*, 4-18.

Maris, E. (2011). Statistical testing in electrophysiological studies. *Psychophysiology, 49,* 549–565. doi: 10.1111/j.1469-8986.2011.01320.x

Maris E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177–190. doi: 10.1016/j.jneumeth.2007.03.024

MATLAB and Statistics Toolbox Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States.

Meredith, M., & Stein, B. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389–391.

Miller, J. (1982). Divided attention: Evidence for Coactivation with redundant signals. *Cognitive Psychology, 14*, (247-279).

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia, 83,* 161-169. doi: 10.1016/j.neuropsychologia.2015.08.011

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain*, *135*(1–2), 55–64. doi: 10.1016/j.pain.2007.05.008

Padmala, S., Lim, S., & Pessoa, L. (2010). Pulvinar and affective significance : responses track moment □ to-moment stimulus visibility, 4(September), 1–9. doi: 10.3389/fnhum.2010.00064

Padmala, S., & Pessoa, L. (2008). Affective learning enhances visual detection and responses in primary visual cortex. *J Neurosci., 28*(24), 6202–6210. doi: 10.1523/JNEUROSCI.1233-08.2008.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a ‘low road’ to ‘many roads’ of evaluating biological significance. *Nat Rev Neurosci,. 11*(11), 773–783. doi: 10.1038/nrn2920

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., … Gray, J. A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc. R. Soc. Lond., 265*, 1809-1817

Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of Facial Expressions and Voices and of their Combination in the Human Brain. *Cortex, 41*(1), 49–59. doi: 10.1016/S0010-9452(08)70177-1

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences, 1*(2), 56-61.

Pöppel, E., Schill, K., von Steinbüchel, N. (1990). Sensory Integration within Temporally Neutral System States: A Hypothesis. *Naturwissenschaften, 77*, 89-91.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience, 15*(8), 520–535. doi: 10.1038/nrn3742

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*, 1964–1973. doi:10.1162/jocn.2007.91213

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., & Anderson, A. K. (2008). Expressing fear enhances sensory acquisition, 11(7), 843–850. doi: 10.1038/nn.2138

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. Nature Neuroscience, 10, 505. doi: 10.1038/nn1864

Thomas, D. R. & Zumbo, B.D. (2012). Difference Scores From the Point of View of Reliability and Repeated-Measures ANOVA: In Defense of Difference Scores for Data Analysis. Educational and Psychological Measurement 72, 37 –43.

Ulrich, R., Miller, J., & Schröter H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods, 39*(2), 291-302. doi: 10.3758/BF03193160

Zhang, D., Liu, Y., Wang, L., Ai, H., & Luo, Y. (2017). Mechanisms for attentional modulation by threatening emotions of fear, anger , and disgust. *Cognitive, Affective, & Behavioral Neuroscience, 17,* 198–210. doi: 10.3758/s13415-016-0473-9

Figure captions.

Figure 1. Redundancy gain in reaction times

Fear produced a significantly larger redundancy gain than disgust; **p<.001

Figure 2. Miller's Race Model Inequality

MSI as defined by Miller's RMI for A) fear showing a significant integration from the 5th to the 55th percentile, and B) disgust that showing no MSI

Figure 3. MSI effects on P100

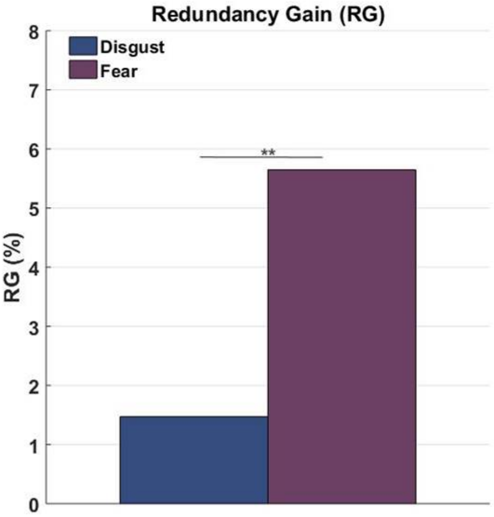
A) PO8 and the topographical maps illustrate the significant increase of the visual P100 amplitude at the bimodal compared to visual condition; B) illustrates the significant CONDITION x EMOTION interaction, where the increase of P100 amplitude at the bimodal condition is amplified for trials of fear compared to disgust

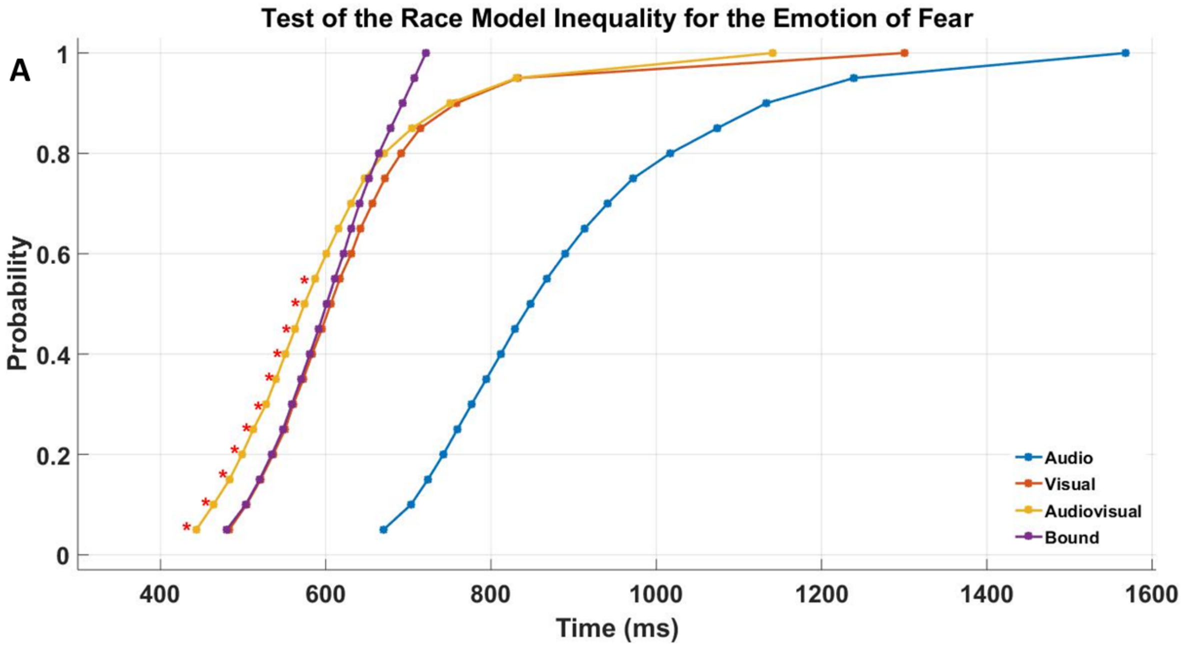
Figure 4. MSI effects on N170

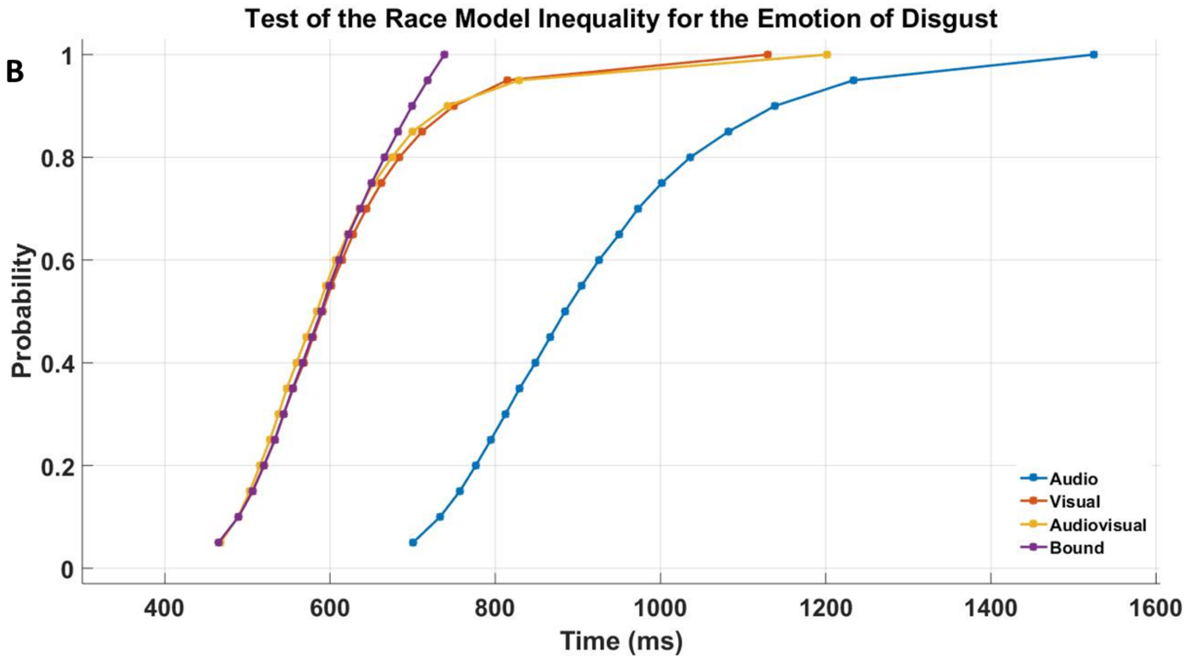
A) PO7 and the topographical maps illustrate the increase of the visual N170 amplitude at the bimodal compared to visual condition; B) illustrates the CONDITION x EMOTION interaction, where the increase of the N170 amplitude at the bimodal condition is amplified for trials of fear compared to disgust

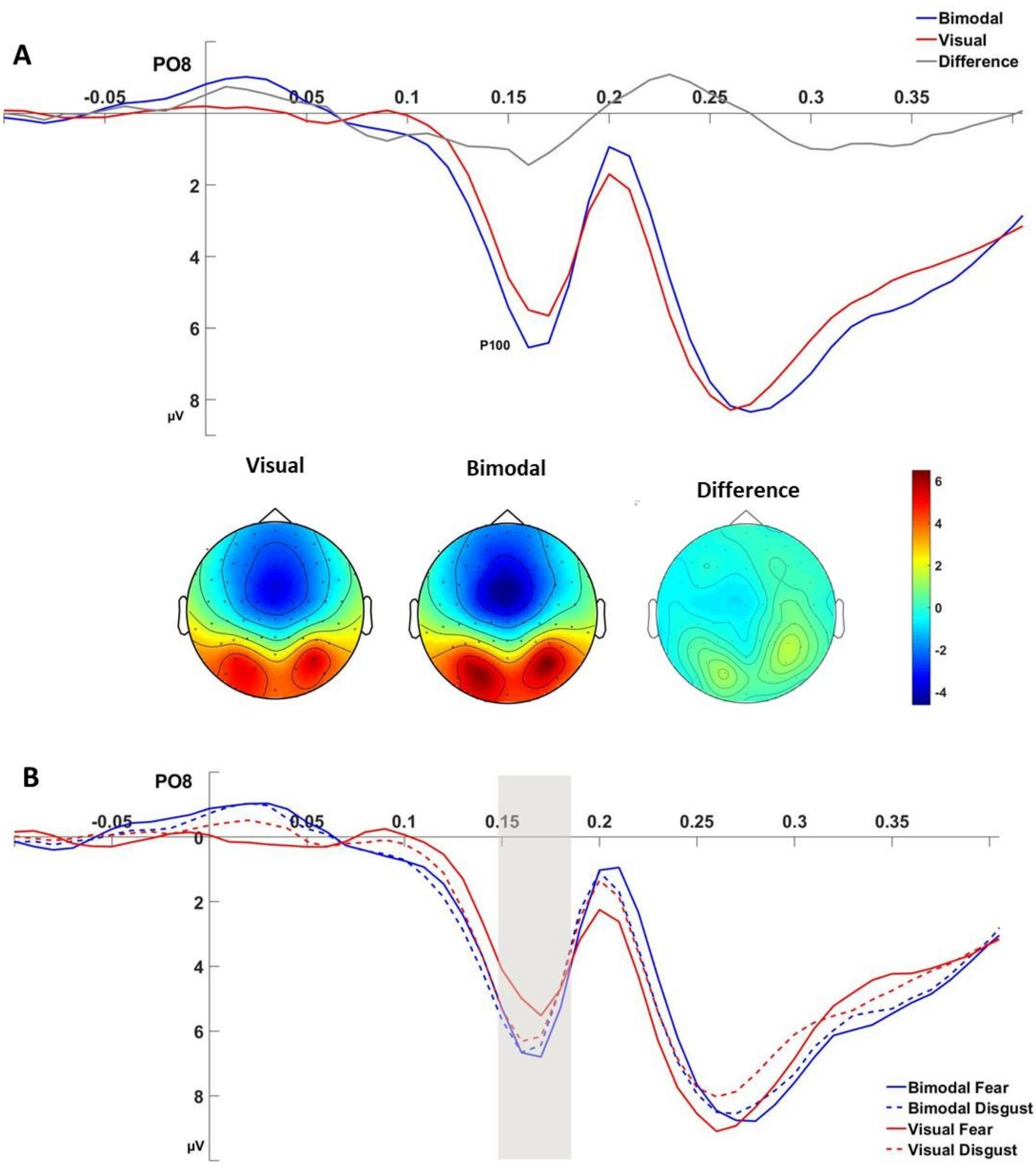
	Accuracy			**Median RTs**		
	Audio	Bimodal	Visual	Audio	Bimodal	Visual
Fear	95.5 *(3.8)*	95.4 *(4.9)*	94.2 *(5)*	842.1 *(154.8)*	574.6 *(113)*	605.44 *(97.3)*
Disgust	95 *(4.6)*	95.3 *(4.8)*	95.5 *(3.5)*	884.3 *(141)*	584.3 *(111.3)*	591.2 *(88.6)*

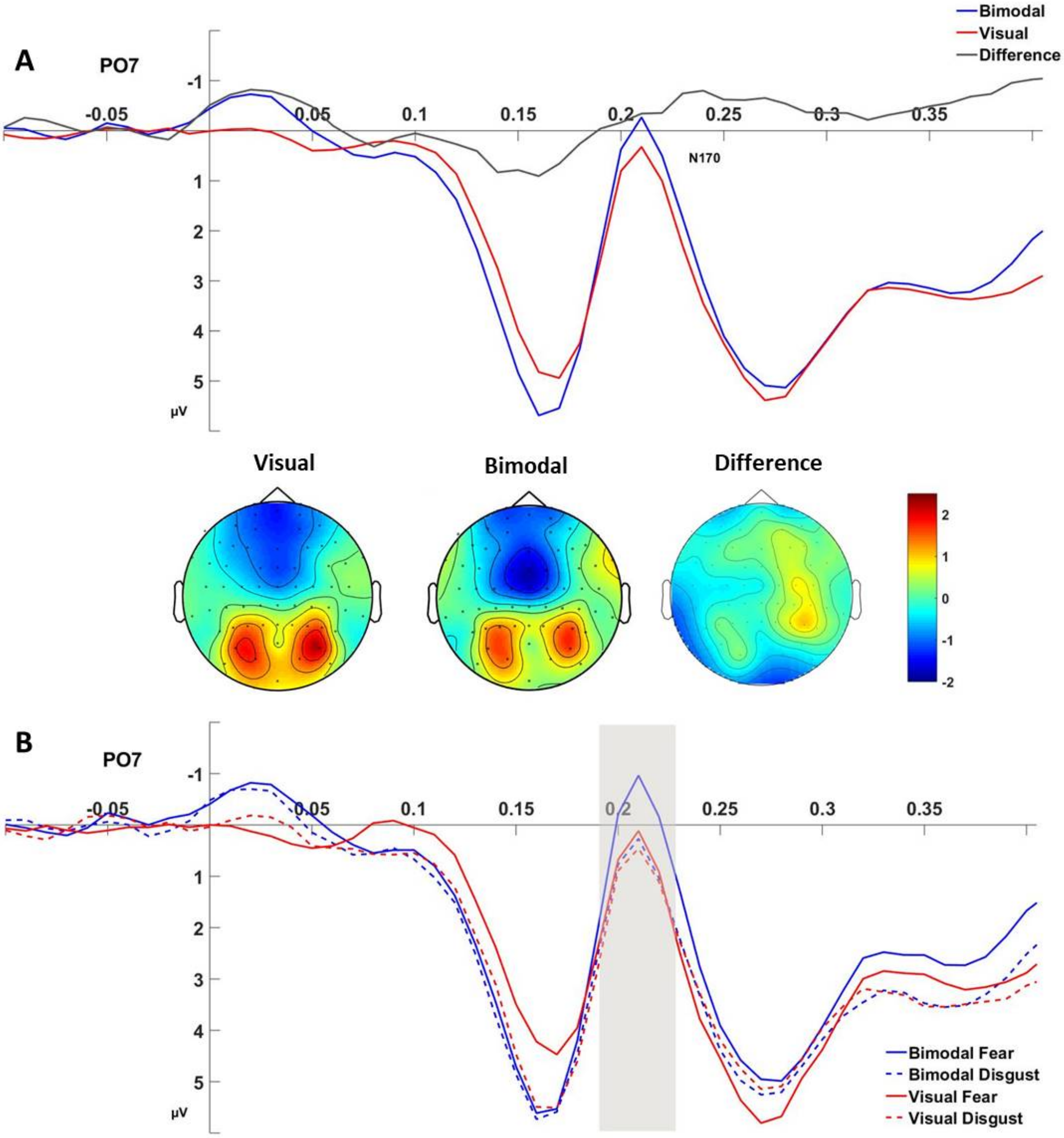
Table 1. Mean Accuracies and median reaction times; standard deviations are denoted parenthetically


Redundancy Gain (RG)
Disgust
Fear
**
RG (%)
0
1
2
3
4
5
6
7
8


Test of the Race Model Inequality for the Emotion of Fear
A
Probability
Time (ms)
0
0.2
0.4
0.6
0.8
1
400
600
800
1000
1200
1400
1600
Audio
Visual
Audiovisual
Bound
*


Test of the Race Model Inequality for the Emotion of Disgust
B
Probability
Time (ms)
0
0.2
0.4
0.6
0.8
1
400
600
800
1000
1200
1400
1600
Audio
Visual
Audiovisual
Bound


A
PO8
Bimodal
Visual
Difference
-0.05
0.05
0.1
0.15
0.2
0.25
0.3
0.35
2
4
6
8
µV
P100
Visual
Bimodal
Difference
6
4
2
0
-2
-4
B
PO8
-0.05
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0
2
4
6
8
µV
Bimodal Fear
Bimodal Disgust
Visual Fear
Visual Disgust


A
PO7
Bimodal
Visual
Difference
-1
-0.05
0.05
0.1
0.15
0.2
0.25
0.3
0.35
N170
1
2
3
4
5
μV
Visual
Bimodal
Difference
2
1
0
-1
-2
B
PO7
-1
-0.05
0.05
0.1
0.15
0.2
0.25
0.3
0.35
1
2
3
4
5
μV
Bimodal Fear
Bimodal Disgust
Visual Fear
Visual Disgust