Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

The Sound of Emotion: Pinpointing Emotional Voice Processing via Frequency Tagging EEG

Silke Vos [1,2,3,*], Olivier Collignon [4,5] and Bart Boets [1,2,3]

1 Center for Developmental Psychiatry, Department of Neurosciences, KU Leuven, Leuven, Belgium
2 Leuven Autism Research (LAuRes), KU Leuven, Leuven, Belgium
3 Leuven Brain Institute (LBI), KU Leuven, Leuven, Belgium
4 Institute of research in Psychology & Institute of Neuroscience, Université Catholique de Louvain, Belgium
5 Center for Mind/Brain Sciences, Università degli Studi di Trento, Italy
* Correspondence: silke.vos@kuleuven.be; Tel.: +32 16 37 76 83

Abstract: Successfully engaging in social communication requires efficient processing of subtle socio-communicative cues. Voices convey a wealth of social information, such as gender, identity and the emotional state of the speaker. We tested whether our brain can systematically and automatically differentiate and track a periodic stream of emotional utterances among a series of neutral vocal utterances. We recorded frequency-tagged EEG responses of 20 neurotypical male adults while presenting streams of neutral utterances at 4 Hz base rate, interleaved with emotional utterances every third stimulus, hence at 1.333 Hz oddball frequency. Four emotions (happy, sad, angry, and fear) were presented as different conditions in different streams. To control the impact of low-level acoustic cues, we maximized variability among the stimuli and included a control condition with scrambled utterances. This scrambling preserves low-level acoustic characteristics but ensures that the emotional character is no longer recognizable. Results revealed significant oddball EEG responses for all conditions, indicating that every emotion category can be discriminated from the neutral stimuli, and every emotional oddball response was significantly higher than the response for the scrambled utterances. These findings demonstrate that emotion discrimination is fast, automatic, and is not merely driven by low-level perceptual features.

Keywords: emotion discrimination; voice; frequency-tagging; EEG

1. Introduction

The We hear sounds every day, everywhere [1]. Being able to discriminate these sounds, contributes to a better understanding of the world around us. The human voice is by far the most socially relevant and familiar sound category for human beings [2]. Besides the specific linguistic content, the human voice offers a lot of socio-communicative information about the speaker. For instance, in a wink, it gives us an idea about the gender, approximate age, and the emotional state of the speaker [3]–[5]. Additionally, when listening carefully, one may even extract more subtle speaker information, such as the speaker's personality (e.g. extravert versus introvert) or the speaker's demographic origin [6], [7]. Efficient processing of all this supra-linguistic information is required to successfully engage in social communication.

1.1. Vocal emotion processing as a gateway to social communication

While zooming in on vocal emotional processing, speech prosody provides important cues about the emotional state of our conversational partners. Similar to the visual face processing domain [8], it has been postulated that a restricted group of so-called "basic" emotions (happy, surprise, angry, fear, sad and disgust) can be universally recognized across different cultures when vocally expressed, even without the presence of linguistic meaning [9]. Supporting this idea of basic emotions, a meta-analysis on the neural

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

correlates of vocal emotion processing revealed that these basic emotions are distinct and characterized by particular patterns of brain activity [10].

The recognition of vocally expressed emotions happens automatically [11]: we cannot inhibit recognizing an emotion in a voice, for instance when talking to someone who recently got fired or, in contrast, who just got a promotion, we can identify the emotional state of this person as sad or happy within a few hundred milliseconds, even without any linguistic context. Emotion recognition also happens extremely fast and based on limited auditory information. An ERP study demonstrated a neural signature of implicit emotion decoding within 200 msec after the onset of an emotional sentence [12], suggesting that emotional voices can be differentiated from neutral voices within a 200 msec timeframe. Explicit behavioral emotion recognition may take a bit longer, ranging from 266 to 1490 msec, depending on the paradigm and the particular emotion [13]–[15]. The fast decoding of emotion prosody is not only found in humans but is also visible in a variety of other animals, which indicates that recognizing emotions from voices is an important evolutionary skill to communicate with conspecific animals [16].

Gating paradigms have indicated that different vocal emotions are recognized within a different time frame (e.g., fear recognition happens faster than happiness), thereby suggesting that the fast recognition relies on emotion specific low-level auditory features [14]. Vocal emotion categories are indeed characterized by particular auditory features. For instance, sad speech is generally lower in pitch, and this is the case across different languages and cultures [17]. A classical, but almost intrinsically paradoxical challenge in vocal emotional neuroscience, is the demonstration that emotion discrimination is not purely driven by low-level acoustic cues. This echoes the broader attempts of demonstrating that (emotional) voice processing and the selective neural activity in the so-called temporal voice areas is not merely determined by particular spectro-temporal acoustic characteristics, often accomplished by a rigorous matching of vocal versus non-vocal low-level cues [18]. Besides determining the basic low-level acoustic cues that characterize and classify vocal emotions, there is evidence that threat related vocal signals mostly attract our attention, even when basic voice acoustics are comparable with non-threat related emotional vocalizations [19]. This indicates that low-level cues alone do not fully capture the experience of the vocally expressed emotions.

The temporal voice areas are located in the middle part of the auditory brain, these areas respond preferentially to voices compared to non-vocal environmental sounds [20]. This selective sensitivity for voices is particularly pronounced in the right hemisphere. Moreover, these temporal voice areas respond stronger to utterances spoken in an emotional rather than neutral tone [3], [18], [21], [22]. The classical rightward lateralization of emotional voice processing was challenged by Kotz et al. (2003)[23] who demonstrated that increasing task demands also resulted in an increasing left lateralization. In terms of lateralization of processing low-level acoustic features, pitch and slowly fluctuating signals have been shown to be processed preferentially at the right side whereas shorter and faster temporal information is typically processed in the left auditory cortex [24]–[26]. Given the critical importance of pitch to differentiate vocal emotion categories, this might explain that the majority of studies observe a right side lateralization for emotional voice processing [27].

As outlined above, it is evident that efficient emotion processing - including vocal emotion processing - is crucial for social functioning. Many psychiatric disorders are characterized by difficulties in social functioning, including emotion processing abilities, with key examples in autism spectrum disorder, schizophrenia, and anxiety and mood disorders (for reviews, see [28]–[31]). Thus, assessing individual differences and deficits in sensitivity for socio-communicative emotional cues is central in clinical practice, but objective and reliable diagnostic instruments are lacking, especially those tapping automatic emotional processing. A series of semi-standardized behavioral socio-cognitive tasks have been developed, assessing emotion recognition abilities for vocal, facial and bodily expressive stimuli (e.g., [32]–[34]). Yet, generally, these tasks do not differentiate sensitively

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

between clinical and neurotypical populations, often because they allow the mobilization of alternative compensatory perceptual and cognitive strategies [35], [36].

Brain imaging studies on the other hand show more robust group differences in vocal emotion processing. The auditory mismatch negativity (MMN), for example, is an event related potential component that reflects the response to an auditory deviant sound. This component is frequently used to investigate group differences in emotion processing. For instance, Schirmer et al. (2005) [37] demonstrated reduced MMN responses to emotionally deviant sounds in men as compared to women, and Lindström et al. (2018) [38] suggested that MMN components could indicate impaired emotional prosody perception in individuals with autism spectrum disorder. However, MMN studies lack high signal-to-noise ratio, thereby necessitating long recording sessions and reducing the utility to characterize performance at the individual subject level [39]. This has clear consequences for research with clinical populations or even infants.

Accordingly, there is a need for instruments that allow objective and robust assessment with high signal to noise responses of automatic and implicit emotion processing abilities, reliable at the individual subject-level, and preferably within a short timeframe. Here, we propose that frequency-tagging EEG in combination with periodic auditory (vocal) stimulation offers this approach, and we present evidence that the brain selectively responds to emotional vocal cues embedded within a stream of neutral vocal utterances.

1.2. Frequency tagging EEG to pinpoint differences in socio-communicative abilities

Recently, it was demonstrated that fast periodic visual stimulation combined with EEG can be used as an implicit neural index of the sensitivity for subtle socio-communicative facial cues, such as facial identity and facial expression [40], [41]. Application of this innovative approach in clinical populations (e.g. autism spectrum disorder and velocardiofacial syndrome), allowed pinpointing subtle but robust deficits in socio-communicative sensitivity that otherwise remained concealed via classical behavioral face processing tasks [42]–[45]. A more recent pioneering study applied this same frequency-tagging EEG approach with auditory stimulation, thereby demonstrating that voices can automatically be differentiated from both non-vocal environmental sounds and music instruments with highly similar low-level features [46]. Proceeding from this seminal study, here, we will extend this frequency-tagging EEG approach and apply it for the first time to investigate vocal emotion processing. In particular, we will characterize the neural signature of automatically detecting periodic emotional vocal utterances among a stream of neutral vocal utterances, and we will explore to what extent this neural discrimination ability is driven by the socio-emotional characteristics of the stimuli or by more basic low-level acoustic differences between the stimulus categories.

To investigate if our brain can systematically track a stream of emotional vocal utterances within a standard stream of neutral vocal utterances, we designed a Fast Periodic Auditory Stimulation (FPAS) paradigm and combined it with scalp EEG recordings. The basic principle of this frequency-tagging approach is that the periodicity of the electrophysiological response on the human scalp corresponds exactly with the periodicity (frequency) of the auditory stimulation. We used an oddball paradigm where standard sounds were presented at a base rate frequency of 4 Hz and oddball sounds were inserted periodically into the sequence every third sound. In particular, neutral voices were presented at 4 Hz base rate and emotional voices (angry, sad, happy, and fearful, in separate paradigms) were presented at 1.333 Hz oddball rate. Whenever a change (i.e., discrimination between the neutral and the emotional utterances) is perceived, in addition to the periodic response to the base rate, a periodic response corresponding to the presentation frequency of the emotional voices (i.e. 4/3=1.333 Hz) is also observed. The main advantages of using this FPAS approach are: (a) The response can be measured implicitly, i.e. without an explicit behavioral task; (b) The response can be identified objectively since it occurs at a predefined frequency; (c) It can be quantified directly by comparing the response at that frequency (signal) with responses at neighboring frequencies (noise); (d)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

The technique is extremely robust, since it is immune to artefacts and yields high signal-to noise ratio (SNR) responses in a short amount of time which makes it suitable for clinical populations (for a review, see [47]

2. Materials and Methods

2.1. Participants

We recruited 20 male participants for this study (mean age = 25.19 years, SD = 4.08, range = 19-34, all right-handed), sample size was based on previous fast periodic auditory stimulation studies (e.g. [46]). We only included male participants to avoid gender effects in the recognition of vocally expressed emotions [48]. All subjects reported intact hearing ability, which was confirmed by pure tone audiometry (average PTA hearing loss below 25 dB HL for every participant). Subjects were Dutch native speakers and received a monetary reward for participating. No one reported any history of psychiatric or neurological disorders. Before the start of the experiment all subjects signed an informed consent form approved by the Medical Ethics Committee UZ/KU Leuven (reference S62969).

2.2. Stimuli: Design of the Emotional Voices and Identity Database (EVID)

For the FPAS trials we created a new, large, and well-controlled voice segments database, incorporating all the stimulus features that are relevant for our research objectives. We aimed for short clips with a recognizable emotional value, while also demonstrating large variability across other sound features such as pitch, harmonic ratio, phonetic content, and speech rate. All voice segments were extracted from the Crowd Sourced Emotional Multimodal Actors Dataset [49], which encompasses audio and video recordings of 13 short sentences, spoken by 48 male and 43 female actors, according to 6 emotional states (neutral, happy, angry, sad, fear and disgust). We extracted 3960 short 250 msec utterances from these emotionally pronounced sentences (20 actors x 33 utterances x 6 emotions). Utterances were cut at the beginning of a randomly chosen phoneme. Thus, depending on speech rate, word length and phoneme position, these 250 ms utterances resulted in words (e.g. get) and non-words (e.g. ge). Each utterance started and ended with a linear fade in and fade out of 10 msec to avoid clipping of the sounds. All utterances were equalized in overall energy (RMS). We validated the stimuli behaviourally in a separate sample of 40 healthy young adults (age = 18 - 35 years old) to examine which stimuli are categorized best in terms of emotion, and we maintained a subset of 500 stimuli that are categorized most consistently. The subset contains a set of 10 speakers (5 female and 5 male speakers), each pronouncing 10 different phonetic utterances of 250 msec with a 10 msec fade in/out according to 5 emotion categories (neutral, happy, angry, sad and fear). Note that these utterances were not the same over all emotions as we selected the utterances with the highest recognition rate. See Figure 1 for the confusion matrix. We refer to this

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

newly designed and validated emotional stimulus set as the EVID (Emotional Voices and Identity Database), which is available upon request by the first or senior author.

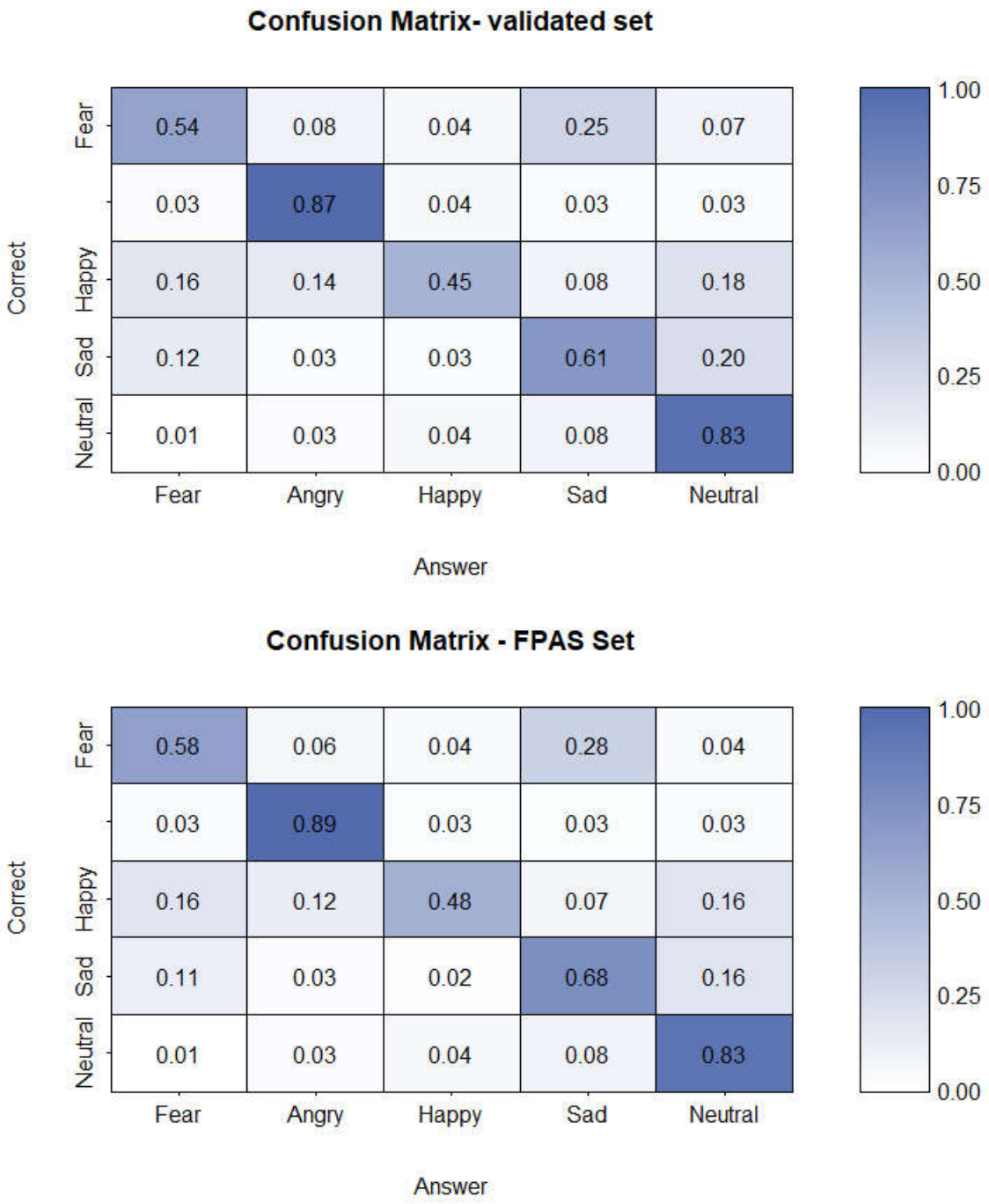


Figure 1. Top: Confusion matrix of the 500 emotional vocal stimuli. The rows indicate the presented emotional stimulus category (correct), the columns indicate the provided response(answer). The numbers indicate the proportional responses, averaged across the 40 participants. The diagonal shows the proportion of correct answers for each emotion. **Below:** Matrix of the stimuli used for the FPAS paradigms

2.3. Procedure and equipment

For each of the emotions (i.e. happy, angry, sad and fear), we created an oddball paradigm where the emotional utterance was periodically presented in a stream of neutral utterances. The 250 msec duration of the utterances naturally leads to the 4 Hz base

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

frequency of the sound presentation, and the emotional utterances were interleaved every third stimulus, leading to an oddball frequency of 1.333 Hz (i.e. NNENNENNE…, see Figure 2A). For every condition (i.e. emotion), we created six sequences, each uttered by a different speaker (including three sequences with a female speaker and three with a male speaker). For each condition we used the speakers with best recognition rate for the emotion in question, see Figure 1 for confusion matrix of the used utterances in this study. Note that in every sequence the same speaker was used for the neutral and for the emotional utterances. The sequences were 64 sec and had a linear fade in/out of 2 sec.

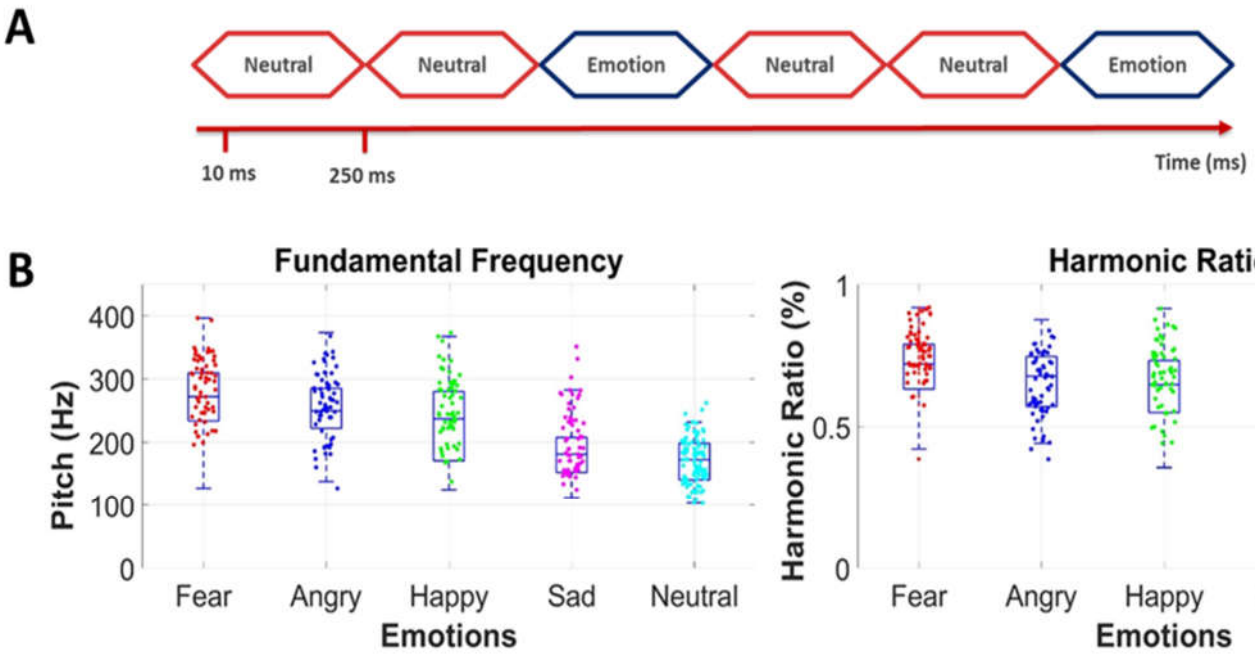
In addition to the four emotion category conditions (happy, angry, sad and fear), we created a scrambled control condition with similar low-level acoustic characteristics but without the emotional content. We scrambled the sounds of the four emotion categories and the neutral emotion category based on the method of Dormal et al. (2018)[50], which results in sounds with equal frequency content and spectral-temporal structure as the original sounds, but with a different harmonicity. This ensures that the emotion category is no longer recognizable in the scrambled sounds, while the low-level acoustic cues are largely preserved. We created six scrambled control sequences with three male and three female speakers, covering the four emotion categories.

Figure 2B provides an overview of the acoustic characteristics of the vocal stimuli included in the experiment, illustrating that all stimulus categories are highly heterogenous in terms of pitch and harmonic ratio, and that the differences within an emotion category are much larger than the average differences between the emotion categories. Pitch is defined as the fundamental frequency (f0) of the utterance, and refers to the perception of the sound as relatively high or low. Here, it has been calculated by means of the MATLAB function pitch(audioIn,fs). Harmonic ratio involves the ratio of the fundamental frequency's power to the total power in an audio fragment and refers to the degree of harmonicity contained in a signal. Here, it has been calculated by means of the MATLAB function harmonicRatio(audioIn,fs). Yet, as expected and in spite of our attempts to induce as much natural variability as possible, neutral and emotional utterances are not perfectly matched for low-level acoustic features. To further investigate the impact of these low-level features on the periodic neural oddball responses, we applied the following procedure: (a) we entered the wav-file of the entire acoustic 6x64 sec sequences in MATLAB and calculated the harmonic ratio using 100 msec rectangular windows with 50 msec overlap and pitch (Normalized Correlation Function for estimation of pitch) with a window length of 52 msec with 42 msec overlap, and (b) we transformed the continuous temporal signal from the temporal to the frequency domain by Fourier transformation to investigate the periodicity of these acoustic features (cf. [51]). As displayed in Figure 2C, one can see that in spite of the massive variation of the heterogenous stimuli, characteristic low-level features were still somehow periodically preserved in the stimulation sequences. Importantly, this low-level acoustic periodicity was not only preserved in all the vocal emotional sequences, but also in the control sequences with the scrambled stimuli.

We used an ActiveTwo Biosemi system with 64-Ag/AgCl electrodes and two additional electrodes as reference and ground electrodes (Common Mode Sense active electrode and Driven Right Leg passive electrode). Sound sequences were created and presented in a random order via a custom-built MATLAB script. Sounds were presented via a calibrated RME Fireface UC with Etymotic Research ER-1 insert earphones to make sure all sounds were presented at an equal intensity of 60 dB SPL. Participants listened to the sound sequences with eyes closed.

To ensure that participants stayed focused on the sound sequences, we included an orthogonal behavioural task which was non-periodic and unrelated to the emotional value of the stimuli. This task involved detecting short 500 msec silence periods in the sounds stream, occurring randomly four times in every sequence (not in the first and last 5 sec, and at least 5 sec apart from each other). Participants had to press a button whenever they detected this silence

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

Figure 2. **2A:** Schematic representation of the paradigm. Showing the base rate frequency of 4 Hz, with emotional stimuli being interleaved every third stimulus, hence at 1.333 Hz. **2B:** Low-level features of the vocal utterances. Low-level features are plotted for every single stimulus of every emotion condition. On the left the pitch (f0) is plotted and on the right harmonic ratio (hr in %). Note the large variability within every stimulus category. **2C:** Periodicity. The periodicity of the low-level features across the entire acoustic sequence. The first box represents a symbolic preview of the first 10s of a sequence and shows the presence of the emotional oddball stimuli in the time domain (sec) and in the frequency domain (Hz), with the clear 4 Hz peak indexing base rate and the 1.333 Hz peak indexing the emotional oddball stimuli (in black). Next, we plotted for all sequences of every emotion category as well as for the scrambled control condition the variability in harmonic ratio and pitch in the time domain and the frequency domain. This analysis reveals that -in spite of inducing a huge amount of acoustic variability- for all conditions (including the scrambled one) the low-level features are periodically preserved in the frequency domain, both at the base and at the oddball rate.

2.4 EEG analysis

2.4.1 Pre-processing

We used the Letswave6 Toolbox running on MATLAB 2019b (the MathWorks) for the EEG analyses. We started with pre-processing the data by applying a fourth-order Butterworth band-pass filter (0.1 -100 Hz) on the segmented data of 68 sec per segment, hence 2 sec before and after sequence onset. Afterwards we down sampled the data to 256 Hz and re-referenced the channels to a common average of all electrodes. Note that there was no need for eye-blink removal as the participants closed their eyes while listening to the sounds.

2.4.2. Frequency-domain analysis

Next, we segmented the pre-processed data again starting after the 2 sec fade-in and ending right before the fade-out, at 59.27 sec leading to an integer number of oddball (1.333 Hz) cycles (15,172 time bins). We averaged the six trials per condition (Fear, Angry, Happy, Sad, and Scrambled condition) for each participant separately in the time domain to reduce EEG activity not in phase with the auditory stimulation (e.g. noise). We transformed these averages into the frequency domain using a fast Fourier transformation (FFT) and the amplitude spectrum was computed with a high spectral resolution (0.0167 Hz).

The base rate of the voices (4 Hz) and the oddball presentation of emotions (1.333 Hz) and their integer multiples (harmonics) are present in the EEG signal. Responses at these frequencies and their harmonics reflect besides the response to the stimulus presentation also the overall noise. Therefore, we used two measures to describe the response in relation to the noise level: signal-to-noise ratio (SNR) and baseline-corrected amplitudes [40], [41]. SNR was computed at each frequency bin as the amplitude value at a given bin divided by the average amplitude of the 20 surrounding frequency bins (i.e. 12 bins on each side, 24 bins, but excluding the 2 directly adjacent bins and the local minimum and maximum). Baseline-corrected amplitude was computed in a similar way, but here we subtracted the average amplitude of the surrounding bins instead of dividing.

Z-score spectra on group-level data were computed to define the harmonics that were significantly above noise level per stimulation frequency ($Z > 1.65$ or $p > .05$). The z-scores were significant until the 2nd harmonic for the base rate (4 Hz) and until the 4th harmonic for the oddball frequency (1.333 Hz). Those harmonics of the oddball frequency that corresponded to the base frequency (3.999 and 7.998 Hz), were excluded thus the neural responses for oddball stimulation were quantified by summing up the baseline-corrected responses for 3 harmonics (1.333 Hz, 2.666 Hz and 5.333 Hz). We used all conditions to determine the number of significant harmonics.

As this is a new paradigm, we wanted to objectively select the regions of interest (ROIs) based on the data of all the subjects. We determined the ROIs separately for the base frequency (4 Hz) and the oddball frequency (1.333 Hz) as we expected different

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

patterns of activation for the different frequencies. We incorporated all conditions including the scrambled one for the ROI delineation. Hence, we calculated the baseline-subtracted amplitude across all subjects for each condition and each electrode, and we summed across the significant harmonics (4 Hz and 8 Hz for the base frequency, and 1.333 Hz, 2.666 Hz and 5. 333 Hz for the oddball frequency). All electrodes for which the baseline-subtracted amplitude of the response was significantly higher than the mean response (Bonferroni corrected) were retained and grouped in an ROI based on their location on the scalp.

2.4.3. Statistical analyses

Separately for the base rate and oddball responses, a series of linear repeated-measures mixed-models (LMM) were calculated. First, we zoomed in on the contrast between emotion conditions versus scrambled condition across all the electrodes included in the significant ROIs. Hence, condition (Fear, Angry, Happy, Sad and Scrambled) and ROI (all significant ROIs) were entered as fixed within-subject factor and participant as random factor. Next, we investigated the lateralization of the neural responses comparing all the emotion conditions. Thus, emotion condition (Fear, Angry, Happy and Sad), ROI (all ROIs including the corresponding contralateral homologue), and emotion x ROI interaction were entered as fixed within-subject factors and participant as random factor. Post-hoc t-tests corrected via Holms correction were calculated to assess the significance of particular contrasts.

3 Results

3.1 Orthogonal Task

We first checked if participants were able to perform the orthogonal task to make sure they were paying attention to the sound streams. The high accuracy of 96.3% (SD = 9.7%) indicated that the participants had no difficulty with the task. Important to note is that there was no significant difference between the conditions in the accuracy of the implicit task, we tested this with a LMM with condition as fixed factor and participant as random factor ($F(4,76) = 0.31$, $p = .86$).

3.2. Region of interests

The explorative investigation of regions of interests resulted in the delineation of 11 significant electrodes for the base frequency (FC6, FT8, Iz, Oz, P10, P7, P9, PO7, T8, TP7, TP8) and 10 significant electrodes for the oddball frequency (F1, Fz, Iz, O1, O2, Oz, P7, P9, PO7, PO8). We divided the significant electrodes in 4 ROIs based on the location of the electrodes, three for the base responses: ROI Left Parietal (LP: P7, P9, PO7, TP7), ROI Medial Occipital (MO: O1, O2, Iz, Oz) and ROI Right Temporal (RT: T8, FC6, FT8); and three for the oddball responses: ROI Left Parietal, ROI Medial Occipital, ROI Medial Frontal (MF: Fz, F1). In addition, to investigate possible lateralization effects and to include all significant electrodes, we also included the corresponding contralateral homologue brain areas in our analyses, thus ROI Right Parietal (RP: P8, P10, PO8, TP8) and ROI Left Temporal (LT: T7, FC5, FT7). Note that O1 and O2 were not significant for the base frequency and TP7 not for the oddball frequency and we still included these electrodes in the ROIs to delimit the number of ROIs needed for the analyses. See Figure 3 for ROI placement for base and oddball separately with contralateral ROIs included.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

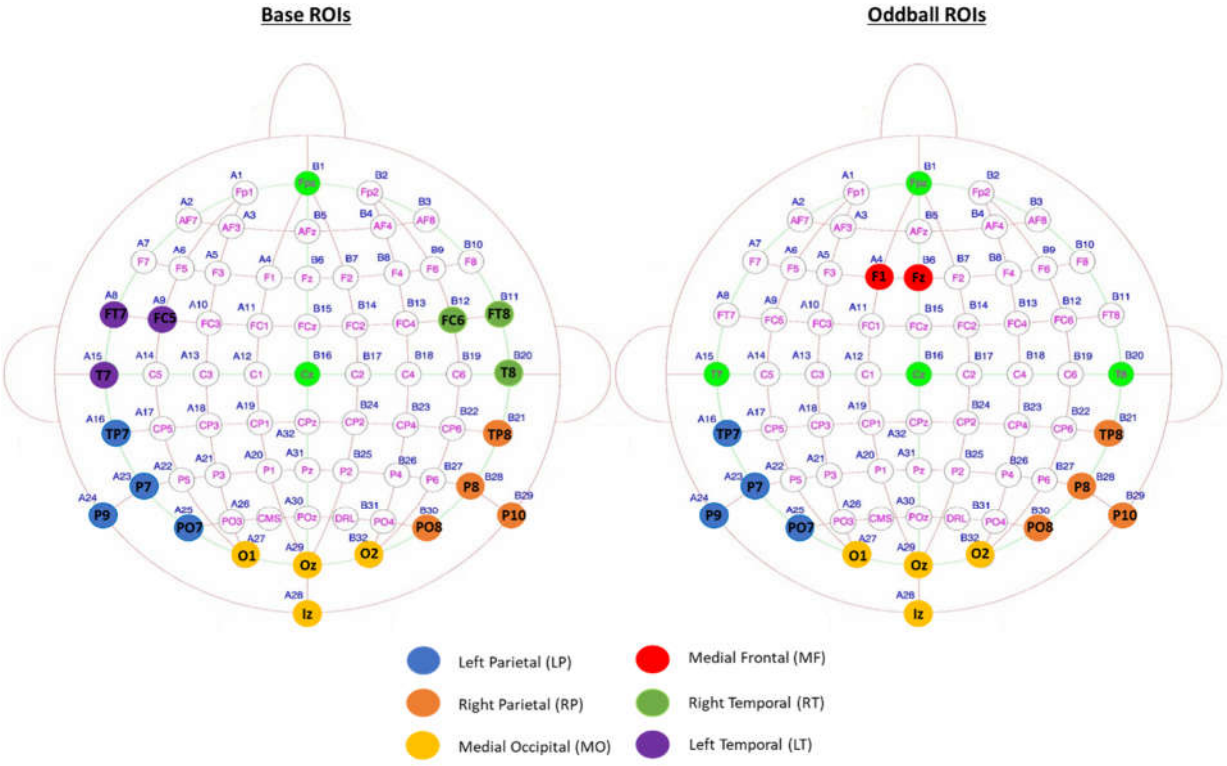


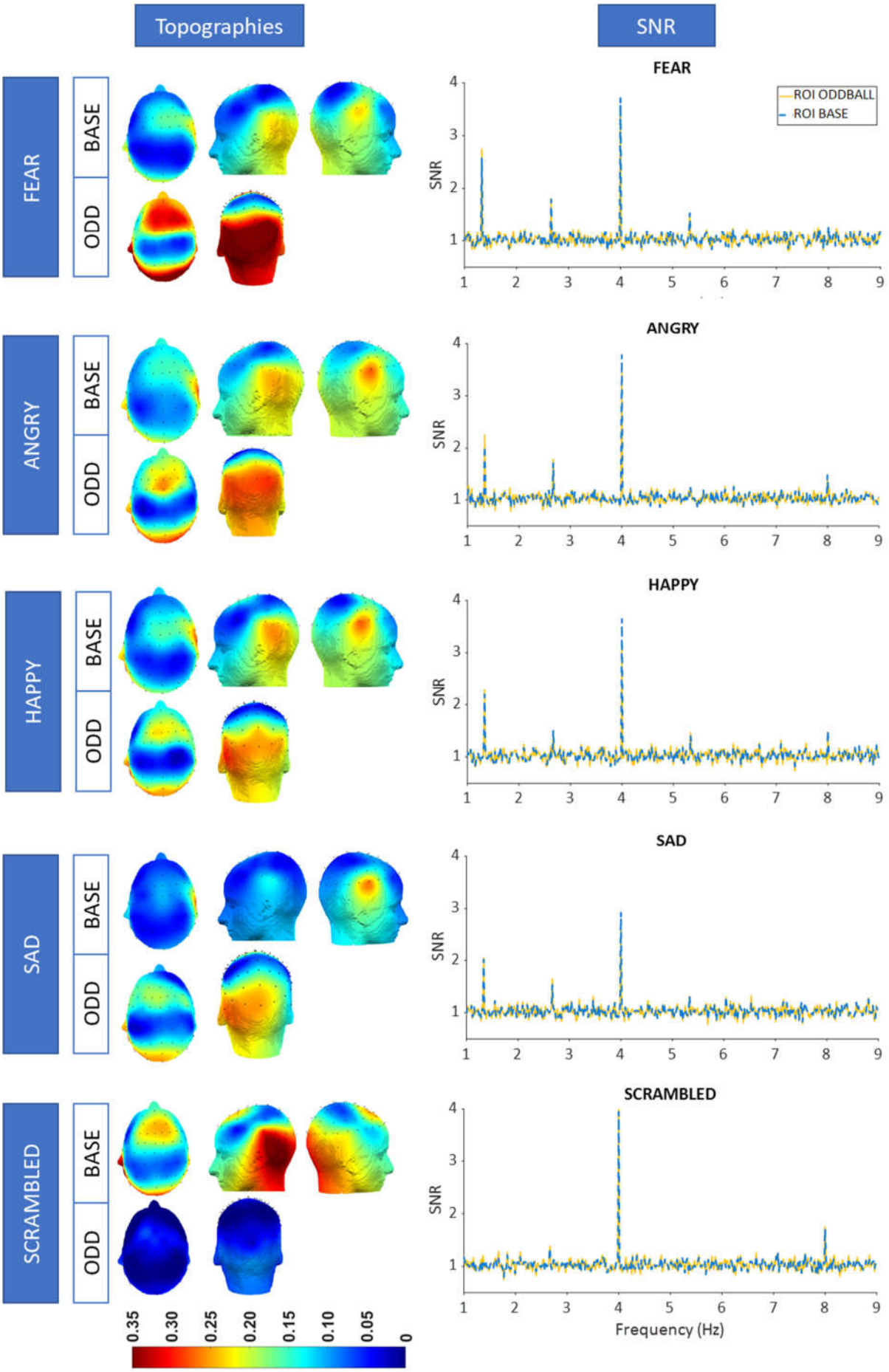
Figure 3. ROIs locations with electrode names for base and oddball separately.

3.3 SNR and topographies

To define the harmonics that were significantly above noise level, we computed Z-score spectra on group-level data for each condition for the base and oddball frequency. We averaged the FFT amplitude spectra across electrodes in the significant regions-of-interest (ROIs) based on the ROI determination and transformed these values into Z-scores. When we look at the signal-to-noise ratio (SNR), we see clear base responses related to the stimulus presentation at the first and second harmonic for every condition ($Z > 1.65$ or $p > .05$). At the oddball frequency we find clear oddball responses at the first, second and fourth harmonic for each of the emotion conditions ($Z > 1.65$ or $p > .05$), but only at the second harmonic for the scrambled control condition. See Figure 4 for the SNR spectra and the topographies of each condition.

Based on visual inspection, for base rate synchronization we see higher activation at the right side of the brain for the emotion conditions but not for the scrambled control condition. For oddball discrimination, the response was more lateralized to the left side of the brain for the emotion conditions, but no clear response was observable for the scrambled condition.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

Figure 4. Left: Topographies of all conditions showing summed baseline-subtracted averages of the significant harmonics, being 4 Hz and 8 Hz for the baseline and 1.333 Hz, 2.666 Hz, and 5.333 Hz for the oddball frequency. **Right:** SNR spectra of all conditions, with the blue spectrum representing the responses in the base frequency ROIs (LT, MO and RT) and the yellow spectrum displaying responses in the oddball frequency ROIs (LP, MO and MF).

3.4 Contrasting emotion-specific responses versus responses for the scrambled condition

First, we compared the emotion conditions with the scrambled condition to investigate to what extent low-level acoustic features versus high-level emotional characteristics explained the oddball effect. Figure 5 displays base rate and oddball rate neural responses for the five conditions averaged across the core ROIs yielding significant responses.

An LMM on the base rate responses with condition, ROI and their interaction as fixed factors and participants as random factor, revealed a significant main effect of condition ($F(4,266) = 16.97$, $p = 2.08e\text{-}12$, partial $\eta2 = 0.20$, 95% CI [0.13, 1]), with post-hoc paired t-testing demonstrating significantly lower responses for the Sad condition as compared to all other conditions ($t(59) > 5.20$, $p < .0001$ for all contrasts). We also found a significant main effect of ROI ($F(2,266) = 6.60$, $p = 0.002$, partial $\eta2 = 0.05$, 95% CI [0.01, 1]) and a significant condition by ROI interaction effect ($F(8,266) = 4.52$, $p = 3.57e\text{-}05$, partial $\eta2 = 0.12$, 95% CI [0.05, 1]). The interaction effect revealed that the amplitudes were equally distributed over the different ROIs for fear, angry and happy ($t(19) < 2.23$, $p > .110$ for all contrasts) but that for the sad condition and the scrambled condition the pattern differed. For the sad condition we found that ROI RT had higher responses in comparison with ROI MO ($t(19) > 3.12$, $p < .017$). The scrambled condition showed a different lateralization and we found lower responses at the right, at ROI RT in comparison with the other ROIs ($t(19) > 5.44$, $p < 7.6e\text{-}05$ for both contrasts).

A similar LMM on the oddball discrimination responses revealed an extreme main effect of condition ($F(4, 266) = 41.28$, $p < 2e\text{-}16$, partial $\eta2 = 0.38$, 95% CI [0.31, 1]), but no effect of ROI ($F(2,266) = 2.26$, $p = .106$, partial $\eta2 = 0.02$, 95% CI [0, 1]) nor condition by ROI interaction effect ($F(8,266) = 0.33$, $p = .95$, partial $\eta2 = 9.84e\text{-}03$, 95% CI [0, 1]). Here, post-hoc testing indicated that the amplitude for the scrambled condition was significantly lower than all emotional conditions ($t(59) > 6.88$, $p < .0001$ for all contrasts), and the amplitude for the fear condition was significantly higher than all other conditions ($t(59) > 3.86$, $p < .001$ for all contrasts).

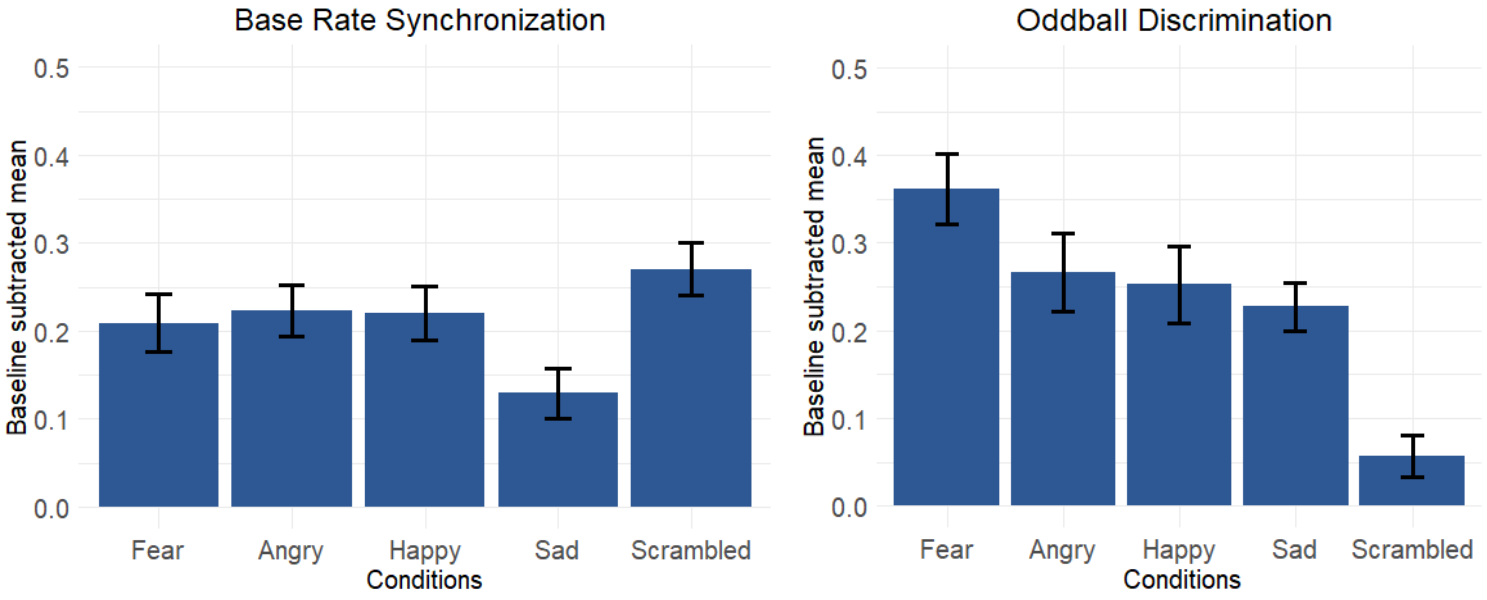


Figure 5. Comparison of the neural responses for the four conditions with emotional utterances and for the scrambled condition. Left: Base rate responses with standard error of the mean as error bar. Summed baseline-corrected amplitudes at significant base rate harmonics (4 Hz and 8 Hz) and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

averaged across LP, MO and RT ROIs reveal that sequences with sad utterances yield lower amplitudes. Right: Oddball responses with standard error of the mean as error bar. Summed baseline-corrected amplitudes at oddball frequencies (1.333 Hz, 2.666 Hz and 5.333 Hz) averaged across LP, MO and MF ROIs reveal that fear discrimination yields the highest amplitudes, and that automatic vocal emotion discrimination is significantly hampered by scrambling the stimuli.

3.5 Investigating lateralisation patterns of emotion-specific responses

To investigate lateralisation patterns of emotion-specific responses we additionally included the contralateral homologue ROIs in our analyses. Figure 6 displays ROI-specific base rate and oddball rate neural responses for the various emotional conditions. An LMM on the base frequency responses with emotion condition (Fear, Angry, Happy and Sad), ROI (all base rate ROIs including the corresponding contralateral homologue), and emotion condition x ROI interaction as fixed within-subject factors and participant as random factor revealed a main effect of condition ($F(3, 361) = 8.44$, $p = 1.96e\text{-}05$, partial $\eta 2 = 0.07$, 95% CI [0.03, 1]) and a main effect of ROI ($F(4, 361) = 12.45$, $p = 1.97e\text{-}09$, partial $\eta 2 = 0.12$, 95% CI [0.07, 1]), but no significant condition by ROI interaction ($F(12, 361) = 1.55$, $p = .10$, partial $\eta 2 = 0.05$, 95% CI [0, 1]). As expected, the main effect of condition was driven by the lower amplitudes for sad as compared to all other emotions ($t(99) > 3.52$, $p < .003$). Post-hoc testing for the main effect of ROI indicated that ROI LT had lower amplitudes than the other significant ROIs ($t(79) > 3.96$, $p < .001$ for all contrasts) and that ROI LP had higher amplitudes than ROI MO ($t(79) = 3.05$, $p = .019$).

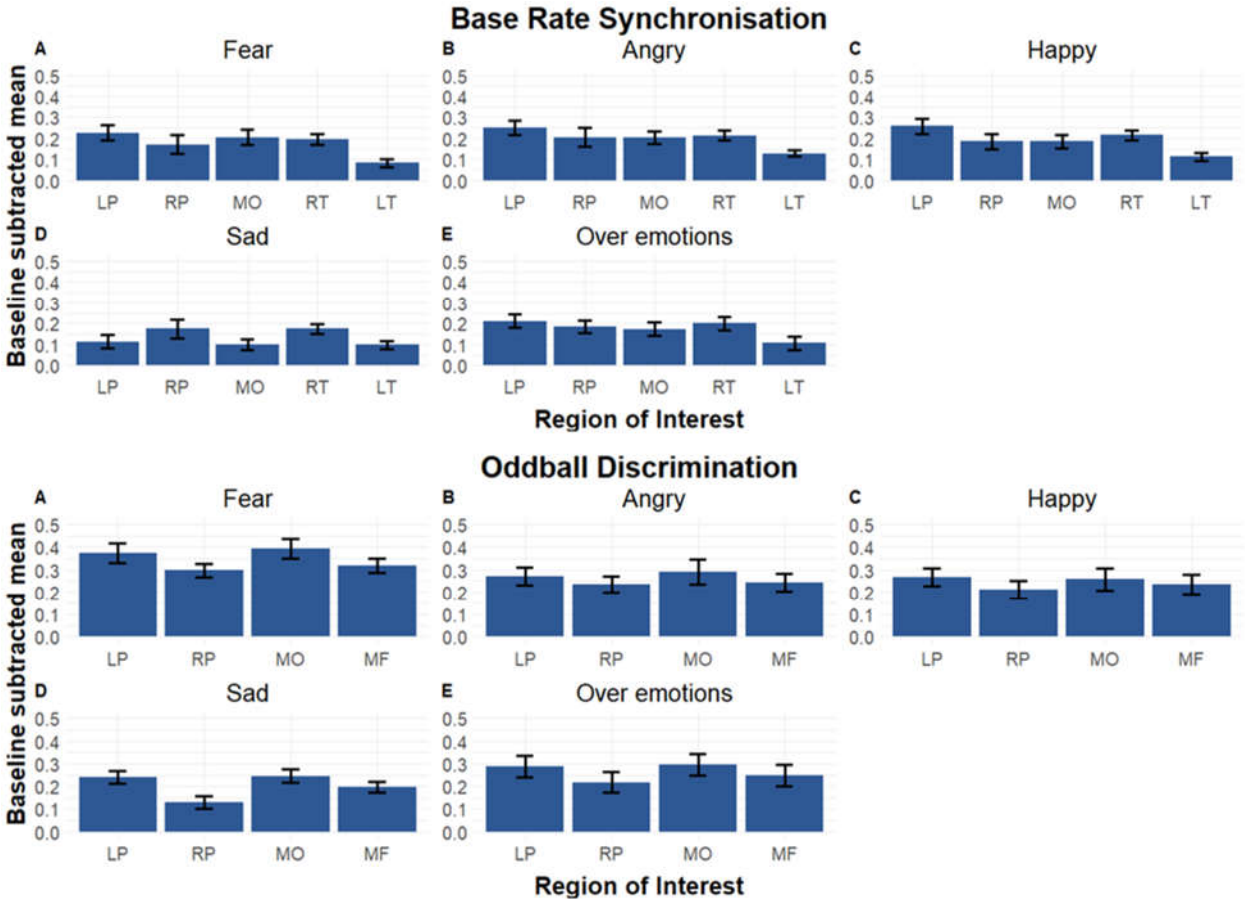


Figure 6. Comparison of the neural responses for the four conditions with emotional utterances as a function of spatial location with standard error of the mean as error bar (ROI). Top: Base rate synchronization. Summed baseline-corrected amplitudes at base rate harmonics reveal that the sad condition yields the lowest responses, and that ROI right temporal (RT) hosts the highest responses. Bottom: Oddball discrimination. Summed baseline-corrected amplitudes at oddball frequencies reveal that fear and angry yield the highest amplitudes and that ROIs left parietal (LP) and medial occipital (MO) show higher activation than the other ROIs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

A similar condition by ROI LMM on the oddball responses revealed a main effect of condition ($F(3, 285) = 7.20$, $p < .0001$, partial $\eta 2 = 0.17$, 95% CI [0.11, 1]),) and a main effect of ROI ($F(3,285) = 19.62$, $p = 1.36e\text{-}11$, partial $\eta 2 = 0.07$, 95% CI [0.02, 1]), but no ROI by condition interaction effect ($F(9, 285) = 0.412$, $p = .92$, partial $\eta 2 = 0.01$, 95% CI [0, 1]). The pairwise comparisons revealed that the amplitudes for fear were higher than any other condition ($t(79) > 4.26$, $p < .0001$ for all contrasts), and angry had higher amplitudes than sad ($t(79) = 2.88$, $p = .015$). Pairwise contrasts for the effect of ROI indicated that ROI LP and ROI MO had both higher amplitudes than ROI RP and ROI MF ($t(79) > 2.75$, $p > .022$).

4. Discussion

We found clear base and oddball responses for every emotion category, indicating that the brain is able to synchronize with the presentation rate of vocal stimuli and to systematically detect and discriminate subtle vocal emotional utterances from neutral utterances. To ensure that this effect was not purely driven by systematic low-level acoustic differences, we preselected highly heterogeneous vocal stimuli with a high variability in pitch and harmonic ratio, which are important low-level features for emotional voice processing. Yet, in spite of this huge random variability, pitch and harmonic ratio still varied in a periodic way in the neutral-emotional sound streams, as indicated in Figure 2C. Against this background, it may have been not too surprising to also observe this same periodicity, including the oddball responses, in the EEG spectrum. To further control the relative importance of low-level acoustic features for automatic emotion processing, we also included a scrambled version of the vocal sound streams. This scrambling procedure preserved the low-level spectro-temporal acoustic structure of the sound but ensured that stimuli were no longer recognizable as voices, let alone that the emotional content would have been identified. Preservation of some of the acoustic structure and its periodicity along the vocal sound stream is again demonstrated in Figure 2C. Yet, crucially and importantly, in contrast with the emotion conditions, for the scrambled control condition we only found an EEG base rate response and no selective oddball discrimination responses for the first harmonic (cf. Figure 4). Accordingly, together, these results clearly indicate that periodicity of low-level acoustic features by itself does not suffice to induce robust oddball EEG responses, but meaningful high-level emotional categories are needed.

For the base rate frequency, we found a main effect of Condition with reduced base responses for the sad condition. It appears that sad utterances are confused and misinterpreted often with neutral utterances (16%) and it might be that habituation occurs more pronounced in the sad condition in comparison with the other conditions due the similarity of the neutral and sad utterances leading to lower responses in the EEG data (Polich, 1989). However, with regard to the confusion matrix of the used stimuli (Figure 1), happy utterances were confused with neutral utterances as much as sad utterances (16%) but did not show reduced base responses in comparison with the other conditions. Although, happy also had the lowest accuracy of all emotions, which is also supported by other vocal emotion studies [15].

While comparing the oddball response between the different emotions, we observed the highest response for the detection and discrimination of fearful and angry voices. This echoes the general observation that threat-related emotions, such as fear and anger, may be important from an evolutionary perspective to survive unknown situations, and may therefore most easily be detected and attract our attention [19]. This finding neatly aligns with a similar observation showing the highest frequency-tagged EEG discrimination responses for visually presented emotional expressions of fearful and angry faces in a continuous stream of neutral faces [44]. In this study, in spite of the difference in modality, a similar pattern of emotion discrimination responses was observed, with fearful and angry expressions eliciting the strongest response, happy expressions an intermediate response, and sad facial expressions the lowest response.

We did not observe a condition (fear, angry, happy and sad) by ROI interaction effect, nor for the base rate responses nor for the oddball responses, suggesting the presence of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

a similar neural activation pattern for all emotions used in the experiment. However, note that EEG, as compared to MRI, is not the most sensitive method to detect small spatial differences in activation patterns. For the base frequency, which indices the periodic presentation of voices, we found a main effect of ROI, revealing a right-side lateralisation of activity at the temporal cortex. Importantly, the scrambled condition, which involved the presentation of non-recognizable artificial sounds, did not display this right lateralisation at base frequency. This echoes the general literature that voice processing, and certainly emotional voice processing, is right lateralised [20], [21]. On the other hand, the absence of a right lateralization for the 4 Hz base rate of the scrambled condition does not corroborate the asymmetric temporal sampling in time theory of Poeppel and colleagues [24], which postulates that slower oscillations (~200 ms) are preferentially processed by the right auditory cortex.

In response to the oddball frequency, we found a different lateralisation pattern, revealing higher amplitudes at the left side of the brain, in particular in the left parietal cortex. The left lateralisation of these emotional voice discrimination responses may be related to the higher difficulty level of the task, as differentiating emotions is harder than simple voice processing, and studies have demonstrated more left lateralized brain activation for tasks that are more difficult [23], [52]. For both the base and the oddball responses we also found activation in posterior occipital regions and even some medial frontal activation for the oddball response. This pattern may originate from activity in auditory cortex and posterior STS projecting towards the posterior and anterior regions of the scalp because of the particular folding of the gyri. However, to pinpoint the exact spatial location and source of these responses, methods with a higher spatial resolution would be required.

5. Conclusions

Overall, we demonstrated that we can track the discrimination and categorization of complex vocal emotional utterances with frequency tagging EEG and that these emotion-selective responses are at least partially independent from low-level acoustic features (see also [46]). This fast, straightforward and double-objective approach offers a unique and powerful tool to quantify the implicit sensitivity for subtle vocal emotional cues at the individual subject-level, without any overt behavioral processing. This opens up the way to apply this paradigm to investigate emotion processing abilities in young children and infants that are unable to understand instructions or provide explicit responses, and to investigate particular clinical populations that are characterized by atypical emotion processing abilities, such as autism spectrum disorder, schizophrenia, frontotemporal dementia, anxiety disorder, etc. Also at a more fundamental level, it paves the way for implementing other complex sound categorization frequency-tagging paradigms (pinpointing for instance vocal identity discrimination), thereby contributing to an advanced understanding of human auditory categorization in general.

Author Contributions: Conceptualization, S.V., O.C and B.B.; methodology, S.V, O.C. and B.B..; software, S.V.; validation, S.V. and B.B.; formal analysis, S.V.; investigation, S.V.; resources, S.V.; data curation, S.V.; writing—original draft preparation, S.V.; writing—review and editing, O.C and B.B..; visualization, S.V.; supervision, O.C and B.B..; project administration, S.V.; funding acquisition, B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Research Foundation Flanders: Excellence of Science EOS, grant number G0E8718N (HUMVISCAT).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Medical Ethics Committee UZ/KU Leuven (reference S62969 26/11/2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

Data Availability Statement: The newly designed Emotional Voices and Identity Database (EVID), created for this study and described in the manuscript, will be made publicly available, as well as the analysis scripts. The EEG data of the study will be available anonymized upon request.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] P. Belin, "Voice processing in human and non-human primates," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 361, no. 1476. Royal Society, pp. 2091–2107, Dec. 29, 2006, doi: 10.1098/rstb.2006.1933.

[2] P. Belin, S. Fecteau, C. Bédard, C. Bé, and C. Bédard, "Thinking the voice: Neural correlates of voice perception," *Trends Cogn. Sci.*, vol. 8, no. 3, pp. 129–135, Mar. 2004, doi: 10.1016/j.tics.2004.01.008.

[3] D. Grandjean *et al.*, "The voices of wrath: brain responses to angry prosody in meaningless speech," *Nat. Neurosci. 2005 82*, vol. 8, no. 2, pp. 145–146, Jan. 2005, doi: 10.1038/nn1392.

[4] D. N. Honorof and D. H. Whalen, "Identification of speaker sex from one vowel across a range of fundamental frequencies," *J. Acoust. Soc. Am.*, vol. 128, no. 5, pp. 3095–3104, Nov. 2010, doi: 10.1121/1.3488347.

[5] R. M. Krauss, R. Freyberg, and E. Morsella, "Inferring speakers' physical attributes from their voices," *J. Exp. Soc. Psychol.*, vol. 38, no. 6, pp. 618–625, Nov. 2002, doi: 10.1016/S0022-1031(02)00510-3.

[6] G. Mohammadi, A. Vinciarelli, and M. Mortillaro, "The voice of personality: Mapping nonverbal vocal behavior into trait attributions," *SSPW'10 - Proc. 2010 ACM Soc. Signal Process. Work. Co-located with ACM Multimed. 2010*, pp. 17–20, 2010, doi: 10.1145/1878116.1878123.

[7] T. Rakić, M. C. Steffens, and A. Mummendey, "When it matters how you pronounce it: The influence of regional accents on job interview outcome," *Br. J. Psychol.*, vol. 102, no. 4, pp. 868–883, Nov. 2011, doi: 10.1111/J.2044-8295.2011.02051.X.

[8] M. Batty and M. J. Taylor, "Early processing of the six basic facial emotional expressions," *Cogn. Brain Res.*, vol. 17, no. 3, pp. 613–620, Oct. 2003, doi: 10.1016/S0926-6410(03)00174-5.

[9] D. A. Sauter, F. Eisner, P. Ekman, and S. K. Scott, "Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 6, pp. 2408–12, Feb. 2010, doi: 10.1073/pnas.0908239106.

[10] K. Vytal and S. Hamann, "Neuroimaging Support for Discrete Neural Correlates of Basic Emotions: A Voxel-based Meta-analysis," *J. Cogn. Neurosci.*, vol. 22, no. 12, pp. 2864–2885, Dec. 2010, doi: 10.1162/JOCN.2009.21366.

[11] M. D. Pell and V. Skorup, "Implicit processing of emotional prosody in a foreign versus native language," *Speech Commun.*, vol. 50, no. 6, pp. 519–530, Jun. 2008, doi: 10.1016/j.specom.2008.03.006.

[12] S. Paulmann and S. A. Kotz, "An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context," *Brain Lang.*, vol. 105, no. 1, pp. 59–69, Apr. 2008, doi: 10.1016/J.BANDL.2007.11.005.

[13] P. Castiajo and A. P. Pinheiro, "Decoding emotions from nonverbal vocalizations: How much voice signal is enough?," *Motiv. Emot.*, vol. 43, no. 5, pp. 803–813, Oct. 2019, doi: 10.1007/s11031-019-09783-9.

[14] F. Falagiarda and O. Collignon, "Time-resolved discrimination of audio-visual emotion expressions," *Cortex*, vol. 119, pp. 184–194, Oct. 2019, doi: 10.1016/j.cortex.2019.04.017.

[15] M. D. Pell and S. A. Kotz, "On the Time Course of Vocal Emotion Recognition," *PLoS One*, vol. 6, no. 11, p. e27256, Nov. 2011, doi: 10.1371/journal.pone.0027256.

[16] E. F. Briefer, "Vocal contagion of emotions in non-human animals," *Proc. R. Soc. B Biol. Sci.*, vol. 285, no. 1873, Feb. 2018, doi: 10.1098/RSPB.2017.2783.

[17] M. D. Pell, S. Paulmann, C. Dara, A. Alasseri, and S. A. Kotz, "Factors in the recognition of vocally expressed emotions: A comparison of four languages," *J. Phon.*, vol. 37, no. 4, pp. 417–435, Oct. 2009, doi: 10.1016/j.wocn.2009.07.005.

[18] Y. Cheng, S. Y. Lee, H. Y. Chen, P. Y. Wang, and J. Decety, "Voice and Emotion Processing in the Human Neonatal Brain," *J. Cogn. Neurosci.*, vol. 24, no. 6, pp. 1411–1419, Jun. 2012, doi: 10.1162/JOCN_A_00214.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

[19] N. Burra, D. Kerzel, D. Munoz Tord, D. Grandjean, and L. Ceravolo, "Early spatial attention deployment toward and away from aggressive voices," *Soc. Cogn. Affect. Neurosci.*, vol. 14, no. 1, pp. 73–80, Jan. 2019, doi: 10.1093/scan/nsy100.

[20] P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, and B. G. Pike, "Voice-selective areas in human auditory cortex," *Nature*, vol. 403, no. 6767, pp. 309–312, Jan. 2000, doi: 10.1038/35002078.

[21] V. Beaucousin, A. Lacheret, M.-R. Turbelin, M. Morel, B. Mazoyer, and N. Tzourio-Mazoyer, "FMRI Study of Emotional Speech Comprehension," *Cereb. Cortex*, vol. 17, no. 2, pp. 339–352, Feb. 2007, doi: 10.1093/CERCOR/BHJ151.

[22] T. Ethofer, D. Van De Ville, K. Scherer, and P. Vuilleumier, "Decoding of emotional information in voice-sensitive cortices.," *Curr. Biol.*, vol. 19, no. 12, pp. 1028–33, Jun. 2009, doi: 10.1016/j.cub.2009.04.054.

[23] S. A. Kotz, M. Meyer, K. Alter, M. Besson, D. Y. Von Cramon, and A. D. Friederici, "On the lateralization of emotional prosody: an event-related functional MR investigation," *Brain Lang.*, vol. 86, no. 3, pp. 366–376, 2003, doi: 10.1016/S0093-934X(02)00532-1.

[24] H. Luo and D. Poeppel, "Cortical Oscillations in Auditory Perception and Speech: Evidence for Two Temporal Windows in Human Auditory Cortex," *Front. Psychol.*, vol. 3, no. MAY, p. 170, 2012, doi: 10.3389/fpsyg.2012.00170.

[25] D. Poeppel, "The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time,'" *Speech Commun.*, vol. 41, no. 1, pp. 245–255, Aug. 2003, doi: 10.1016/S0167-6393(02)00107-3.

[26] R. J. Zatorre, P. Belin, and V. B. Penhune, "Structure and function of auditory cortex: music and speech," *Trends Cogn. Sci.*, vol. 6, no. 1, pp. 37–46, Jan. 2002, doi: 10.1016/S1364-6613(00)01816-7.

[27] R. G. Kamiloğlu, A. H. Fischer, and D. A. Sauter, "Good vibrations: A review of vocal expressions of positive emotions," *Psychon. Bull. Rev.*, vol. 27, no. 2, pp. 237–265, Apr. 2020, doi: 10.3758/s13423-019-01701-x.

[28] J. Edwards, H. J. Jackson, and P. E. Pattison, "Emotion recognition via facial expression and affective prosody in schizophrenia," *Clin. Psychol. Rev.*, vol. 22, no. 6, pp. 789–832, Jul. 2002, doi: 10.1016/S0272-7358(02)00130-7.

[29] M. Jáni and T. Kašpárek, "Emotion recognition and theory of mind in schizophrenia: A meta-analysis of neuroimaging studies," *World J. Biol. Psychiatry*, vol. 19, no. sup3, pp. S86–S96, Nov. 2018, doi: 10.1080/15622975.2017.1324176.

[30] J. M. Leppänen, "Emotional information processing in mood disorders: A review of behavioral and neuroimaging findings," *Curr. Opin. Psychiatry*, vol. 19, no. 1, pp. 34–39, 2006, doi: 10.1097/01.YCO.0000191500.46411.00.

[31] F. Y. N. Leung *et al.*, "Emotion recognition across visual and auditory modalities in autism spectrum disorder: A systematic review and meta-analysis," *Dev. Rev.*, vol. 63, p. 101000, Mar. 2022, doi: 10.1016/j.dr.2021.101000.

[32] T. Bänziger, D. Grandjean, and K. R. Scherer, "Emotion Recognition From Expressions in Face, Voice, and Body: The Multimodal Emotion Recognition Test (MERT)," *Emotion*, vol. 9, no. 5, pp. 691–704, Oct. 2009, doi: 10.1037/a0017088.

[33] T. Bänziger, M. Mortillaro, and K. R. Scherer, "Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception," *Emotion*, vol. 12, no. 5, pp. 1161–1179, Oct. 2012, doi: 10.1037/a0025827.

[34] K. Schlegel and K. R. Scherer, "Introducing a short version of the Geneva Emotion Recognition Test (GERT-S): Psychometric properties and construct validation," *Behav. Res. Methods*, vol. 48, no. 4, pp. 1383–1392, Dec. 2016, doi: 10.3758/S13428-015-0646-4/FIGURES/2.

[35] M. B. Harms, A. Martin, and G. L. Wallace, "Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies," *Neuropsychol. Rev.*, vol. 20, no. 3, pp. 290–322, Sep. 2010, doi: 10.1007/s11065-010-9138-6.

[36] M. E. Stewart, C. McAdam, M. Ota, S. Peppé, and J. Cleland, "Emotional recognition in autism spectrum conditions from voices and faces," *Autism*, vol. 17, no. 1, pp. 6–14, Oct. 2013, doi: 10.1177/1362361311424572.

[37] A. Schirmer, T. Striano, and A. D. Friederici, "Sex differences in the preattentive processing of vocal emotional expressions," *Neuroreport*, vol. 16, no. 6, pp. 635–639, Apr. 2005, doi: 10.1097/00001756-200504250-00024.

[38] R. Lindström *et al.*, "Atypical perceptual and neural processing of emotional prosodic changes in children with autism spectrum disorders," *Clin. Neurophysiol.*, vol. 129, no. 11, pp. 2411–2420, Nov. 2018, doi: 10.1016/j.clinph.2018.08.018.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2022 doi:10.20944/preprints202212.0387.v1

[39] T. J. Mcgee, C. King, K. Tremblay, T. G. Nicol, J. Cunningham, and N. Kraus, "Long-term habituation of the speech-elicited mismatch negativity," 2001, doi: 10.1111/1469-8986.3840653.

[40] M. Dzhelyova, C. Jacques, and B. Rossion, "At a Single Glance: Fast Periodic Visual Stimulation Uncovers the Spatio-Temporal Dynamics of Brief Facial Expression Changes in the Human Brain," *Cereb. Cortex*, vol. 27, no. 8, pp. 4106–4123, Aug. 2016, doi: 10.1093/cercor/bhw223.

[41] J. Liu-Shuang, A. M. Norcia, and B. Rossion, "An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation," *Neuropsychologia*, vol. 52, no. 1, pp. 57–72, Jan. 2014, doi: 10.1016/J.NEUROPSYCHOLOGIA.2013.10.022.

[42] A. Leleu *et al.*, "An implicit and reliable neural measure quantifying impaired visual coding of facial expression: evidence from the 22q11.2 deletion syndrome," *Transl. Psychiatry*, vol. 9, no. 1, p. 67, Dec. 2019, doi: 10.1038/s41398-019-0411-z.

[43] S. Van der Donck *et al.*, "Fast Periodic Visual Stimulation EEG Reveals Reduced Neural Sensitivity to Fearful Faces in Children with Autism," *J. Autism Dev. Disord.*, pp. 1–16, Aug. 2019, doi: 10.1007/s10803-019-04172-0.

[44] S. Van der Donck *et al.*, "Rapid neural categorization of angry and fearful faces is specifically impaired in boys with autism spectrum disorder," *J. Child Psychol. Psychiatry Allied Discip.*, p. jcpp.13201, Jan. 2020, doi: 10.1111/jcpp.13201.

[45] S. Vettori *et al.*, "Reduced neural sensitivity to rapid individual face discrimination in autism spectrum disorder," *NeuroImage Clin.*, vol. 21, p. 101613, Nov. 2019, doi: 10.1016/j.nicl.2018.101613.

[46] F. M. Barbero, R. P. Calce, S. Talwar, B. Rossion, and O. Collignon, "Fast Periodic Auditory Stimulation Reveals a Robust Categorical Response to Voices in the Human Brain," *eneuro*, vol. 8, no. 3, p. ENEURO.0471-20.2021, May 2021, doi: 10.1523/ENEURO.0471-20.2021.

[47] A. M. Norcia, L. G. Appelbaum, J. M. Ales, B. R. Cottereau, and B. Rossion, "The steady-state visual evoked potential in vision research: A review," *J. Vis.*, vol. 15, no. 6, p. 4, May 2015, doi: 10.1167/15.6.4.

[48] A. Lausen and A. Schacht, "Gender differences in the recognition of vocal emotions," *Front. Psychol.*, vol. 9, no. JUN, p. 882, Jun. 2018, doi: 10.3389/FPSYG.2018.00882/BIBTEX.

[49] H. Cao, D. G. Cooper, M. K. Keutmann, R. C. Gur, A. Nenkova, and R. Verma, "CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset.," *IEEE Trans. Affect. Comput.*, vol. 5, no. 4, pp. 377–390, 2014, doi: 10.1109/TAFFC.2014.2336244.

[50] G. Dormal, M. Pelland, M. Rezk, E. Yakobov, F. Lepore, and O. Collignon, "Functional Preference for Object Sounds and Voices in the Brain of Early Blind and Sighted Individuals," *J. Cogn. Neurosci.*, vol. 30, no. 1, pp. 86–106, Jan. 2018, doi: 10.1162/jocn_a_01186.

[51] A. Van Rinsveld, M. Guillaume, P. J. Kohler, C. Schiltz, W. Gevers, and A. Content, "The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged EEG," *Proc. Natl. Acad. Sci.*, vol. 117, no. 11, pp. 5726–5732, Mar. 2020, doi: 10.1073/pnas.1917849117.

[52] A. Schirmer and S. A. Kotz, "Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing," *Trends Cogn. Sci.*, vol. 10, no. 1, pp. 24–30, Jan. 2006, doi: 10.1016/j.tics.2005.11.009.