Timing & Time Perception 4 (2016) 79–98

Emotional Voices Distort Time: Behavioral and Neural Correlates

Annett Schirmer[1,2,3,*], Tabitha Ng[1,2], Nicolas Escoffier[1,2] and Trevor B. Penney[1,2]

[1] Department of Psychology, National University of Singapore, Singapore
[2] Neurobiology and Ageing Programme, National University of Singapore, Singapore
[3] Duke/NUS Graduate Medical School, Singapore

Received 11 August 2015; accepted 1 December 2015

Abstract

The present study explored the effect of vocally expressed emotions on duration perception. Recordings of the syllable 'ah' spoken in a disgusted (negative), surprised (positive), and neutral voice were subjected to a compression/stretching algorithm producing seven durations ranging from 300 to 1200 ms. The resulting stimuli served in a duration bisection procedure in which participants indicated whether a stimulus was more similar in duration to a previously studied 300 ms (short) or 1200 ms (long) 440 Hz tone. Behavioural results indicate that disgusted expressions were perceived as shorter than surprised expressions in both men and women and this effect was related to perceived valence. Additionally, both emotional expressions were perceived as shorter than neutral expressions in women only and this effect was related to perceived arousal. Event-related potentials showed an influence of emotion and rate of acoustic change (fast for compressed/short and slow for stretched/long stimuli) on stimulus encoding in women only. Based on these findings, we suggest that emotions interfere with temporal processes and facilitate the influence of contextual information (e.g., rate of acoustic change, attention) on duration judgements. Because women are more sensitive than men to unattended vocal emotions, their temporal judgements are more strongly distorted.

Keywords

Interval timing, time perception, ERPs, gender differences, affective prosody

1. Introduction

Non-verbal signals are powerful elicitors of emotion. For example, vocal expressions, depending on their acoustic properties and context, can trigger central

* To whom correspondence should be addressed. E-mail: schirmer@nus.edu.sg

 DOI: 10.1163/22134468-00002058

and peripheral nervous system changes in the absence of attentive processing (e.g., Brosch et al., 2008; Grandjean et al., 2005; Schirmer et al., 2005). Furthermore, they can serve as endogenous cues for attention and influence ongoing mental processes such as language and memory (Min & Schirmer, 2011; Schirmer, 2010). Here, we sought to extend current knowledge by examining the role of vocal expressions in time perception, thereby complementing earlier work on speech and timing as reviewed by Schirmer (2004).

Philosophers and psychologists have long been interested in the relation between emotions and time (e.g., James, 1890; Vierordt, 1868). Indeed, research revealed that time is perceived differently in an emotional as compared to neutral context (Droit-Volet & Meck, 2007; Schirmer, 2011). For example, compared to neutral facial expressions, the duration of emotional facial expressions is typically overestimated indicating a speeding up of an internal clock (e.g., Droit-Volet et al., 2010b; Effron et al., 2006; Grommet et al., 2011; Tipples et al., 2013; this issue Droit-Volet et al., 2016; Eberhardt et al., 2016; but see Smith et al., 2011).

To date, few studies of emotions and time have used auditory materials or, more specifically, tackled time perception in the context of interactional sounds such as vocal or musical expression. One example is work by Voyer and colleagues who presented the word 'bower' spoken with an angry, happy, and neutral voice (Fallow & Voyer, 2013; Voyer & Reuangrith, 2015). They found that compared to the neutral condition, durations in the two emotional conditions were underestimated, suggesting that an internal clock slowed down. Musical expressions yielded similar results. In a study by Droit-Volet and colleagues (2010a), participants listened to musical excerpts or to sine wave control stimuli that did not fit Western tonality and were thus perceived as less pleasant/emotional than the musical excerpts. The duration of musical excerpts, when compared to the control stimuli, was underestimated.

We set out to extend this literature and to address three outstanding issues. First, we were interested in understanding the temporal distortions obtained with dynamic interactional sounds. Specifically, we asked whether emotion-induced underestimations could be explained by the acoustic properties of dynamic sounds as well as the acoustic changes introduced by duration manipulations. In past studies, stimulus duration was manipulated by stretching and compressing sounds (Fallow & Voyer, 2013; Voyer & Reuangrith, 2015) such that the rate of acoustic change was greater for short than for long stimuli. This gave listeners a cue to predict stimulus duration before stimulus offset. Moreover, if listeners used this cue they could be expected to judge emotional stimuli as shorter than neutral stimuli because emotionality typically increases the rate of acoustic change (Banse & Scherer, 1996).

Second, we were concerned that research on the influence of emotions on time rarely combined behavioral measures with recordings of neural activity (for recent fMRI studies see Dirnberger et al., 2012; Tipples et al., 2013). Therefore, we examined event-related potential (ERP) components previously implicated in vocal emotional and temporal processing. One such component, the P2, has a positive peak around 200 ms following stimulus onset. The P2 is modulated by the emotionality of vocal expression, although the direction of the effect varies among studies and the reason for the variation is unclear (Iredale et al., 2013; Kokinous et al., 2015; Paulmann & Kotz, 2008; Schirmer et al., 2013a). Additionally, the P2 has been implicated in the temporal comparison of two successive durations. Tones that marked the offset of the second duration elicited a larger P2 the more temporally different (i.e., shorter/longer) this duration was from the first duration (Kononowicz & Rijn, 2014). Thus, given its susceptibility to both emotional and temporal manipulations, the P2 may be a good candidate for revealing the interaction of emotions and time.

Another ERP component of interest is a late negativity, called the N4, that peaks around 400 ms following stimulus onset. The N4 reflects the influence of vocal emotional expressions on language processing. Expressions that conflict with verbal emotion content increase N4 amplitude (Bostanov & Kotchoubey, 2004; Schirmer & Kotz, 2003). Thus, the ERP for a positive word such as 'success' is more negative when heard in a negative as compared to a neutral or positive voice.

The N4 is part of a family of late negative deflections. It is, thus, related to the contingent negative variation (CNV), which emerges at about the same time to a warning stimulus that predicts an imperative stimulus (Tecce, 1972; Walter et al., 1964) or to an imperative stimulus that subjects must time (Macar & Vitton, 1980; Ruchkin et al., 1977). In the context of some timing paradigms, CNV amplitudes are larger for longer than shorter stimuli. Thus, initially, the CNV was thought to reflect the accumulation of temporal pulses or ticks of an internal clock (Macar & Vidal, 2003). More recently, however, response preparation and temporal decision-making are deemed more likely processes underpinning this component (Kononowicz & van Rijn, 2011; Ng, Tobin, & Penney, 2011; van Rijn et al., 2011). Together, N4 and CNV findings suggest that influences of emotions on time may be revealed by late negative aspects of the ERP.

The third issue of interest was the participants' sex. Past research revealed that women are more sensitive than men to vocal expression (Schirmer & Kotz, 2003; van den Brink et al., 2010). For example, when watching a movie and passively listening to an auditory oddball sequence, women are more likely than men to show enhanced processing of emotional relative to neutral oddballs (Schirmer et al., 2005). When asked to judge the lexical or emotional properties of words, women are more affected than men by task-irrelevant vocal expressions (Schirmer et al., 2004). Additionally, women are more likely than men to adjust their attitude toward a word's referent based on the tone of voice in which the word was heard previously (Schirmer et al., 2013a). Together, these and similar findings suggest that task-irrelevant emotional signals are more likely to impact ongoing cognitive processes in women than in men. However, this possibility has rarely been explored in the context of time perception (but see Chambon et al., 2008; Kliegl et al., 2015).

To address the three issues raised above, we presented the syllable 'ah' with a surprised, disgusted, and neutral voice in a temporal bisection paradigm. In line with past research, we stretched and compressed original voice recordings into seven durations and tested the following predictions. First, we expected emotional voices to be judged shorter than same duration neutral voices (Fallow & Voyer, 2013; Voyer & Reuangrith, 2015). Second, if the rate of acoustic change rather than emotionality accounts for this effect, then measures of acoustic change (e.g., variation in pitch) should be greater for emotional compared with neutral voices and explain condition differences in timing. Alternatively, stimulus valence and arousal might predict temporal judgements. Third, we expected emotionality and duration manipulations to modulate and interact in the P2 and/or a late negative potential. Last, in line with earlier work on sex differences, we hypothesized that emotional expressions would be more likely to shape the temporal decisions and ERPs of female as compared to male listeners.

2. Methods

This research was approved by the Institutional Review Board of the National University of Singapore.

2.1. Participants

Twenty-eight participants, half female, were recruited for this research. The average age of females was 22.6 (SD 2.9) and that of males was 22.8 (SD 2.1). Participants self-reported to be free of hearing impairments and neurological conditions. All had normal or corrected-to-normal vision and were right handed. Participants were recruited via campus advertisements and reimbursed for their time at a rate of S$10/hr.

2.2. Stimulus Materials

The stimulus material for the training phase comprised two uniform amplitude 440 Hz sine wave tones of 300 and 1200 ms duration, respectively.

The stimulus material for the test phase comprised a set of vocalizations from five different speakers (two female) who each produced the sound 'ah' with a neutral, a disgusted, and a surprised intonation. On average, neutral sounds were 814 ms long (SD 578), disgusted sounds were 972 ms long (SD 365), and surprised sounds were 582 ms long (SD 181). Sound durations were subjected to an ANOVA with *Emotion* as a between items factor. The result was non-significant [$F(2, 12) = 1.1$, $p = 0.35$]. Apart from duration, we also analyzed pitch standard deviation because visual inspection of sound spectrograms suggested that the main acoustic change was a single rise/fall pattern that extended over the course of the syllable. As expected, an ANOVA with pitch standard deviation as the dependent variable revealed an *Emotion* main effect [$F(2, 12) = 14.4$, $p < 0.001$], indicating greater acoustic change for surprise relative to disgust [$F(1, 8) = 5.1$, $p = 0.05$] and for disgust relative to neutral vocalizations [$F(1, 8) = 6.2$, $p < 0.05$]. The means for these conditions are 112.5 Hz (SD 28.2), 57.4 Hz (SD 46.8), and 5.2 Hz (SD 3.6), respectively.

Using Celemony Melodyne2 (Hoenig & Neubäcker, 2005), the 15 unique vocalizations (five neutral, five disgusted, and five surprised) were compressed and stretched to 300, 378, 476, 600, 756, 952, and 1200 ms duration, yielding a set of 105 vocalizations. The Celemony algorithm was designed to enable duration changes without altering average pitch and short-term spectral features.

The original 15 vocalizations that formed the basis of this study were derived from a larger corpus recorded from 33 speakers asked to say 'ah' with angry, disgusted, fearful, happy, sad, surprised, and neutral intonations. Recordings were made in a soundproof chamber and digitized at a 16 bits/44.1 kHz sampling rate. The corpus was subjected to a rating study with 30 participants (12 female, average age 22.1). After listening to each sound, participants selected one of seven options (i.e., angry, disgusted, fearful, happy, sad, surprised, or neutral) on the screen or entered an expression via the keyboard when they believed none of the provided options was suitable.

Vocalizations with better than 50% accuracy (chance = 12.5%) were subjected to duration manipulation in Celemony Melodyne2 and then evaluated by the present authors. After removing sounds audibly distorted by the manipulation, we identified two emotional and one neutral stimulus subset for which original, unmanipulated sources differed neither in duration nor in expression recognition accuracy (ps > 0.1). The amplitude of each selected sound was normalized in MATLAB using a root-mean-square (RMS) leveler (http://depts.washington.edu/phonlab/resources/rmsLeveler.m), which computes an RMS value for each sound and subsequently normalizes all sounds to the maximum RMS level detected. Acoustic analysis of the manipulated sounds ensured that mean pitch and pitch standard deviation were comparable across the different stimulus durations. Auditory samples are available here: https://dl.dropboxusercontent.com/u/24234612/DurationManipulationSamples_TimingNTimePerception.zip.

2.3. Procedure

Participants were tested individually. After arriving at the lab, they signed a consent form and completed a questionnaire that recorded basic personal data (e.g., sex, handedness, age). To prepare participants for the EEG, a 64-channel cap with empty electrode holders was placed on their head. Electrode holders were filled with an electrolyte gel and electrodes placed according to the modified 10–20 system. Individual electrodes were attached above and below the right eye and at the outer canthus of each eye to measure eye movements. One electrode was attached to the nose for off-line data referencing. The data was recorded at 256 Hz with a Biosemi ActiveTwo system, which uses a common mode sense active electrode for initial referencing. Only an antialiasing filter was applied during data acquisition (i.e., sinc filter with a half-power cutoff at 1/5 the sampling rate). Following the EEG set-up, participants moved into the experimental chamber and sat down in front of a computer screen at a distance of about 1.2 m. Then the experimenter assisted with the insertion of ear-insert headphones for sound delivery.

After reading task instructions on screen, participants completed five blocks comprised of training and test phase. During training, they were presented with ten trials, half of which involved a short and half of which involved a long anchor duration. Each trial began with a fixation cross. After 100 ms, the cross was accompanied by a tone at the offset of which participants pressed one of two buttons indicating whether the tone was short or long. Their response was followed by the words 'Correct' or 'Incorrect', which remained on screen for 1500 ms. The fixation cross then reappeared for an inter-trial interval of random duration between 1000 and 3000 ms. Trials were presented in random order.

During the test phase, a fixation cross was presented on screen without interruption. Vocalizations were played in random order and participants classified each vocalization as being more similar in duration to the long or the short anchor from the training phase. The inter-trial interval was drawn randomly from durations between 1000 and 3000 ms. Each test phase started with a filler trial and was followed by 126 experimental trials. Across the five test phases, there were six presentations for each of the 105 vocalizations and 30 trials for each emotion by duration combination. Response button assignment was counterbalanced across participants.

After completion of the timing task, the experimenter removed the electrodes and cap and participants washed their hair. Once finished, they returned to the experimental room and performed a short sound rating in which the 105 vocalizations presented in the timing task were played again in random order. After each vocalization, participants indicated whether the speaker sounded disgusted, surprised, or neutral. Subsequently, they were given a five-point valence scale ranging from −2 (very negative) to +2 (very positive) and a 5-point arousal scale ranging from 1 (very weak) to 5 (very strong). After entering their ratings of a given vocalization, there was a 1500 ms interval before the next vocalization was presented.

2.4. Data Analysis

We converted the temporal judgements of each participant into the probability of choosing 'long' and fitted this probability to a logistic function using the glm function in R (R Core Team, 2015). We then subjected the resulting fit to the inverse.predict function in the R package chemCal to obtain a predicted duration value for a 0.5 probability of choosing 'long', which indicates the point of subjective equality (PSE). We determined the difference limen (DL) by identifying the predicted duration value for 0.25 and 0.75, subtracting the former from the latter, and dividing the result by 2.

EEG data was processed with EEGLAB (Delorme & Makeig, 2004). The recordings were re-referenced to the nose and subjected to low- and high-pass filtering with a half-power cut-off at 30 and 0.1 Hz, respectively. The transition band was 7.5 Hz for the low pass filter (−6 dB/octave; 221 pts) and 0.1 Hz for the high pass filter (−6 dB/octave; 16501 pts). The continuous data were visually scanned for non-typical artifacts caused by drifts or muscle movements. Time points containing such artifacts were removed. Infomax, an independent component analysis algorithm, was applied to the remaining data, and components reflecting typical artifacts (i.e., horizontal and vertical eye movements and eye blinks) were removed. Back-projected single trials were epoched and baseline-corrected using a 200 ms pre-stimulus baseline and an 800 ms time window starting from stimulus onset. The resulting epochs were again screened visually for residual artifacts. ERPs were derived by averaging individual epochs for each condition and participant.

It was our intention to generate separate ERPs for long and short responses in order to explore temporal decision-making. However, the bisection curves were too steep, making it impossible to match trial numbers for the different durations and the two response types both when considering all durations and when considering duration subgroups (e.g., 600 and 756). Therefore, we generated ERPs based on stimulus duration irrespective of whether participants classified a stimulus as short or long. In this process, we excluded the 300 ms condition because this allowed us to explore ERP effects up to 378 ms that were unaffected by stimulus offsets. Additionally, we combined adjacent duration conditions to create three stimulus sets (short: 378/476, intermediate: 600/756, and long: 952/1200) to simplify the analysis while increasing our signal-to-noise ratio. The comparison of these stimulus sets enabled us to explore effects associated with the rate of acoustic change, which was greatest in the short (compressed) and smallest in the long (stretched) stimulus condition.

We identified the latency ranges of target ERP components and submitted mean voltages from within these ranges to an Analysis of Variance (ANOVA) with *Emotion* (disgusted, surprised, neutral), *Duration* (short, intermediate, long), *Hemisphere* (left, right), and *Region* (anterior, central, posterior) as repeated measures factors and *Sex* as the between subjects factor. The factors *Hemisphere* and *Region* comprised average voltages computed across the following subgroups of electrodes: anterior left, Fp1, AF7, AF3, F5, F3, F1; anterior right, Fp2, AF8, AF4, F6, F4, F2; central left, FC3, FC1, C3, C1, CP3, CP1; central right, FC4, FC2, C4, C2, CP4, CP2; posterior left, P5, P3, P1, PO7, PO3, O1; posterior right, P6, P4, P2, PO8, PO4, O2. This selection of electrodes ensured that the tested subgroups contained equal number of electrodes, while providing a broad scalp coverage that allowed the assessment of topographical effects. If the number of follow-up tests of interactions and main effects exceeded the numerator of the degrees of freedom, test results were corrected for multiple comparisons using the False Discovery Rate (Benjamini–Hochberg procedure).

3. Results

3.1. Behavioral Results

The group average probability of responding 'long', as well as the individual participant PSE and DL values are presented in Fig. 1 for male and female participants separately. As can be seen from this figure, timing responses were overall less accurate and more influenced by emotion in female than in male participants. Moreover, only in female participants, disgust and surprise elicited fewer 'long' responses relative to neutral, pointing to an underestimation of time.

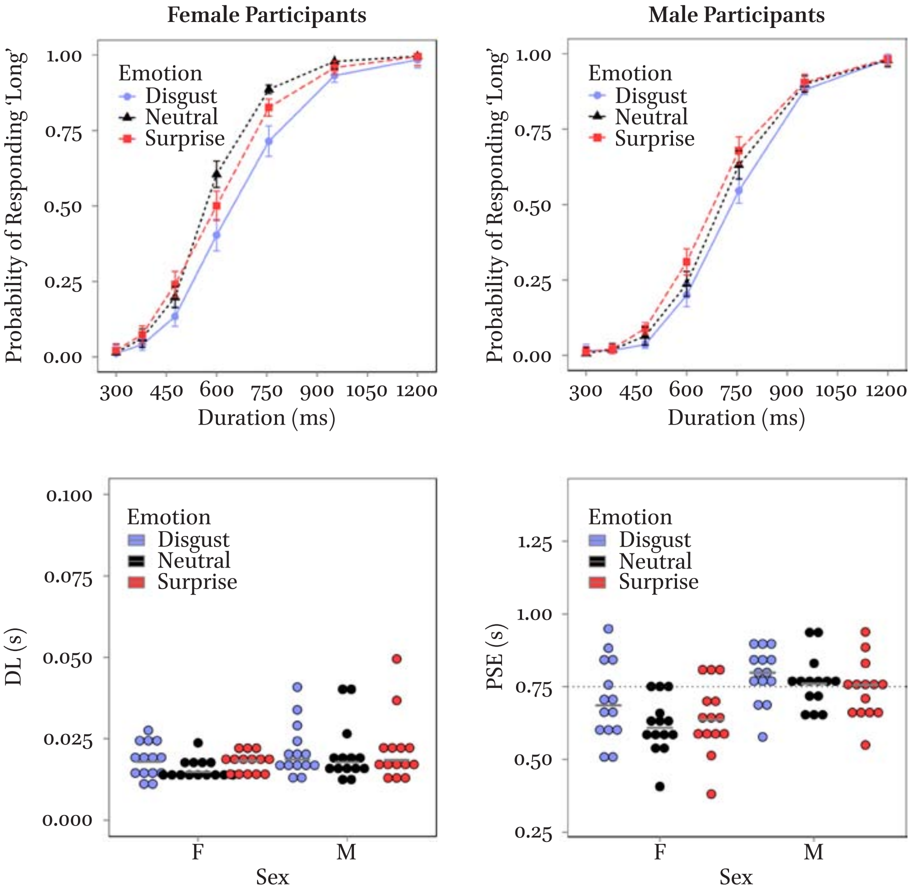

Figure 1. Group mean bisection response functions and individual DLs and PSEs for each emotion condition plotted separately for female and male participants. Error bars reflect the within-subject standard error. The dashed horizontal line in the lower right panel indicates the objective point of equality (i.e., equal arithmetic difference). Grey solid vertical lines in left and right lower panels mark the group median. This figure is published in color in the online version.

To probe these visual impressions statistically, we subjected PSEs to an ANOVA with *Emotion* as a repeated measures factor and *Sex* as a between subjects factor. We found significant effects of *Emotion* [$F(2, 52) = 13.9, p < 0.0001$] and *Sex* [$F(1, 26) = 8.3, p < 0.01$] as well as an interaction [$F(2, 52) = 3.3, p < 0.05$]. In female participants, the *Emotion* main effect was significant [$F(2, 26) = 15.8, p < 0.0001$]. Moreover, follow-up comparisons revealed a larger PSE for the disgust than the neutral condition [$F(1, 13) = 20, p < 0.01$] and for the surprise than the neutral condition [$F(1, 13) = 6.5, p < 0.05$]. Additionally, the PSE was larger for the disgust than the surprise condition [$F(1, 13) = 14, p < 0.01$]. In male participants, the *Emotion* main effect was smaller, but nevertheless significant [$F(2, 26) = 3.7, p < 0.05$]. Although, neither the disgust nor the surprise condition differed significantly from neutral (ps > 0.1), they differed from each other [$F(1, 13) = 14.8, p < 0.01$]. As for females, the PSE was larger in the disgust than the surprise condition.

An ANOVA with DL as the dependent variable revealed a marginal *Emotion* effect [$F(2, 52) = 2.4, p = 0.09$]. All other effects were non-significant (ps > 0.1).

To determine how female and male participants perceived the sounds presented in this study, we subjected emotion recognition accuracy, valence, and arousal from the post-experimental rating to separate ANOVAs with *Emotion* and *Duration* as repeated measures factors and *Sex* as the between subjects factor. For accuracy, this analysis identified an *Emotion* main effect [$F(2, 52) = 12, p < 0.0001$] indicating that neutral expressions were recognized more accurately than disgust [$F(1, 26) = 18.4, p < 0.001$] and surprise expressions [$F(1, 26) = 17, p < 0.001$] which did not differ ($p > 0.3$). A marginal *Emotion* by *Duration* interaction [$F(12, 312) = 1.7, p = 0.06$] hinted that disgust expressions were better recognized at longer durations, whereas surprise expressions were better recognized at shorter durations. However, as can be seen in Fig. 2, all conditions were recognized significantly better than would be expected by chance (i.e., accuracy > 0.33).

An analysis of valence ratings revealed main effects of *Emotion* [$F(2, 52) = 81, p < 0.0001$] and *Duration* [$F(6, 156) = 2.4, p < 0.05$] as well as an *Emotion* by *Duration* interaction [$F(12, 312) = 3.1, p < 0.001$]. Across durations, disgust expressions were more negative than neutral [$F(1, 26) = 6.4, p < 0.0001$] and surprise expressions [$F(1, 26) = 142.4, p < 0.0001$]. Additionally, neutral expressions were less positive than surprise expressions [$F(1, 26) = 18.7, p < 0.001$]. Disgust, but not neutral or surprise expressions (ps > 0.3), showed a *Duration* main effect [$F(6, 156) = 6.1, p < 0.0001$], indicating that longer durations were perceived as more negative.

Analysis of arousal ratings revealed main effects of *Emotion* [$F(2,52) = 264, p < 0.0001$] and *Duration* [$F(6, 156) = 10.8, p < 0.0001$] as well as an *Emotion* by *Duration* interaction [$F(12, 312) = 3.2, p < 0.001$]. Across durations, neutral expressions were perceived as less arousing than disgust [$F(1, 26) = 213, p < 0.0001$] and disgust expressions were perceived as less arousing than surprise expressions [$F(1, 26) = 21.9, p < 0.0001$]. Both disgust [$F(6, 156) = 9.8, p < 0.0001$] and surprise expressions showed a *Duration* main effect [$F(6, 156) = 8.2, p < 0.0001$],

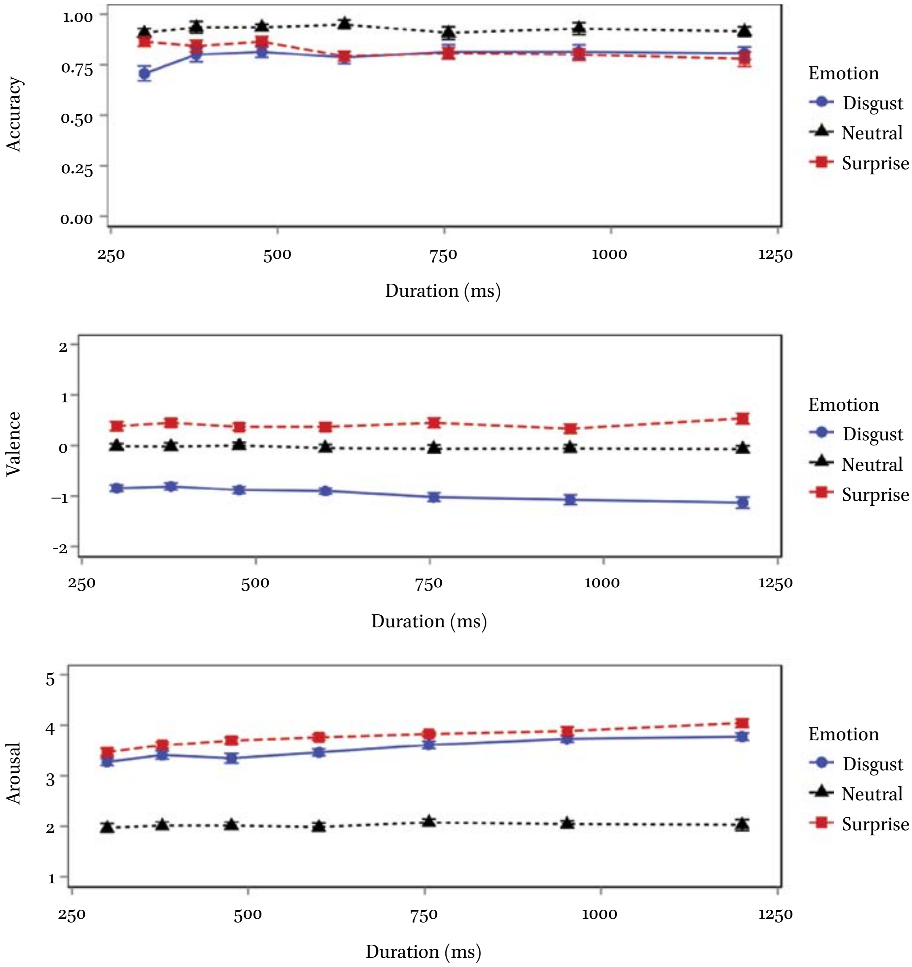

Figure 2. Stimulus rating results for accuracy, valence and arousal. Plotted values are across participant averages. Error bars reflect the within-subject standard error. This figure is published in color in the online version.

indicating that increased durations were associated with greater arousal ratings. The *Duration* effect was not significant for neutral expressions ($p > 0.89$).

To explore whether and how PSE effects may be explained by the role of subjective item characteristics such as valence and arousal, we entered the PSE into two linear regression mixed effects analyses using the lme4 package in R. The first analysis had *Sound Valence*, *Sex* (deviation coded as –0.5 and 0.5), and their interaction as predictors. Subject intercept and *Sound Valence* slope were modelled as the random effect. The *Sound Valence* predictor was derived from the post-experimental ratings as the mean score for each emotion condition and subject. The analysis revealed a significant effect of *Sound Valence* [$t(14.9) = -4.2$, $p < 0.01$; $\beta = -0.039$] and *Sex* [$t(25.3) = -3.2$, $p < 0.01$; $\beta = 0.12$] and a non-significant *Sound Valence* by *Sex* interaction ($p > 0.48$). In both males and females, less positive/more negative valence was associated with a larger PSE.

The second analysis had *Sound Arousal*, *Sex* (deviation coded as –0.5 and 0.5) and their interaction as predictors and subject intercept and *Sound Arousal* slope as the random effect. Again the *Sound Arousal* predictor was derived from the post-experimental ratings as the mean score for each emotion condition and subject. The analysis revealed a significant effect of *Sex* [$t(42.5) = -4.1$, $p < 0.001$; $\beta = -0.21$] and an interaction of *Sound Arousal* and *Sex* [$t(45.1) = 2.1$, $p < 0.05$; $\beta = 0.03$]. The *Sound Arousal* effect was non-significant ($p = 0.14$). Follow-up analyses of the interaction indicated that the *Sound Arousal* effect was significant in women [$t(19.9) = 2.8$, $p < 0.05$; $\beta = 0.027$], but not in men ($p > 0.76$). Women, but not men, showed a larger PSE with increasing stimulus arousal.

To explore the role of pitch standard deviation for temporal decisions, we conducted a binomial regression mixed effects analysis with *Duration*, *Pitch SD,* and *Sex* (deviation coded as –0.5 and 0.5) as well as the interaction of *Pitch SD* and *Sex* as predictors. Subject intercept and *Pitch SD* slope as well as sound intercept and *Pitch SD* slope were modelled as random effects. Participants' responses (0 = short, 1 = long) served as the dependent variable. As expected, the *Duration* effect was significant; the 300 ms condition differed from all other conditions (for all comparisons $\beta > 1.18$, Std Err < 0.27, z-value > 4.15, $p < 0.0001$) in having a smaller number of 'long' responses. Importantly, the *Sex* effect ($\beta = 1.5$, Std Err = 0.43, z-value = 3.6, $p < 0.001$) and the *Pitch SD* by *Sex* interaction were significant ($\beta = -0.005$, Std Err = 0.002, z-value = –2.28, $p = 0.023$). Follow-up analyses indicated that female participants tended to respond 'long' less frequently for stimuli with a greater variation in pitch ($\beta = -0.004$, Std Err = 0.002, z-value = –1.84, $p = 0.066$). A similar effect was non-significant in male participants ($p > 0.76$).

3.2. Electrophysiological Results

As can be seen in Fig. 3, ERPs revealed modulations of the P2 and the following late negative potential (LNP).

To identify a P2 analysis window, we measured peak latencies between 150 and 250 ms for each electrode, condition, and subject. We found an average peak at 201 ms with an SD of 29 ms and centered on the peak a one SD wide analysis window ranging from 186 to 215 ms following stimulus onset. Mean voltages within this window were subjected to an ANOVA with *Emotion*, *Duration*, *Hemisphere,* and *Region* as repeated measures factors and *Sex* as a between subjects factor. The analysis revealed a significant interaction of *Emotion* and *Region* [$F(4, 104) = 7.0$, $p < 0.0001$] as well as *Emotion, Region,* and *Sex* [$F(4, 104) = 3.1$, $p < 0.05$]. These interactions are illustrated in Fig. 4.

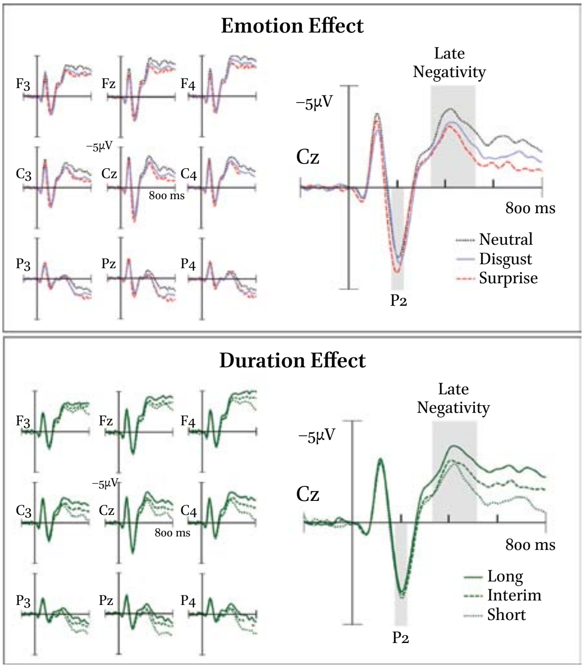

Figure 3. ERP traces illustrating the *Emotion* and *Duration* main effects averaged across male and female participants. Shaded areas indicate the analysis time window for the P2 and the Late Negative Potential, respectively. This figure is published in color in the online version.

Follow-up analyses in female participants indicated that the *Emotion* by *Region* interaction [$F(4, 104) = 7.0, p < 0.0001$] as well as the *Emotion* simple main effects over anterior [$F(2, 26) = 4.7, p < 0.05$], central [$F(2, 26) = 4.6, p < 0.05$], and posterior regions [$F(2, 26) = 4.8, p < 0.05$] were significant. Over anterior [$F(1, 13) = 6.3$, $p = 0.08$; $F(1, 13) = 3.8$, $p = 0.1$] and central regions [$F(1, 13) = 4.5$, $p = 0.08$; $F(1, 13) = 5.8$, $p = 0.08$], surprised voices tended to elicit a larger P2 than neutral and disgusted voices. The P2 to neutral and disgusted voices failed to differ (*p*s > 0.1). Over posterior regions, disgusted voices elicited a smaller P2 than neutral [$F(1, 13) = 10.4, p < 0.05$] and surprised voices [$F(1, 13) = 5.8, p < 0.05$], but follow up analyses revealed no significant effects (*p*s > 0.1).

The LNP had an average latency of 448 ms (SD 94) and was fairly extended in time. Hence, we set our analysis window from 350 to 550 ms. As this window included stimulus offsets from the short duration condition, these stimuli were excluded reducing the *Duration* factor to two levels, intermediate and long. An ANOVA conducted on mean voltages within the analysis window revealed a significant main effect of *Duration* [$F(1, 26) = 9.1, p < 0.01$] as well as significant interactions of *Emotion* and *Region* [$F(4, 104) = 3.8, p < 0.01$], and *Emotion*, *Duration*, *Region* and *Sex* [$F(4, 104) = 4.4, p < 0.01$]. Effects are illustrated in Fig. 5.

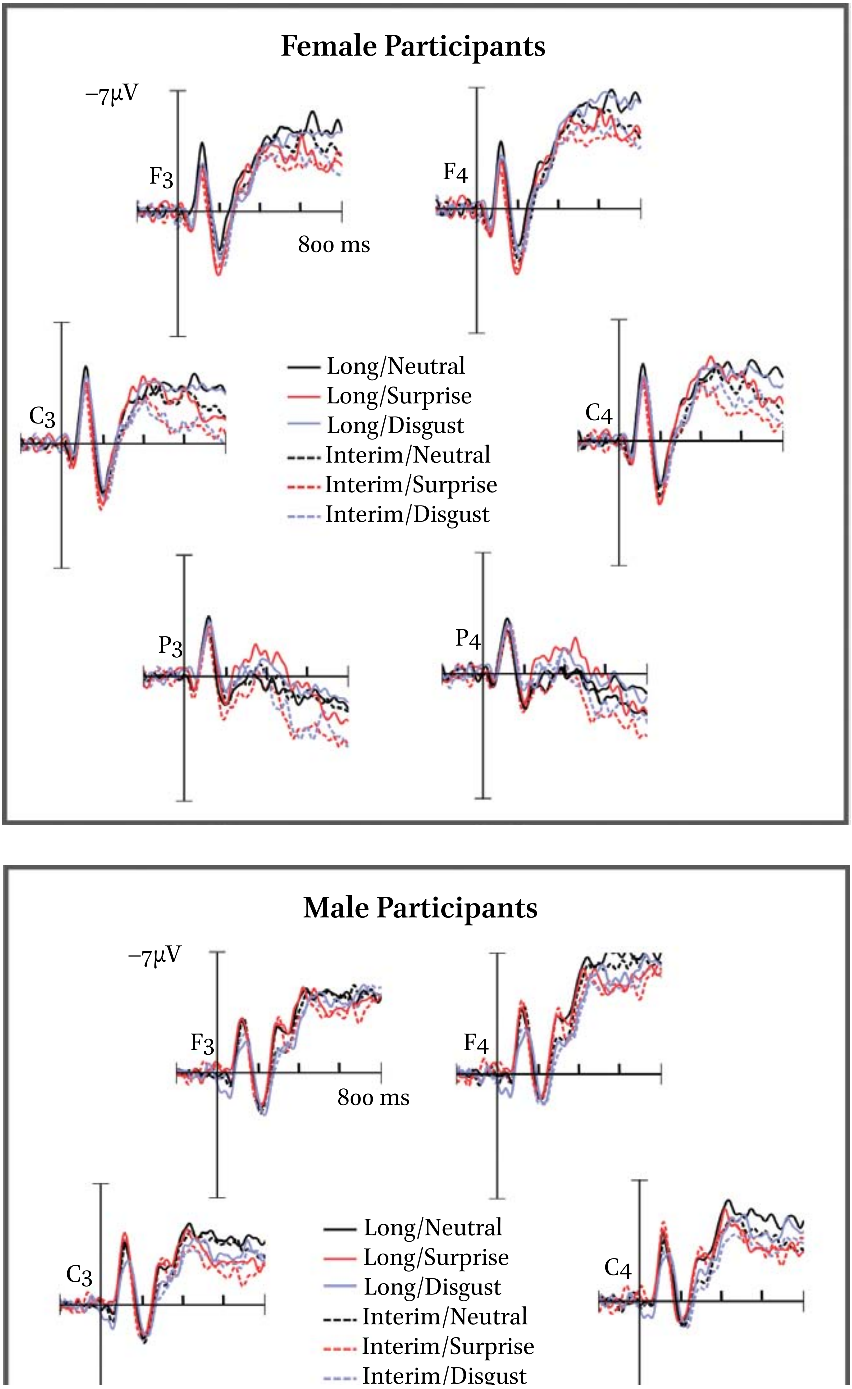

Figure 4. ERP traces illustrating the *Emotion* × *Duration* interaction for female (top) and male (bottom) participants. This figure is published in color in the online version.

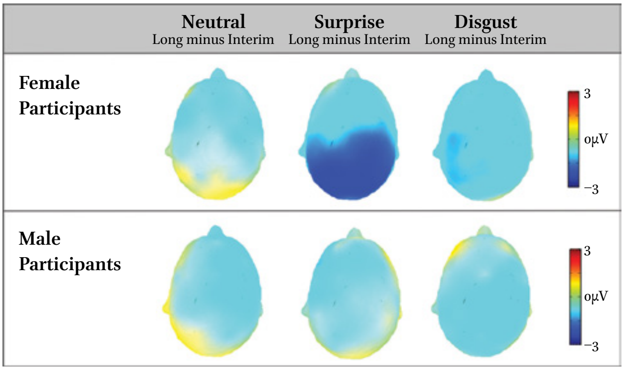

Figure 5. ERP difference maps computed by subtracting the intermediate duration condition from the long duration condition within the LNP time window. This figure is published in color in the online version.

In female participants, the interaction of *Emotion, Duration, and Region* was significant [$F(4, 104) = 4.3, p < 0.01$]. Follow-up analyses indicated that the *Emotion* by *Duration* interaction was significant over posterior [$F(2, 26) = 3.4, p < 0.05$], but not anterior or central electrodes (*p*s > 0.1). Over posterior electrodes, LNP amplitudes were greater for long relative to intermediate durations in the surprise condition [$F(1, 13) = 11.8, p < 0.05$]. Long and intermediate durations elicited comparable amplitudes in the disgust and neutral conditions (*p*s > 0.1). Female participants showed a *Duration* main effect over central electrodes [$F(1, 13) = 13.3, p < 0.01$] whereby long durations elicited a greater LNP than intermediate durations. The *Duration* main effect was non-significant over anterior electrodes (*p*s > 0.1). In male participants, the interaction of *Emotion, Duration, and Region* was non-significant as were the *Duration* main effect and the *Emotion* by *Region* interaction (*p*s > 0.1).

To determine whether differences in overall ERP amplitude between men and women could account for these effects, we repeated the above analyses with normalized values (i.e., mean-centered and divided by SD for males and females separately). This did not change the effects.

4. Discussion

This study explored the role of vocal emotions for the perception of time. In a duration bisection procedure, vocal emotions affected temporal estimates in female and male participants differently. Concurrently recorded ERPs revealed sex differences in the P2 and LNP reflecting an emotion main effect and an emotion by duration interaction, respectively. In the following sections, we discuss these findings in reference to our initial questions concerning the direction of emotion effects on time, their neural correlates, and the role of sex (for evidence from non-human animals see Cheng et al., 2008; Pleil et al., 2011).

4.1. Emotions — Speeding-up or slowing down subjective time?

Past research on the relationship between emotions and time relied on visual and, to a lesser extent, auditory stimulation. Visual studies typically employed facial expressions as well as picture sets like the International Affective Picture System (IAPS; Lang et al., 2008). Similarly, auditory studies employed vocal expressions (Fallow & Voyer, 2013; Voyer & Reuangrith, 2015) or music (Droit-Volet et al., 2010a), but also more general affective sounds such as those from the International Affective Digitized Sound System (IADS; Droit-Volet et al., 2010b; Mella et al., 2011; Noulhiane et al., 2007; Wackermann et al., 2014).

Reported emotion effects on time varied. The majority of visual studies found emotional stimuli to be temporally overestimated relative to neutral stimuli (Droit-Volet & Meck, 2007; Schirmer, 2011), while only few studies found the opposite effect. In the auditory modality, results were fairly evenly divided. Voices or music revealed an underestimation, whereas sounds from the IADS revealed an overestimation. What causes this variability?

In the Introduction to this article, we speculated that sensory properties could be relevant. Specifically, we argued that the compressing/stretching manipulations done for vocal stimuli, but not for visual material or sounds from the IADS, confound stimulus duration with the rate of acoustic change. Temporal decisions could thus be influenced by rate such that 'fast' stimuli are more likely to be judged as shorter than 'slow' stimuli. As emotional voices are often acoustically more variable than neutral voices, this variability rather than emotion may account for differential temporal decisions (Banse & Scherer, 1996).

Our results provide some support for this supposition. We found that unmanipulated emotional expressions had greater pitch change than their neutral counterparts. Thus, for comparable durations, there was more going on in emotional relative to neutral sounds. Additionally, pitch change tended to be negatively associated with 'long' responses in female listeners. Thus, fewer long responses in the emotional relative to the neutral conditions may have resulted from the former having a greater rate of acoustic change (see Lake et al., 2014).

Given this finding, one may rightly ask whether the compression/stretching approach should be avoided in preference for the sound segmentation approach in the study of temporal perception. We believe that the two approaches are equally useful and problematic. Both reflect real-world interactions between stimulus duration and content and both necessarily create a confound. If stimulus content is held constant, acoustic rate must vary. The same content plays out faster in shorter than in longer sounds. If acoustic rate is held constant, stimulus content must vary. Shorter sounds have less content than longer sounds. Thus, whereas compression/stretching manipulations bias 'short' responses to emotional stimuli, the sound segmentation approach can be expected to bias 'short' responses to neutral stimuli.

Notably, however, both approaches differ in their relation to an established tendency to perceive high-content stimuli as longer than low-content stimuli of comparable duration (e.g., Eagleman, 2008; Liverence & Scholl, 2012; Wearden et al., 1999). If this tendency extends to vocal processing, then the compression/stretching approach creates stimuli for which accurate duration judgements correlate negatively with an existing temporal bias (i.e., longer sounds have less content per temporal unit). In contrast, the segmentation approach would create stimuli for which accurate duration judgements correlate positively with an existing temporal bias (i.e., longer sounds have more content). Moreover, the compression/stretching approach may then be considered the stronger test as acoustic rate changes would need to overwrite an existing tendency to equate more content with longer duration. In other words, participants would need to learn within an experimental session that a fast rate of acoustic change (i.e., high content) is associated with short instead of long durations and to allow this knowledge to inform their temporal decisions. As a cautionary note, however, it is unclear whether the tendency to perceive high-content stimuli as long is indeed ubiquitous. In the context of vocalizations, one may venture that it is absent or even reversed as emotional expressions with great acoustic change are often shorter than neutral expressions with little acoustic change (Banse & Scherer, 1996). Listeners may mentally represent such regularities.

Although a high rate of acoustic change may play a role in making vocalizations seem shorter, the present study implies that stimulus emotions are also relevant. We observed a negative relation between valence and the PSE, but a positive relation between arousal and the PSE. This accords with previous evidence highlighting the importance of emotions for time. However, it contradicts a dominant theoretical view of how emotions interact with temporal processes. Specifically, the notion that emotion induced bodily arousal speeds up the ticking of an internal clock (Droit-Volet & Meck, 2007) is at odds with the present effects. To explain this contradiction, we propose that emotions first and foremost disrupt ongoing temporal processes and that, consequently, temporal decisions are inaccurate and prone to extraneous influences. In other words, unreliable temporal record taking falls back on temporal heuristics such that temporal judgments are biased by accessible information. For example, they may be biased by the extent to which a stimulus recruits attention. Positive surprise is arguably more attention capturing and holding than disgust and may hence be perceived as longer (Curtis et al., 2011; Lassalle & Itier, 2013). Although speculative, this interpretation agrees with other attempts to explain temporal distortions (e.g., Schirmer, 2011; Taatgen et al., 2007; Zélanti & Droit-Volet, 2011; this issue Droit-Volet et al., 2016; Halbertsma & Van Rijn, 2016; Lake et al., 2016).

4.2. Neural Correlates for the Influence of Emotions on Time

Looking at ERP correlates, the present study identified emotion effects in an early positive deflection, the P2, and a late negative potential (LNP). The P2 tended to be larger for surprised but was smaller for disgusted relative to neutral sounds. This effect mirrors the valence effect seen for the PSE suggesting that P2 amplitude reflects enhanced attentional engagement with the surprised sounds, but reduced engagement with the disgusted sounds. This interpretation accords with the presumed functionality of both emotions as well as with previous research on their respective roles for attention (Curtis et al., 2011; Lassalle & Itier, 2013). Additionally, it agrees with research linking P2 amplitude to attention (Woldorff & Hillyard, 1991).

Emotion effects in the LNP interacted with the rate of acoustic change. Specifically, surprise evoked LNP differences between the intermediate and long/slow condition over posterior electrodes and thus extended a similar effect over central regions that was independent of emotion. Stimuli that, based on their acoustic rate, could be expected to last longer produced larger LNP amplitudes than stimuli that, based on their acoustic rate, could be expected to end sooner. The absence of this effect earlier in the ERP is likely due to a need for some processing before listeners could represent acoustic rate. Moreover, it indicates that access to such information was slower than access to vocal emotion and, hence, may have been strategic. Notably, the acoustic rate effect resembles previously reported CNV effects that emerge at a similar time with a similar topography (Macar & Vidal, 2003).

The present ERP results match the behavioral results imperfectly, because the two emotion conditions failed to differ consistently from the neutral condition and were seemingly unaffected by stimulus arousal. Moreover, a set of analyses correlating ERP difference scores (disgust/surprise minus neutral) with PSE and DL difference scores was non-significant (omitted from result section in the interest of space). Thus, ERP and behavioral results appear to reflect somewhat different processes. Whereas ERP results are limited to the processing of acoustic rate and necessarily omit the comparison of short and long responses, behavioral results represent the end-product of all stimulus-related processes and are, perhaps, the 'richer' measure. Nevertheless, our ERP results show that emotions interact with the processing of acoustic rate, which is known to play a role in timing (Liverence & Scholl, 2012; Wearden et al., 1999; for a review Eagleman, 2008). Moreover, they indicate that their interaction emerges after an initial emotion analysis reflected by the P2 and for a late negative potential that presumably marks temporal decision-making (Ng et al., 2011; Van Rijn et al., 2011).

4.3. Sex Differences in Emotion Are Relevant for Temporal Processing

The last issue tackled here concerns sex differences. Previous research showed that women and men have different emotion sensitivities (for a review see Schirmer, 2013). For example, there is substantial evidence that social stimuli and the emotions encoded in them are more relevant to female than male individuals (Pasterski et al., 2005; Schirmer et al., 2007; Spreckelmeyer et al., 2009). In line with this, the women, but not the men, in this study showed a significant P2 and LNP effect reflecting the processing of task-irrelevant vocal emotions. Additionally, in women the PSE correlated with both stimulus valence and arousal, whereas in men the PSE correlated with stimulus valence only. These results suggest that, compared to men, women access emotional information more readily from social stimuli (Hung & Cheng, 2014; Schirmer et al., 2005, 2007) and are more likely to integrate this information with ongoing mental processes (Schirmer & Kotz, 2003; Schirmer et al., 2013a,b).

Importantly, these findings should not be taken to suggest that temporal distortions from emotion are always greater in women than in men. As reviewed elsewhere (Schirmer, 2013), there are stimuli more emotionally provoking to men than to women (e.g., erotica, mechanical objects, angry faces). Hence, sensitivity depends on the kind of stimuli that are used for research. The nature of these stimuli and their relevance to each sex must be considered if we are to better understand the relation between emotions and time. Additionally, the present sex effect should be viewed in light of women, but not men, judging durations as consistently shorter than they were. Possibly, ERP and arousal/PSE effects show only in the presence of such temporal underestimation. Future research should replicate their presence in women, but not men, while controlling for sex differences in temporal acuity.

4.4. Conclusions

To conclude, the present study replicates previous findings that emotional voices are perceived as shorter than neutral voices of comparable duration. Additionally, it adds to existing work by (i) establishing a clear link between this phenomenon and perceived stimulus emotionality, (ii) by identifying associated ERP effects in the P2 and LNP, and (iii) by highlighting processing differences between the sexes. Moreover, our behavioral and ERP results offer some suggestions as to why emotions produce temporal distortions and why these distortions differ across studies. Specifically, emotions may disrupt time perception and increase the reliance on contextual cues or temporal heuristics (e.g., acoustic rate, attention).

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Person. Soc. Psychol., 70*, 614–636.

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, *41*, 259–268.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. (2008). Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody. *Cognition*, *106*, 1497–1503.

Chambon, M., Droit-Volet, S., & Niedenthal, P. M. (2008). The effect of embodying the elderly on time perception. *J. Exp. Soc. Psychol.*, *44*, 672–678.

Cheng, R. K., MacDonald, C. J., Williams, C. L., & Meck, W. H. (2008). Prenatal choline supplementation alters the timing, emotion, and memory performance (TEMP) of adult male and female rats as indexed by differential reinforcement of low-rate schedule behavior. *Learn. Mem.*, *15*, 153–162.

Curtis, V., de Barra, M., & Aunger, R. (2011). Disgust as an adaptive system for disease avoidance behaviour. *Phil. Trans. R. Soc. Lond. B. Biol. Sci.*, *366*, 389–401.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Meth.*, *134*, 9–21.

Dirnberger, G., Hesselmann, G., Roiser, J. P., Preminger, S., Jahanshahi, M., & Paz, R. (2012). Give it time: Neural evidence for distorted time perception and enhanced memory encoding in emotional situations. *NeuroImage*, *63*, 591–599.

Droit-Volet, S., & Meck, W. H. (2007). How emotions colour our perception of time. *Trends Cogn. Sci.*, *11*, 504–513.

Droit-Volet, S., Bigand, E., Ramos, D., & Bueno, J. L. O. (2010a). Time flies with music whatever its emotional valence. *Acta Psychologica*, *135*, 226–232.

Droit-Volet, S., Mermillod, M., Cocenas-Silva, R., & Gil, S. (2010b). The effect of expectancy of a threatening event on time perception in human adults. *Emotion*, *10*, 908–914.

Droit-Volet, S., Fayolle, S., & Gil, S. (2016). Emotion and time perception in children and adults: the effect of task-difficulty. *Timing Time Percept., 4*, 7–29.

Eagleman, D. M. (2008). Human time perception and its illusions. *Curr. Opin. Neurobiol.*, *18*, 131–136.

Eberhardt, L. V., Huckauf, A., & Kliegl, K. M. (2016). Effects of neutral and fearful emotional states on duration estimation of neutral and fearful face stimuli. *Timing Time Percept., 4*, 30–47.

Effron, D. A., Niedenthal, P. M., Gil, S., & Droit-Volet, S. (2006). Embodied temporal perception of emotion. *Emotion, 6*, 1–9.

Fallow, K. M., & Voyer, D. (2013). Degree of handedness, emotion, and the perceived duration of auditory stimuli. *Laterality, 18*, 671–692.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat. Neurosci.*, *8*, 145–146.

Grommet, E. K., Droit-Volet, S., Gil, S., Hemmes, N. S., Baker, A. H., & Brown, B. L. (2011). Time estimation of fear cues in human observers. *Behav. Process.*, *86*, 88–93.

Halbertsma, H., & Van Rijn, H. (2016). An evaluation of the effect of auditory emotional stimuli on interval timing. *Timing Time Percept., 4*, 48–62.

Hoenig, U. G., & Neubäcker, P. (2005). Celemony Melodyne Studio User Manual Version 3.2. München, Germany: Celemony Software GmbH.

Hung, A.-Y., & Cheng, Y. (2014). Sex differences in preattentive perception of emotional voices and acoustic attributes. *NeuroReport*, *25*, 464–469.

Iredale, J. M., Rushby, J. A., McDonald, S., Dimoska-Di Marco, A., & Swift, J. (2013). Emotion in voice matters: Neural correlates of emotional prosody perception. *J. Psychophysiol., 89*, 483–490.

James, W. (1890). *The Principles of Psychology*. New York, NY, USA: Henry Holt.

Kliegl, K. M., Limbrecht-Ecklundt, K., Dürr, L., Traue, H. C., & Huckauf, A. (2015). The complex duration perception of emotional faces: Effects of face direction. *Front. Psychol., 6*, 262. doi: 10.3389/fpsyg.2015.00262.

Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Soc. Cogn. Affect. Neurosci.*, *10*, 713–720.

Kononowicz, T. W., & Van Rijn, H. (2011). Slow potentials in time estimation: The role of temporal accumulation and habituation. *Front. Integr. Neurosci.*, *5*, 48. doi: 10.3389/fnint.2011.00048.

Kononowicz, T. W., & Van Rijn, H. (2014). Decoupling interval timing and climbing neural activity: A dissociation between CNV and N1P2 Amplitudes. *J. Neurosci.*, *34*, 2931–2939.

Lake, J. I., LaBar, K. S., & Meck, W. H. (2014). Hear it playing low and slow: How pitch level differentially influences time perception. *Acta Psychol.*, *149*, 169–177.

Lake, J., Meck, W. H., & LaBar, K. S. (2016). Discriminative fear learners are resilient to temporal distortions during threat anticipation. *Timing Time Percept., 4*, 63–78.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8*. Gainesville, FL, USA: University of Florida.

Lassalle, A., & Itier, R. J. (2013). Fearful, surprised, happy, and angry facial expressions modulate gaze-oriented attention: Behavioral and ERP evidence. *Soc. Neurosci.*, *8*, 583–600.

Liverence, B. M., & Scholl, B. J. (2012). Discrete events as units of perceived time. *J. Exp. Psychol. Hum. Percept. Perform.*, *38*, 549–554.

Macar, F., & Vidal, F. (2003). The CNV peak: An index of decision making and temporal memory. *Psychophysiology*, *40*, 950–954.

Macar, F., & Vitton, N. (1980). CNV and reaction time task in man: Effects of inter-stimulus interval contingencies. *Neuropsychologia*, *18*, 585–590.

Mella, N., Conty, L., & Pouthas, V. (2011). The role of physiological arousal in time perception: Psychophysiological evidence from an emotion regulation paradigm. *Brain Cogn.*, *75*, 182–187.

Min, C. S., & Schirmer, A. (2011). Perceiving verbal and vocal emotions in a second language. *Cogn. Emot.*, 1–17.

Ng, K. K., Tobin, S., & Penney, T. B. (2011). Temporal accumulation and decision processes in the duration bisection task revealed by contingent negative variation. *Front. Integr. Neurosci.*, *5*, 77. doi: 10.3389/fnint.2011.00077.

Noulhiane, M., Mella, N., Samson, S., Ragot, R., & Pouthas, V. (2007). How emotional auditory stimuli modulate time perception. *Emotion, 7*, 697–704.

Pasterski, V. L., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Dev.*, *76*, 264–278.

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, *19*, 209–213.

Pleil, K. E., Cordes, S., Meck, W. H., & Williams, C. L. (2011). Rapid and acute effects of estrogen on time perception in male and female rats. *Front. Integr. Neurosci.*, *5*, 63. doi: 10.3389/fnint.2011.00063

R Core Team. (2015). R: A language and environment for statistical computing. (Version 3.2.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Ruchkin, D. S., McCalley, M. G., & Glaser, E. M. (1977). Event related potentials and time estimation. *Psychophysiology*, *14*, 451–455.

Schirmer, A. (2004). Timing speech: A review of lesion and neuroimaging findings. *Cogn. Brain Res., 21*, 269–287.

Schirmer, A. (2010). Mark my words: Tone of voice changes affective word representations in memory. *PloS One*, *5*, e9080. doi: 10.1371/journal.pone.0009080.

Schirmer, A. (2011). How emotions change time. *Front. Integr. Neurosci.*, *5*, 58. oi: 10.3389/fnint.2011.00058.

Schirmer, A. (2013). Sex differences in emotion. In J. L. Armony & P. Vuilleumier (Eds), *The Cambridge Handbook of Human Affective Neuroscience* (pp. 591–610). Cambridge, UK: Cambridge University Press.

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *J. Cogn. Neurosci.*, *15*, 1135–1148.

Schirmer, A., Zysset, S., Kotz, S. A., & Yves von Cramon, D. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, *21*, 1114–1123.

Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *NeuroReport*, *16*, 635–639.

Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Res.*, *1176*, 103–112.

Schirmer, A., Chen, C.-B., Ching, A., Tan, L., & Hong, R. Y. (2013a). Vocal emotions influence verbal memory: Neural correlates and interindividual differences. *Cogn. Affect. Behav. Neurosci.*, *13*, 80–93.

Schirmer, A., Seow, C. S., & Penney, T. B. (2013b). Humans process dog and human facial affect in similar ways. *PLoS One*, *8*, e74591. doi: 10.1371/journal.pone.0074591.

Smith, S. D., McIver, T. A., Di Nella, M. S. J., & Crease, M. L. (2011). The effects of valence and arousal on the emotional modulation of time perception: evidence for multiple stages of processing. *Emotion, 11*, 1305–1313.

Spreckelmeyer, K. N., Krach, S., Kohls, G., Rademacher, L., Irmak, A., Konrad, K., Kircher, T., & Gründer, G. (2009). Anticipation of monetary and social reward differently activates mesolimbic brain structures in men and women. *Soc. Cogn. Affect. Neurosci.*, *4*, 158–165.

Taatgen, N. A., Van Rijn, H., & Anderson, J. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychol. Rev.*, *114*, 577–598.

Tecce, J. J. (1972). Contingent negative variation (CNV) and psychological processes in man. *Psychol. Bull.*, *77*, 73–108.

Tipples, J., Brattan, V., & Johnston, P. (2013). Facial emotion modulates the neural mechanisms responsible for short interval time perception. *Brain Topogr.*, *28*, 104–112.

Van den Brink, D., Van Berkum, J. J. ., Bastiaansen, M. C, Tesink, C. M., Kos, M., Buitelaar, J. K., & Hagoort, P. (2010). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Soc. Cogn. Affect. Neurosci.*, *7*, 173–183.

Van Rijn, H., Kononowicz, T. W., Meck, W. H., Ng, K. K., & Penney, T. B. (2011). Contingent negative variation and its relation to time estimation: A theoretical evaluation. *Front. Integr. Neurosci.*, *5*, 91. doi: 10.3389/fnint.2011.00091.

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen*. Tubingen, Germany: H. Laupp.

Voyer, D., & Reuangrith, E. (2015). Perceptual asymmetries in a time estimation task with emotional sounds. *Laterality*, *20*, 211–231.

Wackermann, J., Meissner, K., Tankersley, D., & Wittmann, M. (2014). Effects of emotional valence and arousal on acoustic duration reproduction assessed via the "dual klepsydra model". *Front. Neurorobot.*, *8*, 11. doi: 10.3389/fnbot.2014.00011

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation : An electric sign of sensori-motor association and expectancy in the human brain. *Nature*, *203*, 380–384.

Wearden, J., Philpott, K., & Win, T. (1999). Speeding up and (…relatively…) slowing down an internal clock in humans. *Behav. Process.*, *46*, 63–73.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr. Clin. Neurophysiol.*, *79*, 170–191.

Zélanti, P. S., & Droit-Volet, S. (2011). Cognitive abilities explaining age-related changes in time perception of short and long durations. *J. Exp. Child Psychol.*, *109*, 143–157.