Vocal emotions influence verbal memory: Neural correlates and interindividual differences

Annett Schirmer · Ce-Belle Chen · April Ching · Ling Tan · Ryan Y. Hong

Abstract Past research has identified an event-related potential (ERP) marker for vocal emotional encoding and has highlighted vocal-processing differences between male and female listeners. We further investigated this ERP vocal-encoding effect in order to determine whether it predicts voice-related changes in listeners' memory for verbal interaction content. Additionally, we explored whether sex differences in vocal processing would affect such changes. To these ends, we presented participants with a series of neutral words spoken with a neutral or a sad voice. The participants subsequently encountered these words, together with new words, in a visual word recognition test. In addition to making old/new decisions, the participants rated the emotional valence of each test word. During the encoding of spoken words, sad voices elicited a greater P200 in the ERP than did neutral voices. While the P200 effect was unrelated to a subsequent recognition advantage for test words previously heard with a neutral as compared to a sad voice, the P200 did significantly predict differences between these words in a concurrent late positive ERP component. Additionally, the P200 effect predicted voice-related changes in word valence. As compared to words studied with a neutral voice, words studied with a sad voice were rated more negatively, and this rating difference was larger, the larger the P200 encoding effect was. While some of these results were comparable in male and female participants, the latter group showed a stronger P200 encoding effect and qualitatively different ERP responses during word retrieval. Estrogen measurements suggested the possibility that these sex differences have a genetic basis.

Introduction

In anger, sadness, exhilaration, or fear, speech takes on an urgency that is lacking from its normal, even-tempered form. It becomes louder or softer, more hurried or delayed, more melodic, erratic, or monotonous. But, irrespective of the type of change, the emotions that break forth acoustically make speech highly salient. Thus, a manager who scolds her personnel, a child who cries for his mother, or an orator pleading with a shaking voice readily capture a listener's attention. How do emotional voices achieve such influence, and is this influence sustained beyond the immediate present?

The first part of this question has been addressed extensively. Using online measures such as scalp-recorded event-related potentials (ERPs) or functional magnetic resonance imaging (fMRI), researchers have identified auditory processing mechanisms that derive emotional meaning from voices quickly and seemingly automatically. Evidence for fast vocal emotional processing has come from Paulmann and others (Paulmann & Kotz, 2008; Paulmann, Seifert & Kotz 2010; Sauter & Eimer, 2010), who compared the processing of emotionally and neutrally spoken sentences that were attended and task-relevant. The researchers found that, relative to neutral speech, emotional speech increased the P200—a centrally distributed positive component that reaches its maximum around 200 ms following stimulus onset. They showed this effect for a wide range of vocal

A. Schirmer · C.-B. Chen · A. Ching · L. Tan · R. Y. Hong
Department of Psychology, Faculty of Arts and Social Sciences,
National University of Singapore, 9 Arts Link,
Block AS4, Level 2,
Kent Ridge 117570, Singapore

A. Schirmer (✉)
Duke/NUS Graduate Medical School, Kent Ridge, Singapore
e-mail: schirmer@nus.edu.sg

emotions, including anger, fear, disgust, happiness, pleasant surprise, and sadness, and they suggested that it reflects early detection of emotional salience (Paulmann & Kotz, 2008; Paulmann et al., 2010).

Other researchers have explored the automaticity of processing emotional voices. For example, Grandjean et al. (2005) conducted an fMRI study that used a dichotic listening task during which participants heard neutral and angry pseudospeech. The participants were instructed to discriminate male from female voices in one ear and to ignore the voices presented to the other ear. Notably, activity of the secondary auditory cortex along the right superior temporal sulcus (STS) was enhanced for angry as compared to neutral pseudospeech, regardless of whether the speech was delivered to the task-relevant or -irrelevant ear.

Schirmer and colleagues (Schirmer, Escoffier, Zysset, Koester, Striano and Friederici 2008b; Schirmer, Simpson and Escoffier 2007; Schirmer, Striano and Friederici 2005) went a step further, to probe differences between the processing of completely unattended emotional and neutral voices. In their experiments, participants engaged in watching a silent subtitled movie while emotionally and neutrally intoned pseudospeech played in the background. In one condition, the speech was primarily neutral, with only a few emotionally spoken deviants. In another condition, the speech was primarily emotional, with only a few neutrally spoken deviants. Using fMRI, Schirmer and colleagues found that, relative to neutral deviants, emotional deviants activated the STS and emotionally relevant structures such as the amygdala and orbitofrontal cortex (Schirmer, Escoffier, Zysset, Koester, Striano and Friederici 2008b). Moreover, in line with the research by Paulmann and colleagues (e.g., Paulmann et al., 2010), they observed emotion effects in the ERP around 200 ms following stimulus onset (Schirmer, Simpson and Escoffier 2007; Schirmer, Striano & Friederici 2005).

Taken together, the above-mentioned findings and related work (de Gelder, Pourtois, & Weiskrantz, 2002; Ishii, Kobayashi, & Kitayama, 2010; Sauter & Eimer, 2010; Wambacq, Shea-Miller & Abubakr 2004) have demonstrated that emotional voices influence listeners by boosting early auditory, and possibly more general (e.g., emotional), processing, regardless of the listeners' concurrent attentional state. However, the extant work is relatively silent on the question of whether emotional voices affect listeners beyond the immediate present. Moreover, while there is ample evidence for the long-term effects of other modes of emotional stimulation, including verbal content (Leclerc & Kensinger, 2011; Schmidt, 2012; Werner, Peres, Duschek & Schandry 2010), faces (D'Argembeau & Van der Linden, 2011; Davis et al., 2011; Righi et al., 2012), and images (Croucher, Calder, Ramponi, Barnard & Murphy 2011; Kaestner & Polich, 2011), as yet little evidence has emerged for long-term effects of the voice.

To the best of our knowledge, only a couple of behavioral studies have tackled this issue. The first study attempted to clarify the role of vocal emotions for implicit verbal memory (Kitayama, 1996). In that study, participants were given a numeric memory span task while listening to distractor sentences of positive or negative verbal content that were spoken in an emotional or a neutral voice. In a subsequent surprise free recall test, the emotionally spoken sentences were remembered better than the neutrally spoken sentences. However, this was found only when the memory span task was demanding. When the memory span task was easy, neutrally spoken sentences were remembered better than emotionally spoken ones.

A second study (Schirmer, 2010) explored the effect of emotional voices on explicit verbal memory as well as the affective connotations of words in semantic memory. Participants were presented with sadly, happily, and neutrally spoken words of neutral content (e.g., the word "table" spoken with a sad voice). The participants were then asked to remember these words for a subsequent visual memory test. In that test, previously studied words were shown together with new words, and participants indicated for each word whether it was old or new. Additionally, they rated the emotional valences of the old and new test words. While the participants' recognition memory was comparable for words previously studied with an emotional or a neutral voice, the responses demonstrated voice-related changes in word valence: Words previously heard with a sad voice were rated more negatively than words previously heard with a neutral voice, and likewise, words previously heard with a happy voice were rated more positively than words previously heard with a neutral voice. Importantly, these effects could be dissociated from voice memory; a person's ability to remember the emotional intonation of the study words failed to predict voice-related changes in word valence.

In sum, research exploring voice effects beyond the immediate present has found evidence for a modulation of verbal memory and word valence. However, the number of studies available is small, and their findings are somewhat contradictory. Moreover, these studies have provided no insights into the biological underpinnings of sustained voice effects. We sought to address these issues and to further explore the influence of vocal emotions on listener memory and attitudes. Following up on existing work (Schirmer, 2010), we employed an old/new word recognition paradigm, in which participants listened to sadly and neutrally spoken study words and later saw these words in a visual test examining word recognition and emotional valence. The following modifications and additional measures were introduced as means to meaningfully extend the current evidence.

First, we modified the interval between the spoken words in the study phase. There is now substantial evidence that

emotions affect memory by modulating stimulus encoding and consolidation (Phelps, 2004). The former mechanism is attention-based, and thus operates relatively quickly. One may speculate that it plays a relatively minor role in an explicit memory task in which both emotional and neutral items are task-relevant. The more important mechanism should be the latter one, which depends on peripheral changes in arousal that feed back to the brain systems involved in stabilizing long-term memory traces (Phelps, 2004). As emotion-induced peripheral changes are relatively sluggish, a quick succession of emotionally and neutrally spoken words, which is typical of many word-based memory studies (Schirmer, 2010), may thus blur emotion effects. We addressed this issue by introducing a temporal delay of 12–15 s between study words, to allow emotion-induced peripheral changes (e.g., heart rate) to return to baseline.

A second modification concerned the assessment of voice-related changes in listener attitudes or semantic memory. Previously, this has been assessed together with word recognition in an old/new test block that followed a study block. Here we sought to replicate this effect and to explore whether it would also show after a somewhat longer delay and after listeners had encountered the words again in a neutral context. To this end, half of the participants in the present study performed a word valence rating task in the old/new test block. Here, as in Schirmer (2010), they performed old/new decisions and word valence judgments together for a given test word. The other half of the participants performed the old/new decisions and word valence judgments in separate blocks; after they had completed old/new decisions for all of the test words, they saw the test words again, and only then rated their valences.

In addition to these modifications, we included one more dependent and one more independent variable. Our additional dependent variable was the ERP, which we recorded during both study and test. Of particular interest were previously identified ERP markers of emotional processing. As we reviewed above, attended emotional vocalizations elicited a greater P200 relative to attended neutral vocalizations (Paulmann & Kotz, 2008; Sauter & Eimer, 2010; Wambacq et al., 2004). We asked whether this effect would predict subsequent differences in verbal memory and word valence. Additionally, we were interested in voice-related ERP differences during the old/new word recognition test. One of the most robust findings within the ERP literature on emotions has been a modulation of a late positive potential (LPP), which peaks around 300 ms or later following stimulus onset, with frontal and parietal subcomponents (Fischler & Bradley, 2006; Olofsson, Nordin, Sequeira & Polich 2008; Schirmer, Teh, Wang, Vijayakumar, Ching, Nithianantham and Cheok 2011b). LPPs have been reported in a number of tasks. LPP latencies are more variable than those of earlier potentials such as the P200, and their peaks extend over a longer duration. Moreover, LPPs are modulated by a range of factors,

including stimulus utility (Johnson & Donchin, 1978), syntactic congruity (Osterhout, McKinnon, Bersick & Corey 1996), similarity to a preceding stimulus (Schirmer, Soh, Penney & Wyse 2011a), familiarity or memory (Johansson, Mecklinger & Treese 2004; Wilding & Rugg, 1996), and of course, emotion. Of primary relevance for the present study were previous reports that the LPP is greater for stimuli with emotional as compared to neutral significance (Fischler & Bradley, 2006; Olofsson et al., 2008; Schirmer, Teh, Wang, Vijayakumar, Ching, Nithianantham and Cheok 2011b). Thus, we reasoned that voice-induced changes in word valence should be reflected by the LPP during visual word processing. Moreover, we expected words previously heard with a sad voice to elicit a larger LPP than words previously heard with a neutral voice.

Finally, we included salivary estrogen as an independent variable in the present study. Although basic emotional processing may be largely comparable across individuals, there are, nevertheless, significant individual differences (Hamann, 2004; Whittle, Yücel, Yap & Allen 2011). One of the factors that seems to contribute to these differences is biological sex or gender. Important for the present research are findings that have suggested a greater sensitivity to vocal emotions in women than in men (Schirmer, Kotz & Friederici 2002; Schirmer et al., 2005; Schirmer, Zysset, Kotz & von Cramon 2004). Moreover, particularly when vocal information is task-irrelevant, women are more likely than men to process and integrate vocal emotions into ongoing mental activities (e.g., language processing). Here, we measured estrogen, the major female sex hormone, to explore its potential role in this effect.

In sum, we set out to investigate the sustained effects of vocal emotional expression on listeners' verbal memory for and affective valuations of interactional content. Moreover, we sought to determine whether such sustained effects could be linked to a prominent ERP marker of vocal emotional encoding, the P200. To this end, we examined P200 differences between sadly and neutrally spoken words. Furthermore, we probed the relationship between these differences and behavioral as well as ERP responses to visual test words that had previously been heard in a sad or a neutral voice. Finally, we wished to characterize potential differences between male and female listeners in vocal emotional processing and to link these differences to the female sex hormone estrogen.

Method

Participants

We recruited 105 participants for this study. Nine of the participants were excluded from the data analysis due to

technical errors during data collection or excessive movement artifacts in the electroencephalography (EEG). Of the remaining participants, 48 were female (mean age = 20.9, SD = 1.4) and 48 were male (mean age = 22.7, SD = 1.2). All of the participants had normal or corrected-to-normal vision and reported normal hearing. They signed informed consent prior to participation in this study.

Materials and design

The word materials for this study were taken from a stimulus set established by Schirmer (2010). It comprised 240 nouns selected from a larger pool of 500 words that referred to relatively neutral everyday stimuli (e.g., "table"). These words were rated by 30 independent rates (15 female, 15 male) on two 5-point scales, one ranging from −2 (very negative) to +2 (very positive) for word valence, and one ranging from 0 (nonarousing) to 4 (highly arousing) for arousal. The 240 selected words had a mean valence of 0.16 (SD = 0.2), a mean arousal of 0.58 (SD = 0.24), and a mean word frequency of 57.2 (SD = 76.5; Kučera–Francis written frequency) as determined by the MRC Psycholinguistic Database (Coltheart, 1981).

Four individuals with drama experience were asked to produce the selected words with an emotional (i.e., angry, sad, or happy) or neutral intonation. Their productions were recorded and digitized at a 16-bit, 44.1-kHz sampling rate. The word amplitudes were normalized at the root-mean square value using Adobe Audition 2.0. A subset of 15 words from each intonation condition and each speaker were subjected to a voice rating study in which 30 raters participated (15 female, 15 male). In this study, the spoken words were presented in random order and rated on a vocal valence scale ranging from −2 (very negative) to +2 (very positive) and on an arousal scale from 0 (not aroused) to 4 (very aroused). Additionally, the participants indicated, for each word, whether its intonation conveyed an angry, sad, happy, or neutral state.

The spoken words selected for the present study comprised 240 neutral words spoken with a sad and a neutral intonation by a female speaker whose speech samples for these intonation conditions were perceived more accurately than the samples of the other speakers. The words were used in an old/new word recognition paradigm, such that every word appeared equally often in each of the experimental conditions (sad/neutral study words and new test words). This paradigm consisted of two study phases comprising 60 words each and two test phases comprising 120 words each. The phases were presented in interleaved fashion (i.e., study–test–study–test).

A study trial began with a white fixation cross that occurred in the center of the computer screen in front of the participant. After 200 ms, a word was presented over headphones. The neutrally spoken words were on average 777.6 ms long (SD = 149), while sadly spoken words were on average 1,132.4 ms long (SD = 245.5). After sound onset, the fixation cross continued for another 2,300 ms. The next trial started after 12, 13, 14, or 15 s, for a quarter of the trials each.

A test trial began again with a 200-ms fixation cross, which was replaced by a word that remained on the screen for 1,000 ms. After the word disappeared, the participants saw the prompt "Old or New?" and had to press the left or the right button of a button response box in order to indicate their choice. Assignment of the left and right buttons to old and new test words was counterbalanced across participants. For half of the participants (i.e., the immediate-rating group), the old/new response brought up a valence scale ranging from −2 (very negative) to +2 (very positive) on the screen in front of them. They now had to indicate the valence of the test word by moving a cursor that was centered at zero to the appropriate point on the scale. For the remaining participants (i.e., the delayed-rating group), the old/new response ended the test trial. The latter group rated the valences of the test words in a similar fashion at the end of the experiment.

Participants could take short breaks after completing a phase in the experiment. They self-initiated the onset of the next phase.

Procedure

Participants were tested individually. After a participant arrived at the lab, he or she was given a written document describing the experimental procedures and asking for a signature if he or she wished to continue with the experiment. After the signature was obtained, participants moved into the experimental room, where they sat down in a comfortable chair and faced a computer monitor at a distance of 60 cm. Each participant then produced the first of three saliva samples by drooling passively through a straw into a plastic test tube. Subsequently, he or she completed a short questionnaire capturing his or her personal information and provided a second saliva sample.

Prior to the start of the experiment, each participant was prepared for the electrophysiological recordings. To this end, a 64-channel EEG cap with empty electrode holders was placed on the participant's head. The electrode holders were then filled with an electrolyte gel and the respective electrodes according to the modified 10–20 system. The individual electrodes were attached above and below the right eye and at the outer canthus of each eye to measure eye movements. Two electrodes were attached on the left and right forearms for heart rate analysis (not reported here), and one electrode was attached at the nose tip for data referencing. The data were recorded at 256 Hz with an

ActiveTwo system from Biosemi (Amsterdam, The Netherlands), which uses a common mode sense active electrode for initial referencing. Only an antialiasing filter was applied during data acquisition (i.e., sync filter with a half-power cutoff at 1/5 the sampling rate).

Following the electrophysiological setup, the participants were instructed about the experimental task. They were told that they would hear a series of spoken words presented over headphones. The task was to listen to these words while fixating the eyes on a fixation cross. The participants were informed that each series of spoken words would be followed by a test phase comprising visually presented old and new words. In this phase, participants had to make old/new decisions and respond by pushing one of two buttons on a response box. After completing two study/test blocks, participants were relieved from the electrophysiological set-up and allowed to wash their face and hair (~5–10 min), and then they produced their last saliva sample. If a participant was assigned to the immediate-rating group, the participant was then dismissed. If assigned to the delayed rating group, the participant sat down again in front of the computer, where he or she now saw all test words a second time and rated their valences, as the participants in the immediate-rating group had done. Participants in the delayed-rating group were not informed about the types of words that would appear on the screen (i.e., that these words comprised old and new words) but were told that their memory of the previous task was no longer relevant.

The two study phases of the experiment lasted about 16 min each. The test phase that followed each study phase was about 10 min long for individuals in the immediate-rating group. The test phases were about 5 min long for individuals in the delayed-rating group. These individuals began the valence rating approximately 20 min after completing the last study phase.

Data analysis

Estrogen Equal-volume aliquots were taken from each of the three samples collected during the experiment. These aliquots were pooled and stored at −80 °C until the data collection for all participants was completed. For analysis, the frozen saliva samples were brought to room temperature, vortexed, and centrifuged at 1,500 × g for 15 min in order to precipitate mucins. The supernatant from each sample was collected and analyzed in triplicate using a high-sensitivity salivary 17β-estradiol enzyme immunoassay kit (Salimetrics, Newmarket, Suffolk, U.K.) according to the manufacturer's instructions. A standard curve was generated by fitting a serial dilution of estradiol standard to a four-parameter sigmoid minus curve using the Softmax Pro software (Molecular Devices, Sunnyvale, CA). The pH level was assessed for each sample and found to be within assay criteria in all cases.

Reference samples with known estradiol levels provided in the kit were analyzed in parallel with actual samples, and the measured values were within the expected range, indicating the accuracy of the results. All measurements were read using the SpectraMax 190 plate reader (Molecular Devices).

Event-related potentials EEG/electrooculography data were processed with EEGLAB (Delorme & Makeig, 2004). The recordings were re-referenced to the nose, and a 0.1- to 20-Hz bandpass filter was applied. At these frequency cutoffs, the voltage gain was reduced by half. The transition bands were 3 Hz for the low-pass filter (rolloff ~100 dB/octave; filter order: 36 pts) and 0.01 Hz for the high-pass filter (rolloff ~100 dB/octave; filter order: 7,680 pts). The continuous data were visually scanned for nontypical artifacts caused by drifts or muscle movements, and time points containing such artifacts were removed. Infomax, an independent-component analysis algorithm implemented in EEGLAB, was applied to the remaining data, and components reflecting typical artifacts (i.e., horizontal and vertical eye movements and eye blinks) were removed. The back-projected single trials were epoched and baseline-corrected using a 450-ms prestimulus baseline and a 1,000-ms time window starting from stimulus onset. The resulting epochs were again screened visually for residual artifacts. Only the ERPs to correctly recognized old words in both study and test were analyzed. These ERPs were derived by averaging individual epochs for each condition and participant. For the study phase, an average of 41 trials ($SD = 8$) per condition entered the statistical analysis. For the test phase, averages of 103 trials ($SD = 13$) for the new condition and 49 trials ($SDs = 8$) each for the sad and neutral old conditions were entered in the statistical analysis. Separate analyses of variance (ANOVAs) for the study and test phases, with trial number as the dependent variable and Voice (sad/neutral), Sex (male/female), and Valence Rating (immediate/delayed) as independent variables, were nonsignificant ($ps > .1$).

We used a series of ANOVA models to test the hypotheses of this study. These models included the factors Region and Hemisphere, comprising 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 would contain equal numbers of electrodes while providing a broad scalp coverage that would enable the assessment of topographical effects. The voltages of electrodes within these regions were averaged and subjected to statistical analysis.

Given our interest in understanding the processing of vocal expressions, we focused this report on main effects and interactions involving the Voice factor.

Results

Estrogen measures

As expected, a one-way ANOVA indicated that estrogen levels were higher in female than in male participants $[F(1, 94) = 5.5, p = .021]$.

Behavioral measures

The behavioral results are illustrated in Fig. 1. Word recognition memory was assessed by computing a d' score. To this end, the normalized probability of false alarms was subtracted from the normalized probability of hits. The resulting value was subjected to an ANOVA, with Voice (sad/neutral) as a repeated measures factor and Sex (male/female) and Valence Rating (immediate/delayed) as between-subjects factors. This analysis revealed a main effect of Voice, indicating that recognition memory was better for words previously heard in the neutral as compared to the sad condition $[F(1, 92) = 6.48, p = .013]$. All of the other effects were nonsignificant ($ps > .1$).

Voice-related changes in word valence were assessed by subjecting the valence rating scores to an ANOVA, with Voice (sad/neutral) as a repeated measures factor and Sex (male/female) and Valence Rating (immediate/delayed) as between-subjects factors. Again, we found a main effect of Voice $[F(1, 92) = 5.41, p = .022]$: Words were rated more negatively if they had previously been studied with a sad as compared to a neutral voice. All other effects were nonsignificant ($ps > .1$).

Event-related potentials

Study phase The ERP results from the study phase are illustrated in Fig. 2. Visual inspection suggested potential changes prior to the onset of the spoken words. These potential changes were likely evoked by the fixation cross, which appeared 200 ms before the spoken words. Following word onset, a sequence of positive and negative deflections occurred. Clear condition differences emerged for the P200, which was greater for sad than for neutral prosody. This effect appeared to be larger in female than in male participants.

To probe these visual impressions, we conducted a series of statistical analyses. Specifically, we first determined the P200 peak latency by selecting the most positive deflection between 150 and 250 ms following word onset and by averaging the time points of this deflection across all channels, conditions, and participants included in the statistical analysis. This revealed an average P200 peak at 206 ms ($SD = 28$ ms) following word onset. To determine whether condition differences would emerge prior to the P200, we used an identical approach to explore the voltage peaks for an early positivity between 0 and 100 ms and an early negativity between 50 and 150 ms. These two components were found to peak at 45 ms ($SD = 31$ ms) and 109 ms ($SD = 31$ ms), respectively.

The Mean voltages in a 10-ms time window centered around the peak of the P200 were subjected to separate ANOVAs, with Voice (sad/neutral), Hemisphere (left/right), and Region (anterior/central/posterior) as repeated measures factors, and Sex (male/female) and Valence Rating (immediate/delayed) as between-subjects factors. This analysis revealed a significant main effect of Voice $[F(1, 92) = 8.65, p = .004]$ and significant interactions of Voice with Region $[F(2, 184) = 5.76, p = .003]$ and with Sex $[F(1, 92) = 3.93, p = .05]$. Follow-up analyses for each level of Region revealed a trend toward an effect of Voice over anterior sites $[F(1, 92) = 3.1, p = .083]$ and a significant Voice main effect over central sites $[F(1, 92) = 4.9, p = .03]$. Across these sites, sadly spoken words elicited a larger P200 than did neutrally spoken words. The Voice effect was nonsignificant over posterior sites ($p > .1$). Follow-

Fig. 1 Bar graph of behavioral measures. The left diagram presents mean d' scores for the emotional and neutral voice conditions in male and female participants separately. The right diagram presents mean valence ratings for the emotional and neutral voice conditions in male and female participants separately. Error bars indicate the standard errors of the means
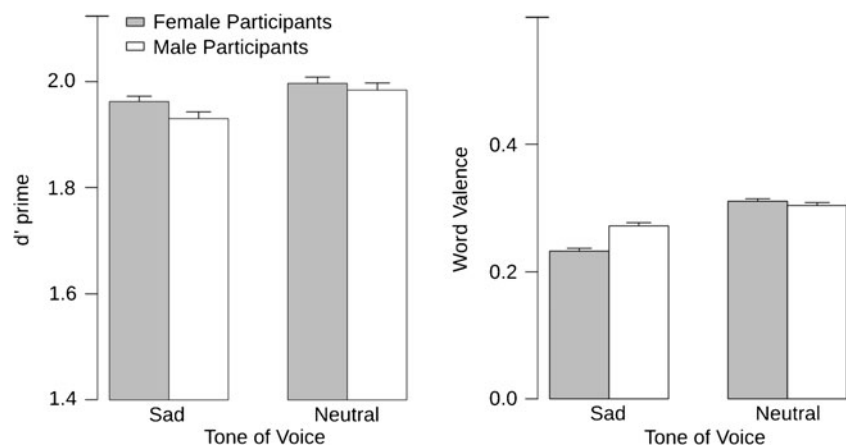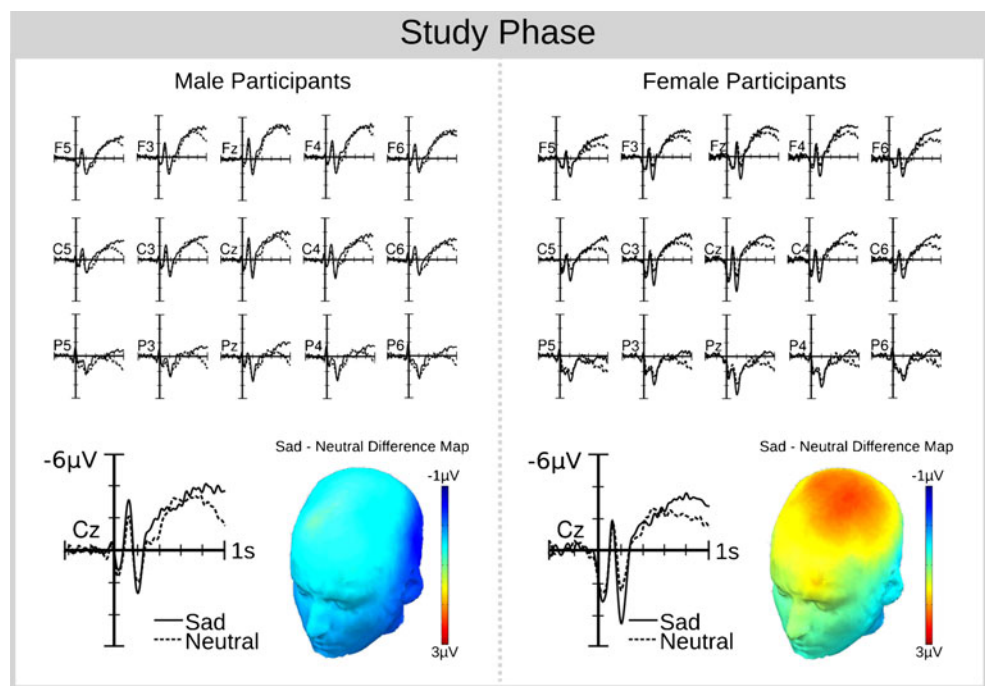
Fig. 2 ERP traces in the study phase. Study phase ERPs are presented for the sad (solid lines) and neutral (dashed lines) conditions. Scalp maps show the differences between these two conditions within the time window used for the statistical analysis (i.e., 201–211 ms)

up analyses for each level of Sex indicated that the Voice effect was significant in female [$F(1, 46) = 18.1$, $p = .0001$] but not in male ($p > .3$) participants. Specifically, only female participants showed a larger P200 for words spoken with a sad as compared to a neutral voice across all regions.

Analysis of the two components preceding the P200 revealed significant main effects of Sex, indicating that the early positivity was greater in women than in men [$F(1, 96) = 4.8$, $p = .031$] and that the early negativity was greater in men than in women [$F(1, 96) = 4.2$, $p = .043$]. Importantly, all effects involving the factor Voice were nonsignificant (ps > .1).

Test phase ERPs for the test phase are illustrated in Fig. 3. Visual inspection of these ERPs revealed an LPP between 400 and 700 ms following word onset that was greater for old than for new words. LPP differences between old words studied with sad and neutral prosody were relatively small and seemed to go in opposite directions for female and male participants. In female participants, the LPP seemed smaller, while in male participants, it seemed greater for the sad as compared to the neutral condition.

To statistically test these visual impressions, we determined the peak latency of the LPP by searching for the voltage maximum between 400 and 700 ms following word onset. The time points associated with the identified voltage maximum were averaged across all conditions, electrodes, and participants included in the statistical analysis. This procedure revealed an average LPP peak at 595 ms (*SD* = 89 ms) following stimulus onset. The mean voltages in a 100-ms time window centered around the identified peak were subjected to an ANOVA with Word (sad/neutral/new),

Hemisphere (left/right), and Region (anterior/central/posterior) as repeated measures factors, and Sex (male/female) and Valence Rating (immediate/delayed) as between-subjects factors. This analysis produced a significant effect of Word [$F(2, 184) = 57.2$, $p < .0001$] and interactions of Word with Region [$F(4, 368) = 29.7$, $p < .0001$] and with Sex [$F(2, 184) = 3.3$, $p = .039$].

A follow-up analysis of the Word effect for each Region indicated that sad and neutral old words elicited greater positivities than did new words over anterior [Fs(1, 92) = 17.2 and 23.2, ps < .0001], central [Fs(1, 92) = 101.1 and 102, ps < .0001], and posterior [Fs(1, 92) = 89.9 and 75.5, ps < .0001] regions. There were no differences between sad and neutral old words (ps > .5).

A follow-up analysis of the Word effect in males was significant [$F(2, 92) = 38.1$, $p < .0001$]: Sad and neutral old words elicited greater LPPs than did new words [Fs(1, 46) = 67.4 and 36.1, ps < .0001], but the amplitudes of sad and neutral words did not differ significantly [$F(1, 46) = 2.5$, $p = .12$]. The Word effect was also significant in female participants [$F(2, 92) = 23.4$, $p < .0001$]. As in male participants, sad and neutral old words elicited greater LPPs than did new words [Fs(1, 46) = 19.2 and 39.1, ps < .0001]. Additionally, words previously heard with a sad voice elicited a smaller LPP than did words previously heard with a neutral voice [$F(1, 46) = 5.2$, $p = .027$].

Correlational analyses

We conducted a series of two-tailed Pearson correlation analyses to explore the relationship between the voice encoding

Fig. 3 ERP traces in the test phase. Test phase ERPs are presented for old words from the sad (black solid lines) and neutral (dashed lines) study conditions, as well as for new words (gray solid lines). Scalp maps show the differences between the sad and neutral conditions within the time window used for the statistical analysis (i.e., 545–645 ms)
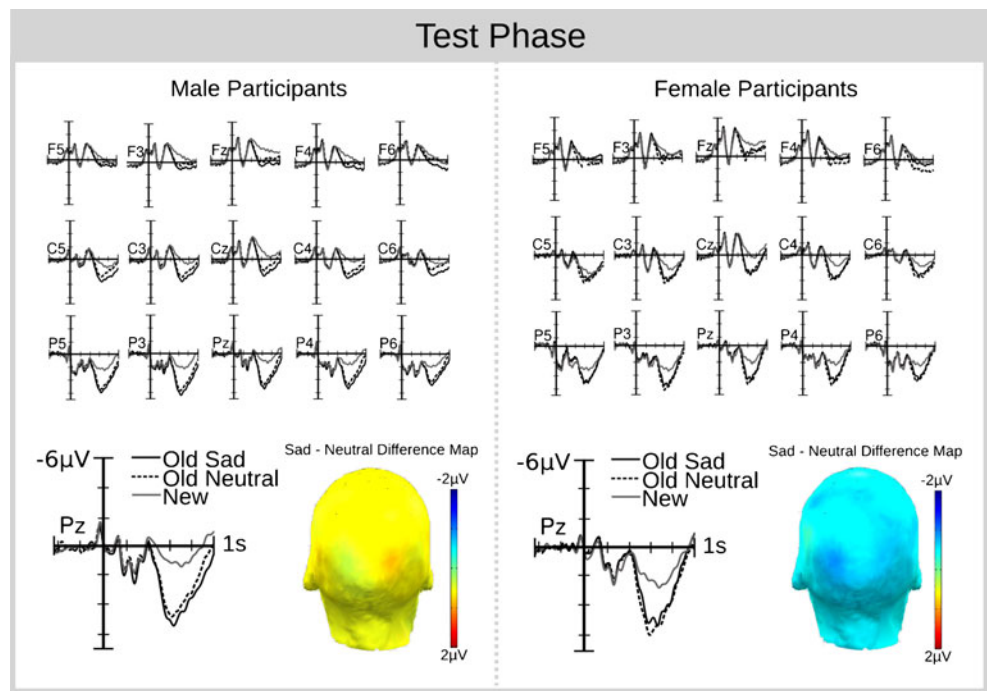
and verbal retrieval effects, as well as a potential role of estrogen in moderating these effects. The results of these analyses are illustrated in Fig. 4.

The first set of analyses assessed the relationship between the P200 encoding effect and behavioral responses to words previously heard with a sad or neutral intonation. We quantified the P200 effect by subtracting the mean voltages in the P200 time window in the neutral condition from those in the sad condition for each channel included in the ERP analysis. We then computed a mean difference score by averaging across these channels for each participant. A correlation of this score and the d' difference between sadly and neutrally spoken words was not significant ($r = .07$, $p = .44$). However, we did find a significant correlation between the P200 effect and the valence rating effect ($r = .27$, $p = .008$). A greater P200 effect during encoding predicted a greater voice-congruent shift in word valence during retrieval. Separate analyses for each sex indicated that this relationship was significant in women ($r = .44$, $p = .003$) and nonsignificant in men ($r = .15$, $p = .31$).

In a second step, we explored the relationship between the P200 encoding effect and the LPP retrieval effect. As with the P200 effect, we quantified the LPP effect as the mean voltage difference between old words previously heard with a sad and a neutral voice. A correlation analysis collapsed across men and women was nonsignificant ($r = -.15$, $p = .14$). Given that the LPP effect was significant in females only, we conducted separate analyses for women and men. In women, the P200 effect correlated negatively with the LPP effect, indicating that greater sensitivity to vocal emotions during study was

associated with smaller LPPs to test words previously heard with sad as compared to neutral prosody ($r = -.29$, $p = .041$). In men, the P200 effect was not associated with the LPP effect ($r = .04$, $p = .78$).

Finally, we assessed the relationship between estrogen and vocal-processing effects. The correlation between estrogen and the P200 effect approached significance. Specifically, higher estrogen tended to be associated with greater P200 differences between sadly and neutrally spoken words ($r = .17$, $p = .097$). Given the sex differences observed for the P200 effect, we conducted separate analyses for female and male participants. These indicated that estrogen predicted the P200 effect in female ($r = .29$, $p = .044$) but not in male ($r = -.01$, $p > .94$) participants. The relationships between estrogen and voice-related retrieval effects (i.e., on d', valence ratings, and LPP) were nonsignificant (all $ps > .1$).

Discussion

The purpose of this study was to examine the encoding of verbal information as a function of tone of voice and to explore voice-related retrieval effects. In line with previous work, we observed increased P200 amplitudes to sadly as compared to neutrally spoken words during encoding. We also identified voice-related retrieval effects in the participants' behavioral responses and ERPs to test words. While the encoding and retrieval effects were related, the relationships were not strictly comparable between male and female participants. In the following
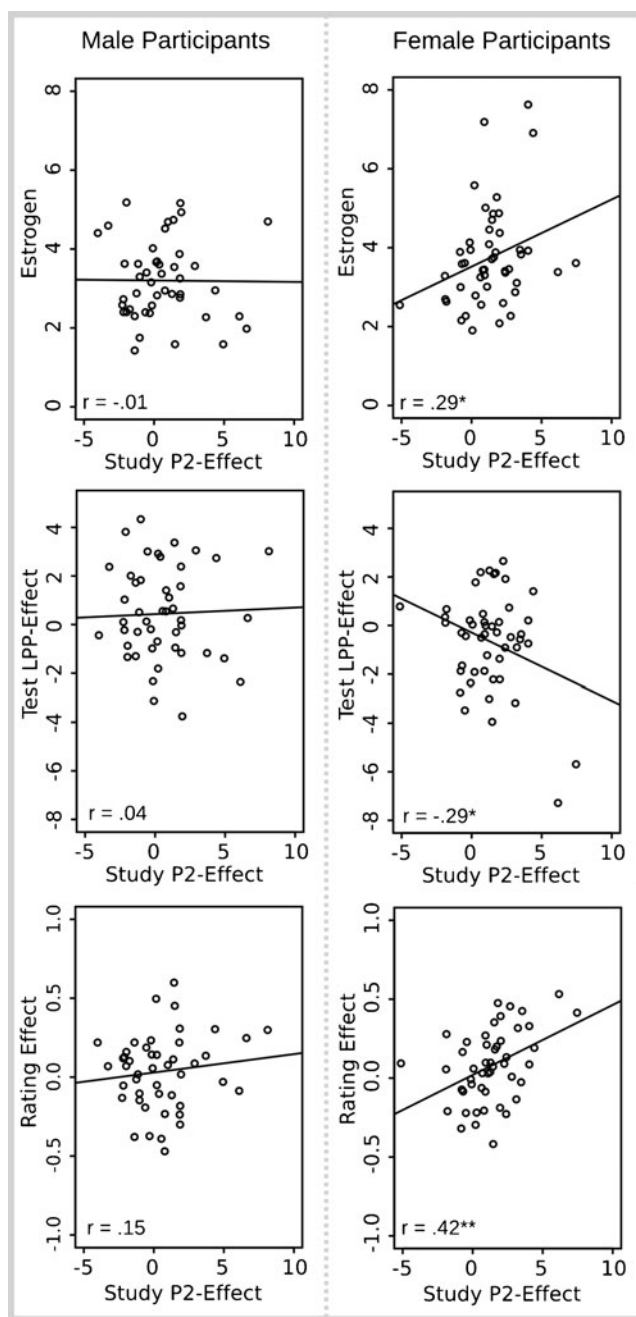
Fig. 4 Scatterplots of the relationship between the study P200 effect and other dependent measures. The illustrated relationships are significant in female but not male participants. The female participants show significant correlation between estrogen and the study P200 effect, as quantified by subtracting the voltage of the neutral condition from the voltage of the sad condition. Additionally, females show a correlation between the study P200 effect and the test late positive potential (LPP) effect, as well as between the study P200 effect and the prosody rating effect, as quantified by subtracting the rated valences of words studied with neutral prosody from the rated valences of words studied with sad prosody. $^*p<.05.$ $^{**}p<.01$

discussion, we will detail these effects and outline how they advance the existing literature on emotional encoding, retrieval, and individual differences.

Emotional encoding

Previous research on vocal encoding had revealed larger P200 amplitudes for emotionally than for neutrally spoken words, highlighting the P200 as a potential marker of emotional processing (Paulmann & Kotz, 2008; Paulmann et al., 2010; Sauter & Eimer, 2010; Wambacq et al., 2004). However, the susceptibility of this component to physical stimulus characteristics raises the possibility that voice-related P200 amplitude differences reflect sensory processing instead. Some have addressed this concern by examining a wide range of vocal emotions that are physically as distinct from each other as they are from a neutral control (Paulmann & Kotz, 2008). The similarity of effects across emotion conditions was taken as evidence against strictly sensory mediation. Nevertheless, sensory mediation could not be completely ruled out, due to the possibility that the various vocal emotions shared acoustic features that were absent in the neutral control. Another attempt to tackle this issue has been to compare emotional and neutral vocalizations with their spectrally rotated analogues (Sauter & Eimer, 2010). However, this approach is also not without problems. By inverting the low- and high-frequency contents, the sound as such is not left intact, as would be the case for inverting a face or other visual object. Rotated sounds are acoustically altered such that critical acoustical information linked to emotions is potentially attenuated or disrupted. Moreover, unlike inverted visual objects, which still maintain their general appearance, these altered waveforms no longer sound human. Thus, acoustic differences between an original and a spectrally rotated sound might account for the observed P200 effects.

Further reasons for concern are the inconsistent P200 effects for nonvocal stimuli, including pictures, faces, written words, and environmental sounds. Specifically, pictures have been shown to elicit an early posterior negativity (EPN) within the P200 time range (Schupp, Junghöfer, Weike & Hamm 2004). This negativity is enhanced for pictures with emotional as compared to neutral content, presumably reflecting enhanced visual processing. A similar EPN effect has also been observed for faces and written words (Herbert et al., 2008; Kissler, Herbert, Winkler & Junghöfer 2009). In contrast, reports of a P200 modulation for these stimuli are rare (Eimer & Holmes, 2007; Herbert, Kissler, Junghöfer, Peyk & Rockstroh 2006). Moreover, unlike the voice findings, visual P200 modulations do not show consistently across a range of emotions, and sometimes they may even entail an emotion-induced amplitude reduction (Sarlo & Munafò, 2010). Finally, research using environmental sounds produced by everyday events or objects has failed to reveal ERP amplitude differences within the P200 time range for neutral and intensity-matched unpleasant sounds (Thierry & Roberts, 2007). Together, these findings have challenged the notion that the P200 could reflect emotional salience.

That the P200 may nevertheless be a marker for emotional processing, perhaps specifically for vocal stimuli, can be inferred from the present study, which replicated the effect of vocal emotions on P200 amplitudes using stimuli and a procedure that differed from those in previous work. Thus, our results confirm the notion that the P200 voice effect occurs relatively independently of acoustic stimulus features and task demands. More importantly, however, the present study showed that P200 modulations during vocal emotional encoding predict subsequent valence ratings of purportedly neutral written words. Thus, it is highly probable that the emotional significance of the speaker's voice is reflected by the P200.

Emotional retrieval

It is widely believed that emotions facilitate the retention of information in memory. Autobiographical memory studies have shown that particularly salient events, such as the terrorist attacks directed at the United States in September 2001, produce more vivid and detailed memories than do mundane events (Sharot, Martorella, Delgado, & Phelps, 2007). Furthermore, research has indicated that emotional items presented to participants in a controlled lab setting also enhance memory. For example, stories, pictures, or video clips with emotional content are remembered better than comparable nonemotional materials (Cahill et al., 1996; Cahill, Prins, Weber & McGaugh 1994; Dolcos, LaBar & Cabeza 2005). Although a similar effect has been reported for verbal information (Kensinger & Corkin, 2003; Phelps, LaBar & Spencer 1997; Windmann & Kutas, 2001), that effect appears to be more variable or "fragile" (Schmidt, 2012), possibly due to the abstract nature of language and to its often casual use, which potentially diminish the associations between words and emotional reactions.

When considering the relationship between emotions and memory, it is important to recognize that an emotional episode typically comprises both emotional and nonemotional or neutral elements. For example, an angry face provides information about an individual's emotional state as well as about his or her identity (e.g., female, age). The sentence "The boy lost his mother" contains several neutral and perhaps positive words, and as a whole it gains negativity only from the negative word "lost." The question thus becomes whether the storage of nonemotional or neutral information benefits or suffers from an emotional context.

Evidence for benefits has come from studies that have examined face identity memory or word memory in the context of neutral and emotional facial expressions (D'Argembeau & Van der Linden, 2011; Davis et al., 2011; Righi et al., 2012) or sentential contexts (Brierley, Medford, Shaw & David 2007; Guillet & Arndt, 2009; Phelps et al., 1997; Schmidt, 2012). Some facial emotion expressions were found to facilitate person or word memory relative to other expressions. However, because different expressions were used and compared (not always including a neutral control), it is currently impossible to infer which emotions reliably enhance memory for to-be-remembered neutral information. Also, emotional sentence contexts were found to facilitate word memory. In the above-mentioned studies, neutral words were remembered better when they had previously been encountered in the context of an emotional as compared to a neutral sentence or word phrase. However, these findings could not always be replicated, and they seem to depend on the order (mixed vs. blocked) in which conditions are presented (Schmidt, 2012). Moreover, quite a number of studies using emotional faces, pictures, or sentences have found no modulation of to-be-remembered neutral content (Johansson et al., 2004; Maratos, Dolan, Morris, Henson & Rugg 2001; Righi et al., 2012; Schmidt, 2012; Wiswede, Rüsseler, Hasselbach & Münte 2006). Additionally, there is evidence that enhanced memory for emotional elements in a visual scene may come at the cost of impoverished memory for neutral elements (Kensinger, Piguet, Krendl & Corkin 2005; Waring, Payne, Schacter & Kensinger 2010).

Extending the latter findings to auditory scenes, the present study revealed poorer memory for neutral words that had previously been heard in an emotional as compared to a neutral voice. Possibly, the emotional voice drew attention to nonverbal aspects of the stimulus (e.g., voice identity), thus leaving fewer resources for the encoding and consolidation of verbal aspects. The present finding is also in line with a previous vocal study in which, in a low-attentional-load condition, poorer memory was found for distractor sentences spoken with an emotional as compared to a neutral voice (Kitayama, 1996). In the latter study, a reversed effect emerged in a high-attentional-load condition, in which attention was less free to dwell on salient but task-irrelevant stimulus features such as vocal expression. Thus, perhaps attentional capacity plays an important role in determining whether and how emotional elements in scenes affect memory for neutral elements.

Apart from investigating vocal influences on long-term memory in the form of recognition success, the present study also explored vocal influences on the affective connotations of words stored in semantic memory. Previous work has already demonstrated that negatively and positively spoken words are subsequently perceived as being more negative and positive, respectively, than neutrally spoken words (Schirmer, 2010). The present study replicated these results, despite extending the average interval between hearing the spoken words and judging their emotional valence. Additionally, we could link the results to a voice-encoding marker in the ERP. Individuals who responded with a greater P200 effect during encoding also showed a greater voice-related shift in word valence during word retrieval. Together, these results highlight vocal

emotions as an important factor in shaping semantic memory, possibly through the activation of voice-specific representations and cross-modal emotion systems (Binder & Desai, 2011).

Individual differences

Although male and female participants were largely comparable in the behavioral effects described above, we nevertheless found a range of differences in their ERPs. First, women and men differed significantly in the P200 effect. While this effect showed regardless of sex over central regions, it was larger and more broadly distributed in women than in men. This suggests that females were more sensitive to emotional variations in speaker tone of voice and were more likely to represent these variations in the brain. As such, these findings extend an existing body of evidence that suggests greater sensitivity to vocal information in female than in male listeners (Schirmer et al., 2002; Schirmer et al., 2005; van den Brink et al., 2012), particularly when this information is task-irrelevant.

In line with these voice-encoding differences, we also observed voice-related retrieval differences between men and women. Retrieval effects were assessed for the LPP, a late positive potential that has reliably been implicated in emotional processing (Fischler & Bradley, 2006; Olofsson et al., 2008; Schirmer, Teh, Wang, Vijayakumar, Ching, Nithianantham and Cheok 2011b). Typically, words, pictures, or other stimuli elicit a larger LPP when they are emotional rather than neutral. Contrary to these results, the present study revealed only a weak tendency toward this pattern in men, and an opposite effect in women. Specifically, women showed significantly smaller LPPs for words previously heard with sad rather than neutral prosody, and this LPP difference correlated negatively with the P200 effect in the study phase.

One possible explanation for these findings is that emotional processing overlapped with other mental operations in the LPP time window. Specifically, memory retrieval operations may have influenced and overwritten the emotional-processing effects in female listeners. It is well established that more familiar or better-remembered items elicit larger LPPs than do items that are less familiar or for which source memory is weaker (Johansson et al., 2004; Wilding & Rugg, 1996). Given that participants remembered words from the neutral condition better than words from the sad condition, females may have emphasized mnemonic over emotional processes during the LPP time frame.

Contrary to our predictions, we found no sex differences in the behavioral measures of this study. One possibility is that the sex differences in the ERPs reflect differential processing mechanisms that ultimately lead to the same behavioral outputs in men and women. Alternatively, however, our behavioral measures may have been less sensitive or accurate than the electrophysiological measures: The participants' behavioral responses in the recognition test were only binary, whereas the ERPs can reflect processing differences continuously. Moreover, memory decisions and rating judgments used to assess changes in word valence are cumulative—including early automatic and higher-order reflective processing. In contrast, ERPs represent mental processes as they unfold in time. Together, these differences may have contributed to smaller and nonsignificant effect sizes for the former than for the latter measure. Notably, the combination of the absence of individual differences in behavior and the presence of such differences in online neuroimaging measures is not uncommon (Schirmer & Kotz, 2003; Schirmer et al., 2004; van den Brink et al., 2012).

When faced with such gender differences, one naturally asks whether they reflect differences in upbringing or genetic differences between men and women. Although recent epigenetic advances have complicated the answers to this traditionally "either–or" question (Weaver et al., 2004; Zhang & Meaney, 2010), we nevertheless sought to uncover some genetic contributions by examining differences in the main female sex hormone, estrogen. While this hormone is produced by different bodily tissues (e.g., the adrenal gland) present in either sex, its main locus of synthesis is the ovaries, which develop in the absence of a Y-chromosome. We found that greater levels of this hormone in female as compared to male participants predicted sex differences in voice encoding. Moreover, women, but not men, showed a significantly positive relationship between estrogen and voice-related modulations of the P200. This finding is largely in line with previous work that has implicated estrogen in social and emotional processing (Gasbarri, Pompili, D'Onofrio, Abreu & Tavares 2008; Schirmer, Escoffier, Li, Li, Strafford-Wilson and Li 2008a; Zeidan et al., 2011). Evidence that links estrogen to the neuropeptide oxytocin (OT), a key player in the processing of socially relevant information such as faces or voices (Averbeck, 2010; Ebstein, Israel, Chew, Zhong & Knafo 2010), offers a potential mechanism underlying these effects. By stimulating the release of OT and the transcription of OT receptor genes, estrogen allows the OT system to become more active (Gabor, Phan, Clipperton-Allen, Kavaliers & Choleris 2012; Lischke et al., 2012). Perhaps due to differences in the significance of both estrogen and OT in male and female bodies, the relationship between estrogen and voice processing is stronger in women than in men (but see Schirmer, Escoffier, Li, Li, Strafford-Wilson and Li 2008a).

Conclusions

The present study extends existing work on voice processing and emotional memory. In it, we have shown that emotionally

Springer

spoken words elicit a greater P200 than do neutrally spoken words and that this encoding difference predicts responses to the same stimuli when they are subsequently encountered in a written form. A greater voice-related P200 effect is associated with greater ERP differences during word retrieval and greater voice-related changes in word valence. Thus, one can conclude that emotional voices, apart from capturing listener attention, produce changes in long-term memory. They affect the ease with which spoken words are later recognized and the emotions that are assigned to the words. Thus, voices, like other emotional signals, affect listeners beyond the immediate present. However, the details of whether and how these effects occur differ between individuals. One of the factors that seems to predict individual differences is estrogen, a sex hormone that regulates the activity of OT, a key player in human social behavior.

References

Averbeck, B. B. (2010). Oxytocin and the salience of social cues. *Proceedings of the National Academy of Sciences, 107,* 9033–9034. doi:10.1073/pnas.1004892107

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15,* 527–536. doi:10.1016/j.tics.2011.10.001

Brierley, B., Medford, N., Shaw, P., & David, A. S. (2007). Emotional memory for words: Separating content and context. *Cognition & Emotion, 21,* 495–521. doi:10.1080/02699930600684963

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., & McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences, 93,* 8016–8021.

Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). Beta-adrenergic activation and memory for emotional events. *Nature, 371,* 702–704. doi:10.1038/371702a0

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505. doi:10.1080/14640748108400805

Croucher, C. J., Calder, A. J., Ramponi, C., Barnard, P. J., & Murphy, F. C. (2011). Disgust enhances the recollection of negative emotional images. *PloS ONE, 6,* e26571. doi:10.1371/journal.pone.0026571

D'Argembeau, A., & Van der Linden, M. (2011). Influence of facial expression on memory for facial identity: Effects of visual features or emotional meaning? *Emotion, 11,* 199–202. doi:10.1037/a0022592

Davis, F. C., Somerville, L. H., Ruberry, E. J., Berry, A. B. L., Shin, L. M., & Whalen, P. J. (2011). A tale of two negatives: Differential memory modulation by threat-related facial expressions. *Emotion, 11,* 647–655. doi:10.1037/a0021625

de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences, 99,* 4121–4126. doi:10.1073/pnas.062018499

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21. doi:10.1016/j.jneumeth.2003.10.009

Dolcos, F., LaBar, K. S., & Cabeza, R. (2005). Remembering one year later: Role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. *Proceedings of the National Academy of Sciences, 102,* 2626–2631. doi:10.1073/pnas.0409848102

Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S., & Knafo, A. (2010). Genetics of human social behavior. *Neuron, 65,* 831–844. doi:10.1016/j.neuron.2010.02.020

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45,* 15–31. doi:10.1016/j.neuropsychologia.2006.04.022

Fischler, I., & Bradley, M. (2006). Event-related potential studies of language and emotion: Words, phrases, and task effects. *Progress in Brain Research, 156,* 185–203. doi:10.1016/S0079-6123(06)56009-1

Gabor, C. S., Phan, A., Clipperton-Allen, A. E., Kavaliers, M., & Choleris, E. (2012). Interplay of oxytocin, vasopressin, and sex hormones in the regulation of social recognition. *Behavioral Neuroscience, 126,* 97–109. doi:10.1037/a0026464

Gasbarri, A., Pompili, A., D'Onofrio, A., Abreu, C. T., & Tavares, M. C. H. (2008). Working memory for emotional facial expressions: Role of estrogen in humans and non-human primates. *Reviews in the Neurosciences, 19,* 129–148. doi:10.1515/REVNEURO.2008.19.2-3.129

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. *Nature Neuroscience, 8,* 145–146. doi:10.1038/nn1392

Guillet, R., & Arndt, J. (2009). Taboo words: The effect of emotion on memory for peripheral information. *Memory & Cognition, 37,* 866–879. doi:10.3758/MC.37.6.866

Hamann, S. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology, 14,* 233–238. doi:10.1016/j.conb.2004.03.010

Herbert, C., Junghöfer, M., Kissler, J., Herbert, C., Junghöfer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading: Event related potentials to emotional adjectives during reading. *Psychophysiology, 45,* 487–498. doi:10.1111/j.1469-8986.2007.00638.x

Herbert, C., Kissler, J., Junghöfer, M., Peyk, P., & Rockstroh, B. (2006). Processing of emotional adjectives: Evidence from startle EMG and ERPs. *Psychophysiology, 43,* 197–206. doi:10.1111/j.1469-8986.2006.00385.x

Ishii, K., Kobayashi, Y., & Kitayama, S. (2010). Interdependence modulates the brain response to word–voice incongruity. *Social Cognitive and Affective Neuroscience, 5,* 307–317. doi:10.1093/scan/nsp044

Johansson, M., Mecklinger, A., & Treese, A.-C. (2004). Recognition memory for emotional and neutral faces: An event-related potential study. *Journal of Cognitive Neuroscience, 16,* 1840–1853. doi:10.1162/0898929042947883

Johnson, R., Jr., & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. *Electroencephalography and Clinical Neurophysiology, 44,* 424–437.

Kaestner, E. J., & Polich, J. (2011). Affective recognition memory processing and event-related brain potentials. *Cognitive, Affective, & Behavioral Neuroscience, 11,* 186–198. doi:10.3758/s13415-011-0023-4

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered

than neutral words? *Memory & Cognition, 31,* 1169–1180. doi:10.3758/BF03195800

Kensinger, E. A., Piguet, O., Krendl, A. C., & Corkin, S. (2005). Memory for contextual details: Effects of emotion and aging. *Psychology and Aging, 20,* 241–250. doi:10.1037/0882-7974.20.2.241

Kissler, J., Herbert, C., Winkler, I., & Junghöfer, M. (2009). Emotion and attention in visual word processing: An ERP study. *Biological Psychology, 80,* 75–83. doi:10.1016/j.biopsycho.2008.03.004

Kitayama, S. (1996). Remembrance of emotional speech: Improvement and impairment of incidental verbal memory by emotional voice. *Journal of Experimental Social Psychology, 32,* 289–308. doi:10.1006/jesp.1996.0014

Leclerc, C. M., & Kensinger, E. A. (2011). Neural processing of emotional pictures and words: A comparison of young and older adults. *Developmental Neuropsychology, 36,* 519–538. doi:10.1080/87565641.2010.549864

Lischke, A., Gamer, M., Berger, C., Grossmann, A., Hauenstein, K., Heinrichs, M., & Domes, G. (2012). Oxytocin increases amygdala reactivity to threatening scenes in females. *Psychoneuroendocrinology, 37,* 1431–1438. doi:10.1016/j.psyneuen.2012.01.011

Maratos, E. J., Dolan, R. J., Morris, J. S., Henson, R. N. A., & Rugg, M. D. (2001). Neural activity associated with episodic memory for emotional context. *Neuropsychologia, 39,* 910–920. doi:10.1016/S0028-3932(01)00025-2

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology, 77,* 247–265. doi:10.1016/j.biopsycho.2007.11.006

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience, 8,* 507–526. doi:10.1162/jocn.1996.8.6.507

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport, 19,* 209–213. doi:10.1097/WNR.0b013e3282f454db

Paulmann, S., Seifert, S., & Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience, 5,* 59–75. doi:10.1080/17470910903135668

Phelps, E. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology, 14,* 198–202. doi:10.1016/j.conb.2004.03.015

Phelps, E. A., LaBar, K. S., & Spencer, D. D. (1997). Memory for emotional words following unilateral temporal lobectomy. *Brain and Cognition, 35,* 85–109. doi:10.1006/brcg.1997.0929

Righi, S., Marzi, T., Toscani, M., Baldassi, S., Ottonello, S., & Viggiano, M. P. (2012). Fearful expressions enhance recognition memory: Electrophysiological evidence. *Acta Psychologica, 139,* 7–18. doi:10.1016/j.actpsy.2011.09.015

Sarlo, M., & Munafò, M. (2010). When faces signal danger: Event-related potentials to emotional facial expressions in animal phobics. *Neuropsychobiology, 62,* 235–244. doi:10.1159/000319950

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience, 22,* 474–481. doi:10.1162/jocn.2009.21215

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., Escoffier, N., Li, Q. Y., Li, H., Strafford-Wilson, J., & Li, W.-I. (2008a). What grabs his attention but not hers? Estrogen correlates with neurophysiological measures of vocal change detection. *Psychoneuroendocrinology, 33,* 718–727. doi:10.1016/j.psyneuen.2008.02.010

Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., & Friederici, A. D. (2008b). When vocal processing gets emotional: On the role of social orientation in relevance detection by the human amygdala. *NeuroImage, 40,* 1402–1410. doi:10.1016/j.neuroimage.2008.01.018

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience, 15,* 1135–1148. doi:10.1162/089892903322598102

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14,* 228–233. doi:10.1016/S0926-6410(02)00108-8

Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research, 1176,* 103–112. doi:10.1016/j.brainres.2007.08.008

Schirmer, A., Soh, Y. H., Penney, T. B., & Wyse, L. (2011a). Perceptual and conceptual priming of environmental sounds. *Journal of Cognitive Neuroscience, 23,* 3241–3253. doi:10.1162/jocn.2011.21623

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

Schirmer, A., Teh, K. S., Wang, S., Vijayakumar, R., Ching, A., Nithianantham, D., & Cheok, A. D. (2011b). Squeeze me, but don't tease me: Human and mechanical touch enhance visual attention and emotion discrimination. *Social Neuroscience, 6,* 219–230. doi:10.1080/17470919.2010.507958

Schirmer, A., Zysset, S., Kotz, S. A., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage, 21,* 1114–1123. doi:10.1016/j.neuroimage.2003.10.048

Schmidt, S. R. (2012). Memory for emotional words in sentences: The importance of emotional contrast. *Cognition & Emotion, 26,* 1015–1035. doi:10.1080/02699931.2011.631986

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2004). The selective processing of briefly presented affective pictures: An ERP analysis. *Psychophysiology, 41,* 441–449. doi:10.1111/j.1469-8986.2004.00174.x

Sharot, T., Martorella, E. A., Delgado, M. R., & Phelps, E. A. (2007). How personal experience modulates the neural circuitry of memories of September 11. *Proceedings of the National Academy of Sciences, 104,* 389–394. doi:10.1073/pnas.0609230103

Thierry, G., & Roberts, M. V. (2007). Event-related potential study of attention capture by affective sounds. *NeuroReport, 18,* 245–248. doi:10.1097/WNR.0b013e328011dc95

van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., & Hagoort, P. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience, 7,* 173–183. doi:10.1093/scan/nsq094

Wambacq, I. J. A., Shea-Miller, K. J., & Abubakr, A. (2004). Non-voluntary and voluntary processing of emotional prosody: An event-related potentials study. *NeuroReport, 15,* 555–559.

Waring, J. D., Payne, J. D., Schacter, D. L., & Kensinger, E. A. (2010). Impact of individual differences upon emotion-induced memory trade-offs. *Cognition & Emotion, 24,* 150–167. doi:10.1080/02699930802618918

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., & Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience, 7,* 847–854. doi:10.1038/nn1276

Werner, N. S., Peres, I., Duschek, S., & Schandry, R. (2010). Implicit memory for emotional words is modulated by cardiac perception. *Biological Psychology, 85,* 370–376. doi:10.1016/j.biopsycho.2010.08.008

Whittle, S., Yücel, M., Yap, M. B. H., & Allen, N. B. (2011). Sex differences in the neural correlates of emotion: Evidence from neuroimaging. *Biological Psychology, 87,* 319–333. doi:10.1016/j.biopsycho.2011.05.003

Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain, 119,* 889–905. doi:10.1093/brain/119.3.889

Windmann, S., & Kutas, M. (2001). Electrophysiological correlates of emotion-induced recognition bias. *Journal of Cognitive Neuroscience, 13,* 577–592. doi:10.1162/089892901750363172

Wiswede, D., Rüsseler, J., Hasselbach, S., & Münte, T. F. (2006). Memory recall in arousing situations—An emotional von Restorff effect? *BMC Neuroscience, 7,* 57. doi:10.1186/1471-2202-7-57

Zeidan, M. A., Igoe, S. A., Linnman, C., Vitalo, A., Levine, J. B., Klibanski, A., & Milad, M. R. (2011). Estradiol modulates medial prefrontal cortex and amygdala activity during fear extinction in women and female rats. *Biological Psychiatry, 70,* 920–927. doi:10.1016/j.biopsych.2011.05.016

Zhang, T.-Y., & Meaney, M. J. (2010). Epigenetics and the environmental regulation of the genome and its function. *Annual Review of Psychology, 61,* 439–466. doi:10.1146/annurev.psych.60.110707.163625