When brain differentiates happy from neutral in prosody?

Xuhai Chen [1, 2], *Yufang Yang* [2]

[1] School of Psychology, Shaanxi Normal University
[2] State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences

shiningoceanchan@gmail.com, yangyf@psych.ac.cn

Abstract

The effect of different intensities of vocal emotion on event related potentials has yet not been studied. We therefore investigated 16 healthy participants with emotion and sound decision on neutral and happy voice which varied continuously in intensity. The result found that neutral and happy voice can be differentiated on P2 component under both explicit and implicit condition. Moreover, the P2 parameters were linear correlated with the rate of happiness, suggesting a graded processing of vocal emotion in early stage. However, the brain distinguished neutral from happy in P3 interval when performing explicit task but exhibit a categorical feature.

Index Terms: emotional prosody, ERP, recognition

1. Introduction

Humans are highly proficient to use emotional cues to communicate. Given that emotional signals provide important information about our environment, emotional information needs to be distinguished rapidly and reliably from other nonemotional stimuli. The differentiation between emotional and neutral facial expression has been well studied. It was reported that the processing of emotional facial expressions occurs faster and shows a differentiation between emotional and neutral faces as early as 120–150 ms after stimulus onset[1]. Moreover, a number of studies have demonstrated an enhanced frontal positivity for emotional as compared with neutral faces, occurring around 150 ms after stimulus onset [2, 3]. In addition, the rapid processing of emotional facial expressions can even occur in the absence of viewers' conscious awareness of the emotional faces[4].

In contrast to the general consensus on the fast differentiation between emotional and neural facial expression, the time course of emotional voice processing remains in controversy. For instance, using an oddball design, Bostanov and Kotchoubey [5] found that context-incongruent vocal stimuli elicited an increased negativity starting 300 ms after stimulus onset, suggesting that the differentiation begins 300 ms. However, studies using similar paradigm found that emotional category changes demonstrated early negative responses around 200 ms [6, 7], indicating that the brain is able to distinguish different vocal emotion within 200 ms. Furthermore, some recent studies reported that ERPs induced by emotional prosodic materials differed from those by neutral materials in P200[1] component[8, 9].

This controversy may result from lack of control of emotion intensity. Previous studies have showed that the brain response is sensitive to the intensity of facial emotion [10, 11] and several studies indicated that vocal emotion differ in emotion intensity and possess different acoustic profiles[12, 13]. Therefore, the intensity must be controlled to elucidate the neural mechanisms related to the processing of vocal emotion. We examined brain responses to happy and neutral voice using electroencephalography by adopting morphed voice of different intensity levels. Then, we identified the association between the ERP parameters and the rate of happiness to further explore the time course of the differentiation between emotional and neutral voice.

2. Method

2.1. Participants

Eighteen right-handed native speakers of Mandarin Chinese were recruited to participate in the experiment. All participants reported normal auditory and normal or corrected-to-normal visual acuity and no neurological or other medical problems. Participants gave written informed consent before the experiment and received monetary compensation for their participation. Two participants were excluded from analysis because of excessive artifacts during the EEG recording session.

2.2. Stimuli

Five interjections (嘿/hei/，噢/o/，哇/wa/，喂/wei/，哟/yo/) produced by a trained actress in happy and neutral prosody were used in the current study. The happy to neutral continua were created for each interjection using STRAIGHT [14], which is a tool for manipulating voice quality, timbre and pitch, in four steps that corresponded to 20/80%, 40/60%, 60/40%, 80/20% happy/neutral. Together with the original happy and neutral voices, thirty voices varying in emotion intensity were used as critical materials. Moreover, the spectral rotated counterparts (see Figure.1) which share low-level acoustic features but without affective properties [15, 16] served as control materials.

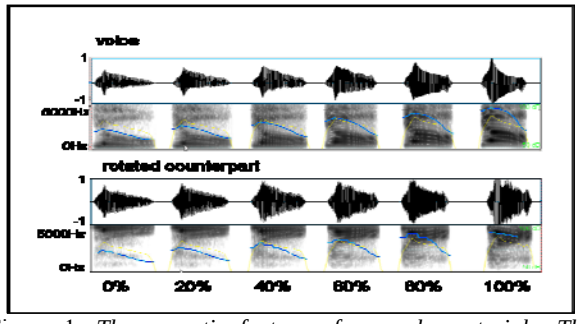

Figure 1: *The acoustic feature of example materials. The dataset consists of oscillogram (up) and voice spectrographs (down) with uncorrected pitch contours (blue line) and intensity contours (yellow line) superimposed.*

2.3. Procedure and design

[1] ERP: event related potentials, a critical technique to study to time course of human mind; P200(P2), P300(P3):ERP components related to specific cognitive processing, for instance, P3 is thought to related to memory updating.

Subjects reclined comfortably at a distance of 100cm from a computer monitor in an acoustically and electrically shielded booth. Stimuli were presented in a pseudo-randomized order in four blocks of 210 trials (each emotional voices and their counterparts repeated 15 times) under each task condition. Each trial began with a fixation cross in the center of the monitor for 300ms, and then the stimulus was presented in stereo via loudspeakers located to the left and right side of the computer at an intensity of 80 dB SPL while the cross remained on the screen. Participants were instructed to respond as accurately and quickly as possible whether the sound expressed happiness (emotion task) or whether the sound was human voice (sound-type task) by pressing the "J" or "F" button on the keyboard after they heard stimuli. The order of the button for "yes" and "no" were counterbalanced across participants. The inter-trial interval was randomized between 800 ms to 1200 ms with the average 1000 ms. Participants were asked to look at the fixation cross and avoid eye movements during sound presentation. Between two blocks participants were given enough time to take a break. Pre-training with 20 practice trials was used before each task in order to familiarize subjects with the procedure. Each subject completed both tasks, with the order of the two tasks counterbalanced across subjects.

2.4. ERP recording

EEG was recorded with 66 Ag-AgCl electrodes mounted in an elastic cap (NeuroScan system). EEG data were referenced online to the left mastoid. Vertical electrooculograms (EOG) were recorded supra- and infra-orbitally at the left eye. Horizontal EOG was recorded from the left versus right orbital rim. EEG and EOG were digitized at 500 Hz with an amplifier bandpass of 0.01–100 Hz including a 50-Hz notch filter and were stored for off-line analysis. Impedances were kept below 5kΩ.

2.5. Data analysis

2.5.1. Behavioral data analysis

The RTs (corrected by 2.5 SD of the mean) for both tasks, ratio of expressing happiness in emotion task, and error rates in sound-type task were calculated for each participant and then analyzed with separate repeated measures ANOVAs with *Intensity level* (0%, 20%, 40%,60%, 80% and 100%), *Sound–type* (human voice and rotated counterpart) as within subject factors. Further, in order to specify the relationship between the emotional intensity and ratio of happiness, a correlation analysis between the emotional intensity and ratio of happiness (Z score) was conducted**.**

2.5.2. ERP analysis

The EEG data were preprocessed and segmented to 900-msepochs time-locked to the onset of the sound stimuli, starting 100ms prior to the stimuli onset. Segments were then averaged for each type of stimuli under both task conditions after baseline correction, low-pass filter and artifact rejection. Then grand average waveforms (see Figure 3, 4) and repeated measures ANOVAs were calculated based on extracted average waveforms. As can be seen, the ERP waveforms started to differentiate at about 200 ms across task requirement only for the voice stimuli, manifested by a frontal peaking but broadly distributed P2 and a posterior peaking P3. According to visual inspection and consecutive 10 ms analysis, the time ranges of 200-280 ms and 300-500 ms were chosen for P2 and P3 respectively.

Mean voltage for P2 and p3 over F3, FZ, F4, C3, CZ, C4, P3, PZ, and P4 was subjected to repeated measures ANOVAs with *Task*, *Intensity level*, *Sound–type, Laterality* (left vs. midline vs. right), *Sagittality* (frontal vs. central vs. posterior region) as within-subject factors. Moreover, the peak latency of the most positive peak of P2 and P3 elicited by voice were determined and subjected to repeated measures ANOVAs with *Task*, *Intensity level* and *electrodes* (F3, FZ, F4, C3, CZ, C4, P3, PZ, and P4) as within-subject factors. Further, in order to test whether the observed ERPs are valid as an index of emotion intensity effect, histograms were drawn based on the original mean ERP amplitudes and peak latencies, and correlations between the ratio of happiness and the amplitude and latency of P2 and P3 (Z scores) were conducted. The degrees of freedom of the F-ratio were corrected according to the Greenhouse–Geisser method in all these analyses.

3. Result

3.1 Behavioral results

Under the emotion task condition, as shown in Figure 2 (A and B), the ratio of expressing happiness increased as the increase of emotion intensity for the human voice while only very low ratio (at most 20%) was found for the rotated counterparts. The RTs for human voice were long for those with medium intensity but short for those with low and high intensity, whereas the RTs for the rotated counterparts were pretty short and had no big variation. These observations were confirmed by the follow statistical analysis. The repeated ANOVA of ratio of happiness revealed significant main effects of *Intensity* [F(5,75)=85.27,P ＜ .001], *Sound–type* [$F(1,15)=57.61$, $P < .001$], and interaction of *Intensity* × *Sound–type* [$F(5,75)=46.96$, $P<.001$].

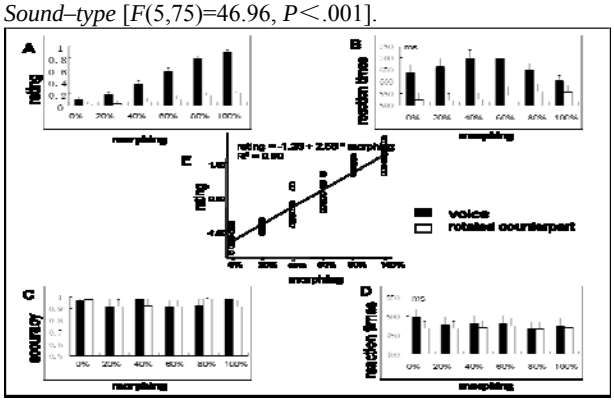

Figure 2: *Behavioral responses. The rate of happiness (A), reaction times in emotion task (B), accuracy (C) and reaction times in sound-type task (D)as a fuction of intensity and sound-type. E: Curve-fitting between rate of happiness and intensityl.*

Further simple tests showed that the ratio of expressing happiness for human voice differentiated from each other (P ＜.001 or P＜.05) while no significant difference was found for the rotated counterparts. The repeated ANOVA of RTs showed significant a main effect of *Sound–type* [$F(1,15)=29.42$, $P < .001$], and interaction of *Intensity* × *Sound–type* [$F(5,75)=8.19$, $P < .01$]. Further simple tests showed that the RTs were significantly shorter for the 100% voice than for 40% and 60% voice (P＜.01), while no other significant difference was found (Ps＞.1). Under the sound-type task condition, as shown in Figure 2 (C and D), the error rates and RTs hardly distinguished across sound-type and emotion intensity. The correlation analysis conducted for the emotional intensity and ratio of happiness (Z score) showed a significant positive correlation between the emotional intensity and ratio of happiness (r=0.946， P＜.0001, see Figure. 2. E), suggesting that the richer of the emotion-related acoustic parameters, the higher probability of rating as happiness.

3.2. ERP results

The repeated measures ANOVAs on the mean amplitude of P2 showed a significant main effect of *Intensity* [$F(5,75)=5.40$, $P<.001$], *Sound–type* [$F(1,15)=5.40$, $P<.05$], and the two way interaction of *Intensity* × *Sound–type* [$F(5,75)=2.37$, $P=.06$]. Further simple tests showed that *Intensity* effect reached significant level only for the human voice [$F(5,75)=6.06$,$P < .001$], and over the frontal [$F(5,75)=8.98$, $P<.001$] and central region [$F(5,75)=6.68$, $P<.001$], with the amplitude more negative going for the 100% than for 0%, 20%, and 40% ($P<.01$ or $P<.05$) and the amplitude more positive going for 0% than for 60%, 80%, and 100% ($P<.01$ or $P<.05$, see Figure 3). It was worth noting that there was no significant effect of task [$F(5,75)=.67$, $P=.80$], suggesting that task relevancy could not affect this early component which distinguishes emotional expressions from neutral ones.

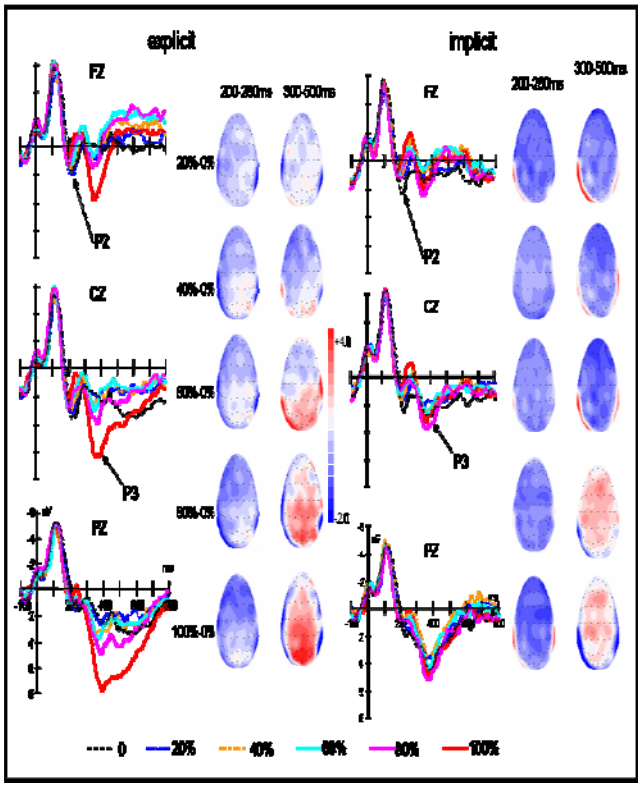

Figure 3: *Grand-average ERP waveforms as a function of intensity and task and topographies of difference curves in selected time periods for human voice. In this figure, as in the following ones, the amplitude (in microvolts) is plotted on ordinate (negative up) and the time (in milliseconds) is on abscissa.*

The analysis of P3 amplitude showed a significant main effect of *Intensity* [$F(5,75)=3.99$, $P < .05$], *Sound–type* [$F(1,15)=21.53$, $P<.001$], and *Task* [$F(1,15)=9.81$, $P<.01$]. Moreover, the two way interaction of *Intensity*×*Sound–type* [$F(5,75)=2.96$, $P < .05$], *Intensity* × *Task* [$F(5,75)=2.68$, $P<.05$], and three way interaction of *Intensity*×*Sound–type*× *Task* [$F(5,75)=5.00$, $P<.01$] and *Intensity*×*Task*×*Sagittality* [$F(10,150)=4.37$, $P<.01$] were all significant. Further simple tests found that *Intensity* effect reached significant level only for the human voice [$F(5,75)=6.11$, $P < .001$] and under emotion task condition [$F(5,75)=5.51$, $P<.001$] over central [$F(5,75)=11.64$,$P < .001$] and posterior electrodes [$F(5,75)=20.39$,$P<.001$], with the amplitude more positive going for the 100% stimuli than the other types of stimuli ($Ps<.001$ or $P<.01$) and more positive going for 80% than 0%, 20%, 40%, and 60% stimuli($Ps<.01$ or $P<.05$).

The peak latency analysis of P2 showed a significant main effect of Intensity [$F(5,75)=7.15$, $P<.001$]. The post hoc comparison revealed the peak latency was significantly shorter for 100% than for 0%, 20% , and 40% stimuli ($Ps<.01$ or $P<.05$), while the peak latency was longer for 0% stimuli than for 100% and 80% stimuli saliently ($Ps<.01$ or $P<.05$). However, the peak latency analysis for P3 revealed no significant difference ($Ps>.1$).

The histograms based on the original ERP mean amplitudes and peak latencies (middle) and the scatter diagram of the correlation analysis (peripheral) were shown in Figure 5. As it showed, the bigger the intensity level, the smaller the P2 amplitude and shorter P2 peak latency, which was confirmed by the significant negative linear correlation between the ratio of happiness and P2 amplitude and peak latency (r=-0.55 and -0.60 , respectively, $Ps < .001$). However, no obvious linear correlation was observed for P3 peak latency but a salient correlation between ratio of happiness and P3 amplitude ($r = 0.383$, $P<.01$).

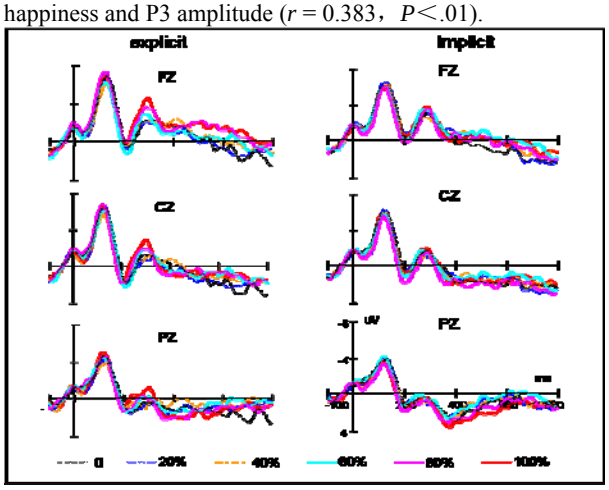

Figure 4: *Grand-average ERP waveforms as a function of intensity and task for rotated counterparts.*

4. Discussion

The present study found that subjects could distinguish the happy from neutral voice very well when performing explicit task. And the rate of happiness linear increased as a function of the intensity level, suggesting a graded processing of neutral and emotional voice, consistent with the studies in facial expression processing [10, 17]. Moreover, given the similar effects were not observed in the acoustic feature controlled rotated counterparts, it is reasonable to conclude that it is the emotionality that contributed to the graded processing but not the low-level acoustic features.

Happy and neutral vocal articulation differentiated from each other in P2 component under both explicit and implicit conditions. These results were consistent with the previous studies indicting that brain can distinguish emotional form neutral voice in P200 component [8, 9, 18]. The neutral voice not only distinguished from 100% happy voice, but also the 60%、80% happy voice, suggesting that human brain can detect the not full blossomed emotional voice. Moreover, this differentiation was not modulated by task demands, implying an automatic processing mechanism. Therefore, in conjunction of the previous studies, the present findings suggested that brain can differentiate emotional information from other non-emotional stimuli within 200 ms. Given that

the timing, polarity, and scalp distribution of this ERP correlate are similar to ERP markers of emotional face processing, a common supramodal brain mechanisms may be involved in the rapid detection of affectively relevant visual and auditory signals.

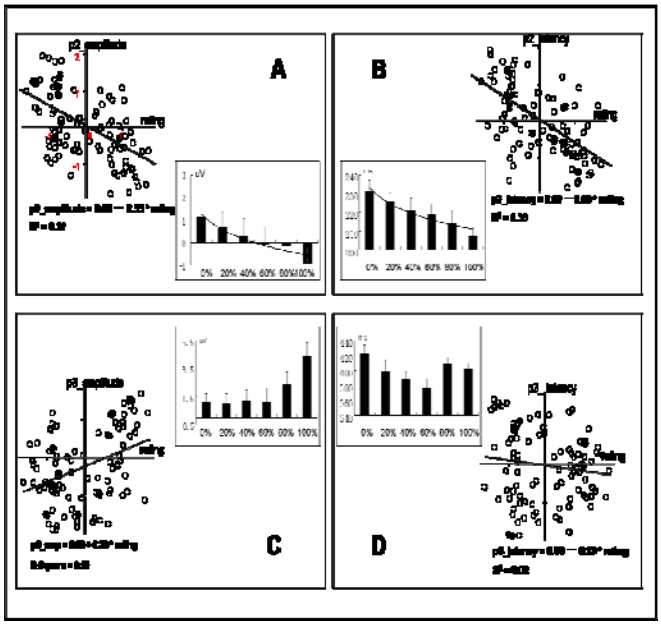

Figure 5: *Histograms based on the original ERP mean amplitudes and peak latencies (middle) and scatter diagram and Curve-fitting of the correlation between the ratio of happiness and P2 and P3 parameters(peripheral).*

More importantly, the P2 component was sensitive to intensity variation. A significant negative linear correlation between P2 amplitude and latency and rate of happiness was observed, that is, the more intense an emotion was expressed, the less positive was the deflection of the P2, and the earlier was the peak latency of the P2. And these effects were not observed in the rotated counterparts which had similar acoustic feature. This result is in consistency with the observation in facial emotion domain, which showed that brain response (N170) is sensitive to the intensity of facial emotion [10, 11]. Taking together, we concluded that the brain response is sensitive to the intensity of emotionality regardless of modality. Further, we speculate that the human brain differentiate emotional signals from neutral ones mainly due to they possess different emotion intensity, in other words, emotional significance.

The two kinds of materials diverged in P3 component under explicit condition besides of P2 component. This result replicated the finding in Wambacq and conleague's study, which suggested that emotional prosody was processed around 160 ms after stimulus onset under non-voluntary processing conditions and around 360 ms under voluntary processing conditions[18]. However, this differentiation only observed between neutral and high intense happy voice (80% and 100%), implying that only the emotion related acoustic feature reached certain level can the brain differentiate it from neutral voice. Moreover, distinct from the P2 parameters showing linear correlation with the rate of happiness, there was no linear correlation between P3 parameters and rate of happiness. These results might suggest that some what categorical processing of vocal emotion in late time course.

5. Conclusions

In short, the current data indicate that the brain can differentiate neutral and happy voice within 200 ms both non-voluntarily and voluntarily, as indexed by the P2 component. Moreover, the P2 component is sensitive to the intensity of the vocal emotion and the linear increasing of P2 amplitude and decreasing of P2 latency might suggest a graded processing of vocal emotion in early stage. However, the brain distinguishes neutral from happy in P3 interval when performing explicit task with a categorical processing feature.

6. References

[1] Eimer, M. and Holmes, A., Event-related brain potential correlates of emotional face processing. Neuropsychologia, 45(1): 15-31, 2007.

[2] Eimer, M., Holmes, A., and McGlone, F., The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. Cogn Affect Behav Neurosci, 3(2): 97-110, 2003.

[3] Ashley, V., Vuilleumier, and Swick, D., Time course and specificity of event-related potentials to emotional expressions. Neuroreport, 15(1): 211-6, 2004.

[4] Kiss, M. and Eimer M., ERPs reveal subliminal processing of fearful faces. Psychophysiology, 45(2): 318-26, 2008.

[5] Bostanov, V. and Kotchoubey, B., Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. Psychophysiology, 41: 259-268, 2004.

[6] Goydke, K.N., et al., Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. Cognitive Brain Research, 21(3): 351-359,2004.

[7] Thönnessen, H., et al., Early sensory encoding of affective prosody: Neuromagnetic tomography of emotional category changes. NeuroImage, 50(1), 250-259, 2010.

[8] Paulmann, S. and S. Kotz, Early emotional prosody perception based on different speaker voices. Neuroreport, 19(2): 209, 2008.

[9] Sauter, D.A. and M. Eimer, Rapid Detection of Emotion from Human Vocalizations. Journal of Cognitive Neuroscience, 22(3): 474-481, 2010.

[10] Utama, N.P., et al., Phased processing of facial emotion: An ERP study. Neuroscience Research, 64(1): 30-40, 2009.

[11] Sprengelmeyer, R. and I. Jentzsch, Event related potentials and the perception of intensity in facial expressions. Neuropsychologia, 44(14): 2899-2906,2006..

[12] Juslin, P.N. and Laukka, Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion, 1(4): 381-412, 2001.

[13] Banse, R. and K. Scherer, Acoustic profiles in vocal emotion expression. Journal of personality and social psychology, 70: 614-636, 1996.

[14] Kawahara, H. and H. Matsui. Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. in 2003 IEEE International Conference on Accoustics, Speech, and Signal Processing, April 6, 2003Hong Kong, Hong kong: Institute of Electrical and Electronics Engineers Inc, 2003.

[15] Blesser, B., Speech perception under conditions of spectral transformation: I. Phonetic characteristics. Journal of Speech and Hearing Research, 15(1): 5-41, 1972.

[16] Chen, X., et al., Event-related potential correlates of the expectancy violation effect during emotional prosody processing. Biological Psychology, 86(3): 158-167, 2011.

[17] Dunning, J.P., et al., In the face of anger: Startle modulation to graded facial expressions. Psychophysiology, 47(5): 874-878, 2010.

[18] Wambacq, I.J.A., Shea-Miller, K.J. and Abubakr, A., Non-voluntary and voluntary processing of emotional prosody: an event-related potentials study. Neuroreport, 15(3): 555-559, 2004.