

语音情绪变化内隐加工的神经生理机制^{*}

陈煦海^{1,2} 杨晓虹² 杨玉芳²

(¹ 陕西省行为与认知心理学重点实验室, 陕西师范大学心理学院, 西安 710062)

(² 脑与认知国家重点实验室, 中国科学院心理研究所, 北京 100101)

摘 要 人类迅速高效识别语音情绪变化的神经生理机制仍是一个未解之谜。本研究试图通过对被试内隐加工语音情绪变化的即时脑电数据做时域和频域分析, 揭示语音情绪变化加工的神经生理机制。结果发现语音情绪变化诱发 N2/P3 复合成分, theta 频段能量增加和试次间相位相干系数(ITC)增大, beta 频段能量降低和 ITC 增大。这些神经生理指标受情绪变化模式的调节, 情绪变化越激烈, 诱发 N2/P3 复合成分的潜伏期越早, theta 能量和相位变化更强烈。这些结果说明语音情绪变化加工是一个变化检测与整合的过程, 该过程主要由 theta 和 beta 频段能量和试次间相位的变化来实现。

关键词 语音情绪; 神经振荡; ERP; ERSP; ITC

分类号 B842

1 导言

口头语言表现情绪主要有两种手段:“说什么”和“怎么说”。前者指语义传达的情绪内容; 后者则指依靠语音的音高、语速、响度和音质等声学特征的不同组合所表达的情绪, 称作语音情绪(Banse & Scherer, 1996)。在交流中, 人们更多依靠语音来判断他人的情绪(Dietrich, Ackermann, Szameitat, & Alter, 2006)。语音情绪变化是生活中常见的现象, 也是凸显情绪的重要手段(Paulmann, Jessen, & Kotz, 2012; Paulmann, Pell, & Kotz, 2008), 人脑能敏感地识别语音情绪的变化并有效加工, 但人们对语音情绪变化加工的神经生理机制还知之甚少。因此, 本研究拟运用脑电技术探索语音情绪变化加工的神经生理机制。

尽管语音情绪变化加工机制的研究还不多见, 但听觉刺激变化加工的研究却颇为丰富。听觉刺激在时间维度延伸, 导致预期式加工是听觉加工的重要特征(Winkler, Denham, & Nelken, 2009), 即听者会根据已听到的刺激对即将出现的刺激形成预期,

并将后来的刺激与预期相对比, 二者相符则顺利整合, 不相符则造成期待违反, 形成变化检测与再整合的过程。对这一机制研究最多的是 oddball 范式, 大概率呈现标准刺激让被试提取规则形成预期, 小概率呈现偏差刺激与预期违反。大量研究发现人脑能够迅速检测这种听觉刺激的变化, 在自动加工时偏差刺激诱发 MMN (Näätänen, 2007), 主动加工时诱发 N2/P3 复合成分(Polich, 2007; Polich & Kok, 1995)。连续性声音刺激的变化加工也表现出相似的机制。如连续语流中, 后续语音与前文句法违反诱发 ELAN 和 P600 (Hahne & Jescheniak, 2001), 而音韵违反诱发 PMN (D'Arcy, Connolly, Service, Hawco, & Houlihan, 2004; Newman, Connolly, & Mcivor, 2003)。音乐旋律中某个音符的音高违反或句法违反也可诱发早期负波 ERAN 和晚正成分 P300 (Magne, Schon, & Besson, 2006; Schön, Magne, & Besson, 2004)。与语言和音乐相似, 语音情绪也是依靠声音为载体, 在时间维度展开, 并且拥有某种固有的发展趋势, 人们可能根据已听刺激对即将听到的刺激形成预期, 表现出预期加工的特点。

收稿日期: 2012-10-22

^{*} 教育部人文社科青年项目(12YJC190003)、陕西省自然科学基金青年项目(2012JQ4010)、中央高校基本科研业务费(GK201102033)、国家自然科学基金(NSFC31070989)资助。

通讯作者: 杨玉芳, E-mail: yangyf@psych.ac.cn

已有一些研究运用语音情绪变化来考察语音情绪加工的时间进程(江爱世等, 2009)。一些研究沿用 oddball 范式, 分别用携带不同情绪的声音为标准刺激和偏差刺激, 结果发现语音情绪偏差刺激在自动加工中诱发 MMN, 但峰值潜伏期不一(Bostanov & Kotchoubey, 2004; Goydke, Altenmüller, Möller, & Mönte, 2004; Thönnessen et al., 2010), 而在主动加工时则诱发 P300 (Thierry & Roberts, 2007; Wambacq, Shea-Miller, & Abubakr, 2004)。另一些研究则用语句情绪韵律为材料, Kotz 和 Paulmann (2007)在语句中交互拼接构建情绪变化, 形成语音情绪发展模式的期待违反, 发现语句中情绪韵律的变化诱发潜伏期约为 350 ms 的正成分——韵律期待违反正波(prosodic expectancy positivity, PEP)。他们随后的研究证明该成分不受情绪类别、注意资源投入程度的影响(Paulmann et al., 2012; Paulmann & Kotz, 2008), 且在基底节受损的患者中仍然存在(Paulmann et al., 2008)。最近 Chen 等人(2011)发现人脑对情绪韵律期待违反的识别可能更早, 表现为早期负波类 N200, 然后是反映对违反再加工并整合的正成分 PEP 及 LPC, 且这些成分会受任务要求和违反特征的调制。显然, 这些研究以情绪变化为手段对语音情绪加工进行了很好的探索, 但其焦点在语音情绪, 未把语音情绪变化本身作为一种现象予以关注。

上述有关声音变化的电生理研究都仅从时域考察变化诱发的事件相关电位(ERP)并推测其背后的认知机制。然而, 在锁时又锁相的 ERP 数据之外, 对脑电(EEG)数据进行频域的神经振荡分析(event related oscillations, ERO), 具体离析各个频段能量变化(事件相关摄动, event related spectral power, ERSP)和相位变化特征(试次间相位相干系数, inter-trial coherence, ITC), 能从神经活动的层面说明语音情绪加工的神经生理机制(Delorme & Makeig, 2004)。已有研究证明不同频段的神经振荡往往和不同的认知活动相关, 一般认为低频振荡(theta 和 alpha 频段)和大范围的神经活动相关, 而高频(如 gamma 频段)振荡和局部的神经活动相关(Başar, Başar-Eroglu, Karakaş, & Schürmann, 2001; Pfurtscheller & da Silva, 1999)。在与本研究内容相关的一些研究中, Fuentemilla 等人(2008)发现前中部分布听觉 MMN 主要源于偏差刺激 theta 能量的增加以及试次间相位同步性增加, 而乳突记录到的颞区 MMN 则只源于试次间相位同步性增加。另外,

Yordanova 等(1998)则发现 theta 能量增加是时域 P300 成分的主要贡献者。事实上, 对奖赏预期的违反也导致 theta 频段的能量增强(Cavanagh, Frank, Klein, & Allen, 2010; Tzur & Berger, 2007), 且 theta 能量增加值和期待违反的凸显性有线性关系(Chen, Yang, Gan, & Yang, 2012; Tzur & Berger, 2007)。因此, 基于如上一些神经振荡理论和研究事实, 可以预期, 从神经振荡特征的角度深入分析语音情绪变化的脑电数据可能进一步揭示语音情绪加工的神经生理机制。

综上所述, 当前有关语音情绪的研究仅以变化为手段来考察语音情绪的加工特点, 未曾把语音情绪变化本身作为一个现象来研究, 而且已有的电生理研究都只做时域 ERP 分析, 没有揭示语音情绪变化加工在频域的神经振荡机制。因此, 本研究拟将语音情绪变化作为关注的焦点, 记录被试加工语音情绪变化的脑电, 从时域和频域挖掘脑电数据, 进一步揭示语音情绪加工的神经生理机制。本研究沿用交互拼接的方法(Chen et al., 2011; Kotz & Paulmann, 2007)构建语句情绪韵律的变化, 同时将凸显程度不同的两种变化模式(中性到愤怒, 愤怒到中性)纳入考察范围, 然后让被试完成与情绪变化无关的语音判断任务, 以期回答如下问题: 情绪无关任务条件下语音情绪变化加工时域和频域的神经生理机制是怎样的? 语音情绪的变化模式对这些机制有什么影响? 本研究预期在完成情绪无关任务时, 语音情绪变化仍能诱发反应期待违反与整合的 ERP 成分, 且受变化模式的调控; 另外, 这种变化加工可能和 theta 能量的增加以及相位同步性变化相关。

2 方法

2.1 被试

在校大学生 15 名, 其中女生 7 名, 年龄在 20 至 26 岁之间, 平均 21.45 岁。所有被试均为汉语母语的右利手者, 身体健康, 没有情感或精神障碍。所有被试均自愿参加实验并获得适量报酬, 正式实验前签署知情同意书。其中一名被试的数据因为伪迹太多而剔除。

2.2 实验材料

以 50 个长度均为 12 个音节的句子为文字材料(例子见图 1)。所有语句均描述一个客观事实, 不含情绪词。请 30 位大学二年级学生就语句的情绪效价评分(1 到 7 点 liket 评分, 7 为表现了积极情绪, 1

为表现了消极情绪, 4 为没有表现明显情绪), 结果 50 个语句的得分均在 3.83 到 4.97 之间, 平均 4.3, 标准差 0.37。请一名有专业表演经验的男演员用中性和愤怒情绪¹录制所有语句, 采样率为 22050 Hz。用 Praat 软件(Boersma & Weenink, 2006)对所有语句完成的声学分析表明: 与经典情绪韵律声学特征的研究结果一致(Banse & Scherer, 1996), 相对于中性韵律, 愤怒的情绪音高更高(197 Hz vs. 132 Hz, $t(49)=28.12$, $p<0.001$), 强度更强(70 dB vs. 63 dB, $t(49)=23.18$, $p<0.001$), 语速更快(206 ms vs. 216 ms/字, $t(49)=5.43$, $p<0.01$), 见图 1B。另请 8 名被试就所有语句韵律情绪进行归类, 被试区分两种韵律的正确率接近 100%。用交互拼接法将原始材料制作成包含情绪变化的新语句: 中性—愤怒、愤怒—中性, 拼接模式见图 1-A。为了避免被试就发生拼接的位置形成固定预期, 分别在第五和第九个音节上拼接。共 200 个语句(两种包含情绪变化的语句和两种没有情绪变化的控制语句)作为刺激材料呈现给被试。

2.3 实验程序

所有材料随机分为 5 个 block, 每个 block 40 句, 含“aa”, “nn”, “an”, “na”各 10 句, 以等概率、伪随机顺序用耳机呈现给被试, 确保同一类别刺激不

会连续呈现 3 次。实验在隔音的小房间进行, 被试端坐在电脑屏幕前 115 cm。先呈现注视点“+”300 ms, 接着注视点变成红色, 耳机呈现语句韵律, 听觉呈现完毕即呈现问号要求被试判断整个句子声音大小是否有变化, 被试双手按键盘上的“F、J”反应, F、J 键代表的含义和反应手在被试间平衡。在整个句子呈现过程中, 要求被试盯着注视点并减少眨眼和头动。判断完毕后间隔 1500 ms 进入下一个 trial。在正式实验前给被试数次练习以熟悉实验程序。

2.4 数据记录与分析

用 NeuroScan 公司放大器及 64 导电帽记录脑电信号 Electroencephalogram (EEG), 采样率 500 Hz, 带宽 0.05 至 100 Hz, 左侧乳突在线参考, 离线分析时从各导联的脑电数据中减去双侧乳突的平均数作再参考。垂直眼电通过左眼上下电极记录, 水平眼电通过两只眼睛外周约 1 cm 处电极记录。GND 电极作为接地, 所有电极的阻抗小于 5 千欧。

首先用 NeuroScan 4.3 软件对数据进行预处理, 包括去除眼电, 去除极端伪迹数据。然后截取关键点(拼接处)之前 1000 ms 和之后 2000 ms 的数据存为 EEG 数据, 剔除判断不正确以及伪迹污染严重

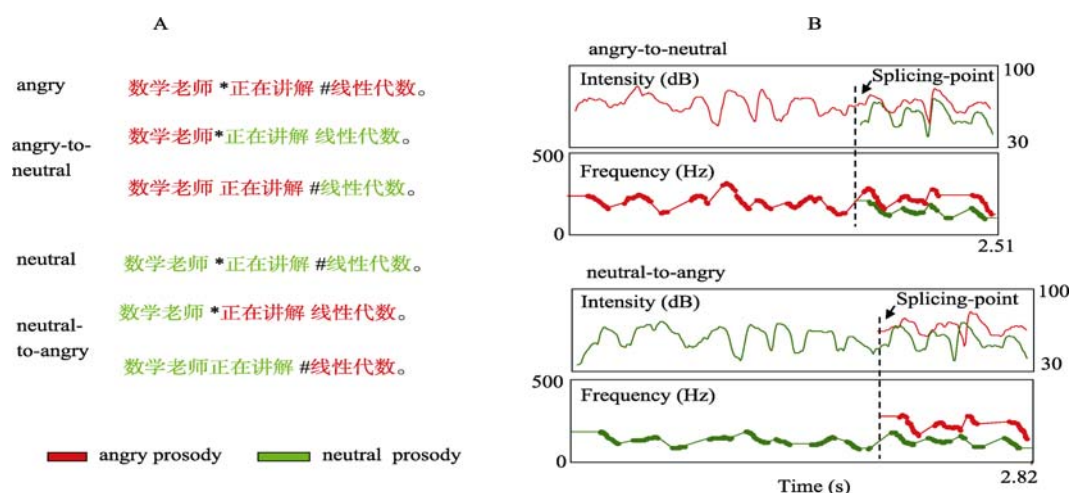


图 1 实验材料示意图。A: 例句及拼接。红色和绿色分别代表愤怒和中性情绪的句子, 在第五和第九个音节上对两种韵律进行拼接, 形成具有情绪变化的韵律“an”和“na”。B: 语句韵律拼接后的声学特征, 中性到愤怒变化的韵律在拼接点处音强、基频增高, 语速增快, 而愤怒到中性韵律声学特征相反。(缩写: aa——全愤怒; an——愤怒-中性; nn——全中性; na——中性-愤怒; 请下载电子版查看彩图, 下同。)

¹ 鉴于前人研究发现不同情绪与中性之间的拼接具有相似的效果(Paulmann and Kotz, 2008), 以及 EEG 实验对时间的限制, 本研究仅在中性和愤怒情绪间进行拼接。

的 trial (电位超过 $\pm 90 \mu V$), 每个条件平均约 40 个 trial 进入最终的时域 ERP 分析和频域 ERSP 和 ITC 分析。ERP 分析中, 首先以句子呈现前 200 ms 做基线校正, 然后以违反点前 200 ms 为基线再做校正, 再分条件平均得到各个被试四种韵律诱发的 ERP, 导出关键时间窗的数据进行统计分析, 平均所有被试各条件下的 ERP 得到总平均图, 本文只呈现关键点前 200 ms 到关键点后 1000 ms 的数据。基于对总平均图的观察和初步分析, 选取如下时间窗作为我们的兴趣窗口: 130~230 ms (N2), 250~450 ms (早正成分)和 450~900 ms (晚正成分 LPC)。

运用 EEGLAB 工具箱对 EEG 数据进行 ERO 分析。该分析用正弦小波与时域数据卷积, 随着频率升高小波周期也升高(Delorme & Makeig, 2004), 先进行单试次分析, 然后再进行多个试次的平均, 最终获得各个条件下的 ERSP 和 ITC。然后以关键点前 200 ms 为基线对关键点后的 ERSP 和 ITC 进行校正。本实验分段的数据包括关键点之前 1000 ms 和之后 2000 ms, 保证感兴趣的时间段(-200 至 1000 ms)在频域处理后不被遗失。为了简明起见, ERSP 和 ITC 数据都只选择 Cz 点作频谱图, 有差异的频段和时程内的全脑数据做地形分布图, 见图 5。

然后以变化(CHANGE: 有情绪变化 vs. 无情绪变化)和关键韵律类型 TYPE: 愤怒 vs. 中性)为被试内因素对正确率数据做重复测量方差分析。以 CHANGE、TYPE 和电极点(ELECTRODE: F5, FZ, F6, C5, CZ, C6, P5, PZ, P6, 前中后和左中右各取一个电极)为被试内因素对 ERP 数据做重复测量方差分析。ERSP 和 ITC 数据先用 EEGLAB 工具箱导出统计差异图, 找到有显著差异的频段和时间窗, 然后导出该范围的数据做统计分析(因素设置跟 ERP 数据相同)。当自由度分子大于 1 时, F 值都进行 Greenhouse-Geisser 校正, P 值进行 Bonferroni 校正。

3 结果

3.1 行为结果

被试对四种韵律反应的正确率都约为 90% (见图 2)。两因素(CHANGE $\times$ TYPE)的重复测量方差分析表明几种韵律彼此之间没有显著差异[CHANGE: $F(1,13)=0.02$, $p=0.90$; TYPE: $F(1,13)=0.71$, $p=0.42$; CHANGE $\times$ TYPE: $F(1,13)=2.23$, $p=0.16$]。由于实验中被试在语句呈现完成后再做反应, 反应时数据有不同延迟, 故未作分析。

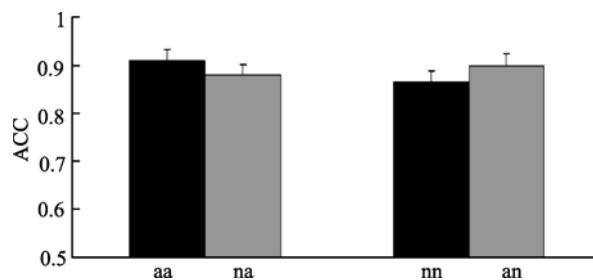


图 2 对四种韵律判断的正确率($M \pm SE$)。

3.2 ERP 结果

四种韵律诱发的 ERP 总平均图见图 3。包含情绪变化的韵律(na 和 an)相对于无变化的韵律(aa 和 nn)都诱发前中部分布的更大波幅的 N2, “na”韵律诱发 N2 的潜伏期相对较早。100 ms 到 300 ms 峰值潜伏期分析验证了这一点, 在中线: Fz, Cz, Pz, 两种韵律变化诱发负波峰值潜伏期分别为 171 ms、206 ms, $t(13)=5.02$, $p<0.001$ 。在该负成分之后, 包含情绪变化的韵律诱发了更大波幅的正漂移。下文分时间窗做具体分析。

3.2.1 N2 分析(130~230 ms)

N2 平均波幅的重复测量方差分析发现显著的 ELECTRODE 效应 [$F(8,104)=4.08$, $p<0.05$, $\eta^2=0.24$] 和 CHANGE 效应 [$F(1,13)=9.51$, $p<0.01$, $\eta^2=0.42$], 其他主效应和交互作用均不显著。两两比较表明, 有情绪变化的韵律诱发 N2 波幅更大, 右中前部 F6、C6 电极 N2 波幅相对较大, 而后部电极 P5、P6、PZ 的 N2 波幅相对较小(见图 3B 地形图)。

3.2.2 早正波分析(250~450 ms)

该时间窗的重复测量方差分析发现显著的 ELECTRODE 主效应, $F(8,104)=5.20$, $p<0.01$, $\eta^2=0.29$, 事后比较发现中部电极的 ERP 相对双侧电极显著更正($ps<0.01$)。还发现显著的 TYPE 主效应 [$F(1,13)=14.19$, $p<0.01$, $\eta^2=0.52$] 和 CHANGE 主效应 [$F(1,13)=6.96$, $p<0.05$, $\eta^2=0.35$]。另外 CHANGE 与 TYPE 有显著的交互作用, $F(1,13)=28.55$, $p<0.001$, $\eta^2=0.69$ 。简单效应分析发现从中性变化到愤怒的韵律相对全无变化愤怒韵律诱发显著更大的正成分, $F(1,13)=21.57$, $p<0.001$; 而相对无转换的中性韵律, “愤怒—中性”韵律在则没有诱发显著更大的正成分, $F(1,13)=0.08$, $p=0.778$ (见图 3 ERP 波形图和地形图)。

3.2.3 晚正波分析(450~900 ms)

该时间窗的重复测量方差分析发现显著的 ELECTRODE 主效应 [$F(1,104)=16.91$, $p<0.001$, $\eta^2=$

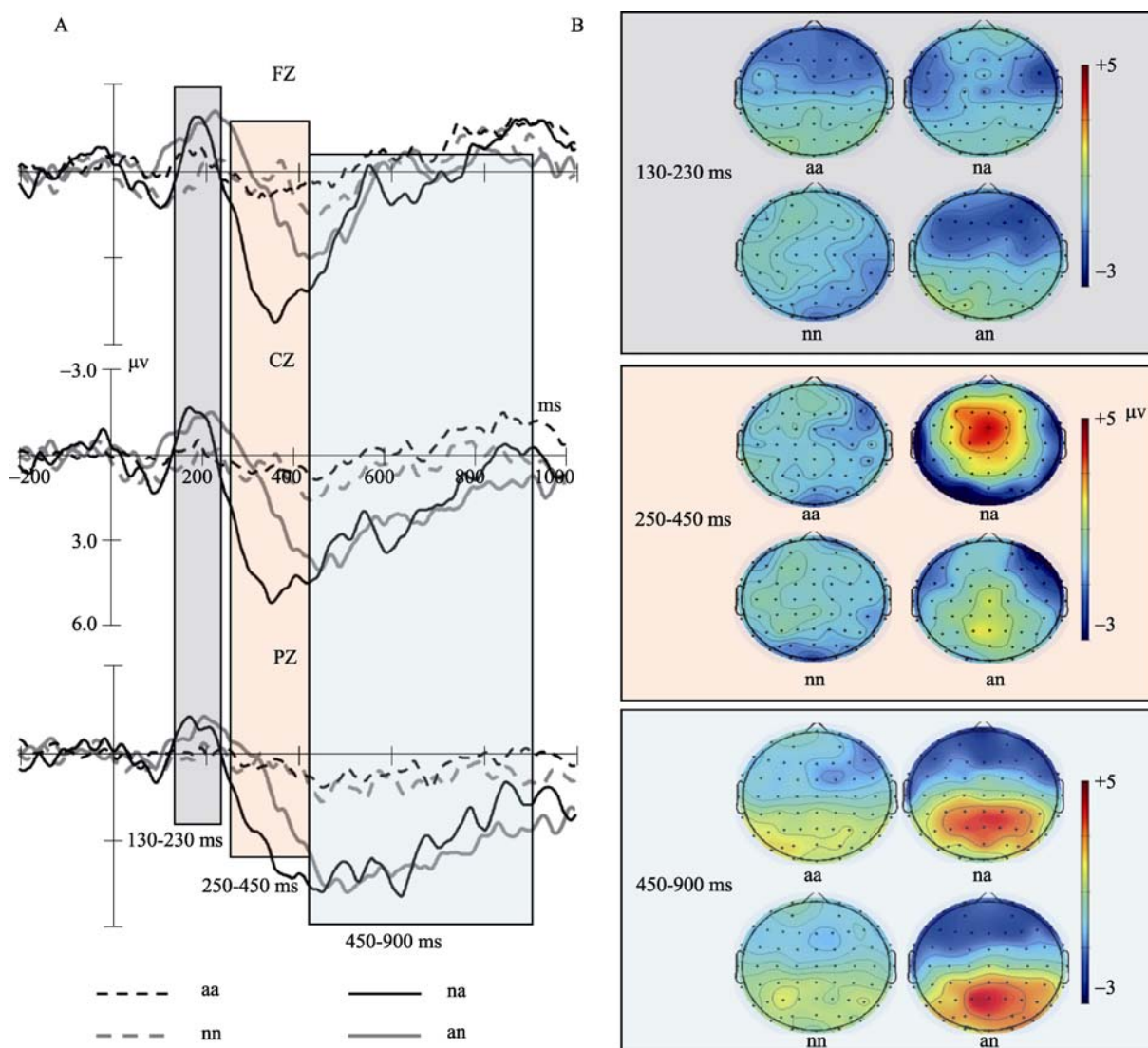


图3 四种韵律诱发的ERP。A: 各个条件在中线3个代表性电极点上的ERPs; B: 四种韵律诱发ERP在兴趣时间窗的地形图。

0.57]、CHANGE 主效应 [$F(1,13)=7.50, p<0.05, \eta^2=0.37$] 以及两因素的交互作用 [$F(8,104)=4.80, p<0.01, \eta^2=0.27$]。简单效应分析表明变化效应出现在中部和后部电极 [C5: $F(1,13)=5.07, p<0.05$, C6: $F(1,13)=3.61, p=0.08$, CZ: $F(1,13)=9.05, p<0.01$, P5: $F(1,13)=11.70, p<0.01$, P6: $F(1,13)=10.40, p<0.01$, PZ: $F(1,13)=22.92, p<0.001$]，且后中部电极 PZ 的幅度值最大 (见图3 ERP 波形图和地形图)。

3.3 ERO 结果

如图4可见, 包含情绪变化的韵律 (na 和 an) 相对于无变化的韵律 (aa 和 nn) 在 theta 和 beta 频段的能量都有显著的变化。下文分频率段做具体分析。

Theta 频段 (4~6 Hz) ERSP 在 200 到 600 ms 较长时程内出现差异, 重复测量方差分析表明 CHANGE

主效应显著, $F(1,13)=11.58, p<0.001, \eta^2=0.27$, 其他条件主效应及交互作用的均不显著, 说明情绪变化相对于无情绪变化的韵律都会引起 theta 的能量的增加。

Theta 频段 (4~6 Hz) ITC 在 0 到 500 ms 时程内出现差异, 重复测量方差分析表明 CHANGE 主效应显著, $F(1,13)=7.84, p<0.05, \eta^2=0.38$, 说明情绪变化相对于无情绪变化的韵律都会引起 theta 频段试次间相位一致性增高。其他条件主效应及交互作用的均不显著。

Beta 频段差异的频率和时程有较大差别, 将两种变化模式分别分析。从中性到愤怒的韵律选择 CZ 和 C6 两个电极点, 取 23 到 25 Hz 在 500 到 700 ms 内能量做重复测量方差分析, 发现有变化的韵

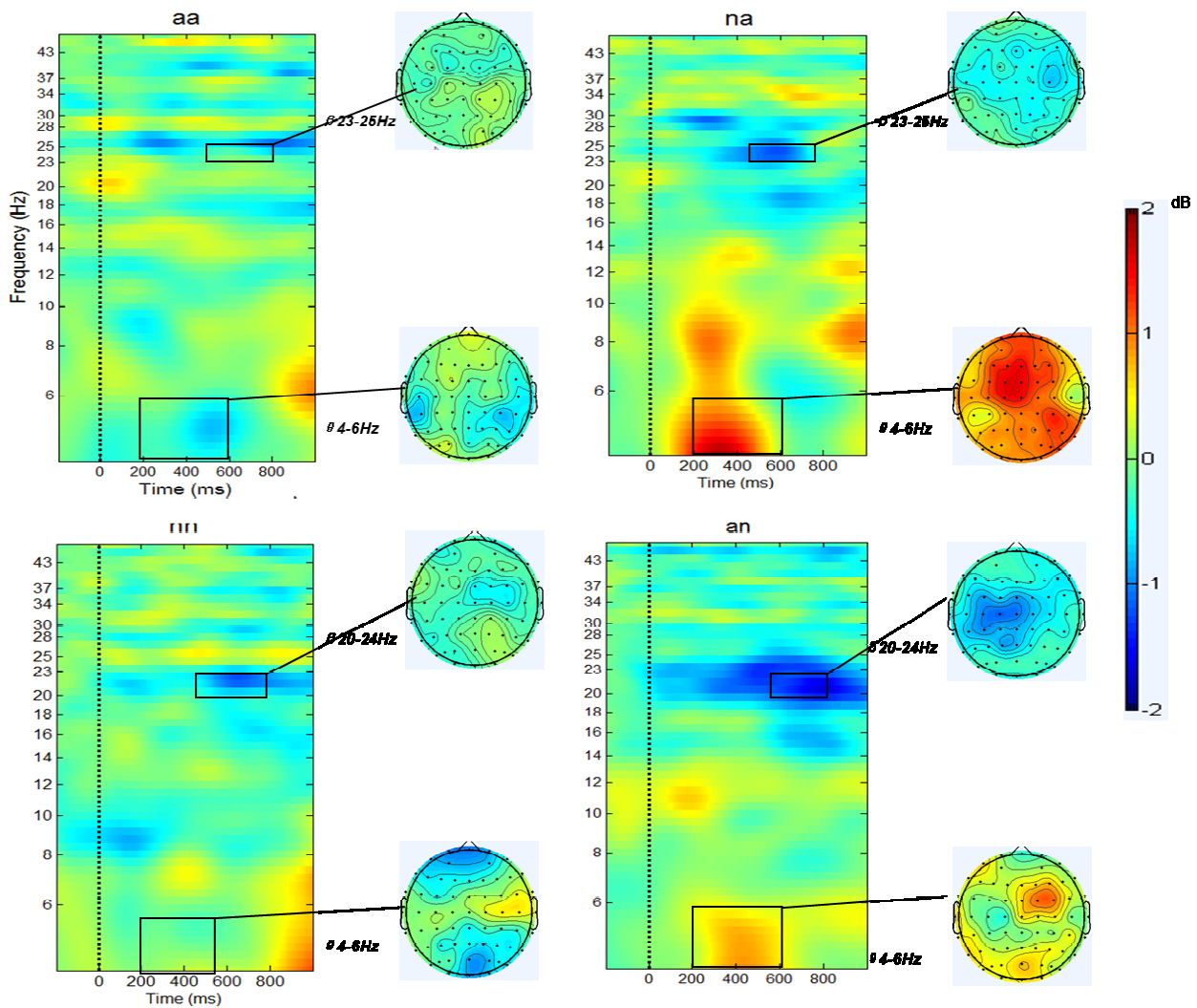


图4 四种韵律诱发的 ERSP。方形图为4种韵律所对应的 ERSP, 竖虚线对应于韵律中的拼接点, 红色代表能量增加, 蓝色代表能量降低, 单位为分贝。方形图右侧为有、无情绪变化韵律之间有显著差异的频段和时程所对应的能量分布地形图。

律能量显著降低, $F(13)=7.26$, $p<0.05$, $\eta^2=0.36$ 。而愤怒到中性的韵律相对无情绪变化的中性韵律 beta 能量降低主要出现在 20~24 Hz 的 450 到 700 ms 时程内, 选择 9 个电极做方差分析发现 CHANGE 主效应边缘性显著, $F(13)=4.41$, $p=0.056$, $\eta^2=0.25$, 进一步分析这种降低主要出现 C3、CZ 和 P5 几个电极。

愤怒变化到中性的韵律相对于无情绪变化的中性韵律 Beta 频段 ITC 差异出现在 19~21 Hz, 300 到 600 ms 的时程内。重复测量方差分析表明 CHANGE 主效应显著, $F(1,13)=8.65$, $p<0.001$, $\eta^2=0.40$, 说明从愤怒到中性变化的韵律在该频段试次相干性增加。而中性到愤怒相对于无变化的愤怒韵

律在 C3 点 24~26 Hz 频段内, 600 至 800 ms 的时程内 ITC 显著增加 $t(13)=2.38$, $p<0.05$ 。

4 讨论

本研究采用交互拼接技术构建语音情绪变化, 记录被试不关注情绪时语音情绪变化诱发的脑电以考查语音情绪变化加工的神经生理机制, 结果发现语音情绪变化诱发反映变化检测与整合的 N2/P3 复合成分, 同时 4~6 Hz 的 theta 能量显著增加, ITC 也显著提高, 而 beta 频段能量在有所下降。这些神经生理特征受情绪变化模式的调节, 激烈的情绪变化检测时间早, theta 能量和相位变化更强烈。下文对上述结果及意义分别进行讨论。

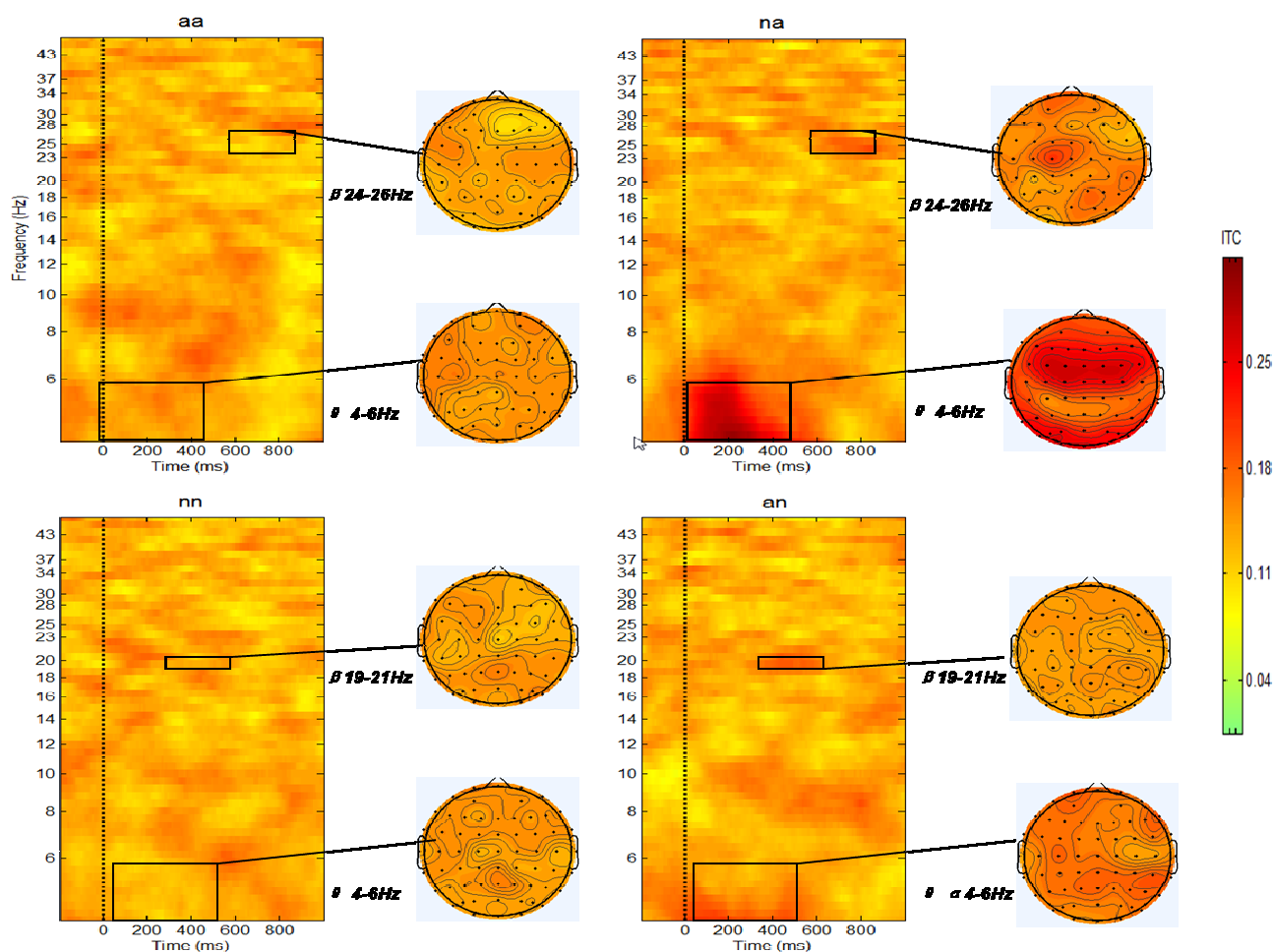


图5 四种韵律诱发的 ITC。方形图为 4 种韵律所对应的 ITC，竖虚线对应于韵律中的拼接点，红色代表试次间相关系数增大，绿色代表降低。方形图右侧为有、无情绪变化韵律之间有显著差异的频段和时程所对应的 ITC 分布地形图。

4.1 语音情绪变化加工的时域特征

当被试不关注情绪变化，仅就情绪韵律的声学特征完成任务时，语音情绪的变化仍然诱发早期负波加晚正波的复合成分。而且该复合成分的特征受到变化模式的调节，变化越激烈早期负波的潜伏期更早，前中部正成分(类 P3a)越明显。这些结果跟 Chen 等人(2011; 2012)的研究一致，跟 Kotz 等人(Kotz & Paulmann, 2007; Paulmann et al., 2012)的系列研究发现也有相似性，即情绪韵律变化诱发 PEP。但是，Kotz 等人的研究仅用了中性到情绪性的变化，情绪凸显性都较高，造成正成分十分明显，而且他们的研究也只报告了该正成分。问题在于，语音情绪变化的检测不可能需要 300ms 以后才被人脑所检测到，因为这跟 oddball 范式研究的结果相左(Goydke et al., 2004; Thönnessen et al., 2010)。因此，结合 Chen 等人(2011; 2012)的研究，本研究认为语音情绪变化诱发的早期负成分(从产生机

理、潜伏期和地形图分布来看，类似于 oddball 范式中的 N200)反应了人脑对语音情绪变化的检测和对新异刺激的注意。该负成分之后的正成分，包括中性到愤怒条件下的稍早的前中部分布正波以及两种转换中均存在的中后部分布的晚正成分 LPC，可能反应了对语音情绪变化的再加工以及重新整合。

事实上，这一推断跟语音和音乐的研究结果相一致。如语句中句法的违反会诱发早期成分 ELAN 加晚期正成分 P600 的复合波，Friederici 等人(Friederici, 2002)曾提出这种组合分别反应了句法违反的检测和再整合。另外，音乐旋律中个别音符音高的违反形成旋律句法的违反也有类似的发现(Koelsch, 2010; Schön et al., 2004)。本文关注的语音情绪也是负载于声音载体之上，表现一种情绪的韵律就有其固有的发展趋势，让听者逐步构建预期，一旦语音情绪发生变化，大脑将迅速检测到该变化并进行再加工。

这种变化检测与整合的现象可能具有进化论层面的意义。事实上, 语音情绪是比语言更为原始的一种交流信号, 在人类和其他物种都广泛使用。生活中不难发现, 人和其他物种都能迅速检测交际对方语音情绪的变化并采用相应的对策。这种推测得到本研究所观察到另一现象的支持, 即语音情绪变化的激烈程度影响情绪变化引起的 ERP 成分的特征。当从中性变成愤怒时, N200 潜伏期更短, 前中部分布的正成分更明显, 而反向变化, 情绪变化不够激烈时, 则 N200 潜伏期稍长, 只有反映后期整合的 LPC。这再次说明语音情绪变化的加工是一个变化检测到再整合的过程, 变化的模式具有调节作用。

4.2 语音情绪变化加工的频域特征

时域分析明晰了语音情绪变化加工是变化检测与再整合的过程, 且人脑对此十分敏感。但人脑是靠何种机制迅速检测这种变化并重新整合? 本研究通过对 EEG 数据的 ERSP 和 ITC 分析, 对这一问题进行探讨。结果表明, 语音情绪变化导致了 theta 频段(4~6 Hz) ERSP 的增加和 ITC 的增加, theta 能量提升发生在变化约 200 ms 之后, 而变化发生之后立即出现 ITC 的提升。此外语音情绪的变化导致了 beta (na: 23~25 Hz; an: 20~24 Hz) ERSP 的降低, 还导致了 beta (na: 24~26 Hz; an: 19~21 Hz) 频段 ITC 增加。

Theta 能量和相位变化的结果跟前人研究结果一致。Fuentemilla 等人(2008)曾经考察了听觉 MMN 的神经生理源, 发现前中部分布的 MMN 主要源于刺激变化引发的 theta 能量和试次间相位同步性的增加。Yordanova 等(1998)发现 theta 能量增加是时域 P300 成分的主要贡献者。另外, 其他一些有关规则违反的研究也发现期待违反引起 theta 频段的能量增强, 能量增幅与违反强度之间存在线性关系(Cavanagh et al., 2010; Tzur & Berger, 2007)。将本研究关注的语音情绪变化置于声音变化研究的大背景之下, 可以很容易找到本研究与前人研究的相似性——都依靠一些刺激(大概率刺激或连续的同类刺激)来提取规则并形成对即将到来刺激的期待, 而随后刺激违反原有规则发展, 构成了期待违反。因此, 基于本研究的结果和已有的研究事实, 可以推知 theta 能量的变化和相位特征的改变是语音情绪变化快速检测并重新整合最主要的神经生理源。

除了 theta 能量和相位的变化之外, beta 频段的

能量和相位在语音情绪变化中也发生了显著的变化。Beta 能量(15~25 Hz)活动被认为和动作控制相关(Pfurtscheller & Lopes da Silva, 1999)。而言语相关研究中, beta 能量的功能还没有一致说法, Bastiaansen 等人认为 16~21 Hz beta 能量的降低可能反应了视觉输入的感知加工(Bastiaansen, van der Linden, Ter Keurs, Dijkstra, & Hagoort, 2005), 而另有研究则发现 13~30 Hz 能量下降和短语的结构违反相关, 认为该频段能量降低反映了语法规则违反的加工(Davidson & Indefrey, 2007)。最近有研究发现与期待不符的韵律结构也导致 beta (20~24 Hz)能量的降低, 可能反应了对这种反常韵律结构的再加工(Luo, Zhang, Feng, & Zhou, 2010)。本研究中, 两种含有情绪变化的韵律都导致 beta 能量的显著降低, 而 ITC 值升高, 出现的时间在较晚的 400 到 500 ms 之后。这跟前人研究结果一致, 可能反应了对韵律情绪期待违反的再加工。

和时域的结果相似, 语音情绪变化的模式对 ERSP 和 ITC 都有影响, 不过这种影响并不显著, 仅表现了一种趋势。这和 Chen 等人(2012)发现的“情绪变化越激烈, 其诱发的能量就越强烈”不完全一致, 可能反应内隐加工与外显加工的差异。beta 频段能量和相位特征在两种情绪变化中没有显著差异。这可能和该频段主要和对情绪变化的再加工有关, 因为在两种变化模式下, 都涉及到对韵律情绪的再加工和重新整合。

4.3 神经震荡分析对脑机制研究的贡献

传统的脑电研究多从时域分析数据, 对同质试次叠加后得到 ERP, 然后根据 ERP 出现差异的成分推测其后的认知加工过程。ERP 的长处在于其高时间精度, 对于探讨认知过程的时间进程有无可取代的优势, 非常适合回答“when”的问题。但是, 认知活动的神经生理基础在于神经细胞的电活动, 神经细胞群的电位能量变化以及相位变化可能才是认知活动的真正基础(Klimesch, Sauseng, Hanslmayr, Gruber, & Freunberger, 2007; Lopes da Silva, 2006)。ERP 只能就锁时又锁相的数据进行分析, 会丢掉不同频段能量变化和相位变化的信息, 因而无从回答认知活动是何以实现(how)的问题。如本研究中, 若只做 ERP 分析, 可知语音情绪变化诱发 N2/P3 复合成分, 进而推知语音情绪变化加工是变化检测与整合的过程, 但无法回答该过程在神经层面是如何实现的。

神经振荡分析避免直接的叠加平均, 而是对单

试次的数据进行小波分析或傅里叶分析,然后再进行多个试次的平均(Delorme & Makeig, 2004),最终获得事件相关的能量变化(ERSP)和相位变化(ITC),能更好地从神经振荡层面回答认知活动何以实现的问题。如本研究中,语音情绪变化的检测与整合就是因为神经细胞 theta 频段能量和 ITC 增高, beta 能量降低与 ITC 增加得以实现的,这就更好地回答了语音情绪变化加工的神经生理基础问题。近年来,从神经振荡层面考察认知活动何以实现的研究越来越多。除了以病人或动物为被试做颅内记录,精细地分析单个神经细胞或神经细胞群的能量及相位变化外,还以正常人为被试,运用脑电图和脑磁图记录数据做神经振荡分析(Luo, Liu, & Poeppel, 2010; Mu & Han, 2010),以解读认知活动的神经生理基础,有力地促进了认知神经科学的发展。

4.4 本研究的局限与未来研究方向

本研究从时域和频域考察语音情绪变化加工的神经生理机制,提出语音情绪变化加工是变化检测与再加工整合的过程,该过程主要依靠 theta 频段能量和相位变化得以实现。但本研究仍存在一些局限。首先,本研究仅考察中性与愤怒之间情绪变化加工的特点,没有涉及积极情绪以及不同情绪之间的转换,因而将本研究的结论推及所有语音情绪变化加工时需要谨慎对待。其次,由于实验时间限制,本研究仅用了一个录音人的语音作为实验材料,也可能限制了本研究的外部效度。另外,本研究虽然尝试回答了语音情绪变化加工神经生理基础中“when”和“how”的问题,但由于脑电研究空间分辨率较低,尚不能很好回答“where”的问题,即这种变化加工的脑区及大脑加工网络。

这些局限都可能是未来研究的方向。首先可以扩展研究的情绪和刺激的多样性,并且逐步引入更具有生态效度的刺激,提高研究的生态效度。其次,未来的研究还应当进一步推进神经振荡分析,对不同脑区之间的相干等特性进行深入分析。最后,也可以引入 fMRI 等研究手段,对语音情绪变化加工的脑区和网络进行深入探讨,进一步揭示语音情绪变化加工的神经生理机制。

5 结论

语音情绪变化的加工是变化检测与再整合的过程,反映为情绪变化诱发的 N2/P3 复合成分以及 theta 和 beta 频段能量和试次间相位同步性的变化。

语音情绪变化加工是通过 theta 频段能量和试次间相位同步性的增加, beta 频段能量的降低和相位一致性增加得以实现的。情绪变化的模式会调控情绪变化加工神经生理特点,变化越激烈,诱发 N2/P3 复合成分潜伏期越早, theta 能量和相位改变也相对强烈。

参 考 文 献

- Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636.
- Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, 39(2-3), 241–248.
- Bastiaansen, M. C., van der Linden, M., Ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, 17(3), 530–541.
- Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer [computer program] (Version 4.2). Retrieved July 2006, from <http://www.praat.org/>.
- Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, 41(2), 259–268.
- Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, 49(4), 3198–3209.
- Chen, X. H., Yang, J. F., Gan, S. Z., & Yang, Y. F. (2012). The contribution of sound intensity in vocal emotion perception: Behavioral and electrophysiological evidence. *PLoS One*, 7(1), e30278.
- Chen, X. H., Zhao, L., Jiang, A. S., & Yang, Y. F. (2011). Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biological Psychology*, 86(3), 158–167.
- D'Arcy, R., Connolly, J., Service, E., Hawco, C., & Houlihan, M. (2004). Separating phonological and semantic processing in auditory sentence processing: A high-resolution event-related brain potential study. *Human Brain Mapping*, 22(1), 40–51.
- Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Research*, 1158, 81–92.
- 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(1), 9–21.
- Dietrich, S., Ackermann, H., Szameitat, D., & Alter, K. (2006). Psychoacoustic studies on the processing of vocal interjections: How to disentangle lexical and prosodic information? *Progress in Brain Research*, 156, 295.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84.
- Fuentemilla, L., Marco-Pallarés, J., Münte, T. F., & Grau, C.

- (2008). Theta EEG oscillatory activity and auditory change detection. *Brain Research*, 1220(1), 93–101.
- Goydke, K. N., Altenmüller, E., Möller, J., & Mönte, T. F. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research*, 21(3), 351–359.
- Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, 11(2), 199–212.
- Jiang, A. S., Chen, X. H., & Yang, Y. F. (2009). The time course of verbal emotional prosody processing. *Advances in Psychological Science*, 17(6), 1109–1115. [江爱世, 陈煦海, 杨玉芳. (2009). 言语情绪韵律加工的时间进程. *心理科学进展*, 17(6), 1109–1115.]
- Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews*, 31(7), 1003–1016.
- Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, 14(3), 131–137.
- Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, 1151(2), 107–118.
- Lopes da Silva, F. H. (2006). Event-related neural activities: What about phase? *Progress in Brain Research*, 159, 3–17.
- Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, 8(8), e1000445.
- Luo, Y., Zhang, Y., Feng, X., & Zhou, X. (2010). EEG oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience*, 169(2), 154–164.
- Magne, C., Schon, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199–211.
- Mu, Y., & Han, S. H. (2010). Neural oscillations involved in self-referential processing. *NeuroImage*, 53(2), 757–768.
- Näätänen, R. (2007). The mismatch negativity: Where is the big fish? *Journal of Psychophysiology*, 21(3-4), 133–137.
- Newman, R., Connolly, J., Service, E., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, 40(4), 640–647.
- Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, 50(7), 1609–1620.
- Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research*, 1217(4), 171–178.
- Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context. *Brain and Language*, 105(1), 59–69.
- Pfurtscheller, G., & da Silva, F. H. L. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology*, 110(11), 1842–1857.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148.
- Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology*, 41(2), 103–146.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341–349.
- Thierry, G., & Roberts, M. (2007). Event-related potential study of attention capture by affective sounds. *Neuroreport*, 18(3), 245–248.
- Thönnessen, H., Boers, F., Dammers, J., Chen, Y. H., Norra, C., & Mathiak, K. (2010). Early sensory encoding of affective prosody: Neuromagnetic tomography of emotional category changes. *NeuroImage*, 50(1), 250–259.
- Tzur, G., & Berger, A. (2007). When things look wrong: Theta activity in rule violation. *Neuropsychologia*, 45(13), 3122–3126.
- 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(3), 555–559.
- Winkler, I., Denham, S., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13 (12), 532–540.
- Yordanova, J., & Kolev, V. (1998). Single-sweep analysis of the theta frequency band during an auditory oddball task. *Psychophysiology*, 35(1), 116–126.

Neurophysiological Mechanism of Implicit Processing of Vocal Emotion Transition

CHEN Xuhai^{1,2}; YANG Xiaohong²; YANG Yufang²

(¹ Key Laboratory of Behavior and Cognitive Psychology in Shaanxi Province, School of Psychology,

Shaanxi Normal University, Xi'an 710062, China)

(² State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)

Abstract

Successful decoding of vocal emotion is critical for social interaction. Specifically, it is adaptively important to detect vocal emotion transition in time, since changes in vocal emotion are important signals for detecting the speakers' emotion change and are common in spoken interactions. Despite the fact that human beings are born with competence to process vocal emotion transition efficiently, the underlying neurophysiological mechanism remains largely unclear.

To answer this question, the present study acquired electroencephalogram (EEG) during implicit processing of two types of vocal emotion transitions (neutral to angry and angry to neutral) and their control unchanged vocal emotion from 15 healthy volunteers. Fifty sentences of neutral content produced by a trained native male actor of Mandarin Chinese in neutral and angry prosodies served as original materials. Then we constructed sentences with vocal emotion transition through cross-splicing.

The participants were required to perform sound intensity change judgment when the EEG was recorded. In addition to the ordinary ERP analysis in time domain, we also conducted the analysis in frequency domain including event related spectral power (ERSP) and inter-trial coherence (ITC) to specify the neurophysiological source of vocal emotion transition processing.

The results indicated that vocal emotion transition elicited N2/P3 complex, as well as theta (4~6 HZ) band power (ERSP) and inter-trial coherence (ITC) increase, irrespective of transition types. In addition, vocal emotion transition induced beta band power decrease and ITC increase. Moreover, these effects were modulated by transition types, specifically, the more intense transition of emotion resulted in shorter latency of N2/P3 complex, and much stronger increase of theta band ERSP and ITC enhancement, and different features in beta band power change.

These findings suggested that the processing of vocal emotion transition is a process consisting of change detection and reintegration, similar to the processing of other auditory materials, such as spoken language and music. Moreover, theta and beta band power change and ITC increase appear to be the main neurophysiological source of the processing of vocal emotion transition. In addition, the present study implies that the analysis in frequency domain can be a useful tool for verifying the neurophysiological mechanism of cognitive processing.

Key words vocal emotion; neuro-oscillation; ERP; ERSP; ITC