

# 情绪语音调节面孔表情的识别:ERP 证据\*

郑志伟 黄贤军\*\*

(首都师范大学心理系,北京,100048)

**摘要** 采用事件相关电位(ERP)技术考察了情绪语音影响面孔表情识别的时间进程。设置效价一致或不一致的“语音-面孔”对,要求被试判断情绪语音和面孔表情的效价是否一致。行为结果显示,被试对效价一致的“语音-面孔”对的反应更快。ERP结果显示,在70-130ms和220-450ms,不一致条件下的面孔表情比一致条件诱发了更负的波形;在450-750ms,不一致条件下的面孔表情比一致条件诱发更正的后正成分。说明情绪语音对面孔表情识别的多个阶段产生了跨通道影响。

**关键词** 情绪语音 面孔表情 跨通道 事件相关电位

## 1 前言

外部刺激的情绪意义在极短的时间内得到有效评估,是个体做出适当的情绪行为反应的重要前提(Phillips, Drevets, Rauch, & Lane, 2003)。在社会交往中,面孔表情是最重要的非言语情绪刺激,现有大量研究考察了面孔表情加工所涉及到的心理和神经机制(Eimer & Holmes, 2007)。但个体还需要持续整合来自多个感觉通道的情绪信息才能准确判断他人的情绪状态,而现有考察情绪背景信息(比如听觉通道的语音信息、音乐信息和视觉情绪信息)在面孔表情识别中的作用的研究却相对较少(de Gelder et al., 2006)。

已有的研究结果表明听觉通道的情绪语音信息能够影响个体对面孔表情的理解。在行为反应上,Ethofer等人(2006)发现,当被试观看伴随有情绪语音的面孔表情时,他们会联合两个通道的信息做出情绪判断。比如,与无语音刺激伴随相比,伴随恐惧语音出现的恐惧面孔的情绪强度会被评价为更强。另外,与不一致刺激对相比,具有一致情绪内容的语音-面孔刺激对会导致面孔表情分类的正确率的增加和反应时的减少(Pell, 2005)。

此外,研究者还使用事件相关电位技术(ERP)考察听觉通道的情绪信息影响面孔表情识别的时间动态性。Logeswaran 和 Bhattacharya (2009)发现中性面孔在高兴音乐作为启动项的条件下,比在悲伤音乐作为启动项的条件下诱发走向更负的、额中区分布的N1成分,表明具有情绪意义的音乐可以自上

而下的影响中性面孔的早期加工。Paulmann 和 Pell (2010)考察了情绪语音对面孔表情识别的跨通道影响,结果显示,与情绪语音效价不一致的面孔表情比一致条件下的面孔表情诱发了位于头皮顶区走向更负的N400成分。N400是刺激意义加工的指标(Brown, Hagoort, & Chwilla, 2000),因而她们认为该实验效应反映了情绪语音背景对面孔表情的意义评估的影响。

除听觉信息外,视觉情绪背景信息也会影响面孔表情的加工。Werheid, Alpay, Jentzsch 和 Sommer (2005)的面孔启动研究的ERP结果显示,与启动面孔效价不一致的面孔表情比效价一致的面孔表情在100-200ms诱发了更负的波形,研究者认为反映了面孔早期加工的促进作用;与启动面孔效价不一致的目标面孔在500-600ms诱发了走向更正的顶枕区分布的LPC(late positive component)。该效应可能反映了面孔效价发生改变导致面孔与个体的相关性增大,使得个体投入更多的注意资源用于加工面孔表情。这一解释与情绪图片加工的研究一致,研究表明情绪图片要比中性图片诱发波幅更正的LPC,研究者认为更正的LPC反映了情绪刺激具有更强的、内在的动机意义,因而吸引了更多的注意资源(Schupp et al., 2000)。

综上所述,情绪背景信息对面孔表情识别的影响可能发生在多个阶段,如早期的知觉加工,中期的意义评估以及后期的加工。本研究主要关注的是情绪语音信息对面孔表情识别的影响。虽然对这一问题的研究已有一定的成果,但仍然存在一些问题。

\* 本研究得到国家自然科学基金(31100816)、教育部人文社科青年基金项目(09YJCLX021)和高等学校博士学科点专项基金(20101108120005)的资助。

\*\* 通讯作者:黄贤军。E-mail: hxj\_cnu@yahoo.cn

一方面,以往研究所使用的语音材料为从具有情绪语调的句子中截取的语音片段(Paulmann & Pell, 2010),情绪语音的声学变化是随言语内容而不断扩展的,从情绪语句中截取的语音片段缺乏连贯性,不能很好的体现说话人的情绪状态。另一方面,关于情绪语音信息能否影响面孔表情的早期和后期加工,目前还没有相关证据。

本研究采用具有高时间分辨率优势的ERP技术和情绪启动范式考察情绪语音调节面孔表情识别的时间进程。语音材料为具有情绪语调的中性词。由于日常生活中个体需要关注语音和面孔两方面信息,为了让被试注意语音和面孔表情,采用了效价匹配任务。研究假设为,如果情绪语音调节面孔表情的早期阶段的加工,效价不一致的面孔表情可能要比效价一致的面孔表情诱发更负的N1成分;如果情绪语音调节面孔表情情绪意义评估阶段的加工,与情绪语音效价不一致的面孔表情可能会诱发更大的类似于N400的负波;在后期加工阶段,与语音效价不一致的面孔表情可能会诱发更正的LPC。

## 2 研究方法

### 2.1 被试

16名在校大学生参加了实验,男女各8人。被试年龄范围为17~25岁,右利手,听力、视力或者矫正视力正常。被试实验完毕获少量报酬。

### 2.2 刺激材料

从《现代汉语词典》(第5版)中选出中性双字名词120个,对这些候选词在7点量表上进行熟悉度评定(1表示非常不熟悉,7表示非常熟悉)和效价评定(1表示非常不愉快,7表示非常愉快)。评定结果显示平均熟悉度为6.32( $SD=0.76$ ),平均效价值为4.08( $SD=1.21$ )。将这120个双字词随机分成两组,将这两组词汇随机分配给两名播音员(一名男性,一名女性),每组词汇中有一半用高兴语调表达,另一半用愤怒语调表达,同时进行声音采集。语音是单声道16位采样,采样率为44100Hz。录音完毕后对语音材料进行剪切,获得60个高兴语音词和60个愤怒语音词。对情绪语音词进行评定,高兴语音词和愤怒语音词的平均识别率均在98%以上(高兴语音词98.6%;愤怒语音词99.4%)。情绪语音词的平均时长为596ms( $SD=118ms$ )。

从中国化面孔情绪图片系统(CFAPS)(王妍,罗跃嘉,2005)选取40张黑白情绪面孔,其中男性面孔(高兴10张,愤怒10张)和女性面孔(高兴10张,愤怒10张)各半。高兴面孔的平均效价值为6.57,平均唤醒度为5.82;愤怒面孔的平均效价值为2.69,平均唤醒度为6.28。高兴面孔与愤怒面孔的效价值差异显著( $p < .001$ ),唤醒度差异不显著( $p > .05$ )。

以上材料组成了240个语音-面孔对,启动项为情绪语音,目标项为面孔表情。其中包括120个效价一致的刺激对(高兴-高兴和愤怒-愤怒各半)和120个效价不一致的刺激对(愤怒-高兴和高兴-愤怒各半)。每个语音词重复呈现2次,每张面孔重复呈现6次。每一对语音-面孔对应的性别一致。

### 2.3 实验程序

正式实验包括6组,每组包含40个试次,其中效价一致刺激对和效价不一致刺激对各半。实验过程中,被试坐在电屏蔽实验室里,与计算机屏幕距离为50cm。每个试次开始时会在屏幕中央出现一个注视点,注视点呈现500ms后开始呈现情绪语音,语音结束后呈现500ms的面孔表情,面孔视角为 $4.99^\circ \times 6.84^\circ$ ,语音与面孔之间的时间间隔(ISI)为1000ms,屏幕空白2000~2200ms之后开始下一试次(如图1所示)。要求被试在面孔呈现后尽可能快且准确地判断语音与面孔之间的情绪效价是否一致,按键反应。被试被告知在实验过程中控制身体移动和眨眼。组顺序和按键顺序在被试之间进行了平衡。

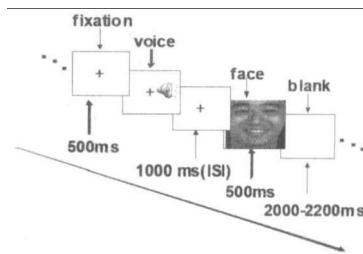


图1 实验流程图

### 2.4 脑电记录与分析

采用Neuroscan ERPs记录与分析系统,根据国际10-20系统扩展的64导Ag/AgCl电极帽记录EEG,左眼眶上下记录垂直眼电,双眼外侧记录水平眼电。参考电极置于左侧乳突,接地点在FPz和Fz的中点,A/D采样频率为500Hz,滤波带通为0.05~40Hz,电极与头皮之间的阻抗小于5kΩ。对记录获得的连续EEG进行离线分析处理,排除眼动伪迹,其它在±75μV之外的伪迹在迭加的过程中被剔除。以左右乳突的代表平均为参考电压进行修正。只叠加反应正确的试次中面孔表情诱发的ERP波

形。分析时程为 -100~800 ms, 面孔呈现前 100 ms 的脑电为基线。

通过对总平均波形图的观察, 确定 ERP 分析的时窗为 70~130 ms, 220~450 ms 和 450~750 ms。根据头皮分布与电极之间的关系以及前人研究 (Righart & de Gelder, 2007), 不同时窗的分析选取了不同的电极位置。时窗 70~130 ms, 计算的电极为位于额中区的 F3/4, FC3/4, C3/4, Fz/FCz/Cz; 时窗 220~450 ms, 计算的电极为位于中央顶区的 C3/4, CP3/4, P3/4, Cz/CPz/Pz; 时窗 450~750 ms, 计算的电极为位于顶枕区的 P3/4, PO3/4, O1/2, Pz/POz/Oz。对每一个时窗的平均波幅进行 2(效价一致性)  $\times$  2(面孔效价)  $\times$  9(电极位置) 的重复测量方差分析。显著性水平设在 0.05, 统计分析结果选用 Greenhouse-Geisser 纠正法。

### 3 结果

#### 3.1 行为结果

表 1 各个条件下的平均反应时和正确率

效价一致性	面孔类型	反应时/ms (SD)	正确率/% (SD)
一致	高兴面孔	743(194)	96.8(3.5)
	愤怒面孔	852(256)	89.8(8.0)
不一致	高兴面孔	908(216)	92.5(6.7)
	愤怒面孔	933(260)	96.8(3.7)

表 1 显示了剔除反应时小于 300 ms 和大于 2000 ms 的试次(占所有试次的 1.5%) 的结果, 分别以反应时和准确率为指标进行 2(效价一致性)  $\times$  2(面孔效价) 的重复测量方差分析。结果显示, 在反应时上, 效价一致性的主效应显著, 被试对效价一致条件比效价不一致条件的反应更快,  $F(1, 15) = 31.406, p < .001$ 。效价一致性和面孔效价的交互作用不显著。在准确率上, 效价一致性主效应不显著。效价一致性与面孔效价的交互作用显著,  $F(1, 15) = 16.50, p < .01$ , 简单效应检验结果显示, 在高兴面孔水平上, 不一致条件比一致条件的正确率更低,  $F(1, 15) = 6.32, p < .05$ ; 在愤怒面孔水平上, 不一致条件比一致条件的正确率更高,  $F(1, 15) = 12.52, p < .01$ 。

#### 3.2 ERP 结果

图 2 左侧显示了情绪效价一致和不一致条件下面孔表情诱发的 ERP 的总平均图, 阴影部分显示了两种条件具有差异的时间窗口。右侧显示的是三个时窗内, 面孔表情在不一致条件和一致条件下的 ERP 差异波的地形图。

在 70~130 ms, 面孔诱发的 ERP 波形对应于分布

在头皮额中区的 N1 成分。2(效价一致性)  $\times$  2(面孔效价)  $\times$  9(电极位置: F3/4, FC3/4, C3/4, Fz/FCz/Cz) 重复测量方差分析结果显示, 情绪一致性的主效应显著,  $F(1, 15) = 10.737, p < .01$ , 情绪不一致条件下的面孔表情比情绪一致条件下的面孔诱发的波幅更负。效价一致性与面孔效价、电极位置之间的交互作用以及三重交互作用均不显著,  $ps > .05$ 。

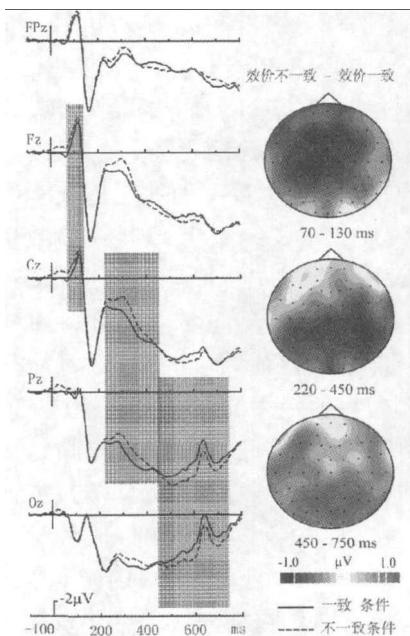


图 2 两种条件下面孔表情诱发的 ERP 总平均图及地形图

在 220~450 ms, 面孔诱发的 ERP 波形对应于分布在头皮中央顶区的 N400 成分。2(效价一致性)  $\times$  2(面孔效价)  $\times$  9(电极位置: C3/4, CP3/4, P3/4, Cz/CPz/Pz) 重复测量方差分析结果显示, 效价一致性与面孔效价之间的交互作用显著,  $F(1, 15) = 11.437, p < .001$ , 进一步的简单效应检验显示, 在高兴面孔水平上, 效价不一致条件比效价一致条件的电压值更负,  $F(1, 15) = 17.43, p = .001$ ; 在愤怒面孔水平上, 则不存在效价一致性效应,  $F(1, 15) = .33, p = .576$ 。效价一致性与电极位置之间的交互作用不显著, 效价一致性、面孔效价以及电极位置之间的三重交互作用也不显著,  $ps > .05$ 。

在 450~750 ms, 面孔诱发的 ERP 波形对应于分布在头皮顶枕区的 LPC。2(效价一致性)  $\times$  2(面孔效价)  $\times$  9(电极位置: P3/4, PO3/4, O1/2, Pz/POz/Oz) 重复测量方差分析结果显示, 效价一致性的主效应达到显著水平,  $F(1, 15) = 7.702, p < .05$ , 效价不一致条件下面孔表情比效价一致条件下的面孔表情诱发波幅更正的 LPC。效价一致性与面孔效价、电极位置之间的交互作用均不显著,  $ps > .05$ 。

## 4 讨论

在反应时上,效价一致条件要比效价不一致条件更快,表明在效价一致条件下的情绪语音促进了情绪面孔表情的识别。在正确率上,情绪效价的一致性效应受到面孔效价的调节:在高兴面孔水平上,一致条件比不一致条件更高;而在愤怒面孔水平上,一致条件则比不一致条件的正确率更低。这一结果与 Paulmann 和 Pell(2010)的行为结果相似,不同效价面孔表情的效价一致性效应的趋势相反。启动研究中除了存在行为结果不一致现象之外,还存在行为结果与 ERP 结果的不一致现象,即行为结果上不存在启动效应,ERP 结果上却存在该效应(Kotz, 2001)。Zhang, Lawson, Guo 和 Jiang(2006)认为行为指标对决策过程以及任务反应策略的敏感度较高,因而本研究中准确率上的不一致可能与个体在执行任务过程中所使用的反应策略有关。

ERP 结果显示,在 70 – 130ms,与情绪语音效价不一致的面孔表情诱发了波幅更负的、位于头皮额中区的 N1 成分。由于这一外源性成分受到注意因素的调节(Vogel & Luck, 2000),因而本研究中的效价不一致的面孔表情可能要比效价一致的面孔表情吸引更多的注意资源,以用于随后的更加精细的加工。这一额区分布的 N1 成分的变化可能反映了个体在面孔识别过程中自上而下的注意监控机制(Logeswaran & Bhattacharya, 2009)。本研究表明在面孔表情的早期知觉阶段,情绪语音就可以对其产生影响。

与 Paulmann 和 Pell(2010)的“语音 – 面孔”启动研究结果一致的是,本研究中的效价不一致条件下的高兴面孔表情在 220 – 450ms 诱发波幅更负的 N400 成分。N400 成分是意义加工的指标,研究显示与句子背景不一致的词汇通常会比与句子背景一致的词汇诱发波形更负的 N400(Kutas & Hillyard, 1980)。情绪启动的 ERP 研究中也发现了 N400 效应。比如,Zhang, Li, Gold 和 Jiang(2010)发现与背景图片效价不一致的情绪词可以诱发波幅更大的 N400,作者认为该成分反映了意义整合的困难程度。由此可见,N400 不仅仅敏感于词汇信息在言语背景中意义的整合,也敏感于情绪刺激与情绪背景信息的相互作用。本研究中的 ERP 效应与上述研究获得的 N400 效应相似,均反映了情绪刺激的意义在情绪背景中的整合,因而进一步提供了情绪语音与面孔表情识别在意义评估阶段相互作用的证据。

本研究得到的最后一个 ERP 效应是效价不一致条件下的面孔比一致条件下的面孔在 450 – 750ms 之间诱发波幅更正的 LPC。这一结果与之前的情绪启动研究结果一致(Werheid et al., 2005; Zhang et al., 2010)。图片加工研究中,情绪图片比中性图片诱发更大波幅的 LPC(Schupp, Junghffer, Weike, & Hamm, 2003)。研究者认为情绪刺激与个体的相关性更大,因而吸引了更多的注意资源,且刺激的唤醒度越高,吸引的注意资源越多。在本研究中,情绪效价不一致的面孔表情可能具有更大的个体相关性,进而导致唤醒度的增加。由此可见,在社会情境中变化的情绪刺激要比没有变化的情绪刺激被理解为更具相关性(Werheid et al., 2005)。

## 5 结论

综上所述,本研究通过设置效价一致和不一致的“语音 – 面孔”对,发现情绪语音促进面孔表情的识别,被试在一致条件下的反应更快,并且在 70 – 130ms、220 – 450ms 和 450 – 750ms 三个时间窗内,出现了情绪效价一致性效应,以上结果说明情绪语音可以在面孔表情识别的早期知觉阶段、意义评估阶段及后期阶段调节其加工。

### 参考文献

王妍,罗跃嘉.(2005).大学生面孔表情材料的标准化及其评定.《中国临床心理学杂志》,13,396–398.

Brown, C., Hagoort, P., & Chwilla, D. (2000). An event – related brain potential analysis of visual word – priming effects. *Brain and Language*, 72, 158 – 190.

de Gelder, B., Hanneke, K. M. M., Righart, R., van den Stock, J., van de Riet, W. A. C., & Tamietto, M. (2006). Beyond the face: Exploring rapid influences of context on face processing. *Progress in Brain Research*, 155, 37 – 48.

Eimer, M., & Holmes, A. (2007). Event – related brain potential correlates of emotional face processing. *Neuropsychologia*, 45, 15 – 31.

Ethofer, T., Ander, S., Erb, M., Droll, C., Royen, L., Saur, R., et al. (2006). Impact of voice on emotional judgment of faces: An event – related fMRI study. *Human Brain Mapping*, 27, 707 – 714.

Kotz, S. A. (2001). Neurolinguistic evidence for bilingual language representation: A comparison of reaction times and event – related brain potentials. *Bilingualism: Language & Cognition*, 4, 143 – 154.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203 – 205.

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters*, 455, 129 – 133.

Paulmann, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: How much is enough? *Cognitive, Affective, & Behavior Neuroscience, 10*, 230–242.

Pell, M. D. (2005). Nonverbal emotion priming: Evidence from the "Facial Affect Decision Task". *Journal of Nonverbal Behavior, 29*, 45–73.

Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry, 54*, 504–514.

Righart, R., & de Gelder, B. (2007). Impaired face and body perception in developmental prosopagnosia. *Proceedings of National Academy of Sciences of the United States of America, 104*, 17234–17238.

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology, 37*, 257–261.

Schupp, H. T., Junghffer, M., Weike, A. I., & Hamm, A. O. (2003). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science, 14*, 7–13.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*, 190–123.

Werheid, K., Alpay, G., Jentzsch, I., & Sommer, W. (2005). Priming emotional facial expressions as evidenced by event-related brain potentials. *International Journal of Psychophysiology, 55*, 209–219.

Zhang, Q., Lawson, A., Guo, C. Y., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin, 71*, 316–323.

Zhang, Q., Li, X. H., Gold, B. T., & Jiang, Y. (2010). Neural correlates of cross-domain affective priming. *Brain Research, 1329*, 142–151.

## Emotional Voice Modulates the Recognition of Facial Expression: Evidences from an ERP Study

Zheng Zhiwei, Huang Xianjun

(Department of Psychology, Capital Normal University, Beijing, 100048)

**Abstract** Continuous integration of information from multiple sensory inputs is very important for the daily life of human beings. But the mechanisms underlying the interaction of cross-modal stimulus processing failed to draw sufficient attention, especially when it comes to the cross-modal interaction of the stimulus containing emotional significance. This study aimed to investigate the neural mechanism of the interaction of emotional voice and facial expression.

Event-related potentials (ERP) technique and cross-modal priming paradigm were used to explore the influence of emotional voice on the recognition of facial expression. The materials consisted of 240 prime-target pairs using voices as primes and facial expressions as targets. Neutral semantic words were spoken with happy or angry prosody and followed by congruous or incongruous facial expressions. The participants were asked to judge the consistency between the valence of emotional voice and facial expression, during which ERPs were recorded. Each trial began with a central fixation cross presented for 500 ms. Then, the priming stimulus (emotional voice) was presented through headphones. The central fixation cross displayed on the screen until the target (facial expression) was presented. The inter-stimulus-interval (ISI) is 1000 ms. The facial expression was presented for 500 ms, followed by a black screen for 2000–2200 ms. After the presentation of facial expression, participants were instructed to indicate the consistency of valance between the emotional voice and facial expression by pressing a mouse button as quickly and accurately as possible.

The results were analyzed by repeated measures ANOVA. The response time (RT) results showed that participants responded more quickly to the congruous trials than the incongruous trials. It suggested the existence of the priming effect of emotional voice on recognition of emotional facial expression. The analysis of ERPs waveforms indicated that emotional voice modulated the time course of processing of facial expression. At the time window of 70–130 ms and 220–450 ms, facial expressions evoked more negative waveforms in incongruous trials than in congruous trials. At the time window of 450–750 ms, facial expressions evoked more positive late positive component (LPC) in incongruous trials than in congruous trials. The ERPs results suggested that emotional voice influenced the processing of emotional facial expression at the early perception stage, the emotional significance evaluation stage and the subsequent decision-making stage.

This study demonstrates that emotional voice can influence the processing of facial expression in a cross-modal manner. Also, it provides converging evidence for the interaction of multi-sensory inputs.

**Key words** emotional voice, facial expression, cross-modal, event-related potential (ERP)