

成人依恋风格对情绪语音声学特征的作用*

胡 涵¹ 顾文涛^{1,2†}

(1 南京师范大学 文学院 南京 210097)

(2 麦吉尔大学 沟通科学与障碍学院 加拿大蒙特利尔 H3A1G1)

2023 年 4 月 9 日收到

2023 年 6 月 25 日定稿

摘要 为探究说话人的依恋类型(安全型、超脱型、专注型、恐惧型)对情绪语音产出的影响,设计了符合语法规则但是无意义的伪句,招募了有恋爱经验的被试,采用阈下词汇启动范式激活依恋系统后,被试观看 4 种基本情绪(开心、愤怒、悲伤、恐惧)的诱发影片,用体验到的情绪向想象中的恋爱伴侣说出这些句子。对递归特征消除算法筛选出的每句 14 个声学参数做半参数重复测量多元方差分析,结果显示依恋类型和情绪类别的主效应显著、交互效应不显著;聚集性分层聚类分析发现,在声学特征空间中,超脱型和专注型距离最近,而安全型则远离其他类型;有监督分类发现,14 个声学参数可有效区分 4 类依恋类型;特征重要性分析发现,韵律参数对分类的贡献较大;累积局部轮廓分析发现,4 类人群间基频特征的差异在各种情绪上基本一致,但是音质特征的差异受到情绪类别的影响。研究揭示了依恋类型对情绪语音声学特征的作用,验证了不同依恋类型在情绪调节策略上的差异,为个性化人机语音交互技术的发展提供了科学依据。

关键词 成人依恋类型,情绪语音,声学特征,情绪调节策略,有监督分类

PACS: 43.66, 43.70

DOI: 10.12395/0371-0025.2023051

CSTR: 32049.14.11-2065.20243051

The effects of adult attachment style on acoustic characteristics of emotional speech

HU Han¹ GU Wentao^{1,2†}

(1 School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210097)

(2 School of Communication Sciences and Disorders, McGill University, Montreal, Canada H3A1G1)

Received Apr. 9, 2023

Revised Jun. 25, 2023

Abstract This study explored the effect of a speaker's attachment style (secure, detached, preoccupied, and fearful) on acoustic characteristics of emotional speech. Grammatical but meaningless pseudo sentences were designed, and participants with romantic relationship experiences were recruited. After activating their attachment systems with the subliminal lexical priming paradigm, the participants watched the videos that tended to evoke one of the four basic emotions (happiness, anger, sadness, and fear). Following this, they expressed these sentences with the corresponding emotions to their imagined romantic partners. Based on 14 acoustic parameters per utterance selected using the recursive feature elimination algorithm, semi-parametric repeated measures multivariate analysis of variance shows significant main effects of attachment style and emotion type, but does not show a significant interaction effect between them. Agglomerative hierarchical cluster analysis shows that, in the acoustic space, "dismissing" and "preoccupied" are the closest, while "secure" is the farthest from other attachment styles. Supervised classification algorithms effectively differentiate the four attachment styles based on 14 acoustic parameters, with prosodic parameters contributing more in terms of feature importance analysis. Furthermore, accumulated-local profiles analysis indicates that the variations in fundamental frequency characteristics among the four attachment styles remain basically consistent across the four emotions, but the differences in timbre and voice quality characteristics are influenced by emotion type. In summary, this study unveils the impact of attachment style on emotional speech and confirms the variations in emotion regulation strategy among individuals with four attachment styles. This

* 国家社会科学基金重大项目(13&ZD189)及国家留学基金资助

† 通讯作者: 顾文涛, wtgu@njnu.edu.cn

provides a scientific foundation for the development of personalized human-machine speech communication technologies.

Keywords Adult attachment style, Emotional speech, Acoustic characteristics, Emotion regulation strategy, Supervised classification

引言

情绪在人际沟通和人机交互中都扮演着重要角色,而情绪语音作为情绪信息的重要载体,受到语音学、心理学和人工智能领域的广泛关注。目前,已有数量可观的研究揭示了典型情绪和语音特征之间的稳定联系^[1-2]。但是,研究结论并不一致,这可能由多方面因素造成^[3-4],其中说话人的个体差异是一个重要原因,而人格特质便属此列。依恋理论^[5]为探究人格特质和情绪之间的关系提供了可靠的理论框架。本文拟通过考察不同依恋类型个体表达4种基本情绪(开心、愤怒、悲伤、恐惧)时的语音声学特征,揭示人格特质对情绪语音表达的影响,为个性化人机语音交互技术的发展提供科学依据。

Bowlby^[5]提出的依恋理论,核心假设为:个体在生命早期与抚育者的互动经验,会逐渐内化为有关自我(是否有价值)与他人(是否可信赖)的心理表征(也称内部工作模式),解释并指导着个体在浪漫关系中的情绪和情绪调节^[5-6]。个体依恋风格的定义方法,有类型法和维度法。Bartholomew等^[7]的类型法,是根据内部工作模式的差异,将依恋风格划分为4种类型,即安全型、超脱型、专注型、恐惧型,后3类都可归为非安全型。Brennan等^[8]的维度法,则是在依恋回避、依恋焦虑这两个连续维度上定义依恋风格,依恋回避是有关他人的表征、表示个体对亲密关系感到舒适与依赖的程度,而依恋焦虑是有关自我的表征、表示个体对被抛弃与被拒绝感到担忧的程度。依恋类型和依恋维度的对应关系如下:安全型对应较低的依恋回避和依恋焦虑值,超脱型对应较高的依恋回避值、较低的依恋焦虑值,专注型对应较低的依恋回避值、较高的依恋焦虑值,恐惧型对应较高的依恋回避和依恋焦虑值。

在Bowlby^[5]的依恋理论中,情绪调节的地位尤为重要。黄于飞等^[9]系统回顾了不同依恋类型情绪调节策略的选择偏好,指出安全型个体倾向于使用更具建设性和灵活性的情绪调节策略(如认知重评),超脱型个体倾向于采用保持认知距离和情感疏离等下调情绪的策略(如表达抑制)。专注型个体采用的情绪调节策略,争议颇大,有学者认为这类人群倾向于采用强化情绪表达的策略吸引对方的注意(也称

过度激活)^[10],也有学者认为这类人群倾向于采用表达抑制策略^[11]。有关恐惧型个体情绪调节策略的研究较少,Winterheld^[12]指出恐惧型个体兼具安全型和超脱型的情绪调节策略。Walker等^[13]利用元分析技术对26项研究成果整合后发现,依恋回避和依恋焦虑维度均与情绪控制能力呈显著负相关($r = -0.21$; $r = -0.16$)。综上,由于不同依恋类型在情绪调节策略上各有偏好,而情绪调节策略直接影响情绪输出的结果^[14],所以依恋类型理应对情绪表达产生影响。

语音承载着丰富的情绪信息,已有大量声学 and 感知实验的研究揭示了各类基本情绪(如开心、愤怒、悲伤、恐惧)的语音特征或声学线索^[12,15]。Johnstone和Scherer^[16]总结了各类基本情绪的语音表现:开心的基频均值、基频范围、能量较大;愤怒和恐惧的基频均值、基频范围、能量较大,语速较快;悲伤的基频均值、基频变异、能量较小,语速较慢。但是,这些映射关系并不稳定,正如文献[1]的表2显示,开心的高频能量、愤怒的基频均值、悲伤的基频均值和高频能量都未获得一致认可;Scherer^[17]也发现实验研究得到的语音特征规律与理论预测存在诸多不吻合之处。

近年的研究表明,嗓音音质也是情绪语音的重要线索。Gobl和Chasaide^[18]让被试对不同发声类型的合成语音做情绪评价,发现气嗓音(breathy voice)与悲伤、亲密情绪相关,紧嗓音(tense voice)与愤怒、开心情绪相关,嘎裂声(creaky voice)与无聊情绪相关。Yanushevskaya等^[19]也支持该结论,但是Murray和Arnott^[20]的结论正好相反,认为愤怒和开心情绪对应着气嗓音,而悲伤情绪对应着共振性强的发声类型(常态或紧嗓音)。可见,特定情绪具有特定语音特征的假设并非普遍成立,存在其他潜在因子调节着两者的关系。

Johnstone和Scherer^[16]认为,情绪语音的研究结果不一致,被试间的个体差异和文化差异是重要原因之一。Bryant^[3]对情绪语音的文化差异做了系统的回顾和梳理。但是,说话人特质对情绪语音的调节作用,很少有研究涉及。成人依恋风格属于人格特质的一种,而且与情绪调节策略有密切联系,因此依恋类型可能是调节情绪和语音间映射关系的重要因素。

目前仅有为数不多的研究考察了说话人依恋类型和语音特征之间的关系。Spinelli 等^[21]通过考察消极和积极词汇的语音韵律特征,发现相比于中性词语,安全型个体表达消极或积极词汇时都有更大的基频变异和更慢的语速,超脱型个体表达消极词汇时显示出较大的基频变异和较快的语速、表达积极词汇时却表现出较小的基频变异和中等的语速。Monti 等^[22]考察了依恋维度与歌手嗓音特征间的关系,发现依恋回避和依恋焦虑值均可正向预测基频扰动和振幅扰动,说明高回避和高焦虑个体的声带振动更加不规则。胡涵和顾文涛^[23]研究了依恋类型对亲密话语的韵律和嗓音特征的影响,发现高回避个体的语音特征传达出强势和无聊等态度(基频均值、谐波差、谐噪比较小)。可见,语音特征可作为反映依恋类型的有效指标。

然而,上述研究多是直接考察个体依恋类型对语音声学特征的影响,没有考虑作为纽带的情绪的中介作用。胡涵和顾文涛^[23]初步探索了亲密话语,但尚未区分亲密关系中表达不同情绪的话语,还不能直接验证依恋类型对情绪调节策略的作用。因此,本文拟进一步分析不同依恋类型个体在几种基本情绪下的语音声学特征,揭示成人依恋风格对情绪语音的作用,为依恋理论的情绪调节假设提供语音声学数据的支持,为个性化人机情感语音通信的发展奠定基础。

1 方法

1.1 实验材料

(1) 情绪影片

采用影片诱发情绪。在中国情绪影像材料库(CEVS)^[24]中选取诱发开心、愤怒、悲伤、恐惧情绪的影片各 5 部。招募 15 名来自某高校的研究生(3 男 12 女;年龄: $M = 25.33$, $SD = 2.82$, 在本文中 M 与 SD 分别表示变量的均值与标准差)评估影片。要求被试每看完一部影片,根据自身体验到的情绪,在开心、愤怒、悲伤、恐惧、中性 5 个选项上分别用 5 度制打分(1~5 表示从不符合逐渐过渡到符合),将获得最高分的选项作为该被试的选择。最终,根据被试选择与目标情绪一致的人数,筛选出每种情绪排名前 3 的影片(选择与目标情绪一致的人数均为 12~15 人),作为情绪诱发材料。开心、愤怒、悲伤、恐惧情绪的平均得分为 4.16 ($SD = 1.09$), 4.29 ($SD = 0.87$), 4.31 ($SD = 0.93$), 4.27 ($SD = 0.96$)。Friedman 秩

和检验的结果显示,筛选出的每类情绪影片在 5 个情绪选项上的得分均存在显著差异 ($ps < 0.001$)。经 Holm 修正的 Demsar 事后检验显示,筛选出的每类情绪影片在目标情绪选项上的得分均显著高于其他情绪选项 ($ps < 0.05$)。

(2) 语料设计

为了排除语义信息的干扰,设计了符合语法规则但是语义无效的伪句,作为情绪表达的语料。伪句的设计规则参考文献^[25],首先构造符合语法和语义规则的自然语句,然后使用随机选取的汉字组成的假词替换自然句中的实词,但不替换功能词以保留语法信息。招募 15 名有语言学背景的研究生(与影片评估完全相同的被试)对 40 个伪句文本做自然度评估,采用 5 度制计分(1~5 表示由不自然逐渐过渡到自然)。最终,筛选出自自然度高于 3.5 分的 15 个伪句作为目标句 ($M = 4.09$, $SD = 0.36$),见附录 A。其中 27% 为 7 字句, 33% 为 8 字句, 13% 为 10 字句, 27% 为 11 字句。Kruskal-Wallis 秩和检验发现,不同长度的伪句在自然度上的差异未达显著 ($\chi^2 = 5.83$, $df = 3$, $p = 0.12$)。

1.2 测量工具

亲密关系体验量表(ECR)中文版^[26]。ECR 用于测量被试的依恋回避和依恋焦虑值,将其代入判别公式可计算出依恋类型。该量表包含依恋回避分量表(18 题;如:我倾向于不跟恋人过分亲密)和依恋焦虑分量表(18 题;如:我有点担心会失去恋人)。被试根据依恋经验在 7 分制量表上评价各题项的描述与自身感受的相符程度(1~7 表示由完全不符合逐渐过渡到完全符合)。最终各自计算两个分量表 18 个题项的均值作为依恋回避和依恋焦虑值。ECR 具有良好的信效度^[26],本研究中依恋回避、依恋焦虑分量和总量表的内部一致性系数 McDonald ω 分别为 0.89, 0.83, 0.86。

背景信息采集。背景信息的搜集参考文献^[23],包括性别、年龄(单位:岁)、恋爱次数(单位:次)、是否正在恋爱、与伴侣年龄差的绝对值(单位:月)、是否为异地恋、恋爱阶段(采用 5 度制计分,1~5 表示从平淡逐渐过渡至热烈)。

1.3 被试选取

从某高校的大学生中招募被试,采用了以下标准筛选:(1)有恋爱经历,(2)熟练地使用普通话进行交流,(3)无言语听力疾病史,(4)近期与伴侣之间无争吵和分手经历。最终,选取了符合标准的 48 名被

试, 被试信息详见表 1。其中, 13 男 35 女; 年龄 $M = 24.10$ ($SD = 3.08$); 62% 正在恋爱, 其中异地恋占 60%; 依恋回避和依恋焦虑的均值分别为 2.97 ($SD =$

0.74) 和 3.52 ($SD = 0.68$), 两者的相关性未达显著 ($r = 0.05$, $p = 0.757$); 由依恋维度值计算可得, 安全型有 18 名、超脱型和专注型各有 7 名、恐惧型有 16 名。

表 1 不同依恋类型被试的基本信息比较

	背景变量	总计 $N = 48$	安全型 $N = 18$	超脱型 $N = 7$	专注型 $N = 7$	恐惧型 $N = 16$	四类比较的 p 值
离散变量的分布	性别						0.391
	女	35 (73%)	15 (83%)	6 (86%)	4 (57%)	10 (62%)	
	男	13 (27%)	3 (17%)	1 (14%)	3 (43%)	6 (38%)	
	恋爱次数						0.505
	1	18 (38%)	7 (39%)	3 (43%)	4 (57%)	4 (25%)	
	2	18 (38%)	8 (44%)	1 (14%)	2 (29%)	7 (44%)	
	3	8 (17%)	3 (17%)	1 (14%)	1 (14%)	3 (19%)	
	4	4 (8%)	0 (0%)	2 (29%)	0 (0%)	2 (12%)	
	是否正在恋爱						<0.001
	否	18 (38%)	1 (6%)	3 (43%)	3 (43%)	11 (69%)	
是	30 (62%)	17 (94%)	4 (57%)	4 (57%)	5 (31%)		
是否异地恋						0.489	
否	12 (40%)	7 (41%)	3 (75%)	1 (25%)	1 (20%)		
是	18 (60%)	10 (59%)	1 (25%)	3 (75%)	4 (80%)		
连续变量的均值(标准差)	年龄	24.10 (3.08)	23.78 (2.02)	24.43 (3.51)	23.71 (0.95)	24.50 (4.40)	0.878
	年龄绝对差值	6.67 (6.39)	7.29 (6.79)	4.50 (4.43)	6.75 (5.62)	6.20 (8.17)	0.748
	恋爱阶段	2.90 (0.80)	2.76 (0.83)	3.25 (0.96)	3.50 (0.58)	2.60 (0.55)	0.248

采用 Kruskal-Wallis 秩和检验与 Fisher 精确检验, 分别考察这些连续和离散的背景变量在 4 种依恋类型中的分布, 发现除“是否正在恋爱”外其他变量的差异均不显著 ($ps > 0.1$), 说明背景变量在各依恋类型间几乎均衡分布。后续分析将首先检验“是否正在恋爱”对语音特征的影响, 以排除可能的干扰。

1.4 实验过程

语音采集在专业录音棚内完成。使用 Philips 220SW 显示屏 (分辨率: 1280×1024) 呈现影片和语料。语音采集设备为 H4N 便携式录音机和 AKG 头戴式传声器, 每位被试的语音独立保存为 WAV 文件 (采样率 44.1 kHz, 精度 16 bit)。使用 E-prime 3.0^[27] 设计实验程序。实验分为 3 个阶段: 实验准备、正式实验、问卷填写。

在实验准备阶段, 主试将被试引入录音棚, 安置在距显示器约 70 cm 的座椅上, 并调整座椅高度以确保被试平视屏幕中央。用头戴式传声器采集语音信号, 被试嘴部距传声器约 10 cm。被试充分熟悉语料并静息平复心情后, 逐句说出所有目标句, 作为中性句。

在正式实验阶段, 按照 4 种情绪分为 4 个组块, 每个组块包含 3 部诱发相应情绪的影片, 这 3 部影片各自对应 5 个目标句 (15 个目标句划分为 3 个子集)。4 个组块的顺序随机, 每个组块内 3 部影片的

播放顺序随机, 每部影片后 5 个目标句的呈现顺序随机, 影片与目标句子集的对应关系也在被试和组块间随机。实验采用阈下词汇启动范式激活被试的依恋系统^[28], 这里“阈下”是指低于能够引起机体感觉的最小刺激量, 该范式可有效防止个体因意识到词汇代价而干扰到情绪。每个组块内的实验流程如下: 首先在屏幕中央呈现 500 ms 的注视点, 接着呈现 500 ms 的网格图, 然后随机呈现 20 ms 的启动词 (“分离”或“失败”), 并再次呈现 500 ms 的网格图。启动词前后呈现的网格图是掩蔽刺激, 用于避免视觉暂留效应。接下来播放影片, 每部影片播放完毕后依次呈现 5 个目标句, 要求被试根据从影片体验到的情绪, 想象在与恋爱伴侣对话, 说出目标句。若被试说错或者不流利, 则要求其即刻补录。影片和目标句呈现之前都先呈现 500 ms 的注视点。15 个目标句采集完毕后, 该组块结束, 被试充分休息后, 进入下一个组块。

在问卷填写阶段, 被试填写基本信息和 ECR 量表。正式实验结束后填写量表, 一是为了避免主试在录音前知晓被试的依恋风格后产生诱导, 二是为了防止被试填写完 ECR 量表后产生情绪波动。实验持续约 1 个小时。

1.5 声学参数

语音声学特征受到依恋风格和情绪类别的直接影

响。参照前人研究^[1,2,16,21,22,23], 本文使用 Praat 6.2.12^[29], 对每个目标句提取 9 个韵律参数和 17 个音质参数。

(1) 韵律参数

基频 (F0): 周期性复合声波的最低频率, 对应于声带振动的频率。使用自相关算法于浊音段每隔 5 ms 提取一个基频值, 由此计算出全句的基频均值 (F0-M)、标准差 (F0-SD)、最小值 (F0-MIN)、最大值 (F0-MAX)。单位为半音, 由 Hz 到半音的转换公式为: $12\log_2(F0/F0_{ref})$, 为减小性别差异的影响, 基于以往文献的建议^[30], $F0_{ref}$ 男性取 55 Hz, 女性取 64 Hz。

音强 (INT): 表示声音的强弱, 单位为 dB。每隔 5 ms 提取一个音强值, 由这些音强值计算出整句的音强均值 (INT-M)、标准差 (INT-SD)、最小值 (INT-MIN)、最大值 (INT-MAX)。

语速 (SPRATE): 表示每秒产生的音节数量, 单位为音节/秒。

(2) 音质参数

谱矩 (SPEC): 表示频谱能量的分布。每隔 10 ms 提取 25 ms 窗口内语音信号的前 4 阶谱矩, 包括谱重心 (SPEC-COG)、谱离散度 (SPEC-SD)、谱偏度 (SPEC-SKE)、谱峰度 (SPEC-KUR)。最终计算出全句的均值。谱重心、谱离散度的单位为 Hz, 谱偏度、谱峰度无量纲。

谐波差 (HD): 表示频谱的倾斜程度, 可反映嗓音类型, 单位为 dB。将每个浊音段等分为 3 个时段, 在每个时段中心点, 以 25 ms 为窗长, 提取经过共振峰修正^[31-32]的谐波差值, 包括低频段谐波差值 H1-H2、H2-H4, 高频段谐波差值 H2K-H5K, 以及第一谐波与前 3 个共振峰的差值 H1-A1、H1-A2、H1-A3。最终, 计算所有时段 HD 的均值作为整句的 HD。

语音不规则性参数: 谐噪比 (HNR)、平滑倒谱峰值系数 (CPPS)、基频抖动 (Jitter)、振幅抖动 (Shimmer)。HNR 表示语音中周期成分与噪声成分的比例, CPPS 指示平滑后倒谱的峰值与峰值下回归线 (倒谱能量关于倒谱的回归) 的差异, Jitter 和 Shimmer 指示连续周期间基频和振幅的差异。HNR 和 CPPS 的提取, 是将每个浊音段等分为 3 个时段, 在每个时段中心点, 以 25 ms 为窗长, 分别计算 CPPS 以及 4 个 HNR: 0~0.5 kHz (HNR05)、0~1.5 kHz (HNR15)、0~2.5 kHz (HNR25)、0~3.5 kHz (HNR35)。Jitter 和 Shimmer, 则是对每个浊音段计算 1 个数值。最终, 计算全句各段的均值作为整句的特征参数。HNR 和 CPPS 的单位为 dB, Jitter 和 Shimmer 以百分数 (%) 表示。

1.6 统计分析

为避免与情绪无关变量的干扰, 参考文献 [25] 的方法, 采用式 (1) 对提取的声学参数 X^{ij} (i 为句子编号, j 为被试编号) 做人际归一化:

$$X_{norm}^{ij} = (X^{ij} - X_{ref}^j) / X_{ref}^j, \quad (1)$$

其中, X_{ref} 为参考值, F0 和 INT 相关参数 (F0-SD、INT-SD 除外) 取每位被试所有中性句 F0-MIN 和 INT-MIN 的均值, 其余参数则取每位被试所有中性句该参数的均值。

将人际归一化后的声学参数, 进一步标准化, 转换为 z 值。然后, 按照以下步骤, 使用 R 4.1.1^[33] 做统计分析:

(1) 特征选择: 首先, 剔除存在多重共线性的参数 SPEC-SKEW、HNR05、HNR15、HNR25 ($r \geq 0.90$); 然后, 对剩余的 22 个参数通过递归特征消除 (RFE), 去除冗余特征。

(2) 多元差异性分析: 首先, 采用半参数多元方差分析 (SPMAV) 检验“是否正在恋爱”对声学特征的影响; 然后, 运用半参数重复测量多元方差分析 (SPRMMAV) 考察依恋类型与情绪类别对多维声学特征的主效应及交互效应。

(3) 聚类分析: 利用聚集性分层聚类分析 (AHCA), 以系统树的形式呈现 4 种依恋类型人群的情绪语音声学特征距离。

(4) 有监督分类: 利用 5 种机器学习算法探究 4 种依恋类型在情绪声学特征空间上的决策边界。

(5) 分类模型解释: 通过特征重要性分析和累积局部轮廓 (ALP) 分析, 解释分类模型结果。

2 结果

2.1 特征选择

递归特征消除 RFE, 包括 4 个步骤: (1) 选择模型种类; (2) 基于所有特征构建模型, 根据重要性分数对特征做降序排列; (3) 遍历预先设定的特征数量集合, 从排好序的特征集中选取相应数量的特征构建模型, 并计算拟合优度指标; (4) 选取拟合优度最佳的特征子集做后续分析。

具体来说, 以依恋类型为分层因子, 采用分层随机抽样将数据集 (2837 个样本: 安全型占 37%, 超脱型和专注型各占 15%, 恐惧型占 33%) 划分为 70% 训练集 (1988 个样本) 和 30% 测试集 (849 个样本)。在训练集上采用袋状决策树算法, 基于 22 个声学参数拟合 4 种依恋类型的分类模型, 将特征数量集合设

为 2~22 (步长为 2) 的向量, 采用 10 折交叉验证法计算识别率作为拟合优度指标。结果显示, 最优特征子集的特征数量为 14 (识别率为 80%), 包括 F0-M、F0-MIN、F0-MAX、INT-M、INT-MAX、SPRATE、H1-H2、H2-H4、H1-A2、H1-A3、H2K-H5K、CPPS、Jitter、Shimmer。基于上述特征做后续分析。

2.2 多元差异性分析

使用 SPMAV 检验“是否正在恋爱”对语音声学特征的影响。由于表 1 显示变量“是否正在恋爱”的分布在各依恋类型之间有显著差异, 因此以“是否正在恋爱”为自变量、14 个声学特征为因变量, 对每种情绪分别做 SPMAV。统计量为 Modified Anova-Type Statistic (MATS), 适用于数据不满足正态性和同质性的情况^[34]。采用 Wild Bootstrap 法做 1000 次迭代计算 p 值, 结果显示“是否正在恋爱”在 4 种情绪上的主效应均不显著 ($MATS > 8.7, ps > 0.05$), 于是可排除该变量对实验结果的干扰。

采用 SPRMMAV 考察依恋类型和情绪类别对声学特征的主效应和交互效应。4 类依恋类型为组间变量, 4 种基本情绪为组内变量, 14 个声学特征为因变量。统计量与 p 值的计算方法同 SPMAV。结果显示, 依恋类型和情绪类别的主效应均显著 ($MATS = 245.15, p = 0.043$; $MATS = 83.56, p < 0.001$), 两者间的交互效应不显著 ($MATS = 42.32, p = 0.276$)。

2.3 聚类分析

由于依恋类型和情绪类别的交互效应不显著, 所以可以不区分情绪而对所有语音数据做 AHCA, 考察 4 种依恋类型人群之间的语音声学距离。基于 14 个声学特征构建欧氏距离矩阵, 使用 Ward 最小方差法作为连接标准。AHCA 的结果以系统树的形式呈现, 树的高度表示欧氏距离大小, 越先连接的节点之间距离越小。图 1 的 AHCA 结果显示, 超脱型与专注型率先聚合, 然后再与恐惧型聚合为非安全型人群, 与安全型对立。可见, 在语音声学特征空间上, 超脱型和专注型距离最近, 而安全型则远离其他类型。

2.4 有监督分类

采用情感分类研究中 5 种常用的有监督的机器学习算法, 在情绪语音的多维声学特征空间上进一步考察 4 类依恋类型人群的决策边界。这里的机器学习算法包括线性判别分析 (LDA)、支持向量机 (SVM)、随机森林 (RF)、极致梯度提升决策树 (XGBoost)、多层感知机 (MPM)。

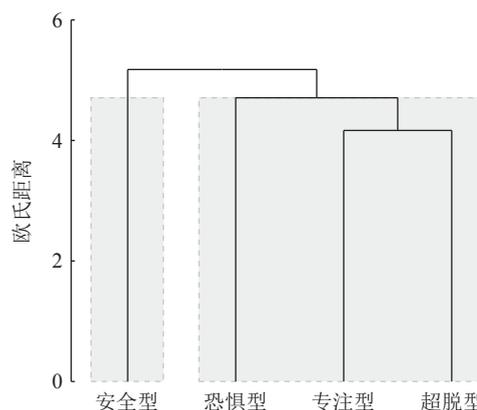


图 1 基于 4 类人群声学距离的系统树

训练集和测试集的划分方法同 2.1 节。训练集用于拟合模型和调整超参, 测试集用于检验模型的预测效力。采用 5 重 10 折交叉验证法 (Cross Validation) 调参以避免模型过拟合。由于 4 类人群的占比不均衡, 训练模型时采用增采样法 (Upsampling) 平衡占比, 同时利用准确率 (ACC) 和 4 类人群的 F1 分数衡量模型的预测效力。以 14 个声学特征为预测变量, 依恋类型为响应变量, 基于所有情绪的数据合集与每种情绪的数据子集分别构建模型。

5 种机器学习算法的分类预测结果见表 2。选取 4 类人群中占比最大的先验概率 (安全型 37%) 作为机会水平。基于所有情绪的数据合集建模, 5 种算法的识别率均显著高于机会水平 (Binomial Test: $ps < 0.001$), 其中 RF 的预测效果最佳 (ACC: 0.83; F1: 0.70~0.88)。基于每种情绪的数据子集分开建模, 除了愤怒情绪时的 LDA 算法以外, 所有算法的识别率均显著高于机会水平 ($ps < 0.05$)。可见, 基于情绪语音的声学数据可以有效地区分 4 类依恋类型的人群。

2.5 分类模型解释

基于预测性能最佳的 RF 模型, 在测试集上采用特征重要性分析与累积局部轮廓分析两种方法, 进一步解释结果。

(1) 特征重要性分析

采用置换法, 以置换给定特征前后的交叉熵之比作为该特征的重要性分数。置换 100 次后, 结果如图 2 所示, F0-M、F0-MAX、INT-M、INT-MAX 在 4 种情绪上对区分依恋类型的贡献度均位居前 5, 而 H2K-H5K (开心)、SPRATE (愤怒)、H2-H4 (悲伤)、CPPS (恐惧) 仅在特定情绪上对分类有较大的贡献。

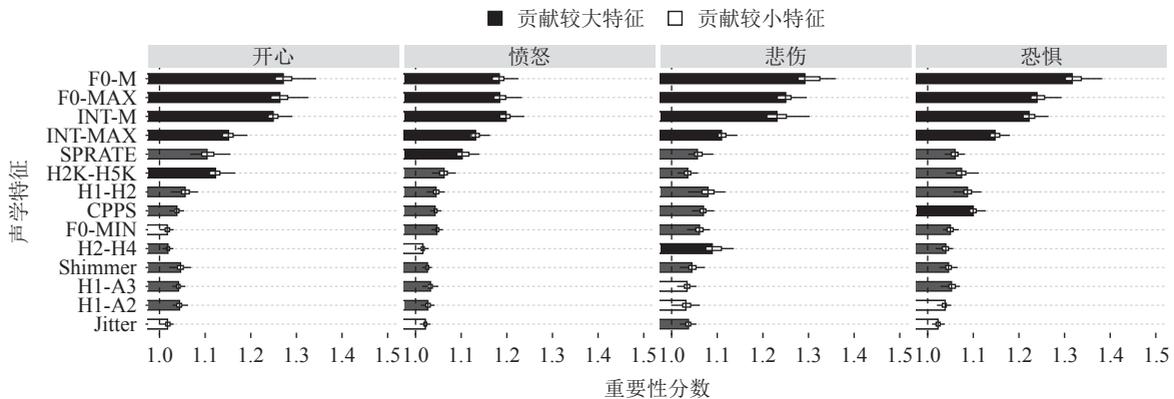
(2) 累积局部轮廓分析

结合特征重要性分析结果和先验知识, 选取了 10 个特征做 ALP 分析, 以考察声学特征对依恋类型

表 2 机器学习算法的分类预测结果

		LDA	SVM	RF	XGBT	MPM
所有情绪	ACC (95%CI)	0.48 (0.45, 0.52)	0.81 (0.78, 0.83)	0.83 (0.80, 0.85)	0.82 (0.79, 0.84)	0.82 (0.79, 0.85)
	F1	0.52; 0.40; 0.49; 0.49	0.87; 0.76; 0.73; 0.80	0.88; 0.80; 0.70; 0.84	0.87; 0.79; 0.73; 0.81	0.86; 0.78; 0.74; 0.83
开心	ACC (95%CI)	0.47 (0.40, 0.54)	0.72 (0.65, 0.78)	0.78 (0.72, 0.84)	0.80 (0.74, 0.85)	0.69 (0.62, 0.75)
	F1	0.54; 0.32; 0.54; 0.43	0.82; 0.69; 0.65; 0.63	0.84; 0.76; 0.71; 0.76	0.84; 0.79; 0.72; 0.78	0.76; 0.62; 0.68; 0.62
愤怒	ACC (95%CI)	0.41 (0.34, 0.48)	0.66 (0.59, 0.72)	0.74 (0.68, 0.80)	0.73 (0.67, 0.79)	0.66 (0.59, 0.72)
	F1	0.51; 0.30; 0.45; 0.35	0.75; 0.52; 0.48; 0.69	0.83; 0.72; 0.60; 0.73	0.82; 0.71; 0.58; 0.72	0.76; 0.55; 0.57; 0.64
悲伤	ACC (95%CI)	0.51 (0.44, 0.58)	0.78 (0.72, 0.84)	0.74 (0.67, 0.79)	0.74 (0.67, 0.79)	0.75 (0.68, 0.80)
	F1	0.55; 0.49; 0.49; 0.51	0.78; 0.78; 0.63; 0.86	0.74; 0.80; 0.68; 0.72	0.74; 0.86; 0.64; 0.72	0.76; 0.78; 0.60; 0.78
恐惧	ACC (95%CI)	0.48 (0.41, 0.55)	0.74 (0.67, 0.79)	0.76 (0.70, 0.82)	0.72 (0.66, 0.78)	0.67 (0.61, 0.74)
	F1	0.55; 0.39; 0.40; 0.50	0.78; 0.72; 0.67; 0.72	0.83; 0.64; 0.64; 0.78	0.78; 0.62; 0.60; 0.75	0.74; 0.57; 0.61; 0.68

注: ACC为准确率, 4个F1值分别对应安全型、超脱型、专注型和恐惧型。灰色单元指预测效果最佳的模型。



注: 条形图表示特征重要性分数的均值; 箱线图表示特征重要性分数的分布; 垂直虚线表示置换给定特征前后交叉熵不变时的重要性分数。

图 2 特征重要性分析结果

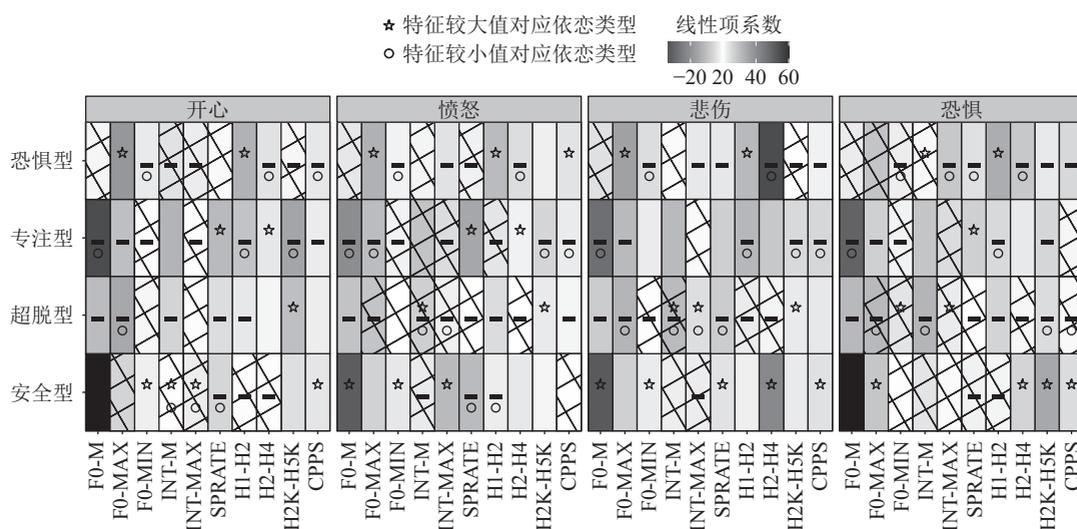
决策概率的预测作用。首先, 选取以上 8 个重要性分数高的特征 (包括 F0-M、F0-MAX、INT-M、INT-MAX, 以及 SPRATE、H2-H4、H2K-H5K、CPPS)。其次, 根据已有知识, H1-H2 与嗓音类型直接相关, 因此将 H1-H2 也纳入分析^[35]。并且, 近期研究发现基频范围 (F0-RANGE) 对区分依恋类型有重要作用^[23], 因此将 F0-MIN 也纳入, 从而结合 F0-MAX 可以体现基频范围的影响。

ALP 分析通过计算 Ceteris-paribus 曲线的条件均值获得声学特征与依恋类型决策概率间的非线性关系。为分解出其中的线性趋势, 运用正交多项式回归 (OPR) 拟合 ALP 曲线以提取线性项分量。为满足 OPR 对数据的要求, 使用合理化反正弦变换 (Rationalized Arcsine Transform) 将值域为 [0, 1] 的概率值转换为值域为 [-20, 120] 的合理化反正弦单位值。每个特征以 $M \pm SD$ 为界划分为较低、中等、较高区间, 分别计算 4 类人群在较低和较高区间内具有最大决策概率的比例, 选取比例最大的人群作为与高/低特征值对应的依恋类型。为避免多次建模增加 I 型错误, 采用 Holm 法修正 p 值, 显著水平设为 0.01。

ALP 的结果如图 3 所示。各依恋类型与其他类型相比较, 特点如下: 安全型人群, F0-M 在 4 种情绪上均最高, SPRATE 在开心和愤怒情绪上最慢、在悲伤情绪上最快, H1-H2 在愤怒情绪上最小, H2-H4 在悲伤和恐惧情绪上最大, CPPS 在愤怒以外的情绪上均最大; 超脱型人群, F0-MAX 在愤怒以外的情绪上均最低, F0-MIN 在恐惧情绪上最高, INT-M、H2K-H5K、CPPS 在恐惧情绪上最低, 该类人群的决策概率与 F0-M 在所有情绪上均呈显著负相关、与各频段 HD 仅在恐惧情绪上呈显著负相关; 专注型人群, F0-M 在所有情绪上均最低, SPRATE 在悲伤以外的情绪上均最快, 至少有两个频段的 HD 在所有情绪上均最小, CPPS 仅在愤怒和悲伤情绪上最小; 恐惧型人群, F0-MAX 和 F0-MIN 在恐惧以外的情绪上分别最大和最小, H1-H2 与 H2-H4 在所有情绪上分别最大和最小, CPPS 仅在愤怒情绪上最大。

3 讨论

本文基于 RFE 算法筛选出 14 个声学特征, 通过多元差异性分析、分层聚类分析、有监督分类 (含模



注: 灰度表示对 ALP 曲线做 OPR 拟合的线性项系数的绝对值; 一代表系数为负值; ☆代表语音特征较大 ($> M + SD$) 时对应的决策概率最高的依恋类型; ○代表语音特征较小 ($< M - SD$) 时对应的决策概率最高的依恋类型; 网格表示线性项系数不显著。

图3 ALP 分析结果

型解释), 系统考察了 4 类依恋类型人群在表达开心、愤怒、悲伤、恐惧这 4 种基本情绪时语音声学特征的分布。

多元差异性分析发现, 4 类依恋类型在情绪语音声学特征空间中的分布存在显著差异; 分层聚类分析发现, 在声学特征空间中, 超脱型和专注型的距离最小, 而安全型与其他类型之间的距离最大, 这与 Bartholomew 等^[7]安全型自成一类、其余 3 类归为非安全型的观点一致; 基于 14 个声学特征的 RF 算法发现, 在每种情绪上均可有效区分 4 类依恋类型 (识别率 > 0.70), 说明 4 类人群在情绪语音声学特征空间中存在清晰的决策边界。

为进一步解释 RF 模型的结果, 运用 ALP 分析考察了各声学特征与 4 类人群决策概率间的线性关系。Laukka 等^[2]基于唤醒度、强度、愉悦度定义情绪, 其中唤醒度和强度与 F0-M 呈显著正相关、与 SPRATE 呈显著负相关。Xu 等^[36]则从交际功能角度, 基于尺寸投影 (Size Projection)、动态性 (Dynamicity)、可闻度 (Audibility)、关联度 (Association) 定义情绪: 尺寸投影表示语音特征与身体尺寸的投射关系, 如较低/高的 F0、粗糙/规则的噪音投射成大/小身体尺寸, 旨在驱逐/吸引或控制/安抚听者, 传达威胁/友好、强势/屈从等信息; 动态性表示发声的活力 (发声的强与弱的控制能力), F0 和共振峰动态范围越大, 动态性越高; 可闻度表示话者声音能够传递的最远距离, INT 越大、噪音越规则, 可闻度越高; 关联度则表示声音的关联运用^[36]。

安全型的 F0-M 在所有情绪上均最高, SPRATE

在开心和愤怒上最慢, CPPS 在除愤怒以外的情绪上均最大。根据 Laukka 等的理论框架, F0-M 高、SPRATE 慢, 表明安全型个体总的来说有最高的情绪唤醒度和强度, 这与文献^[37]的结果一致。灵活的情绪调节策略与积极的他人表征是这类个体情绪表达更具开放性的主要原因。根据交际功能的理论框架, 最高的 F0、最大的 CPPS (对应规则性最强的噪音) 分别暗示了安全型个体表达愤怒以外的情绪时尺寸投影最小、可闻度最高。由于安全型个体期待自身的情感诉求获得回应, 所以他们通过传达友好顺从的态度和提高可闻度的方式来吸引听者的注意。

超脱型的 F0-MAX 在除愤怒以外的情绪上最低、F0-M 在所有情绪上较低, 表现出较低的情绪唤醒度和强度^[2], 与这类个体表达抑制的策略偏好一致^[38]; 而且, F0-M 低与身体尺寸大相联系, 传达出威胁、强势的态度, 这与超脱型个体消极的他人表征有关, 以达到自我保护的目的。此外, 超脱型个体在恐惧情绪上, 有最窄的 F0-RANGE (F0-MIN 最高、F0-MAX 最低)、最小的 HD (与嘎裂声的使用有关: HD 值越小, 越有嘎裂倾向^[39]) 和 CPPS, 体现出最低的动态性和可闻度^[36], 从而反映了最低的发音努力 (根据 Effort Code 假设^[40], F0-RANGE 越窄, 则发音努力越少), 同样说明超脱型个体在恐惧情绪下有抑制表达的倾向。

专注型的 F0-M 在所有情绪上均最低, SPRATE 在除悲伤以外的情绪上均最快, 体现出最低的情绪唤醒度和强度^[2]。愤怒和悲伤情绪上的 HD 和 CPPS 最小, 表现出规则性低的噪音, 这与 Monti 等的结果

一致^[22],暗示专注型个体表达此类情绪时传达出强势、威胁等信息,并表现出最低的可闻度^[36,41]。值得注意的是,研究结果表明了专注型个体表达抑制的策略偏好,他们消极的自我表征促使其选择这一策略以掩饰内心的自卑与焦虑。

恐惧型的 F0-RANGE 在恐惧以外的情绪上均最宽 (F0-MAX 最高、F0-MIN 最低),显示出最高的动态性^[36]; CPPS 在愤怒情绪上最大,显示出最高的可闻度。这些都表明恐惧型个体表达情绪时发音更加努力,从而挑战了 Winterheld^[12] 的观点 (恐惧型个体的情绪调节策略兼具安全型和超脱型的特点)。实验结果提示,恐惧型个体的情绪调节策略更接近安全型的开放的认知重评,而不是超脱型的表达抑制。

4 结论

研究初步揭示了 4 类依恋类型人群的情绪语音在声学特征空间的分布模式。主要结论如下: (1) 不同依恋类型人群的语音在多维声学特征空间上的分布存在差异,且差异模式不受情绪类别影响;其中,超脱型和专注型的声学距离最近,而安全型则远离其他类型。(2) 选取的 14 个声学特征能够较好地区分 4 类人群,其中 RF 算法的预测性能最佳 (识别率 0.83),且韵律特征对分类的贡献较大。(3) 4 类人群在 F0 (含 F0-M、F0-MAX、F0-MIN) 上的差异模式在各种情绪之间基本一致,在 HD、CPPS 上的差异模式则受到情绪类别的调节。

基于情绪语音的多维声学特征,本文还推断出各个依恋类型的情绪调节策略:安全型倾向于更具开放灵活性的策略,超脱型、专注型都倾向于采用表达抑制策略,而恐惧型的情绪调节策略与安全型更加接近。从而,本文为依恋系统的情绪调节理论提供了进一步的证据。

关于依恋类型对情绪语音产出的作用,后续研究将在以下三方面进一步探索: (1) 设计真实的情侣交际场景,激活依恋系统,诱发有关情绪,采用更符合语言事实的语料,以提高实验的生态效度; (2) 验证情绪调节在依恋类型和语音之间的中介作用,可运用路径分析考察依恋类型经过情绪调节对语音特征的间接影响; (3) 增加样本和特征数量,借助深度学习技术提高模型的预测性能。

参 考 文 献

- 1 Banse R, Scherer K R. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.*, 1996; **70**(3): 614–636
- 2 Laukka P, Juslin P, Bresin R. A dimensional approach to vocal

- expression of emotion. *Cogn. Emot.*, 2005; **19**(5): 633–653
- 3 Bryant G A. Vocal communication across cultures: Theoretical and methodological issues. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 2022; **377**(1841): 20200387
- 4 Marrero Z N, Gosling S D, Pennebaker J W, et al. Evaluating voice samples as a potential source of information about personality. *Acta Psychol.*, 2022; **230**: 103740
- 5 Bowlby J M. Attachment and loss: Attachment (vol. 1). London: Random House, 1969
- 6 Mikulincer M, Shaver P R. Attachment in adulthood: Structure, dynamics, and change. 2nd edition. New York: The Guilford Press, 2016
- 7 Bartholomew K, Horowitz L M. Attachment styles among young adults: A test of a four-category model. *J. Pers. Soc. Psychol.*, 1991; **61**(2): 226–244
- 8 Brennan K A, Clark C L, Shaver P R. Self-report measurement of adult attachment: An integrative overview. In: Simpson J A, Rholes W S (Eds.), Attachment theory and close relationships. New York: The Guilford Press, 1998: 46–76
- 9 黄于飞, 史攀, 陈旭. 依恋对情绪调节过程的影响. *心理科学进展*, 2022; **30**(1): 77–84
- 10 Mikulincer M, Shaver P R. Attachment orientations and emotion regulation. *Curr. Opin. Psychol.*, 2019; **25**: 6–10
- 11 Troyer D, Greitemeyer T. The impact of attachment orientations on empathy in adults: Considering the mediating role of emotion regulation strategies and negative affectivity. *Pers. Individ. Dif.*, 2018; **122**: 198–205
- 12 Winterheld H A. Calibrating use of emotion regulation strategies to the relationship context: An attachment perspective. *J. Pers.*, 2016; **84**(3): 369–380
- 13 Walker S A, Double K S, Kunst H, et al. Emotional intelligence and attachment in adulthood: A meta-analysis. *Pers. Individ. Dif.*, 2022; **184**: 111174
- 14 Gross J J. Emotion regulation: Current status and future prospects. *Psychol. Inq.*, 2015; **26**(1): 1–26
- 15 Ilie G, Thompson W F. A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Percept.*, 2006; **23**(4): 319–330
- 16 Johnstone T, Scherer K R. Vocal communication of emotion. In: Lewis M, Haviland J (Eds.), Handbook of Emotions. New York: The Guilford Press, 2000
- 17 Scherer K R. Acoustic patterning of emotion vocalizations. In: Frühholz S, Belin P (Eds.), The Oxford handbook of voice perception. Oxford: Oxford University Press, 2018
- 18 Gobl C, Chasaide A N. The role of voice quality in communicating emotion, mood and attitude. *Speech Commun.*, 2003; **40**(1-2): 189–212
- 19 Yanushevskaya I, Ni Chasaide A, Gobl C. Voice parameter dynamics in portrayed emotions. In: Manfredi C (Ed.), Models and Analysis of Vocal Emissions for Biomedical Applications. Florence: Firenze University Press, 2009
- 20 Murray I R, Arnott J L. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J. Acoust. Soc. Am.*, 1993; **93**(2): 1097–1108
- 21 Spinelli M, Fasolo M, Coppola G, et al. It is a matter of how you say it: Verbal content and prosody matching as an index of emotion regulation strategies during the Adult Attachment Interview. *Int. J. Psychol.*, 2019; **54**(1): 102–107

- 22 Monti E, Kidd D C, Carroll L M, *et al.* What 's in a singer 's voice: The effect of attachment, emotions and trauma. *Logoped. Phoniatr. Vocol.*, 2017; **42**(2): 62–72
- 23 胡涵, 顾文涛. 个体依恋风格对亲密话语韵律及嗓音特征的影响. *声学学报*, 2022; **47**(2): 276–286
- 24 徐鹏飞, 黄宇霞, 罗跃嘉. 中国情绪影像材料库的初步编制和评定. *中国心理卫生杂志*, 2010; **24**(7): 551–555
- 25 Liu P, Pell M D. Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behav. Res. Methods*, 2012; **44**(4): 1042–1051
- 26 李同归. 成人依恋的测量: 亲密关系经历量表(ECR)中文版. *心理学报*, 2006; **38**(3): 399–406
- 27 Psychology Software Tools, Inc. E-Prime 3.0. <https://support.pst-net.com/>, 2016
- 28 Mikulincer M, Gillath O, Shaver P R. Activation of the attachment system in adulthood: Threat-related primes increase the accessibility of mental representations of attachment figures. *J. Pers. Soc. Psychol.*, 2002; **83**(4): 881–895
- 29 Boersma P, Weenink D. Praat: Doing phonetics by computer. Version 6.2.12. <http://www.praat.org/>, 2022
- 30 刘艺, 荣蓉. 汉语学习者陈述句音高节高的声学实验分析. *语言教学与研究*, 2014(5): 35–41
- 31 Hawks J W, Miller J D. A formant bandwidth estimation procedure for vowel synthesis. *J. Acoust. Soc. Am.*, 1995; **97**(2): 1343–1344
- 32 Iseli M, Shue Y L, Alwan A. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *J. Acoust. Soc. Am.*, 2007; **121**(4): 2283–2295
- 33 R Core Team. R: A language and environment for statistical computing. <https://www.R-project.org/>, 2021
- 34 Friedrich S, Pauly M. MATS: Inference for potentially singular and heteroscedastic MANOVA. *J. Multivar. Anal.*, 2018; **165**: 166–179
- 35 Esposito C M. An acoustic and electroglottographic study of White Hmong tone and phonation. *J. Phon.*, 2012; **40**(3): 466–476
- 36 Xu Y, Kelly A, Smillie C. Emotional expressions as communicative signals. In: Hirst D, Hancil S (Eds.), *Prosody and iconicity*. Amsterdam: John Benjamins, 2013
- 37 Little L M, Nelson D L, Wallace J C, *et al.* Integrating attachment style, vigor at work, and extra–role performance. *J. Organ. Behav.*, 2011; **32**(3): 464–484
- 38 杨青青, 胡娜, 陈旭, 等. 恋人亲密情景下的回避型与安全型依恋个体情绪调节电生理差异. *心理学报*, 2018; **50**(3): 306–316
- 39 Keating P A, Garellek M, Kreiman J. Acoustic properties of different kinds of creaky voice. The 18th International Congress of Phonetic Sciences, Glasgow, UK, 2015
- 40 Gussenhoven C. *The phonology of tone and intonation*. Cambridge: Cambridge University Press, 2004
- 41 耿浦洋, 顾文涛, 曹文. 汉语态度语音的韵律, 嗓音和调音运动特征. *语言研究集刊*, 2020; **26**: 322–341

附录 A

表 A1 纳入正式实验的 15 个目标伪句

目标句	自然度均值	自然度标准差	句长 (音节数)
你们泡到羊表里	3.60	1.59	7
他们踩到道杯中	3.73	1.33	7
他们往车桌里排	3.67	1.23	7
我们到尾边打澡	4.13	1.06	7
你角里铺着个豆仔	4.07	1.33	8
你们在光海上跳球	4.40	0.83	8
你在排房上晒春某	4.53	0.64	8
他的发签上沾了麻	4.27	1.10	8
他的胎上有个表票	3.80	1.47	8
他在坊门里唱了一个绳	4.20	1.01	10
她从包盒里拿出一个诗	4.60	0.63	10
你们到线桥上切了一个棋	3.93	1.28	11
他们从角客里偷了一个盘	4.73	0.59	11
她们到枕尺里救了一个盐	4.00	1.51	11
我们去心沙里丢了一个发	3.73	1.62	11