

人形机器人面部情感表达技术研究综述

柯显信, 李玄, 宋骏骐

引用本文

柯显信, 李玄, 宋骏骐. [人形机器人面部情感表达技术研究综述](#)[J]. 计算机科学, 2026, 53(7): 1-8.

KE Xianxin, LI Xuan, SONG Junqi. [Research on Facial Emotion Expression Technologies for Humanoid Robots](#) [J]. Computer Science, 2026, 53(7): 1-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于OCC模型的E-learning系统情感建模](#)

Emotional Modeling in an E-learning System Based on OCC Theory

计算机科学, 2010, 37(5): 214-218.

[基于非线性状态空间模型的情感模型研究](#)

计算机科学, 2008, 35(12): 178-182.

人形机器人面部情感表达技术研究综述

柯显信 李 玄 宋骏骐

上海大学机电工程与自动化学院 上海 200444

摘 要 人形机器人面部情感表达技术是实现自然人机交互的核心支撑,近年来在医疗陪护、教育服务与社交娱乐等领域展现出广泛的应用前景。回顾了该领域的发展历程,按照技术能力演进将其划分为基础情感表达、情感交互增强与情感共鸣个性化3个阶段,系统探讨了各阶段在仿生机械结构设计、多模态情感交互系统构建方面的关键突破。重点分析了面部动作编码系统、柔性仿生皮肤材料、深度学习与多模态大模型等在情感建模中的应用,进一步探讨了多模态时空对齐机制在表情、语音与动作协同优化中的作用。最后,展望了未来的研究方向,包括高精度驱动与传感融合、多模态大模型集成、个性化与文化适应性增强,以及恐怖谷效应的缓解策略,为推动人形机器人面部情感表达技术的智能化和自然化发展提供了理论支持。

关键词: 人形机器人;面部情感表达;仿生机械结构;情感建模;多模态时空对齐;恐怖谷效应

中图分类号 TP242

Research on Facial Emotion Expression Technologies for Humanoid Robots

KE Xianxin, LI Xuan and SONG Junqi

College of Mechanical Engineering and Automation, Shanghai University, Shanghai 200444, China

Abstract Facial emotional expression in humanoid robots represents a cornerstone technology for enabling naturalistic human-robot interaction, exhibiting broad application prospects in sectors such as medical support, educational services, and social entertainment. This paper reviews the developmental trajectory of the field, delineating its evolution across three stages based on technological capability: foundational emotional expression, enhanced affective interaction, and personalized emotional resonance. It systematically examines the key breakthroughs within each stage concerning the design of biomimetic mechanical structures and the construction of multimodal affective interaction systems. The review places particular emphasis on the application of the facial action coding system, flexible biomimetic skin materials, deep learning, and multimodal large models (MLMs) in affective modeling. Furthermore, it investigates the critical role of multimodal spatio-temporal alignment mechanisms in the coordinated optimization of facial expressions, vocalizations, and physical actions. Finally, the paper outlines future research trajectories, including the fusion of high-precision actuation and sensing, the integration of multimodal large models, the enhancement of personalization and cultural adaptability, and strategies for mitigating the uncanny valley effect. This work provides a theoretical foundation to advance the intelligent and naturalistic development of facial emotional expression technology in humanoid robots.

Keywords Humanoid robot, Facial emotional expression, Biomimetic mechanical structure, Affective modeling, Multimodal spatio-temporal alignment, Uncanny valley effect

人形机器人的面部凭借高拟真度特征与动态表情模拟能力,正逐步突破人机情感交互的技术边界,已经在医疗陪护^[1]、教育服务^[2]和社交娱乐^[3]等领域展现出重要的应用价值,并成为人工智能与人机交互领域的前沿研究方向。本文在系统梳理国内外研究历程的基础上,重点综述了仿人表情机器人的结构设计、情感控制两大核心方面。最后,结合多学科融合的趋势,对未来技术挑战与发展方向做出展望,为提升人形机器人面部情感表达的自然度与社会融贯性研究提供参考。

1 国内外研究历程

机器人面部情感表达技术的发展可以追溯到20世纪,最

初的研究人员专注于搭建基础系统与探索基础功能。日本早稻田大学率先取得突破,他们开发的拟人头眼机器人能展现6种自然表情,并实现类似人类的眼球运动^[4];美国麻省理工学院同期也推出了具备眨眼、微笑功能的实验机型^[5]。进入21世纪后,人工智能和材料科学飞速发展,机器人的表情控制和外观拟人化技术也随之取得了显著进展。中国在这一领域的研究取得了重要突破,清宝引擎机器人的“清宝”机器人和EX机器人公司的“小柒”机器人^[6]等,在情感建模与实现、表情识别与再现等方面取得了重大的研究成果。随着人形机器人面部研究的不断深化,它的应用领域也从娱乐、教育领域迅速扩展,展现出了广阔的发展前景。

根据技术能力突破、交互模式演进与应用场景扩展3个

核心标准,可将人形机器人面部情感表达技术的发展划分为3个阶段。

第一阶段:基础情感交互阶段(20世纪末—2005年)。此阶段的核心技术标志是基础感知与计算情感模型的结合,其目标是实现对人类社交信号的感知与响应,奠定了人机情感交互的基础。研究重点在于通过集成摄像头、麦克风等基础传感器,并构建初步的计算情感模型,使机器人能够对外部刺激(如声音、运动)做出预设的表情反应。日本早稻田大学的WE-3RII^[7]能够展现6种基本表情并追踪目标;而麻省理工学院的Kismet^[5]机器人则是这一时期的里程碑,它能够通过识别并模仿看护者的情感线索,进行类似婴幼儿的社交互动,验证了情感交互的可行性。

第二阶段:多模态融合与动态交互阶段(2005—2015年)。此阶段的核心技术标志是先进的多模态感知与动态内部状态模型的引入。随着传感器技术和计算能力的发展,机器人开始具备更强的环境理解和主动交互能力。技术重点从简单的刺激反应转向对多模态信息的深度融合。例如,MIT的Nexi^[8]机器人集成了3D环境识别与触觉反馈;早稻田大学的WE-4RII^[9]则引入混沌神经网络,试图

构建更智能的内部情感状态模型,实现主动交互。交互模式从单向或简单的反应式,演变为有一定语境理解能力的双向交流;应用场景也开始从实验室走向教育、医疗等领域的初步探索。

第三阶段:AI驱动的个性化共情阶段(2015年至今)。此阶段的核心技术标志是人工智能(特别是深度学习与大语言模型)与先进仿生材料的深度融合。其目标是实现与人类的情感共鸣和个性化、自适应的交互。Hanson Robotics的Sophia机器人^[10]利用其专利皮肤材料“Frubber”和先进的AI算法,能够生成超过60种细腻的微表情。Engineered Arts公司的Ameca机器人^[11]则以极其流畅、自然的表情和动作震惊业界,展现了多模态AI与先进机电系统无缝集成的成果。中国的EX-Robots公司也推出了“小柴”等一系列超写实机器人,集成了多模态大模型、智能表情系统和仿生皮肤技术,实现了自然流畅的情感交互。在此阶段,交互模式进化为多模态协同与情感共鸣,应用场景深化至自闭症干预^[12]、高端陪护^[13]等对情感交互要求极高的社会服务领域。

表1依据技术代际,系统性地梳理了这3个阶段中具有代表性的仿人面部表情机器人及其技术特性。

表1 国内外代表性表情机器人的技术特性与实现功能

Table 1 Technical characteristics and functions of representative emotive robots in China and abroad					
阶段	名称	研究机构	年份	技术特性	实现功能
机械驱动与基础表情阶段	Kismet ^[5]	麻省理工学院	1999	语音合成系统能够生成类似婴儿的语音;内置情感移情系统,能模仿人类的情感反应	可以识别情感意图并反馈;能像婴儿一样通过与看护人的互动来学习社会行为
	WE-3RII ^[7]	日本早稻田大学	1999	改变眼球控制参数追踪目标;根据目标位置和亮度调整面部表情	实现6种基本表情;三维目标识别和追踪
	KOBIAN-R ^[14]	日本早稻田大学	2012	整合表情与动作,达成情感表达全身协调;支持机器人在交互中稳定运动	实现基本6种表情;通过面部与身体协同表达情感
	SHFR-III ^[15]	上海大学	2015	采用FPGA作为控制器;建立SOPC多路舵机控制系统	实现了常见的8种基本表情与头颈协调运动;能够与人类进行自然的对话
多模态感知与交互增强阶段	WE-4RII ^[9]	日本早稻田大学	2004	将混沌神经网络与联想记忆相结合,实现了更加智能化的行为控制;多模态协调控制表达情感	通过面部表情来表达情感;具备视、触、听和嗅觉;具有主动交互和行为决策能力;能够通过记忆检索表现出不同的情感表达
	SAYA ^[16]	东京理工大学	2006	开发一种能与人类进行情感交流的交互式通信系统;通过类似人类肌肉的McKibben气动执行器表达面部表情	具备一定的语言理解和表达能力;通过精确控制24块人造肌肉来创造逼真的面部表情
	H&F robot-III ^[17]	哈尔滨工业大学	2008	具有语音及口形控制系统	语音与口形同步;可进行面部表情识别
	佳佳 ^[18]	中国科学技术大学	2016	采用硅胶皮肤与微型电机驱动的机械结构,实现高仿真面部表情及基础多模态感知	通过语音交互与表情反馈提供基础问答、信息播报
AI驱动的个性化共情阶段	Albert Hubo ^[19]	汉森机器人公司	2009	使用“Frubber”皮肤材料实现面部皱纹的微妙变化	具有高度仿真的面部表情和语音交流能力
	Affetto ^[20]	日本大阪大学	2011	气动驱动装置控制表情;采用了基于衰减波合成的动态表情生成系统	能够根据内部情感状态实时生成面部表情;实现了更加丰富和细腻的表情变化
	FACE ^[21]	意大利比萨大学	2012	利用HEFES模块生成基于FACS的面部表情	能做出基本表情组合而成的复杂表情
	Sophia ^[10]	汉森机器人公司	2015	采用先进的人工智能算法驱动表情和对话;具有机器学习能力	可实现62种面部表情;具有表情展示与互动功能;可进行语音交互
	Alice ^[22]	谢里夫理工大学	2016	通过Kinect传感器和模糊C均值算法进行捕捉和情感识别	能根据用户情感状态的变化,动态调整自身的情感状态和表情
	蒋莱莱 ^[23]	EX-robot	2019	外形通过3D扫描仪采集真人数据,再进行数字化建模;采用仿生皮肤和复合材料制作	上百种精细表情;具有极高的仿真度;打破“你问我答”模式,实现了自然流畅的情感交互
	Ameca ^[11]	Engineered Arts公司	2022	使用Mesmer技术进行仿生设计;搭配Tritium操作系统进行智能响应和云端交互;多模态AI集成	具有极高的表情与动作仿真度;能感知周围物体,感知入侵行为并做出反应;可进行艺术与语言创作
	小柴 ^[6]	EX-robot	2024	集多模态大模型、智能表情系统、一体化关节与仿生皮肤技术、智能协同控制能力于一体	具备智能交互问答、表情互动、动作演示、场景适应性服务及角色扮演与定制等多功能

由上述分析可以得出:人形机器人面部情感表达能力的技术实现,主要依赖于面部仿生机械结构设计与多模态情感交互系统的构建与优化的协同优化。前者通过基于仿生原理

的面部机械结构设计,为机器人提供了模拟人类微妙面部表情的物理基础;后者通过多模态情感交互系统的构建与优化,使机器人能够融合多种传感信息,实现情感理解和反馈。

2 面部仿生结构设计

在人形机器人的研究中,面部仿生结构设计是实现情感表达的关键技术之一。面部仿生结构的主要目标是通过合理的机械结构和驱动方式,模拟人类面部肌肉的复杂运动,从而实现逼真的表情表达。

2.1 面部动作编码系统

面部动作编码系统(Facial Action Coding System, FACS)是由心理学家 Ekman 等^[24]于 1978 年提出的一个标准化的面部表情描述系统。基于对面部解剖学的系统研究,他们根据人类面部肌肉的运动规律,将面部肌肉的运动分解为 46 个基本的动作单元^[25](Action Units, AU),每一个 AU 对应一块或一组面部肌肉的运动(如 AU1 对应眉毛内侧抬起, AU12 对应嘴角上扬)。不同 AU 的组合及强度变化,可以展现所有可能的面部表情。人的面部表情主要由 24 个动作单元控制生成,而仅有 14 个 AU 与基本表情相关。

在机器人领域, FACS 的价值在于它提供了一个将抽象、连续的人类情感参数化和离散化的理论框架。它将复杂的表情生成问题转换为更易于工程实现的、对特定 AU 组合进行控制的问题。这为研究者提供了如下便利。

1) 标准化设计:在设计机器人面部机械结构时,可以将驱动器的布局和运动方向直接映射到关键的 AU 上,确保了设计的科学性和系统性。

2) 量化控制:通过控制每个 AU 的激活强度(通常为一个 0 到 1 的浮点数),可以实现从微表情到夸张表情的连续、细腻控制。

3) 跨平台复现: FACS 提供了一种通用的表情语言,使得在一个机器人上开发的表情模型,理论上可以迁移到另一个同样基于 FACS 设计的机器人上。

同时,将 FACS 应用于机器人面临着严峻的实时性挑战:简单地激活 AU 组合会导致切换动作生硬,实现表情平滑过渡需要对多个驱动单元进行毫秒级的协同控制,这对计算能力和控制算法提出了极高的要求。

近年来,随着人工智能技术的发展, FACS 的应用也从简单的 AU 静态映射转向了动态优化。例如,通过强化学习(RL)训练一个策略网络,该网络以目标情感和机器人当前表情状态为输入,实时输出最优的 AU 权重组合^[26]。这种方法能够自主学习表情间的平滑过渡路径,显著降低了手动调节的复杂性,并有效抑制了因驱动器响应不一致而产生的延迟和抖动。此外, Auflem 等^[27]提出了一种 AI 辅助的 AU 评估与控制系统,通过摄像头采集机器人面部动作并使用视觉识别模型实时反馈其 AU 表达强度,从而实现驱动系统对 AU 组合效果的自适应优化。该方法不仅规避了传统驱动表情映射不一致的问题,也为 FACS 在不同机器人平台中的实时部署与性能控制提供了可验证路径,有效提升了面部表情控制的通用性与自然度。

2.2 面部机械结构设计

仿生机械结构是实现 FACS 所描述的 AU 动作的物理基础。其核心目标是设计一套驱动系统,使其能够精确、快速地牵动面部皮肤上的特定控制点,以复现肌肉运动。

按照机械结构的设计理念和技术复杂度的不同,可以将机器人的面部机械结构设计分为“机械式”和“仿生式”两种^[15]。机械式主要侧重于传统的机械结构和驱动方式,主要通过机械结构来实现面部表情的变化,如连杆、齿轮、舵机等。这种设计通常依赖于传统的机械传动方式,通过电机驱动连杆等部件,控制面部各部位的运动,如早稻田大学研制的 KOBIAN 机器人^[14](见图 1(a)),其机械结构采用模块化设计,分为眉毛机构、眼部机构和嘴部机构,各个机构都添加了控制单元。KOBIAN 的模块化机械结构虽成本低、易维护,但其表情细腻度受限。

仿生式则更注重模拟人类面部肌肉的运动和表情变化,采用仿生学原理,模拟人类面部肌肉和皮肤的运动方式,使用软材料、柔性驱动器等先进技术,使机器人的面部表情更加自然、逼真。图 1(b)所示为 Hanson Robotics 生产的 Sophia^[10],它采用基于“柔性驱动-肌肉模拟”结构,即利用柔性驱动元件(如气动人工肌肉、形状记忆合金或电致变形材料等)来模拟人脸肌肉的收缩和舒张,从而实现细腻而自然的面部表情变化,但面临“恐怖谷”风险。

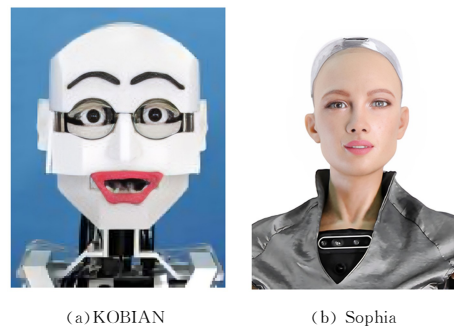


图 1 典型仿人面部表情机器人

Fig. 1 Typical humanoid facial expression robots

当前机器人面部应用最广泛的驱动方式为多模态混合驱动系统,其通过机电-气动协同控制实现表情动态调节。典型技术路径为:基于微型舵机阵列的刚性驱动负责眉弓、颧肌等大位移区域运动,气动人工肌肉(PAM)等柔性驱动模块控制唇部、眼睑等精细表情单元,形成刚柔耦合的仿生驱动架构。表情控制采用情感计算-运动学映射-力位混合控制三级体系^[28]:首先通过深度神经网络(DNN)解析情感语义,继而基于 FACS 系统建立表情参数化模型,最终通过阻抗控制算法协调多驱动单元,实现毫秒级动态表情生成。

2.3 皮肤设计

仿生皮肤是情感表达的最终载体,其材料特性和传感能力直接决定了表情的真实感和交互的深度。这一层的设计至关重要,因为它直接影响机器人的外观和表情自然度。

目前,主流的皮肤材料主要有两种。硅胶因具有良好的加工性能、易成型性和对传感器的基础兼容性,被广泛应用于表情机器人的面部皮肤制作。其制作工艺相对简单,能够满足基本的外观需求,但标准硅胶本身是绝缘材料,其机械性能与真实皮肤仍有差距,难以自然形成复杂的褶皱和微表情。相比之下, Hanson Robotics 研制的“Frubber”皮肤材料^[19]以其卓越的生物机械保真度脱颖而出。它具有肉质般的弹性质地,能够模仿真实的人类肌肉组织,实现极为自然的皮肤

褶皱,使仿真度得到了极大提升^[29]。图2所示为两种材料制成的机器人。

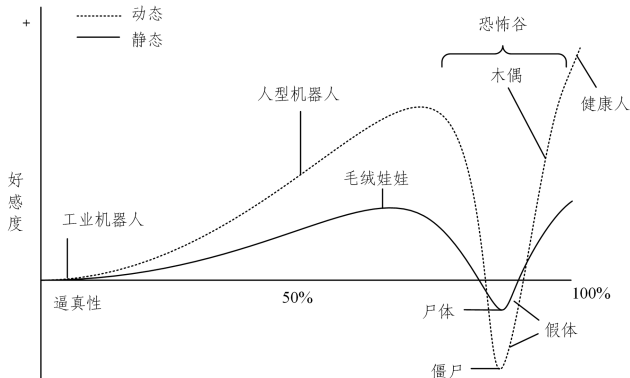


(a) 硅胶

(b) Frubber

图2 硅胶和“Frubber”^[30]材料制成的机器人Fig. 2 Robots made of silicone and “Frubber”^[30] material

然而,随着机器人外观和行为越来越接近人类,出现了一个严峻的挑战,即当这种静态仿真度达到某个临界点时,人类会突然产生强烈的反感或不舒适感,这一现象被称为“恐怖谷效应”(见图3)。这种效应的根源往往在于,机器人高度仿真的外观与其不够自然的动态表情之间存在失配。

图3 恐怖谷理论^[31]Fig. 3 Uncanny valley theory^[31]

为从算法层面消解恐怖谷效应,研究正转向表情平滑过渡算法。例如,利用生成对抗网络(GANs)或变分自编码器(VAEs)学习人类表情数据库中不同情感状态之间的平滑过渡路径^[32]。这些模型可以在潜在空间中进行插值,生成一系列中间表情帧,从而使机器人从“微笑”到“惊讶”的转变不再是两个离散状态的突变,而是一个连续、自然的过程。此外,基于衰减波合成的动态表情生成系统将单个面部动作建模为衰减波,通过波的叠加产生平滑、连续的表情过渡,也能有效避免机械感^[33]。

在应对了外观和动态表情的挑战之后,实现真正智能的仿生皮肤的核心挑战已从单纯的机械驱动转向了更深度的传感器集成问题,这也是当前该领域面临的根本性技术瓶颈。

1)导电性实现与传感器实现的权衡。为了让皮肤“有感觉”,需要在硅胶基体中添加导电填料(如碳纳米管、石墨烯、银颗粒),制成导电硅胶复合材料^[34]。当皮肤受压或拉伸时,内部导电网络发生改变,导致电阻变化,从而实现触觉的感知。然而,这引入了新的权衡:导电填料的增加虽然会提升导电性,但通常会牺牲材料的机械性能,导致其变脆或变硬,偏离了仿生皮肤的初衷。

2)信号串扰。在需要集成多种传感器的多模态电子皮肤中,不同传感器的电信号可能会相互干扰,导致数据损坏和读数不准确,这极大地限制了感知能力的复杂度和可靠性。

3)绝缘表面层问题。在导电聚合物复合材料(如导电硅胶)的表面,往往会自然形成一个纳米级的绝缘氧化层。这个薄层会严重阻碍传感器与电极之间的有效电接触,导致测量信号极其不稳定,重复性差,这也是许多电子皮肤研究难以被复现的根源之一。

4)机械失配与耐久性。传感器与柔软皮肤基底之间的杨氏模量差异会导致界面应力集中,在外力作用下容易发生分层或脱落。此外,软体材料的磨损和蠕变问题,意味着基于数据驱动的AI模型需要频繁地对传感器进行重新校准,这对于要求长期稳定运行的机器人系统而言是一个巨大的实际障碍。

综上所述,仿生皮肤技术已经从一个纯粹的材料外观问题,演变为一个涉及导电性、传感器设计、信号处理和长期可靠性的深度交叉学科挑战。为了应对这些挑战,各国研究人员正在积极探索新的解决方案。例如,Xu等^[35]开发了一种自供电的仿生电子皮肤,其可以全面感知液滴的二维滑动行为,并在指示器上实时直观地反馈,这一策略突破了电子皮肤感知液滴的局限性,大大缩小了人造电子皮肤与人体皮肤在感知功能上的差距。

3 多模态情感交互系统的构建与优化

多模态交互融合技术是人机情感交互的核心技术。该系统通过整合情感状态建模、多模态感知融合与动态交互控制,构建从环境感知到情感决策的闭环框架。其以深度学习为基础,结合跨模态同步机制,实现情感表达的精准映射与自适应调节。通过优化情感语音、面部表情及肢体动作的协同作用,可显著提升交互的自然性与可信度,从而推动人形机器人在医疗陪护、社交服务等场景的实用化突破。

3.1 情感计算模型

人形机器人的情感表达是基于人为设计和构建的情感模型来实现的,通过形式化描述与数学建模对人类情绪进行界定,从而为构建机器人心理模型提供理论依据,还通过整合传感器、算法和硬件平台,实现对情感状态的实时捕捉、处理和表达,以提升机器人在人机交互中的情感智能水平。为实现高效的情感数学模型,通常需要借助于维度空间分析、统计学方法等对情感特征进行量化和形式化描述^[36]。目前,人工情感领域的研究成果已广泛应用于各类机器人系统,显著提升了机器人的自然情感交互能力。例如,SoftBank Robotics推出的Pepper^[37]机器人可通过识别人脸表情、声音和肢体语言,实时调整自身行为以回应用户的情绪。

在情感计算领域,多个典型的情感模型被提出,以不同的方式对人类情感进行建模和分析。日本早稻田大学的WE-4R^[38]机器人将情感空间分为高兴、激励和肯定三维空间,从而根据内外环境刺激相应地改变其情感状态。比较典型的情感模型是Mehrabian等^[39]提出的PAD模型,将情感分为愉悦度、唤醒度和支配度3个维度,并通过将所有情感映射到情感空间中的不同坐标点来量化PAD情感模型。该模型在情

感识别和情感分析中得到了广泛应用。Huang 等^[40]提出了基于博弈论的认知情感交互模型,该模型在 PAD 情感空间的基础上,通过分析参与者的情感输入和人机交互关系,建立情感生成过程的数学模型,并采用嵌入博弈的子博弈完美均衡策略作为机器人的最优情感选择策略,减少了机器人对外界情感刺激的依赖并提高了参与者的参与度。Liu 等^[41]提出了一种基于 AU 的连续域面部表情生成模型 AU-FEDS,并将其应用于开发的情感机器人中。该模型能够准确地表达丰富且自然的面部表情,并通过与机器人的物理面部运动相结合,实现了更真实的人类情感交互。Mishra 等^[42]利用大型语言模型(如 GPT-3.5),将情绪评估解释为对话中的情绪识别任务,通过对话历史实时预测机器人当前轮次应表现的情绪,并使用面部表情来表示预测的情绪。

3.2 面部表情合成

面部表情是重要的非语言行为表达方式,可以有效地将自身的感受和情绪传达给他人。Mehrabian 在其关于矛盾情感信息传递的研究中指出,55%的信息是通过肢体语言传达的,38%的信息是通过语气传达的,仅 7%的信息是通过文字传达的^[43]。尽管该结论的普适性存在争议,但它深刻揭示了非语言线索在情感交流中的压倒性重要性。因此,面部表情作为非语言行为,对于沟通至关重要,即使在不同的文化或语言之间,也可以通过面部表情来传达基本情绪。

面部表情合成技术是一种通过计算机生成机器人面部表情的技术,它通过结合机器人技术、计算机视觉和人类情感理论等技术理论,创造出可以显著增强人机交互的动态表情,并使这些面部表情看起来自然、真实,还能够根据情境和情感状态进行变化。传统的面部表情合成技术主要依赖人们手动定义的关节配置和参数化模型,通过几何特征、表情动作单元和特征点映射等方法生成表情,但在细节处理和适应性方面存在局限。如今,数据驱动的方法已基本占据主导地位。现代的面部表情合成技术结合了先进的机械设计、柔性仿生材料、多模态驱动系统和机器学习等技术,实时感知并动态生成适配人类情感的自然表情。现代系统将机器人关节配置转换为捕捉基本人类特征的简化面部图像,基于 FACS 的框架再将基本动作单元映射到物理动作。例如,ExGenNet 等框架使用深度生成模型与基于 CNN 的表情分类器相结合,自动学习每个表情的一系列关节配置^[44]。有些方法超越了静态映射,采用了动态合成技术,如日本大阪大学等^[45]联合提出的基于衰减波合成的动态表情生成系统,使用“波形运动”的技术将单个面部表情(如眨眼或细微的肌肉抽搐)建模为衰减波,当这些波叠加时,它们会在不同的表情之间产生平滑、连续的过渡,使机器人的脸看起来更自然,减少机械感。

3.3 情感语音合成与多模态交互融合

3.3.1 情感语音合成技术演进

情感语音合成是赋予机器人拟人化语音表达能力的关键技术,通过声学参数调制与韵律模型构建,将文本或情感状态映射为具有特定情感色彩的语音信号。早期的情感语音合成主要依赖于统计建模方法,如隐马尔可夫模型和高斯混合模型,通过建模不同的语音参数来生成带有情感的语音。随着深度学习的发展,深度神经网络、长短期记忆网络等被应用于

情感语音合成,显著提高了语音的自然度和情感表达能力。近年来,深度学习端到端模型在情感语音合成中得到了广泛应用,它可直接从文本生成高保真情感语音,结合对抗训练或扩散模型进一步优化情感表达的细腻度^[46]。当前,研究者开始关注如何合成具有混合情感的语音,以更真实地反映人类复杂的情感状态,这需要更精细的情感建模和控制策略。例如,Zhou 等^[47]提出的基于序列到序列模型的情感语音合成框架,能够合成混合情感并在运行时控制情感渲染,以提高合成语音中的情感可理解性。

目前,情感语音合成技术已经在机器人中得到了广泛应用。Liu 等^[48]提出了一种基于循环一致生成对抗网络的情感语音转换方法,它将潜在的情感分为自然的和转化的语音。基于这种方式的 Erica 机器人可以在情感陪伴中分享相似的情感体验,更好地传递低沉、鼓舞人心的情感^[49]。Lee^[50]采用可生成情感语音的日本 TTS 引擎 AItalk,赋予了机器人 NAO 情感表达能力,并通过手动修改声音特征来控制情绪。

3.3.2 多模态时空对齐机制

在多模态人机交互中,实现情感语音、面部表情、肢体动作和生理信号的无缝同步对于提升机器人的情商和交互流畅性至关重要。多模态时空对齐机制确保不同模态协调工作,从而产生连贯自然的情绪反应。为了有效地模拟人类情绪,机器人必须整合来自语音、视觉和运动等各种来源的感官输入,包括语音的频谱与韵律特征、基于面部动作编码系统的面部动作单元以及肢体动作的轨迹与姿态等。

实现多模态时空对齐面临着严峻的技术挑战。其主要挑战之一是确保不同通道之间的时间一致性,因为语音随着时间的推移按顺序展开,但面部表情和手势通常会滞后或在不同的时间尺度上发生。此外,不同模态的数据在表示上存在异构性(如像素、波形、关节角度),语义上也不对称,这使得构建统一的表示空间和捕捉其深层关联变得极为困难。

为应对这些挑战,现代计算框架正以大语言模型为认知核心。情感语言生成是实现对齐的关键前置模块,基于 GPT-4 等模型,系统可通过情感状态向量动态生成语义与情感一致的文本内容。例如,Meta 的 BlenderBot 3^[51]在对话生成中引入情感分类器,优先选择与目标情感匹配的词汇,并通过强化学习与语音合成模块联合优化,确保文本情感强度与语音韵律参数的动态适配。

在将生成的语言与物理动作对齐方面,研究者开发了先进的技术。一种前沿方案是 Google DeepMind 团队提出的 Q-Transformer^[52]框架,它通过整合 Transformer 架构与强化学习,在机器人多模态交互任务中实现了跨模态协同优化,其核心创新在于将连续的动作空间(如关节角度)离散化为一系列“动作词元”。通过这种方式,Transformer 的注意力机制可以在决定下一个动作的同时“关注”到语言指令、视觉信息和机器人当前状态,从而将语音、视觉与机械动作进行时空对齐,并通过 Q-Learning 优化生成时序上协调且能成功完成任务的完整动作序列。

另一种更工程化的实现路径体现在 Pepper 机器人上,其核心是情感引擎和 NAOqi 操作系统 SDK。Pepper 的情感引擎会根据识别到的用户情绪,调用预设脚本来协同调整语音

和表情。开发者可通过 SDK 对语音进行精细的参数化协同控制^[53]。例如,在文本转语音指令中直接嵌入标签来修改韵律:“\\vct=85\\”表示将语速调整为 85%(更慢);“\\rspd=90”表示将音调降低到 90%(更低沉)。在执行这些语音指令的同时,系统可以并行调用动画任务,控制眼部 LED 灯变为代表同情的柔和蓝色,并执行轻微的点头动作,从而通过精确的参数化脚本实现语音与表情的协同。

为了进一步完善多模态情绪表达,生成对抗网络在合成和增强一致性方面发挥着关键作用。条件生成对抗网络预测和完善跨模态情绪表达,确保面部动作和手势与口语情绪语境相符;变分自动编码器通过在学习到的情感状态之间进行插值来实现细微的情绪表达,减少突然的转变或不匹配^[54]。通过整合这些生成技术,机器人可以减少语音、表情和手势之间的语义不一致,使情绪反应更加真实。多模态时空对齐机制的实现方式如图 4 所示。

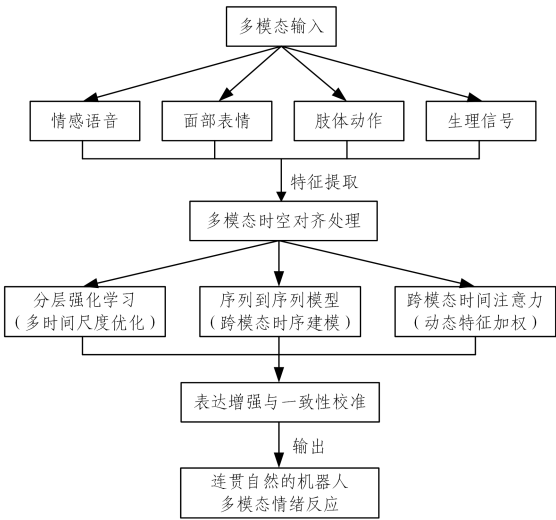


图 4 多模态时空对齐机制的实现

Fig. 4 Implementation of multimodal spatio-temporal alignment mechanism

多模态时空对齐机制已成功应用于各种应用,例如 Pepper 和 Erica 等情感助手利用多模态同步富有同理心的言语、表情和手势增强用户参与度。Aly 等^[55]提出了一种通过多模态机器人行为适应视频情感内容的方法,以引起人类用户的情感反应。他们使用了表情、手势和语音合成技术,并通过实验验证了多模态机器人的行为可以显著提高交互情感内容的表现力和清晰度。未来的研究旨在开发用于情绪适应、实时多模态情绪生成和跨文化情感建模的自我监督学习,使人机交互从脚本响应推进到真实和情境感知的情绪行为。基于这些进步,机器人可以与人类用户建立更深层、更有意义的情感联系。

4 结论与展望

本文系统综述了人形机器人面部情感表达技术的发展历程、核心技术与未来趋势。通过将技术演进划分为以核心技术为标志的 3 个阶段,揭示了该领域从简单的机械模仿向复杂的智能共情演变的清晰脉络。研究表明,当前的技术前沿正由两大引擎共同驱动:一是物理实现的突破,特别是以智能

仿生皮肤为代表的传感技术^[56];二是计算框架的革新,以大语言模型为核心的认知架构正在统一过去分散的情感计算模块^[57],并通过先进的多模态时空对齐技术,实现表情、语音和动作的高度协同。

然而,要实现机器人与人类社会真正的情感共融,当前的技术路径仍面临诸多未解难题,这些难题构成了该领域未来研究的核心议程。

1)物理实现的根本瓶颈:高保真、低成本、可大规模制造且内置高密度传感网络的电子皮肤仍是重大挑战。解决导电性与机械性能的权衡、信号串扰、绝缘表面层以及长期耐久性问题,是决定机器人获得真正“触觉”的关键。

2)计算框架的深层困境:多模态大模型的集成带来了计算复杂性与文化偏见的双重困境。前者对资源受限的机器人平台构成巨大压力,亟需在模型压缩和边缘计算部署上取得突破。后者则更为棘手,其中,跨文化情感模型的适配性是当前最严峻的未解难题之一^[57]。由于高质量、跨文化的情感交互标注数据极其稀缺,现有模型大多基于西方文化数据训练,难以泛化至全球多样的文化背景,这可能导致机器人在特定文化中做出不恰当甚至冒犯性的情感表达。

3)人机交互的终极挑战:消解“恐怖谷”效应不仅是为了提升外观逼真度,更是解决多模态失配的算法难题。实现所有表达通道在时序上的完美同步与语义上的和谐一致,即“形神兼备”,是通往自然人机交互的必经之路,但目前仍缺乏鲁棒的解决方案。

4)社会与伦理对齐的挑战:技术的飞速发展远超伦理与法规的构建速度。如何应对“情感欺骗”的风险,是一个具体挑战。例如,Hanson Robotics 的 Sophia 机器人就曾因其看似自主的情感对话在很大程度上是预设脚本而引发广泛批评,许多 AI 专家指出这种呈现方式可能误导公众,造成“情感欺骗”^[58]。为了应对此类风险,具体的伦理约束正在形成。例如,欧盟的《人工智能法案》(AI Act)明确禁止在工作场所和教育机构使用 AI 系统来推断情绪,除非是出于医疗或安全目的^[59]。这标志着为机器人情感表达技术设立明确的法律和伦理边界已成为全球共识。未来的研究必须采用可解释人工智能(XAI)和差分隐私等技术^[60],在保证交互效果的同时,确保决策的透明度和用户数据的安全。

综上所述,人形机器人面部情感表达技术的根本目的并非创造人类情感的替代品,而是构建增强社会连接、提供有效情感支持的桥梁。唯有直面并系统性地解决上述在物理实现、计算框架、交互体验和社会伦理等技术路径中的未解难题,在追求机器人“形神兼备”的道路上审慎前行,方能最终实现机器人与人类社会的和谐共生。

参 考 文 献

[1] ESFANDBOD A, ROKHI Z, MEGHDARI A F, et al. Utilizing an emotional robot capable of lip-syncing in robot-assisted speech therapy sessions for children with language disorders[J]. International Journal of Social Robotics, 2023, 15(2): 165-183.

[2] IVANOV S, WEBSTER C, SEYITOĞLU F. Humans and/or robots? Tourists' preferences towards the humans-robots mix in

- the service delivery system[J]. *Service Business*, 2023, 17(1): 195-231.
- [3] PODPEČAN V. Can you dance? A study of child-robot interaction and emotional response using the NAO robot[J]. *Multimodal Technologies and Interaction*, 2023, 7(9): 85.
 - [4] TAKANISHI A, HIRANO S, SATO K. Development of an anthropomorphic head-eye system for a humanoid robot-realization of human-like head-eye motion using eyelids adjusting to brightness[C]// *Proceedings 1998 IEEE International Conference on Robotics and Automation*. IEEE, 1998: 1308-1314.
 - [5] BREAZEAL C L. *Sociable machines: expressive social exchange between humans and robots*[M]. Cambridge, MA: Massachusetts Institute of Technology, 2000.
 - [6] WANG W. Top 10 Innovative Humanoid Robot Products of 2024 Revealed [N]. *China Electronics News*, 2025-01-17(6).
 - [7] TAKANISHI A, TAKANOBU H, KATO I, et al. Development of the anthropomorphic head-eye robot WE-3RII with an autonomous facial expression mechanism [C] // *Proceedings 1999 IEEE International Conference on Robotics and Automation*. IEEE, 1999: 3255-3260.
 - [8] ALLMAN T. *Nexi robot, the (A Great Idea)* [M]. Norwood House Press, 2009.
 - [9] MIWA H, ITOH K, MATSUMOTO M, et al. Effective emotional expressions with emotion expression humanoid robot WE-4RII[C]// *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*. IEEE, 2004: 2203-2208.
 - [10] FORTUNATI L, SORRENTINO A, FIORINI L, et al. The rise of the roboid[J]. *International Journal of Social Robotics*, 2021, 13(6): 1457-1471.
 - [11] SHARMA N, PANDEY J K, MONDAL S. A review of mobile robots: applications and future prospect[J]. *International Journal of Precision Engineering and Manufacturing*, 2023, 24(9): 1695-1706.
 - [12] YOUSIF M. Humanoid robot enhancing social and communication skills of autistic children: review[J]. *Artificial Intelligence & Robotics Development Journal*, 2021, 1(2): 80-92.
 - [13] ABDOLLAHI H, MAHOOR M H, ZANDIE R, et al. Artificial emotional intelligence in socially assistive robots for older adults: a pilot study[J]. *IEEE Transactions on Affective Computing*, 2023, 14(3): 2020-2032.
 - [14] KISHI T, OTANI T, ENDO N, et al. Development of expressive robotic head for bipedal humanoid robot[C]// *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012: 4584-4589.
 - [15] HE M, BI J, WANG Y J, et al. A Review of Emotional Expression Technology in Humanoid Robots [J]. *Journal of Chongqing University of Technology(Natural Science)*, 2022, 36(11): 183-192.
 - [16] HASHIMOTO T, HITRAMATSU S, TSUJI T, et al. Development of the face robot SAYA for rich facial expressions[C]// *2006 SICE-ICASE International Joint Conference*. 2006: 5423-5428.
 - [17] WU W G, SONG C, MENG Q M. Research and Experiment of Voice and Lip-shaped System of Android Head Robot “H&frobot-III” [J]. *Journal of Machine Design*, 2008(1): 15-19.
 - [18] CHEN W L, HUANG J H, YANG Z P. Research and Development of AI Autonomous Guided Tour Robot Based on 3D Vision [J]. *Computer Knowledge and Technology*, 2024, 20(5): 95-98.
 - [19] OH J H, HANSON D, KIM W, et al. Design of android type humanoid robot albert HUBO[C]// *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006: 1428-1433.
 - [20] ISHIHARA H, YOSHIKAWA Y, ASADA M. Realistic child robot “affetto” for understanding the caregiver-child attachment relationship that guides the child development[C]// *2011 IEEE International Conference on Development and Learning(ICDL)*. 2011: 1-5.
 - [21] MAZZEI D, ZARAKI A, LAZZERI N, et al. Recognition and expression of emotions by a symbiotic android head[C]// *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2014: 134-139.
 - [22] GHORBANDAEI POUR A, TAHERI A, ALEMI M, et al. Human-robot facial expression reciprocal interaction platform: Case studies on children with autism[J]. *International Journal of Social Robotics*, 2018, 10(2): 179-198.
 - [23] SUN J L. Manufacturing the Future: Integration and Innovation of Technology and Livelihood [J]. *Contemporary Television*, 2020(2): 45-47, 98.
 - [24] EKMAN P, FRIESEN W V. Facial action coding system[M]// *Environmental Psychology & Nonverbal Behavior*. 1978.
 - [25] LI Y, ZHU L, ZHANG Z, et al. Humanoid robot heads for human-robot interaction: a review[J]. *Science China Technological Sciences*, 2024, 67(2): 357-379.
 - [26] HUSSAIN N, ERZIN E, SEZGIN T M, et al. Training socially engaging robots: modeling backchannel behaviors with batch reinforcement learning[J]. *IEEE Transactions on Affective Computing*, 2022, 13(4): 1840-1853.
 - [27] AUFLEM M, KOHTALA C, JUNG M F, et al. Facing the FACS-Using AI to Evaluate and Control Facial Action Units in Humanoid Robot Face Development[J]. *Frontiers in Robotics and AI*, 2022, 9: 887645.
 - [28] ZHANG D, PENG J, JIAO Y, et al. ExFace: expressive facial control for humanoid robots with diffusion transformers and bootstrap training[J]. *arXiv:2504.14477*, 2025.
 - [29] SNYDER D E, SAPPER E D. See me, feel me, touch me, heal me: a contextual overview of conductive polymer composites as synthetic human skin[J]. *Journal of Composites Science*, 2022, 6(5): 141.
 - [30] HABIB A, DAS S K, BOGDAN I C, et al. Learning human-like facial expressions for android phillip K. Dick[C]// *2014 IEEE International Conference on Automation Science and Engineering(CASE)*. IEEE, 2014: 1159-1165.
 - [31] MORI M, MACDORMAN K, KAGEKI N. The uncanny valley [from the field][J]. *IEEE Robotics & Automation Magazine*, 2012, 19(2): 98-100.
 - [32] TANG H, SEBE N. Facial expression translation using landmark guided GANs[J]. *IEEE Transactions on Affective Computing*, 2022, 13(4): 1986-1997.

- [33] ISHIHARA H, HAYASHI R, LAVIEILLE F, et al. Automatic generation of dynamic arousal expression based on decaying wave synthesis for robot faces [J]. *Journal of Robotics and Mechatronics*, 2024, 36(6): 1481-1494.
- [34] KIMIZUKA K, AZHARI S, TOKUNO S, et al. Haptic in-sensor computing device made of carbon nanotube-polydimethylsiloxane nanocomposites[J]. *arXiv*: 2406. 03958, 2024.
- [35] XU Y, SUN Z, BAI Z, et al. Bionic e-skin with precise multi-directional droplet sliding sensing for enhanced robotic perception [J]. *Nature Communications*, 2024, 15(1): 6022.
- [36] CHEN G W, ZHANG P Z, WANG T, et al. Overview of Multimodal Sentiment Analysis [J]. *Journal of Communication University of China (Natural Science Edition)*, 2022, 29(2): 70-78.
- [37] PANDEY A K, GELIN R. A mass-produced sociable humanoid robot; Pepper; the first machine of its kind[J]. *IEEE Robotics & Automation Magazine*, 2018, 25(3): 40-48.
- [38] MIWA H, ITOH K, ITO D, et al. Design and control of 9-DOFs emotion expression humanoid arm [C] // *IEEE International Conference on Robotics and Automation (ICRA '04)*. 2004. IEEE, 2004: 128-133.
- [39] MEHRABIAN A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament[J]. *Current Psychology*, 1996, 14(4): 261-292.
- [40] HUANG H C, LIU N, HU M, et al. A Game-Based Cognitive Emotional Interaction Model for Robots [J]. *Journal of Electronics and Information Technology*, 2019, 41(10): 2471-2478.
- [41] LIU X, NI R, YANG B, et al. Unlocking human-like facial expressions in humanoid robots; a novel approach for action unit driven facial expression disentangled synthesis[J]. *IEEE Transactions on Robotics*, 2024, 40: 3850-3865.
- [42] MISHRA C, VERDONSCOT R, HAGOORT P, et al. Real-time emotion generation in human-robot dialogue using large language models[J]. *Frontiers in Robotics and AI*, 2023, 10: 1271610.
- [43] MOSTAFAVI M, PORTER M D. How emoji and word embedding helps to unveil emotional transitions during online messaging[C] // *2021 IEEE International Systems Conference (syscon)*. 2021: 1-8.
- [44] RAWAL N, KOERT D, TURAN C, et al. ExGenNet: learning to generate robotic facial expression using facial expression recognition[J]. *Frontiers in Robotics and AI*, 2022, 8: 730317.
- [45] ISHIHARA H, HAYASHI R, LAVIEILLE F, et al. Automatic generation of dynamic arousal expression based on decaying wave synthesis for robot faces [J]. *Journal of Robotics and Mechatronics*, 2024, 36(6): 1481-1494.
- [46] MANDEEL A R, AL-RADHI M S, CSAPÓ T G. Speaker adaptation experiments with limited data for end-to-end text-to-speech synthesis using Tacotron2[J]. *Infocommunications Journal*, 2022, 14(3): 55-62.
- [47] ZHOU K, SISMAN B, RANA R, et al. Speech synthesis with mixed emotions [J]. *IEEE Transactions on Affective Computing*, 2023, 14(4): 3120-3134.
- [48] LIU S, CAO Y, MENG H. Emotional voice conversion with cycle-consistent adversarial network[J]. *arXiv*: 2004. 03781, 2020.
- [49] FU C, DENG Q, SHEN J, et al. A preliminary study on realizing human-robot mental comforting dialogue via sharing experience emotionally[J]. *Sensors*, 2022, 22(3): 991.
- [50] LEE J. Generating robotic speech prosody for human robot interaction; a preliminary study [J]. *Applied Sciences*, 2021, 11(8): 3468.
- [51] SHUSTER K, XU J, KOMEILI M, et al. BlenderBot 3: a deployed conversational agent that continually learns to responsibly engage[J]. *arXiv*: 2208. 03188, 2022.
- [52] CHEBOTAR Y, VUONG Q, HAUSMAN K, et al. Q-transformer: scalable offline reinforcement learning via autoregressive Q-functions[C] // *Conference on Robot Learning*. PMLR, 2023: 3909-3928.
- [53] GARCIA S, GOMEZ-DONOSO F, CAZORLA M. Enhancing human-robot interaction; development of multimodal robotic assistant for user emotion recognition[J]. *Applied Sciences*, 2024, 14(24): 11914.
- [54] HUANG W C, HSUEH Y L. An emotional dialogue system using conditional generative adversarial networks with a sequence-to-sequence transformer encoder[J]. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2024, 23(12): 164: 1-164: 23.
- [55] ALY A, TAPUS A. Multimodal adapted robot behavior synthesis within a narrative human-robot interaction[C] // *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2015: 2986-2993.
- [56] YAN Z, SONG Y, ZHOU R, et al. Facial expression realization of humanoid robot head and strain-based anthropomorphic evaluation of robot facial expressions[J]. *Biomimetics*, 2024, 9(3): 122.
- [57] WANG C, HASLER S, TANNEBERG D, et al. LaMI: large language models for multi-modal human-robot interaction[C] // *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 2024: 1-10.
- [58] INDURKHYA B. Ethical aspects of faking emotions in chatbots and social robots[C] // *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 2023: 1719-1724.
- [59] RUDSCHIES C, SCHNEIDER I. The long and winding road to bans for artificial intelligence; from public pressure and regulatory initiatives to the EU AI act[J]. *Digital Society*, 2025, 4(2): 57.
- [60] DORAFSHANIAN M, AITSAM M, MEJRI M, et al. Beyond data collection: safeguarding user privacy in social robotics [C] // *2024 IEEE International Conference on Industrial Technology (ICIT)*. 2024: 1-6.



KE Xianxin, born in 1973, postdoctoral, associate professor. His main research interests include intelligent robot development, artificial intelligence applications, multimodal emotional state recognition and emotional interaction.