

AI 合成主播与真人主播的语言审美比较研究

李滨惠

(吉林艺术学院, 吉林 长春 130000)

摘要:随着“智媒”技术的迅速迭代, AI 合成主播在新闻播报、综艺节目、电商直播等场景中广泛出现, 他们正在重塑传统的播音主持格局。相较于之前技术路线与产业影响的研究, AI 合成主播与真人主播在“语言审美”差异的梳理仍然有限。而本文在系统的梳理过相关文献之后, 引入具身传播、媒介可供性与有声语言审美等相关理论, 对 AI 合成主播和真人主播在声音形态、语言风格、情感表达与价值意蕴等维度进行比较分析, 揭示出 AI 合成主播在可控性、效率与标准化方面的优势, 以及在气息韵律、情感质感和文化意蕴方面的不足。在未来的新闻传播实践中, 可能会出现“人机协同”的主播格局, AI 合成主播会承担结构化、常规化的信息播报, 真人主播则通过具身在场和个体风格来维系语言审美高地。

关键词: AI 合成主播; 真人主播; 语言审美; 具身传播; 播音主持

一、引言

人工智能驱动的虚拟数字人已经成为媒体转型中的显眼符号。目前多家主流媒体先后推出了 AI 合成主播, 将他们用在新闻快讯、数据播报、专题解说等场景当中。如今大部分技术公司可以在短时间内为不同频道“定制”虚拟主持形象, 并配合语音合成和表情驱动, 实现全天候播出。这种新型主播形式的出现, 不仅缓解了人力紧张、制作周期长等问题, 还引发了“机器人是否会取代真人”的担忧。在传播学界, 关于 AI 主播的研究已经形成了若干个方向, 有人关注技术路径和平台实践, 有人讨论人机关系和伦理边界, 还有人会从受众感知视角检验其可信度和接受度^[2]。并且在关于播音主持语言艺术、有声读物播读以及“中华声音”审美传统的研究中积累深厚^{[7][8][9]}。如果将两条研究线索合在一起, 我们能够看到一个尚未被充分展开的问题: 从语言审美意义角度来讲, AI 合成主播与真人主播究竟有何差异? 这里所说的“语言审美”, 不仅指声音是否悦耳、吐字是否清楚, 更包括节奏组织、情感渲染、人格呈现以及声音背后的价值立场。就像新闻节目中的一句“我们一起守望这座城”, 综艺现场的一句即兴调侃, 都不只是技术层面的发声, 而是审美和意义同时发生的瞬间。AI 合成主播的加入, 让“谁在说话”变得更加复杂, 是算法、机构还是一个被设计出来的“人格”? 而本文将采用纯文献的研究方法, 在现有的研究基础上进行文献梳理与观念整合。目的不在于简单判定“谁更好”, 而是揭示两类主播各自的审美优势与限制, 为后续的人机协同实践提供一个较为清晰的观察框架。

二、理论基础与分析维度

(一) 智媒传播与“类主体”

在移动互联网和算法推荐主导的信息环境中, 传播早已不局限于“人对人”的线性模式了。有的研究者指出, 在以大数据和人工智能为支撑的智媒系统中, 技术不再只是中性的通道, 而是带着偏好参与内容的筛选、组合与生成, 呈现出“类主体”的特征^{[5][6]}。在主播的语境下, 这也意味着 AI 不只是在后台辅助剪辑和写稿, 而是以“说话者”的形式出现在屏

作者简介: 李滨惠 (2001-), 女, 硕士研究生, 研究方向为播音与主持艺术。

幕上,与真人主播共同构成的新的主体结构。本研究并不试图深入展开算法机制,而是将这一判断作为前提,AI合成主播已经成为观众眼中的“角色”,而不只是“工具”。因此,对其语言审美的分析,应当像分析真人主播一样,从声音、文本和身份三者之间的关系入手。

（二）具身传播视角

具身传播的理论时刻提醒我们,语言不是悬浮在空气中的抽象符号,而是镶嵌在身体动作、空间环境和情绪状态之中的整体行为。真人主播的声音离不开呼吸节奏、姿态变化和目光交流,这也正是这些看似“无意”的细节构成了观众感受到的“在场感”^[1]。当虚拟形象代替身体出现在屏幕上时,具身结构被打散并重新组装,虽然技术可以模拟嘴型和眨眼,可以让虚拟人物做出标准化的手势,但是却难以再现紧张时略微发抖的声音、或是突然加重的某个词语、因情绪激动而拉长的停顿。这些细小的变化才是构成真人语言审美的“活性”。

（三）有声语言审美传统

从本土文化脉络上来看,广播电视语言的背后有着漫长的有声传统。戏曲、说唱、评书和地方曲艺都强调“口传心授”的艺术经验,形成了强调气韵、注重节奏和追求“声情并茂”的审美观念^[8]。在现代媒介环境中,有声语言的评价逐渐被细化,针对有声读物播读的研究提出了以吐字清晰、音色纯度、节奏层次和情感表达为核心的多维度指标,这也说明了语言审美既可以被整体感知,也可以被拆分成若干个可观察、可训练的要素^[7]。而在此基础上,本文会将比较维度概括为:声音形态:音色、清晰度、高低强弱;语流节奏:语速、停连、重音与句群组织;情感表达:声音中的情绪起伏、态度温度;主体形象:声音背后呈现的人格感、可信度和价值立场的这四种维度。

三、AI合成主播的语言审美特征

（一）声音高度“整洁”，但缺少“毛边”

AI合成主播的声音首先就会给人一种“整洁”的印象。普通话标准、共鸣均匀、背景干净,很少出现气息声和口误。对于新闻快讯、数据播报这类强调信息清晰度的内容,这种声音无疑是合格的甚至是理想的。然而这也会存在一些问题,比如过分“整洁”的声音往往也缺少“毛边”。真人主播偶尔的换气声、轻微破音或短暂停顿,会让观众感到“有人在说话”。这些不完美之处在AI合成语音中几乎被完全抹平,听感上一切都“刚刚好”,所以就会显得缺少生命力。从审美的角度来看,AI合成主播目前主要奉行的是“技术美学”,它会尽量减少噪音和偶然性,以稳定、均匀的语速作为主要目标。在资讯密集型节目中,这样的声音有它一定的合理性,但如果节目需要营造情绪氛围,或者希望通过声线体现人物的性格和命运,这种高度“抛光”的音色就会略显单薄。

（二）节奏依赖算法，难以随场而变

AI合成主播的语速和节奏通常是由系统根据句长和标点自动设定。一般而言,合成系统会保持中等偏快的语速,以此适应移动端用户短时收看的习惯。这种节奏对于简短、结构清晰的新闻信息是有效的,但如果涉及到情节推进或情绪递进的内容时,问题就会暴露出来。例如,在报道灾害、追思或庄严仪式时,真人主播往往会本能地放慢语速,加长停顿,用较低的音高和收束的语势来表达敬意和哀悼,而AI合成主播如果没有专门设置过相应模板,往往会仍按一般新闻节奏“平推”过去,给人一种“没进入情绪”的感觉。即便技术可以通过“情感标签”人为调整语速和语调,但这种调整也会带有明显的“外在设计”痕迹,缺少从情境中自然生长出来的连贯感。

（三）情感表达主要依赖视觉和设定

当观众谈及“喜欢某个主播”，往往说的是“感觉他/她很真诚”“和我们站在一起”。这种好感不只是来自词语的内容，更来自声音里的态度和情绪。在AI合成主播那里，人格感和情绪主要通过视觉形象和后台设定完成的。角色被赋予姓名、年龄、爱好和性格标签，但这些设定并未真正进入到语音系统的内部。而此时，观众看到的是一个表情生动、衣着考究的虚拟人物，但听到得却是一段几乎可以用于任何角色的标准播音腔。从语言审美的角度来看，这种“视觉人格—语言中性”的分离，会让观众在新鲜感过去后产生距离感，形象很精致，声音却像某种“公共模板”。如果未来要提升AI合成主播的语言审美水平，它的关键不在于再增加多少表情细节，而在于如何让语音模块真正“长出”与具体角色和情境相匹配的声线和节奏。

（四）主体形象容易被“机构化”

AI合成主播的幕后是技术团队和机构决策。谁来写稿、谁来设定话语风格、谁来控制播出场景，决定了它在屏幕上“怎样说话”。从这个角度上来说，AI合成主播更像是机构话语的“化身”。这种特性具有双重效果，一方面，它确保了机构可以精确控制每一句话的措辞和语气，减少个人发挥带来的不可控风险；另一方面，也削弱了观众对“主播个人”的情感认同空间。观众很难把一个明显由系统生成的声音当作有血有肉的“新闻人”，因此在面对敏感议题或价值冲突时，观众才会更愿意听到真人主播的声音。

四、真人主播的语言审美特征

（一）身体与声音共同构成“现场感”

真人主播的声音从来都不是孤立存在的。微微前倾的上身、握稿的手势、与摄影机短暂的对视，都在悄悄的影响他说出下一句话的方式。呼吸稍重，声音就会带一点哽咽，身体紧绷，语速往往会不自觉加快。对于熟练的专业人士而言，这些变化并非完全无意识的，而是他与稿件的内容和节目的整体氛围相互调节的结果。也正是因为这种“身体—声音—空间”连在一起的状态，才会让观众在听到真人主播说话时，会不自觉地感觉“他就在这里”。在一场重大的直播中，观众往往会对主播的某个眼神、某句略带颤抖的话印象深刻，这种记忆本身就是一种审美经验。

（二）个人经历沉淀出的语言气质

许多主播在长期工作中逐渐形成了独特的语言气质。有的擅长平静而坚定的叙述，有的善于用轻松幽默化解紧张气氛，有的在文化类节目中展现出温雅、克制的风格。这些气质并非单靠课堂训练就能获得，而是与其成长背景、阅读经验、性格特征和职业选择相互缠绕的结果。观众之所以会说“听这个人的声音就安心”，往往正是因为他在反复接触过程中，把主播声音与某种稳定的人格特质联系在一起。这种“人格化声音形象”，是AI合成主播目前尚难以企及的。

（三）对文本的“二度创作”

在现代传媒机构中，主播很少是“照本宣科”的朗读者。即便在高度规范的新闻联播里，他也要根据当天的议题轻重和段落结构，在语势、停顿和重音上做出细微的安排。有经验的主播在拿到稿件后，往往会先快速划分语群，标记情绪转折点和核心观点，再通过排练找到合适的声调走向。这种“二度创作”的过程，本质上就是一种审美劳动。主播不是简单把文字转换成声音，而是在短时间内完成对内容的理解、筛选和情绪定位，并用自己的声音去传

递这一判断。这也是为什么在突发事件中，观众通常更愿意听到熟悉主播的解说——不仅因为他“专业”，更因为他们相信这个人会“把握好分寸”。

（四）现实压力下的审美隐忧

当然，真人主播也并非没有审美隐忧。在碎片化平台竞争的压力下，很多节目需要频繁更新，因此主播每天就要完成大量的录制和直播任务，他们不可能每一次都有时间能够进行充分的创作准备。每当时间紧张，就会容易催生语流模板和表达套路，久而久之，语言审美也会走向“自动驾驶”。此外，机构为了“品牌统一”，往往对声音风格制定较严格的规范，要求主播“保持某种形象”。这种规范虽然会有利于机构运营，却也会压缩个体风格的生长空间，使部分主播陷入在“统一标准”与“自我表达”之间进退两难的境地。

五、比较分析与人机协同的审美路径

（一）从“谁替代谁”到“各司其职”

如果仅从技术指标比较，AI 合成主播完全可以在清晰度、连贯性上达到甚至超过一般水平的真人播报。但如果从语言审美的整体体验来看，两者呈现的是互补关系。AI 合成主播擅长的是结构化、重复性和低情绪负荷的内容，真人主播擅长的是高复杂度、高情绪密度和需要价值判断的内容。因此，更合理的想象不是“机器最终会把人挤出去”，而是“在同一节目结构中安排不同类型的声音来完成不同任务”。例如，在财经频道，一天内频繁更新的行情播报可以交给 AI 合成主播处理，而涉及政策解读、人物故事和专题观察的栏目则由真人主播承担；在大型晚会或大型新闻直播中，AI 可以用于实时生成多语种字幕和辅助说明，而把最关键的控场和表达交给经验丰富的主持人。

（二）技术参与后的语言审美重构

技术的介入并不意味着大众审美的下降。关键在于，是否允许审美标准在新条件下重新生长。对 AI 合成主播而言，未来的提升方向不应只停留在“更像真人”，而是探索“什么样的声音是适合机器承担的审美角色”。例如，可以为数据可视化内容设计具有“冷静理性”气质的声线，为科普动画配置略带童趣的声音，为某些文化类 IP 打造具有独特节奏感的虚拟主播。在这些场景中，机器声音的“非人性”反而可能会成为一种风格资源。对真人主播而言，技术参与带来的挑战在于如何避免被迫“跟着模板跑”。一方面，主播需要理解 AI 系统的工作逻辑，学会在选题策划和脚本撰写阶段就要与技术同事沟通，把合成声音和真人声音的分工，考虑进节目的结构设计里。另一方面，主播也要坚持对语言质量和审美细节的专业要求，在有限的时间内尽可能守住“每一段关键话语都经得起回听”的底线。

（三）教育与培训的调整方向

在播音主持教育中，面对 AI 合成主播的快速发展，如果我们还仍然只强调传统的发声方法和播音技巧，我们将难以应对新的行业生态。更为重要的是，我们在训练基本功的同时，也要引导学生从一开始就要把“技术理解力”和“审美判断力”视为职业能力的一部分。我们可以在课程里尝试引入虚拟主播制作、语音合成调参等实践内容，让学生亲自体验“当自己的声音被数字化之后会怎样”。同时，也要引导他们反思，哪些表达是只有真人才能完成的？哪些部分是可以交给机器的？这样一来，未来走上岗位的主播才不会把 AI 视为单纯的威胁，而是能够有意识地设计人机分工，主动维护和拓展自己的审美空间。

六、结语

AI 合成主播的出现，使“声音是谁的”这个问题变得前所未有的复杂。先进的技术不

仅带来了新的可能，也带来了审美和伦理上的诸多不确定，我们要通过对现有研究的梳理与比较。例如：AI 声音在清晰度、稳定性和复制能力方面具有天然优势，却无法承载复杂的情绪和人格气质；而真人声音的背后却承载着身体、经验和价值判断。哪怕在高强度劳动和标准化要求的挤压下，他们仍然是公共传播中最具信任度和温度的表达方式。所以我们与其把未来想象成“机器取代人”的单线叙事，不如把它理解为“人机协作下语言审美的再分配”。在这一过程中，技术设计者、媒体机构、教育者和主播本人都需要参与进来，共同回答一个看似简单、实则关键的问题，在充满算法和数字人的媒体环境里，我们希望听到的是怎样的声音？只要这个问题被认真讨论，AI 合成主播和真人主播就都有可能在新的传播格局中找到合理的位置，共同构成一个更加多样和细腻的声音世界。

参考文献：

- [1] 於春. 传播中的离身与具身：人工智能新闻主播的认知交互[J]. 国际新闻界, 2020.
- [2] 周勇. 建构与驯化：人工智能主播的技术路径与演化逻辑[J]. 现代传播, 2022.
- [3] 邵鹏. AI 虚拟主播与主持人具身传播[J]. 中国广播电视学刊, 2021.
- [4] 郑子睿. AI 主播的平台实践、技术演绎与伦理审视[J]. 现代传播, 2023.
- [5] 方兴东. 从 TikTok 到 ChatGPT：智能传播的演进机理与变革路径[J]. 新闻与传播研究, 2023.
- [6] 杨保军. 论人主体新闻与智能体新闻的关系[J]. 新闻战线, 2022.
- [7] 曾志华. 中国有声读物播读评价体系构建研究[J]. 当代传播, 2020.
- [8] 祝捷. 中华有声语言审美范畴溯源——基于一种媒介考古的分析[J]. 中国语文, 2021.
- [9] 王秋硕. 智媒时代播音主持艺术的创作嬗变与价值考究——以“AI 合成主播”为例[J]. 今传媒, 2022.

A Comparative Study on the Language Aesthetics of AI Synthetic Streamers and Real Streamers

LI Binhui

(Jilin University of Arts, Changchun, Jilin 130000, China)

Abstract: With the rapid iteration of "smart media" technology, AI synthetic anchors have widely emerged in scenarios such as news broadcasts, variety shows, and e-commerce live streaming, and they are reshaping the traditional landscape of broadcasting and hosting. Compared with previous studies on technical routes and industrial impacts, the analysis of the differences in "language aesthetics" between AI-generated live-streamers and real live-streamers remains limited. After systematically reviewing relevant literature, this paper introduces theories such as embodied communication, media affordance, and the aesthetics of audible language to conduct a comparative analysis of AI synthetic anchors and real anchors in terms of voice form, language style, emotional expression, and value implications, revealing the advantages of AI synthetic anchors in controllability, efficiency, and standardization. And there are deficiencies in terms of breath rhythm, emotional texture and cultural connotation. In the future practice of news dissemination, a "human-machine collaboration" pattern of anchors may emerge. Ai-synthesized anchors will undertake structured and regular information broadcasting, while real anchors will maintain the high ground of language aesthetics through their presence and individual style.

Keywords: AI synthetic streamer; Real live-streamer; Language aesthetics; Embodied communication; Broadcasting and hosting