

DOI: 10.16538/j.cnki.fem.20240815.301

“声”入人心: 声音的营销效应、机制与自动提取技术

冉雅璇, 董林康, 黄雨婷, 向力子

(中南财经政法大学 工商管理学院, 湖北 武汉 430073)

摘要: “未见其人, 先闻其声”——声音作为一种重要的沟通工具, 在营销中无处不在。相比其他营销沟通工具(如文字、图片、气味等), 声音具有渠道广、影响隐性且深入、营销目的不易察觉等优势, 引起学界和业界的关注。理解营销中声音有何具体作用及其理论机制, 是有效利用声音的关键前提, 是推动声音研究的重要议题。文章系统地介绍了声音研究的发展脉络, 并揭示了声音研究三大主题(包括感官营销、多模态和人工智能)的关键词。声音会使个体产生一系列情绪和认知反应, 促进购买和服从行为, 达成营销效果。神经机制、注意控制理论、预测加工理论和联想心理学四个解释机制是认识声音营销效应的关键视角。文章归纳和对比了声音的分析方法, 梳理了其自动化提取和分析流程。最后, 文章提出了四大未来议题, 包括人工智能声音、中国本土声音、AIGC 多模态声音研究和自动提取技术的开发。

关键词: 声音; 声音营销效应; 感官营销; 自动提取技术; 非结构化数据

中图分类号: F270 **文献标识码:** A **文章编号:** 1001-4950(2024)11-0085-18

一、引言

自古以来, 声音就被认为具有奇妙的作用。例如, 声音能够反映自然规律——“小楼一夜听春雨, 深巷明朝卖杏花”(陆游《临安春雨初霁》); 再如, 声音可以承载思想与情感——“喜见儿童色, 欢传市井声”(王安石《冬至》)。在商业实践中, 声音通过多种渠道进行传播, 能够产生巨大的营销效应。首先, 广告是声音最常见的载体和渠道。无论是在过去的广播广告、电视广告, 还是互联网时代的短视频广告中, 声音都传递着信息和情感。“今年过节不收礼, 收礼就收脑白金”——广告语将语言文案与声音巧妙协调, 以形成声情并茂的效果, 进而更有效地向消费者传达产品信息; 华为的《Dream It Possible》和蜜雪冰城的《蜜雪冰城甜蜜蜜》这类广告曲则将品牌要素融入音乐之中, 为品牌增添了新的记忆点, 与消费者建立更独特的情感连结; 类似于宝马汽车发动机的特殊轰鸣声、可口可乐清脆悦耳的开罐声, 许多企业通过在广告中添加产品音

收稿日期: 2024-03-30

基金项目: 国家自然科学基金面上项目(72272152); 中央高校基本科研业务费项目(2722022BY013)

作者简介: 冉雅璇(1991—), 女, 中南财经政法大学工商管理学院副教授;

董林康(2000—), 男, 中南财经政法大学工商管理学院硕士研究生(通讯作者, 1801690594@qq.com);

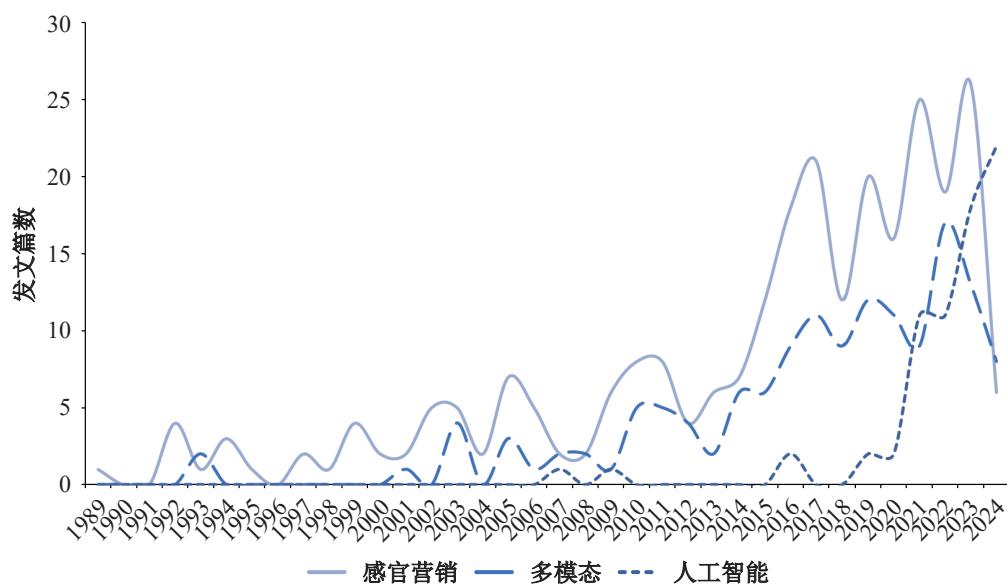
黄雨婷(2000—), 女, 中南财经政法大学工商管理学院硕士研究生;

向力子(2002—), 女, 中南财经政法大学工商管理学院硕士研究生。

效,为消费者提供更加真实沉浸的视听体验,进而强化消费者对产品或品牌的记忆。其次,销售和服务的过程也存在声音的传播。优衣库的线下门店会播放两种音乐,轻快舒适的流行歌曲用于营造氛围,而轻音乐则用来帮助店里的顾客缓解购物疲惫;电商主播的直播间也常常会播放时尚动感、富有节奏的背景音乐,以丰富消费者的体验,激发购物欲望。再者,新技术的发展为声音开辟了新的传播渠道。AI语音助手、服务机器人在工作和生活中得到了广泛应用,消费者甚至可以为自己的语音助手选择喜欢的声音,如动漫角色、明星偶像等,以获得更好的交互体验。语音合成技术的不断升级,推动了虚拟偶像、虚拟代言人、服务机器人等新业态的发展,为传统营销场景注入了新的声音。

随着声音在营销中的广泛运用,其相关效应也越来越受到学界关注。本文通过Web of Science(WOS)核心数据库和中国知网,于2024年7月16日对1989—2024年35年来有关“声音的营销效应”的文献进行检索和筛选,具体步骤如下:(1)英文文献使用“AB=(voice OR auditory OR audition OR sound OR music OR audio OR vocal) AND AB=(marketing OR consumer OR customer)”的关键词组合进行检索,中文文献使用“(摘要:声音+音频+音乐+人声+语音+噪音)(精确)AND(摘要:营销+消费者)(精确)”的关键词组合进行检索;(2)保留英文和中文学术期刊,且期刊为SSCI或CSSCI期刊;(3)主题领域为商业、管理、信息管理、心理学等领域;(4)通过文章摘要对无关文献进行剔除,例如消费者的意见(voice)等。最终获得476篇文献用于回顾和分析(如图1),结果发现:在文献数量上,关于声音营销效应的研究总体呈现逐年上升的趋势;在研究主题的演变上,早期主要聚焦于声音要素本身及其在感官营销中的作用,随后逐渐扩展到同时探讨声音和其他感官交互的多模态研究,而近来则进一步关注声音与人工智能技术的结合与应用。整体来看,这些研究不仅验证了声音在营销中的重要作用,同时也揭示了声音相比于其他营销沟通工具(如文字、图片等)的独特优势:(1)与气味、味道、触感的近距离感官营销方式相比,声音依赖于远距离的听觉感官,既可以线下使用,也可以线上使用,具有广泛的传播渠道(Zhou等,2024)。(2)尤其是,声音通常作为一种辅助营销工具(如背景音),其影响更为隐性,即消费者会在无形中受到声音的影响(钟科等,2016;Lowe等,2019)。(3)鉴于声音的隐性影响特征,消费者也更难以察觉声音的说服目的性,消费者的反应通常是直觉的、生理的(Moreno-Lobato等,2023)。(4)研究指出,在人脑的感官记忆功能上,声像记忆(echoic memory)在脑海中保留的时间比图像记忆(iconic memory)要长,因此利用声音来营销能产生更为持续的影响,特别是在形成短期记忆后,大脑还会反复回顾该声音,即产生“耳虫效应”(Baddeley等,2009)。

综上,声音在营销中已经得到长期的、广泛的运用,并发挥着独特的作用,但现有研究中仍存在诸多局限和挑战。其一,尽管越来越多的研究对声音的营销效应进行了探索,但相关研究较为分散,不同学者关注的声音特征和作用类型不同(Spence和Keller,2024),因此需要系统地理解声音的营销效应。其二,声音研究的相关理论视角涉及生物学、神经科学、物理学、心理学等多个领域(Zwicker和Fastl,1990),因此了解声音效应的内在机制有助于我们更好地把握声音在营销中所发挥的作用。特别是,随着计算神经科学的兴起,对声音效应机制的认识有助于优化声音的自动化处理和分析(刘扬等,2013),进而推动声音大规模分析的发展。其三,尽管声音在营销实践中应用广泛,但相较于文本、图片等其他非结构化数据研究的蓬勃发展,声音研究仍处于起步阶段,原因主要在于声音分析技术上的困难(Wang等,2024a);在文献回顾过程中我们发现,现今营销研究大多使用简单的实验控制方法来探索声音营销效应的因果关系(Bigand等,2005;Chattopadhyay等,2003;Moreno-Lobato等,2023),仅有少数研究对海量声音



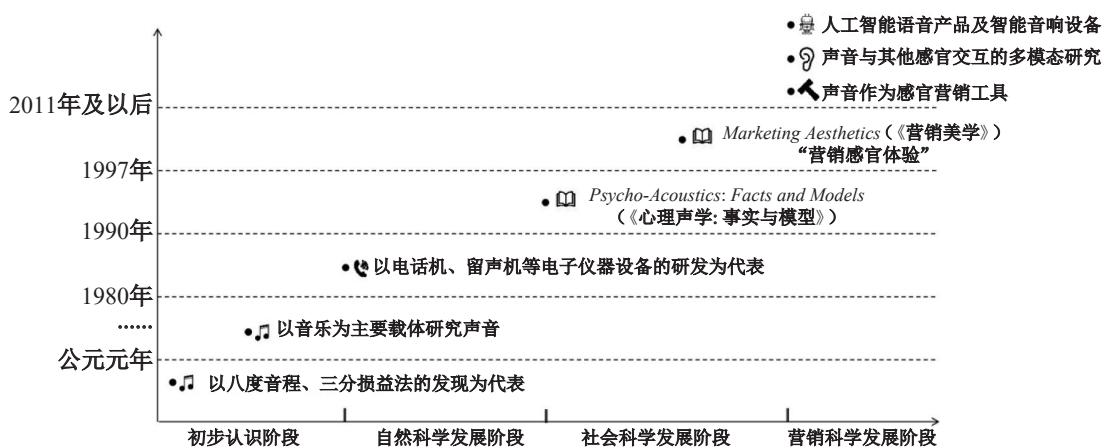
资料来源：本文作者绘制。

图 1 声音的营销效应研究主题发文情况

数据进行了分析与提取(Wang等, 2021),这与当下的大数据现实不匹配。因此,本研究在重新梳理现有研究的基础上,结合ABC态度模型从情绪、认知和行为三个方面来归纳声音的营销效应,剖析其机制的解释视角,并系统总结声音数据的提取技术和步骤,最后提出可能的未来研究方向。

二、声音营销效应的研究脉络

声音的研究主要经历了四个阶段(如图2所示)。第一,初步认识阶段。这一阶段开始于公元前500年左右,毕达哥拉斯发现了人们听起来最和谐的八度音程,战国时期的《吕氏春秋》也记载了可以让乐声听起来和谐的三分损益法,这些研究成果都帮助了人们更好地认识声音并运用声音。



资料来源：本文作者绘制。

图 2 声音研究发展阶段图

第二,自然科学发展阶段。19到20世纪,贝尔、爱迪生先后发明了电话机、留声机,声音研究进入快速发展阶段。随着科技的发展进步,电声换能器和电子仪器设备被运用到声学研究中,这些仪器进一步降低了研究声音物理特性的难度,促进了声音研究在自然科学领域的兴起,催生出建筑声学、电声学和超声学等分支学科,现代声学体系逐步形成。

第三,社会科学发展阶段。随着自然科学领域对声音物理特性研究的不断深入,学者们的视线逐步转移到不同物理特性声音的社会作用。1990年,*Psycho-Acoustics: Facts and Models*(《心理声学:事实与模型》)一书诞生,标志着声学进入社会科学领域,该书描述了人类听觉系统处理声音的心理过程,引发后续研究对声音的社会意义和功能的关注([Zwicker和Fastl, 1990](#))。这一时期的研究发现,声音可以反映个人特质和情感。例如,说话声音更大的人被视为是更自信的([Kimble和Seidel, 1991](#)),说话低沉的人则被认为是更值得信赖的、有能力的([Oleszkiewicz等, 2017](#))。

第四,营销科学发展阶段。1997年,伯德•施密特在*Marketing Aesthetics*(《营销美学》)一书中,提出了“感官营销体验”的说法,强调通过听觉、视觉、触觉等感官体验可以影响消费者对品牌的态度。声音作为人类感官体验的重要来源之一,受到了营销研究的重视。我们首先梳理了营销领域聚焦于声音研究的418篇英文文献(具体流程见引言),并对这些文献中的1884个关键词进行人工编码,发现营销领域的声音研究主要经历了三个主题:感官营销的声音研究(启蒙阶段)、多模态的声音研究(发展阶段)以及和人工智能相关的声音研究(智能化阶段)。其中,与感官营销相关的关键词共966个,与多模态相关的关键词共593个,与人工智能相关的关键词共340个。随后,再按照同样的步骤对58篇中文文献进行梳理,并对179个关键词进行人工编码,发现了与英文文献相似的三阶段发展趋势。详细的统计结果见表1^①。

表1 营销领域声音研究文献关注的主题及关键词

阶段	主题	关键词
启蒙阶段	感官营销	music(76), brand(57), sensor marketing(25), advertising/advertisement(22), consumer/customer behavior(17), background music(15), marketing(11), atmosphere(10), sound symbolism(9), arousal(6)
发展阶段	多模态	music(19), visual/eye-tracking(19), consumer/customer behavior(18), multisensory(16), sensory marketing(11), advertising(9), taste(5)
智能化阶段	人工智能	artificial intelligence(28), voice assistant(17), consumer/customer behavior(13), human likeness(8), smart speakers(6), trust(6), social presence(6)

在启蒙阶段,相关研究聚焦于感官营销中的声音。[Bhatia等\(2021\)](#)回顾了1979—2020年间与感官营销有关的文献,发现由于消费者变得更加重视体验和感官信号的整体接收,与感官营销有关的研究出现了井喷式增长。其中,得益于机器学习方法在声音分析中的运用,研究人员可以从营销场景中丰富的环境音、人声和音乐等信息中提取更多有用的信息进行分析。研究表明,声音通过脑区加工显著影响消费者的认知和情感,进而影响判断、决策和评价等行为。譬如,超市播放节奏欢快的歌曲,会让消费者感觉时间过得更快,从而加速购物进度,帮助商超缓解节假日的人流压力;反之,如果超市播放慢节奏的音乐,则会让消费者停留更长时间,花费更多([Ahlbom等, 2023](#))。

在发展阶段,营销场景中的声音作为多模态问题研究的一部分得到大量讨论,其中研究最多的是“视—听”交互、“视—味”交互和“听—味”交互([张全成等, 2017](#))。例如,[Klein等\(2021\)](#)让294名平均年龄为23岁的被试倾听不同复杂程度的音乐,并利用量表衡量被试对看到的不同图

^①经统计,中文文献关键词无明显集中趋势,因此本表格仅展现了英文关键词。

片的喜爱程度,结果表明音乐改变了视觉复杂性与图像喜好之间的倒“U”形关系,即简单的音乐刺激可能会降低对复杂视觉图像的喜爱程度。de Paulo等(2022)考察了声音对味觉的影响,他们召集消费者在相同的场景中听不同类型的音乐,并就自己对不同啤酒产品的感受进行评分。他们发现在播放流行音乐时,消费者认为啤酒口感更轻盈,而当播放软摇滚时,消费者认为啤酒的口感更苦涩。

在智能化阶段,得益于人工智能技术的发展,以Alexa、Siri、小爱同学等为代表的智能语音产品得到了广泛的运用。虽然现有研究关注到了以声音为主的人工智能语音助手的作用(Ijiga等,2024),以及声音在人机互动中的核心角色(Hasan等,2021),但是人工智能的声音特征及其作用机制还有待进一步探讨。

三、声音的营销效应

(一)情绪反应:声音影响情绪唤起和情绪效价

作为一个重要的环境因素,声音不仅会影响情绪的唤起水平,还会影响情绪的效价。其中,情绪唤起用于衡量感知情绪从冷静到兴奋的程度,很容易受到声音的调动,且仅需要250毫秒的音乐就能唤起情绪反应(Loui等,2013)。Bigand等(2005)让受过音乐训练和未受过音乐训练的听众一起聆听同样的音乐选段,并按照“感动”程度进行分组,受过音乐训练和未受过音乐训练的听众给出了相似的回答,证实了声音对情绪唤起的普遍性。另外,能唤起情绪反应的声音具有相似的声学特征。Hofbauer和Rodriguez(2023)通过回顾笑声中的情感线索相关研究,总结出能够引起高水平情绪唤起的声音通常具有更短的持续时间、更短的停顿、更快的语速、更强烈的情感、更剧烈的变化、更高的基频变化、更清晰的共性。

声音还会影响情绪效价,即影响情绪的消极或积极程度。其中,最为突出的影响来自音乐调式和节奏。Morreale等(2013)操纵了28段小提琴音乐选段,要求被试以7分制对情绪效价进行评分,结果证实,在节奏和调式表达的情绪不一致的情况下,节奏中的情绪线索更容易影响未受过专业音乐训练的普通人的情绪效价。一般来说,大调、快节奏音乐唤起的是更为积极的情绪,而小调、慢节奏音乐则相反。Biswas等(2019)在咖啡馆播放不同音量的音乐,通过一周的观察和调研发现,相对低音量(55分贝)的音乐比高音量(75分贝)的音乐更能放松消费者的紧张情绪,增加对健康产品的购买。声音通常会同时影响情绪唤起和情绪效价。例如,Lai和Chiang(2012)通过线上实验得出相似的结论:在消费者进行网络购物的过程中,浏览一段时间后出现的背景音乐会诱发更高的愉悦感、唤醒度和更积极的行为意向。Roschk等(2017)对1982—2016年间的66篇与音乐对店铺氛围塑造相关的文献进行的元分析,也得出了类似的观点。

(二)认知作用:声音改变信息加工和信息推断

声音可以通过语音信息分析通路、情感信息分析通路和声音信息分析通路在大脑中被解码处理,使得人们产生对营销信息、具体情感和说话人及说话内容特质的认知(Belin等,2011)。

在语音信息分析通路中,声音特征会影响听者理解和记忆信息的过程。比如广告旁白的语速过慢或过快,都会导致观众在解码过程中遇到障碍,不能完整处理广告信息,从而降低广告效果(Oleszkiewicz等,2017;Anikin,2020)。Chattopadhyay等(2003)的实验证明,只有在旁白语速适中时(4.6音节/秒),广告才能吸引更多的注意力,增加观看者对广告信息的记忆,进而提升品牌和产品知名度,积极影响购买意愿。除此之外,广告中声音变化的增加也能够吸引注意力,加深听众对产品的印象(Chang等,2023)。有趣的是,Huang和Labroo(2020)进一步发现,较

高音调的声音会让消费者有“向上看”的倾向,当屏幕上方显示具有道德含义的信息时,人们也能更快地识别出来。在此逻辑的基础上,他们揭示了高音调的声音能够唤醒消费者更高的道德感知,从而推动消费者选择低卡路里的健康食物。

声音特征还会影响人们对声音中具体情感的认知和识别,比如人们会认为较大的音调变化常常与惊喜、幸福、愉快等情绪联系在一起(Scherer和Ekman,2014)。因此,营销人员能够操纵不同音调的音乐,调动消费者的记忆,使得他们产生相对积极的情绪,从而影响产品偏好和购买行为(Moreno-Lobato等,2023)。

音量、音调、语速等声音特征还会影响听者对说话人身份特征的推测。通常情况下,人们会认为说话音量更大和音调更低的导购员是更加真诚的,因此更愿意接受他们推荐的产品(Kimble和Seidel,1991)。音调特征还会影响听者对说话人的能力认知。Wang等(2021)使用自动语音挖掘技术,分析了众筹网站Kickstarter的视频声音,发现音调中表现出的专注、低压、情绪平稳的特质能够让听众推测说话者是有能力的,进而提高众筹产品的说服效果。

(三)行为反应:声音助推消费者行为

声音的不同特征能够传递许多语义内容之外的信息,改变消费者的情感和认知,提升说服效果,最终驱动行为反应的产生。

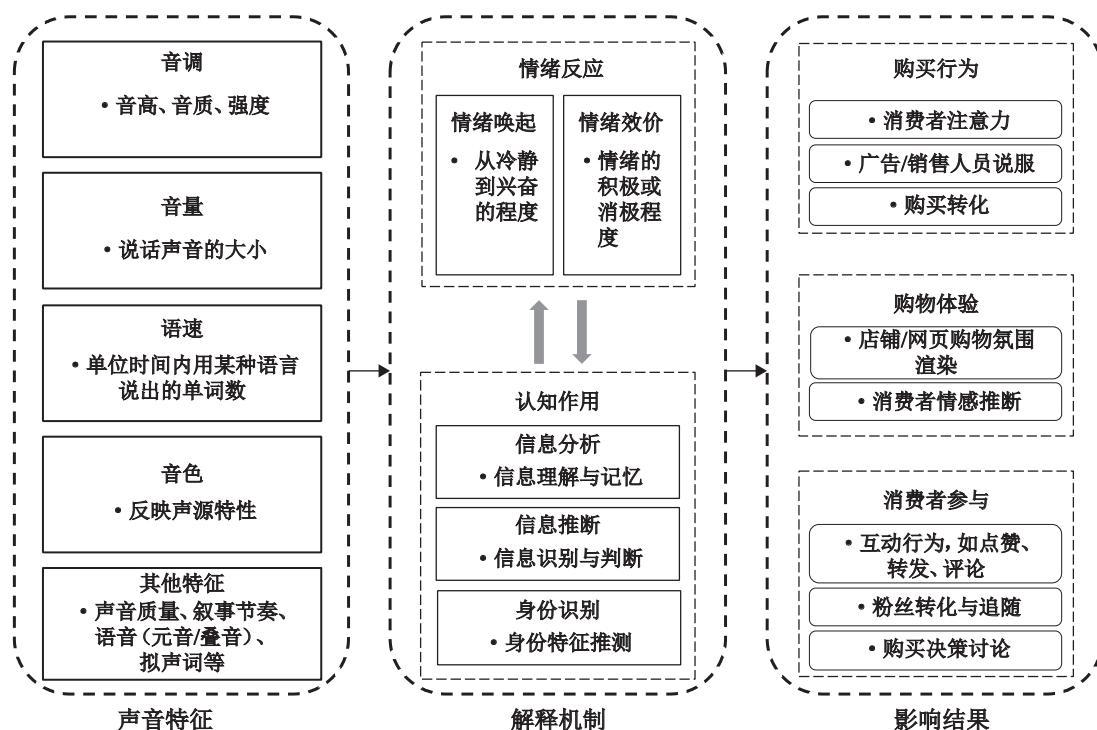
首先,在沟通过程中,营销人员可以通过改变声音的特征,来增强说服,促进购买行为的转化。第一,音高会影响听者对说话人的信任。Oleszkiewicz等(2017)通过实验证实了低音高与感知信任的正向关系,即有较低音高的说话人更容易获得信任。第二,语速也会对声音的说服力度产生直接影响。如前文所述,语速过快或过慢,都会对听者对信息的理解和记忆产生负面影响,不利于说服。Rodero(2017)通过实验分析了不同语速对生理唤起、情绪效价、注意力吸引以及听者对声音信息的识别和回忆的影响,发现每分钟180词、中等语速的广告效果最好,说服力度最强。而Wang等(2024b)从中国一个新兴的数字口译平台上收集数据分析发现,语速正向影响消费者对口译服务的购买,而音量和音高负向影响购买。第三,声音质量会影响听者对说话人的喜爱程度。声音质量是对一个人声音的整体衡量,结合了音调、响度和音色等声音特征,人们可以通过改变声音质量来传递某种特定情绪或者含义,而对于声音质量更高的说话者,人们会更倾向于对其介绍的产品或广告有积极态度,进而产生更强的购买意愿和行为(Rodero,2020)。第四,人类声音的说服效果显著高于人工智能声音。Rodero(2017)的实验表明,同样的语义内容被人工声音和人类声音分别传递出来时,人类声音被视为是更清晰、更自然、更愉快、更有活力、更有说服力的,且更有利于信息传递的有效性以及听者的注意和回忆。但是,学者们也关注到,人工智能生成的声音具有更多的创造性和艺术性,可以创作出自然界没有的声音(Ijiga等,2024)。

其次,购物场景中的音乐声音可以对购物体验和行为产生影响。总的来说,音调较高的音乐作为一种环境线索,可以营造更加积极的购物氛围,提升消费者对店铺的整体评价,进而延长购物时间,增加开销金额(Ahlbom等,2023)。此外,音乐还可以提升消费者对产品的态度,这个效应在工作日更容易被观察到,因为人们在工作日会感到更加疲惫,因此更容易被音乐调动情绪,进而增加购买(Ahlbom等,2023)。总的来说,增加音乐且使用合适的音乐,可以将购物转化为独特和愉快的体验,能够在产品销售、顾客满意度和顾客忠诚度方面给营销人员带来积极的结果。

最后,声音的合理运用还可以促进消费者参与。例如,Fu等(2024)基于Bilibili网站的视频数据,通过自动语音识别分析和文本分析方法发现,当视频的音量变化较大时,观众会认为该视频传递的情感更丰富,会更多地参与视频内容并点赞,甚至加关注成为视频创作者的粉丝。

同时,在类似的多模态感官背景下,声音可以通过提供与文字、画面互补的信息,吸引注意力、调动积极情绪来促进互动。再如,Al-Qershi等(2022)利用微软视频索引器对众筹网站上的视频进行提取分析,发现当语音内容与画面文字相符合时,观众更愿意参与到众筹项目中去,这对众筹项目的成功有较大的帮助。除此之外,音乐还可以作为“社交润滑剂”,在推动合作等亲社会行为方面发挥突出的作用。Lang等(2016)先将被试暴露在有节奏音乐、无节奏音乐和白噪音三种情况下,随后让他们参与五项需要人际协调的实验,最终发现有节奏的音乐会激活“人际联结”,使人们更愿意和同伴、朋友做出相同的决策。

综上,我们提出了一个有关声音营销效应的整合研究框架,如图3所示。具体而言,个体接收到的声音,会首先通过人体的听觉系统,形成不同特征的声音信号(如音调、音量等),进而使个体产生情绪反应,同时形成听觉理解,最终影响个体的购买行为、购物体验和参与行为等。



资料来源:本文作者绘制。

图3 声音的营销效应作用路径图

四、声音营销效应的解释机制

(一)神经机制视角

声音的营销效应源于人体的听觉系统,并在大脑各个区域的协同作用下得以产生。具体而言,听觉系统分为外周系统与中枢系统。外周系统由外耳、中耳、内耳和听觉神经组成,主要负责接收声音并进行预处理,随后将声音信号传递给大脑。中枢系统则专注于声音的神经处理,声音信号在传导过程中会经过多个听觉中枢的初步加工,最终在大脑皮层中由最高级听觉中枢进行精细加工并形成听觉(Plack, 2008; Zwicker和Fastl, 1990)。而大脑的不同区域承担着不同的功能,它们会在接收到声音信号刺激后做出相应的反应。基于此,本研究从声音的接收、处理和反应三个环节来揭示声音营销效应的神经机制。

第一,声音的接收发生于听觉外周系统。首先外耳负责收集声音,随后中耳进行声音的传

导,最终内耳耳蜗的听觉感受器将声音从机械振动信号转化为神经信号,从而完成对声音的接收。紧接着,这些声音信号将沿着听觉神经继续传导至各级听觉中枢进行处理(Plack,2008)。

第二,声音的处理可以细分为预处理和神经处理两个阶段。预处理在外周系统中进行,主要涉及对声音的频率、强度等物理特性的初步识别;而神经处理则由中枢系统主导,这一过程始于较低级的听觉中枢对声音信号的初步加工,随后大脑听觉皮层进行更为高级的处理,最终形成完整的听觉感知(Zwicker和Fastl,1990)。以音乐感知为例,声音振动信号在耳蜗处即开始被转化为神经信号,其频率得到初步识别,随后经过多级中枢的精细加工,大脑听觉皮层完成音高编码,并进一步对连续时间内的多个音高进行整合处理,从而构建出对和弦乃至旋律的完整感知。尽管不同类型的声音在处理过程上存在差异,但总体上均遵循从简单到复杂、从低维到高维的处理过程(侯建成和刘昌,2008)。

第三,对声音的反应主要依赖于大脑各个区域的协同作用。具体而言,大脑皮层作为高级神经中枢,主导着人类的复杂情绪功能与高级认知功能,并在其他脑结构的参与下实现对声音的反应(Vuust等,2022)。例如,大脑中的杏仁核和伏隔核等区域掌管着大脑的奖赏、情绪、成瘾、快乐、恐惧等功能,特定的声音会促使伏隔核(NAC)和腹侧被盖区(VTA)释放多巴胺,进而引发情绪反应(Menon和Levitin,2005);再如,在形成对声音的理解后,颞叶、额叶等多个脑区会唤起相关的联想与记忆,进而影响认知乃至后续的行为反应(Juslin和Västfjäll,2008)。

(二)注意控制理论视角

注意控制理论揭示了人类在处理感官刺激时的两类调控模式。一种为自下而上的调控(bottom-up),又称为外源性注意控制或“刺激—驱动”模型,反映了具有物理凸显性的刺激对注意选择的影响;另一种为自上而下的调控(top-down),又称为内源性注意控制或“目标—驱动”模型,反映的是个体基于目标和意图进行注意选择的控制方式(Failing和Theeuwes,2018)。

注意调控模式主要影响声音的处理过程,虽然声音信号进入大脑后能形成听觉,但只有那些被注意到的声音才会进一步得到处理和引发反应(Baddeley等,2009)。自下而上的“刺激—驱动”模型是注意控制的一种更为基础的方式。当声音在听觉皮层被编码后,具有听觉凸显性的声音特质更容易吸引注意力,从而促使大脑投入更多的认知资源对其进行深入加工(Kaya和Elhilali,2017)。例如,视频中突然变化的声音能够迅速吸引消费者的注意力,进而增强信息的说服效果(Chang等,2023)。相比之下,自上而下的“目标—驱动”模型则是由大脑主动参与的另一种注意控制方式,它强调听觉的选择性。大脑带着特定的目标去听声音,会增加焦点目标与背景之间的对比度,从而聚焦并优先处理特定的声音,分配更多注意力给这些声音特征。正是通过这种方式,个体能够在嘈杂的环境中准确识别并听到自己关心的声音,即“鸡尾酒会效应”(Kaya和Elhilali,2017)。

注意调控模式不仅会影响声音处理的焦点,还会影响声音处理的效率。在“自下而上”的调控模式下,个体从更低级、更具体的层次开始处理声音信息,逐步向更高级、更整体或更抽象的层次进行信息编码和传递;而在“自上而下”的调控模式下,由于大脑其他区域的积极参与,声音信息的编码和传递过程会更为迅速,从而能够更高效地形成听觉理解和认知(Pardi等,2020)。

总的来说,注意控制理论为营销中声音的注意捕捉与认知加工机制提供了深刻的见解。一方面,该理论揭示了何种类型的声音特征更能有效吸引并维持消费者的注意力,这是很多广告声音研究的理论基础(Chattopadhyay等,2003);另一方面,该理论也构成了认知负荷理论的重要基石,在此框架下,当声音更能吸引消费者注意力时,消费者会主动调配更多的认知资源来

优化声音信息的处理过程,进而引发更积极的反应(Chang等,2023)。

(三)预测加工理论视角

预测加工理论侧重于解释大脑知觉系统对感官刺激的推断。该理论的基本观点是,大脑本质上是一个预测机器,它会基于先前的经验,在当前的感官刺激下生成对后续感官刺激的预测(Clark,2013)。大脑会将生成的这些预测与随后接收到的感官刺激进行匹配和比较,以检验预测的准确性。当预测出现误差时,大脑会努力调整以减少这种误差(Friston,2010)。

大脑在处理声音时也会进行预测加工,当预测准确时,预测结果有助于加快声音的编码过程、形成听觉理解,这也是人类能快速理解声音的关键所在(Wei等,2023)。同样,预测结果也会影响个体对声音的反应,例如,音乐心理学学者在针对音乐的预测编码模型(predictive coding of music)中指出:当听到的音乐超出预测或者符合预测时,个体就会产生积极的情绪;反之,当听到的音乐不如预测时,则会产生消极的情绪。此外,个体还会通过调整认知和行为的方式来使预测误差最小化,例如改变对音乐的预测,或随着音乐打节拍(Vuust等,2022)。总的来说,预测加工是个体对声音产生情绪、认知和行为反应的重要机制。

在声音的营销效应研究中,预测加工理论主要体现在三个方面:第一,该理论被视为声音诱发情绪反应的核心机制之一(Juslin和Västfjäll,2008);第二,该理论是众多跨模态一致性研究的重要理论基础,例如,广告视听一致性、声音—形象一致性会引发消费者愉悦体验的一个重要原因就在于,其他感官刺激也参与到对后续声音的预测中,并通过预测结果和后续声音的比较来引发反应(张全成等,2017;Spence和Keller,2024);第三,该理论还揭示了声音处理和反应上个体差异的根源,即不同的个人经历会导致个体在声音预测能力和风格上的差异(如,音乐家的预测比普通人的预测要好),进而影响声音处理和行为反应(Vuust等,2022)。

(四)联想心理学视角

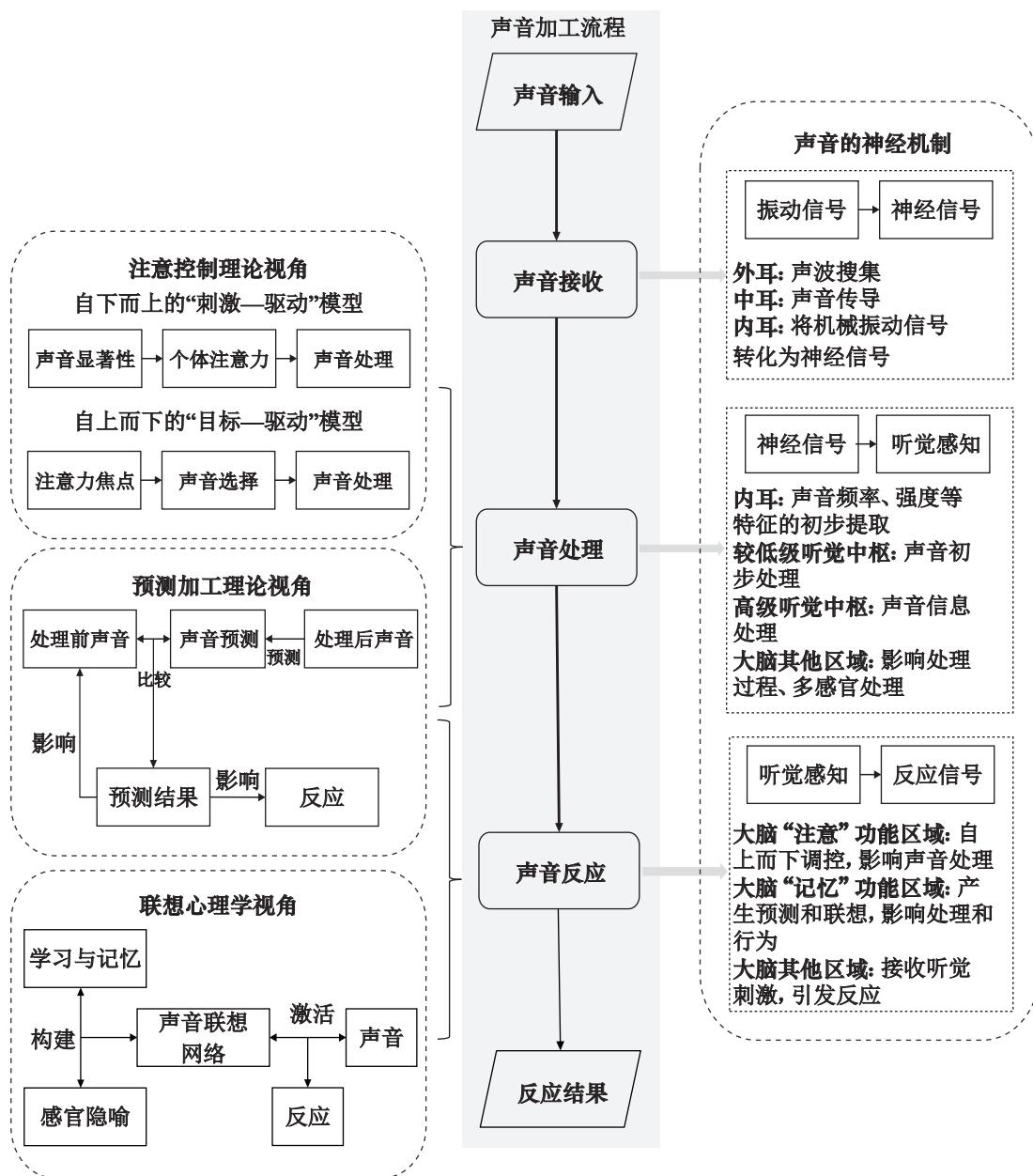
联想心理学的核心观点认为,一切复杂的心理现象都是通过联想而产生的。其中,联想网络记忆模型(associative network memory model)揭示了联想和记忆的关系,指出记忆是由一系列的节点和联结所构成的网络。节点是一个个单独的概念,而联结反映了概念之间联系的强度。生成新的节点就是“记忆”,由一个节点而联系到另一个节点则是“联想”(Hopfield,1982)。而感官刺激通过“联想”引发反应的过程,存在联想网络的建立和激发两个阶段。

声音联想网络的建立一般有两个来源:一方面,当某种声音与某些事物频繁共现时,人们会建立相应的联想。例如,消费者听到钢琴的声音,会联想起音乐厅、黑领带西服等钢琴相关元素。另一方面,个体可能会将声音和另一种感官模态进行隐喻性联想,如个体常常将明亮的颜色与响亮的声音而不是柔和的声音联系起来(Puligadda和Vanbergen,2023)。当声音联想网络建立后,声音的再次输入会激活基于联想网络的对应节点。此时,个体会将声音与特定的人、物和情景联系起来,进而产生一系列情绪与记忆,并进一步形成一定的认知推断和行为反应(Juslin和Västfjäll,2008)。

特质推断理论、图式理论以及感官隐喻都是在联想心理学视角下形成的。当声音与某个具体特征建立联系时,消费者能够利用这些联系通过声音来推断声源的特质(Kimble和Seidel,1991)。在品牌图式的构建中,融入声音元素能够形成独特的品牌声音联想网络,从而深化品牌在消费者心中的印象,这种效果往往会超越单纯的视觉元素所起的作用(Puligadda和Vanbergen,2023)。此外,感官之间的相互作用,比如自然界中大发声体通常发出比小发声体更低沉的声音,这一现象促使人们将声音与物体的大小相匹配,形成听觉对视觉的间接影响(张全成等,2017),其本质也是联想在发挥作用。

(五)四种解释机制的整合

如图4所示,神经机制、注意控制理论、预测加工理论和联想心理学四个视角共同揭示了声音营销效应的基本过程。其中,神经机制揭示了个体对声音的接收、处理和反应的完整生理过程,注意控制理论展现了处理声音的两类不同机制,预测加工理论揭示了声音感知偏误的自我调控机制,涉及声音处理和反应过程,联想心理学则解释了个体对声音产生反应的重要方式。四个视角各有侧重,能够对声音营销效应研究的众多机制和理论进行解释,它们相互之间也存在紧密的联系。



资料来源:本文作者绘制。

图 4 声音营销效应的解释机制

一方面,注意控制的两类方式贯穿声音处理的全过程,而预测加工主要在自上而下的调控模式下起作用。连接耳朵和大脑听觉皮层的听觉神经有两种,一种为传入神经,一种为传出神经。声音被耳朵接收,转化为神经信号后,沿传入神经传导到大脑皮层并接受处理,这是自下而上的注意控制方式;而自上而下的注意控制方式,则是大脑发出指令,选择性地处理特定声音,此时指令信号会沿着传出神经传达到各个听觉中枢,影响其加工处理声音的过程(Plack, 2008; Zwicker和Fastl, 1990)。之所以在“自上而下”的调控模式下,声音处理的效率更高,其重要原因在于人脑的预测加工在发挥作用。对于后续声音输入的预测结果也会随传出神经进行传递,且当预测准确时,声音处理区域会被激活,其处理速度会提高(Wei等,2023)。

另一方面,预测加工和联想在神经机制上有相似之处。预测的产生不仅基于当前的声音信息,还会结合个体的经验、学习和记忆(Vuust等,2022)。同样,声音引发的联想也是激活的记忆中的声音联想网络。可见,预测和联想的产生都需要大脑的记忆功能发挥作用。而两者的不同之处在于,预测是基于当前的声音对未来将要听到的(还未听到的)声音进行“想象”,而联想更多是将当前的声音与记忆中的(已经历过的)特定对象联系起来。

五、声音分析方法与提取技术

声音分析方法与提取技术经历了一个快速发展的过程,主要可以分为以下三个阶段(如表2):(1)简单定量阶段。早期研究者主要使用计算器、秒表和分数表等客观技术(Black, 1961),可以直接测量音调、响度、语速等基础声音特征。(2)定性编码阶段。后来,有研究者采用人工编码的定性方法探索高阶声音特征,如徐震南等(2023)通过对抖音视频是否使用方言进行手动数据收集与特征标记,来分析方言使用对视频点赞、评论数的影响。其优势在于能够识别声音质量、风格等高阶特征,但人力工作量大,可处理的数据量级相对较低。(3)算法模型阶段。近年来,随着机器学习和自然语言处理方法的兴起,许多计算机学者开始探索以算法模型为基础的声音分析技术,使声音的提取和分析技术得到了显著的改善,如Fu等(2024)使用MoviePy包从Bilibili视频中提取音频。除了软件与现成编程包外,一系列更高阶的算法(如深度学习)也被运用到声音分析领域(徐婕和肖莉,2022),其中以CNN为基础的神经网络算法在声音分析领域发展尤其迅猛、应用范围广。综上,我们根据声音分析方法与提取技术的发展阶段以及方法实现的难易程度,结合以往文献,总结了与声音特征提取相关的主要分析方法及其优劣势以及在营销中的应用场景(见表2)。

声音分析与提取主要分为两个阶段,共五个步骤(见图5)。

第一个阶段是声音表征,包括数据分类、数据预处理、方法选择三个步骤,这一阶段可以输出声音的基础特征。

(1)数据分类。非结构化数据一般分为图片、文本、声音三类,声音数据可以直接获取(如使用录音机和麦克风记录声音)以及间接获取(如从视频数据中提取音频)。明确数据类别后需要对其进行规范的存储,研究者可以选用丰富多样的音频格式^①进行数据预处理,如无压缩的WAVE(.wav)音频格式^②、AIFF(.aif)音频格式^③。以分析一段视频为例(见图5),视频数据中有许多图片帧构成的图片数据集、音频数据,以及包括标题、互动信息和语音内容在内的文本数据。数据分类需要用软件如FFmpeg、VLC媒体播放器等从视频中提取出音频,并转换成与声音

^①音频格式是指要在计算机内播放或处理音频文件,对声音文件进行数模转换的过程。音频格式最大带宽是20 000Hz,速率介于40~50kHz之间,采用线性脉冲编码调制PCM,每一量化步长都具有相等的长度。

^②WAVE格式采用PCM(pulse-code modulation,脉冲编码调制)对音频进行编码,而PCM遵循RIFF(resource information file format,资源信息文件格式)规格,允许把额外的用户信息(如时间码定位信息)嵌入并保存在其中。

^③为苹果电脑系统设计。

表2 声音特征及其分析方法

发展阶段	分析方法	声音特征	优势	劣势	应用场景	代表文献
简单定量阶段	计算器、秒表	特定声音的简单计数、时长等基础特征	简单	人力工作量大	线下购物行为分析	Black, 1961
定性编码阶段	人工编码	特定声音的简单计数、时长,以及复杂的音色、风格、情感特征	简单,可解释力强	人力工作量大、数据量级有限	方言相关研究、声音事件识别	徐震南等, 2023
算法模型阶段(软件与现成编程包)	QA5软件	响度、音调、语速等基础特征	操作简单、易上手	数据量级受限	用户行为二手数据分析	Wang等, 2021
	Praat软件					Shang和Liu, 2022
	Librosa库	梅尔频率倒谱系数、谱质心、过零率、声谱衰减、色度频率	Python工具易上手、可视化	需要预先做好识别与清洗,并结合其他工具使用	用户行为可视化	刘德文等, 2022
算法模型阶段(以CNN为基础的神经网络架构)	CNN	梅尔频率倒谱系数及其他自定义特征	自动提取特征	声音神经网络方法起源于计算机视觉,对音频信号的时频特性处理能力有限	智能客服	Jia, 2022
	卷积循环神经网络(CRNN)	响度、音调、语速等基础特征,以及激动、伤心、愤怒等情感特征	自动提取特征		顾客通话监测、声音情感分析	Choi等, 2017
	神经网络混合架构(如CNN和Bi-GRU)	旋律特征、风格特征(蓝调、古典、乡村、迪斯科、嘻哈、爵士、金属、流行、雷鬼和摇滚)	识别能力强	复杂	声音事件识别、声音流派分类	Ashraf等, 2023
	有监督模型(混合卷积-LSTM模型)	旋律特征	处理和记忆时间序列数据、具有较好的泛化能力	参数较多	声音情感分析、风格分类	de Benito-Gorron等, 2019

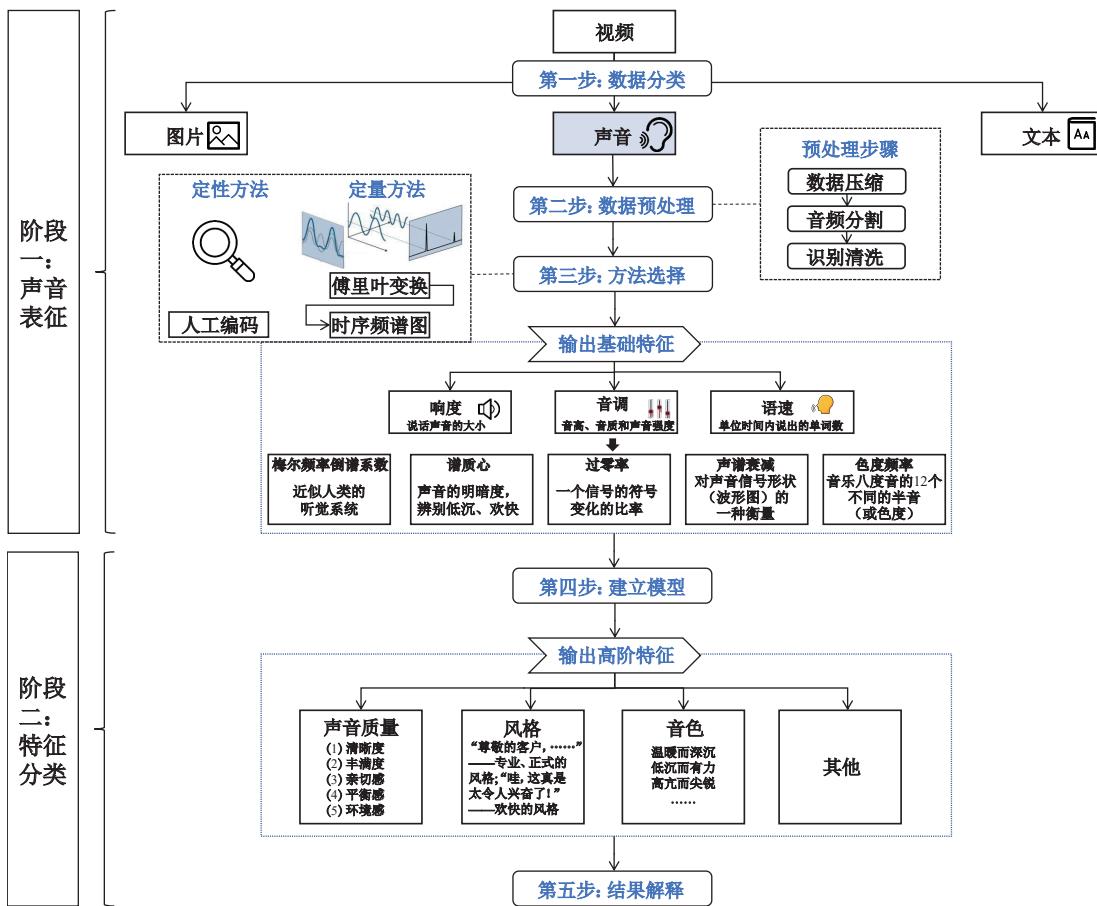
资料来源:本文作者总结。

分析软件适配的音频格式。

(2)数据预处理。分割(segment)是音频数据分析中的一个重要的预处理步骤。主要借鉴图片数据预处理方式,将音频数据转换为特征向量、频谱图等量化形式,再通过特征学习方法识别音频内容,将一个嘈杂而冗长的音频信号分割成短小、均匀的段落。目前有许多成熟的预处理方法可以使用,例如,Venkatesh等(2022)通过模仿图像视觉领域广泛采用的“你只看一次”(YOLO)目标检测算法,提出了一种名为“你只听一次”(YOHO)的音频分割方法。在第一步中得到的音频文件需要经过严谨的预处理,比如若想继续从这段视频(见图5)中提取出某几段旋律,在这里可以先用剪辑软件手动编码,将符合研究目的的音频分割出来并清除杂音(如环境噪音、BGM)。

(3)方法选择。研究者应根据研究目的,选择恰当的表征方法。研究者可以直接使用已经训练好模型的声音分析软件提取特征。例如,Shang和Liu(2022)使用Praat软件(v.5.3.85)量化声音的基础特征,结合实验法对比不同性别的声音吸引力差异。除了使用软件方法进行提取,算法模型同样被广泛应用。Jia(2022)使用机器学习方法提取梅尔频率倒谱系数(MFCC)^①作为声音特征值,利用卷积循环神经网络(CRNN)和双向长短期记忆(Bi-LSTM)网络分别提取声音序列特征和语境信息,并融合这些基础特征进行情绪分类,以提高情绪(高阶特征)识别的准确性。通过这些方法,研究者可以直接提取出常用的声音特征。

①一种更符合人耳听觉特性的语言识别算法。



资料来源：本文作者绘制。

图 5 声音的自动分析与提取流程

第一阶段的三个步骤适用于声音数据的基础特征提取，而对于涉及情绪、认知、行为等的潜变量，需要对声音进行特征分类，也就是声音分析的第二个阶段，包括建立模型和结果解释两个步骤。这一阶段对于营销研究而言至关重要。

(4)建立模型。研究者将第一阶段的基础特征输入到模型中以进一步分析更加复杂的任务。比如，特定的识别任务可以选择有监督学习，de Benito-Gorron等(2019)提出有监督的混合卷积-LSTM模型，识别出声音中的特殊事件，如“人群”“欢呼”或“孩子们的喊叫声”，并验证了其在声音识别任务中的优异表现；Ashraf等(2023)提出了梅尔频谱(Mel-spectrogram)^①的CNN和Bi-GRU混合架构，并证明了其在声音流派分类(蓝调、古典、乡村、迪斯科、嘻哈、爵士、金属、流行、雷鬼和摇滚)上的优势。例如，对于这段视频中的音频文件，可能还存在旋律、配乐、鼓点以及语气等需要识别，这些精准的识别任务无法通过简单的人工剪辑实现，而是依赖于更高阶的算法模型来增强这些特殊声音事件的发生。

(5)结果解释。通过上述步骤可以实现多种声音分析任务，在这一步研究者需要通过变量设置、多方法结合等将先前提取的特征与营销相结合。除了分割与识别任务，还有声音的情感分析。大多数传统音乐情感分类方法都侧重于声音中文本的提取，进而用文本的情感特征表示声音的情感特征。但是，文本数据只能获得部分特征，结合特定的声音情感分析，更能体现结果

^①梅尔频谱是音频频率转换为梅尔刻度的频谱，梅尔刻度刻画了人们对频率的非线性感。

的稳健性和可解释性(Al-Qershi等,2022)。近年来,越来越多的研究人员开始关注多模态融合技术,这也为结果解释提供了新的思路(Nestor等,2023)。以上述视频为例(见图5),结合视频的语言文本内容与其对应的声音,可以更好地探究视频所传达的信息与情绪。此外,还可以通过分析各个时点的弹幕,判断整个视频的高潮点,进而通过提取高潮点的声音来更精准地了解粉丝的喜好与互动特征。

六、未来研究展望

(一)发掘传统声音与人工智能声音的不同,聚焦时代新声

随着在线购物与智能设备的发展,越来越多的消费者会在购物过程中使用Siri、Alexa等人工智能语音助手,比如:在购物、查询信息时询问语音助手,获取信息;在家居设备上安装智能设备,实现声控互动,打造智慧家庭;在车载导航中选择自己喜欢的明星语音包,使得行车导航过程更具有趣味性;等等。其中,大多数语音助手采用人工智能合成的声音,一些学者认为技术合成的声音与人类真实声音相比显得不自然,缺乏韵律感(Rodero,2017),不利于提升用户使用体验。尽管也有研究证明人工智能声音可以驱动创意表达的范式转变(Ijiga等,2024),但现有研究大多只是简单对比人工智能声音与传统人声,缺乏对其特殊性的深入研究,也没有研究系统分析消费者对人工智能合成声音与人类声音的偏好与反应。基于此,如何设计人工智能的声音特征,了解消费者对人工智能声音与真人声音的反应差异,放大人工智能声音便利性的优势并进一步挖掘其独特优势,是未来可以深入挖掘的研究方向,具有较强的实践意义和现实需求。

(二)开展扎根中国本土的声音营销效应研究,传播中国声音

目前关于营销中声音的研究,大多基于西方语言文化背景,聚焦英语的文化环境和语言环境。已有研究表明,中文和英文在语言结构、语言文化乃至大脑加工方式上都存在差异(Ge等,2015),那么它们所带来的心理反应和行为效应是否也存在区别?该问题仍处于“黑箱”之中。因此,未来研究可以基于该话题进行深入,一方面,探讨中国本土营销实践中的声音问题,比如销售人员话语中的汉语独有特质所带来的影响、中西方背景音乐的效果差异等。另一方面,分析中国本土文化下的声音问题。中国地大物博、民族繁荣,几千年的历史底蕴为我们留下了许多诗词歌赋,其韵律、平仄也是激发消费者感受的重要因素,需要进一步探索。此外,中国具有繁多的地区方言,承载着民族的历史与文化,随着近年来经济的复苏,许多方言和民俗在旅游营销、广告设计、直播电商中得到广泛应用,比如短视频平台上风靡的歌曲方言版、品牌的民俗音乐广告、方言直播带货等。这些营销中的声音已经在互联网上引发了巨大的关注,但其效应和边界还未得到学界的广泛关注(徐震南等,2023),因此未来研究可以进一步探索方言在营销中的应用,推动声音营销效应研究的本土化、中国化。

(三)推动大数据时代的AIGC多模态研究,洞察“画外之音”

如果说智能手机的变革带来了用户生成内容(user generated content,UGC)的快速发展,那么人工智能技术的变革将会带来人工智能生成内容(artificial intelligence generated content,AIGC)的火爆应用。GPT-2发布于2019年,被许多人认为是第一个大型语言模型,有15亿个参数,训练成本估计为5万美元,大语言模型和多模态模型正变得越来越大和昂贵(Nestor等,2023)。如果说ChatGPT的横空出世是AIGC时代的起点,2023年12月谷歌多模态大模型Gemini 1.0版^①的推出便是多模态AIGC进入新阶段的标志。2024年2月16日,OpenAI发布首个AI文字生成视频模型Sora,使用者只需用文字描述一个场景,Sora便可生成一个时长最高可达一分钟

^①Gemini是一个全能的模型,它能结合文字、图像和声音信息,更准确地判断用户的兴趣,从而提供更符合用户需求的推荐内容。

的视频。文本不是唯一的数据类型,还有图像、视频、音频等;2024年5月,OpenAI推出GPT-4o^①,与现有模型相比,GPT-4o在视觉和音频理解方面非常出色,特别是具备了“看”世界的能力。如果能把这些不同的数据类型都整合起来,模型的能力就会更强,更能理解和生成复杂的内容,这也就是多模态的内涵。因此,对于管理领域的学者来说,数据不仅仅是单一的文字、图片、视频,更是大量的多模数据的结合。斯坦福《2023年AI指数报告》指出,多模态技术向我们展示了令人印象深刻的能力,并已开始在现实世界中得到广泛部署,这种技术仍随着大语言模型(以文本为研究对象)而被人们熟知,声音与其他模态的交互研究目前还处于起步状态。在未来,研究内容不仅仅是视觉上的文本与图像、听觉上的声音,更是多感官的组合匹配。因此,随着AIGC更接近人类的认知和沟通方式,对声音线索的研究也不会再局限于提取声音本身,能够同时支持文本、音频、视频的分析技术也是未来值得关注的领域。

(四)关注声音提取与分析的新方法,紧跟前沿声音

声音自动提取与分析技术方面问题一直被认为是声音研究受阻的核心原因(Wang等,2024a)。虽然学者们注意到声音在营销场景中的普遍性,从销售人员、客户服务代表到医生和政客,几乎人人都需要用声音来交流,但是相比于蓬勃发展的文本、图片等非结构化数据分析技术,声音分析技术仍然是一个亟待探索的领域。尤其地,当较为简便的自动声音提取软件得到广泛应用时,营销领域也仅有Wang等(2021)在音频挖掘中使用了QA5自动语音分析软件,因此存在一定的技术应用不足问题。虽然音频分析算法的发展比文本、图像分析算法慢得多,但在计算机科学领域,已有各种各样较高精度的卷积神经网络模型进行自动音乐标注、音频指纹识别、音频推荐、音频分类、情感分析等声音挖掘任务(de Benito-Gorron等,2019;Choi等,2017)。因此,关注声音提取与分析方法的发展与应用,提出并优化具有场景性、可解释性的声音算法,有助于营销领域对声音线索的理论探索。

主要参考文献

- [1]侯建成,刘昌.语言与音乐活动的脑机制研究述评[J].心理科学进展,2008,16(1): 50-58.
- [2]刘德文,高维和,闵凉宇.声音特征和文本策略的说服效应研究[J].管理学报,2022,19(9): 1373-1381,1408.
- [3]刘扬,张苗辉,郑逢斌.听觉选择性注意的认知神经机制与显著性计算模型[J].计算机科学,2013,40(6): 283-287.
- [4]徐婕,肖莉.一画胜千言:图像数据在营销领域的应用[J].外国经济与管理,2022,44(9): 51-69.
- [5]徐震南,陈可,郭晓凌,等.“声”临其境:营销沟通中的方言积极效应[J/OL].南开管理评论,<http://kns.cnki.net/kcms/detail/12.1288.f.20230216.1751.014.html>,2023-02-21.
- [6]张全成,赖天豪,杨宇科,等.消费者的多感觉交互:表现、形成机制及研究展望[J].外国经济与管理,2017,39(7): 80-90.
- [7]钟科,王海忠,杨晨.感官营销研究综述与展望[J].外国经济与管理,2016,38(5): 69-85.
- [8]Ahlbom C P, Roggeveen A L, Grewal D, et al. Understanding how music influences shopping on weekdays and weekends[J]. Journal of Marketing Research, 2023, 60(5): 987-1007.
- [9]Al-Qershi O M, Kwon J, Zhao S N, et al. Predicting crowdfunding success with visuals and speech in video ads and text ads[J]. European Journal of Marketing, 2022, 56(6): 1610-1649.
- [10]Anikin A. The link between auditory salience and emotion intensity[J]. Cognition and Emotion, 2020, 34(6): 1246-1259.
- [11]Ashraf M, Abid F, Din I U, et al. A hybrid CNN and RNN variant model for music classification[J]. Applied Sciences, 2023, 13(3): 1476.
- [12]Baddeley A, Eysenck M W, Anderson M C. Memory[M]. New York: Psychology Press, 2009.
- [13]Belin P, Bestelmeyer P E G, Latinus M, et al. Understanding voice perception[J]. British Journal of Psychology, 2011, 102(4): 711-725.
- [14]Bhatia R, Garg R, Chhikara R, et al. Sensory marketing—A review and research agenda[J]. Academy of Marketing Studies

^①GPT-4o里的“o”是omni的缩写,也就是“全能”的意思,指接受文本、音频和图像任意组合的输入,并生成文本、音频和图像输出。

Journal, 2021, 25(4): 1528-1558.

- [15]Bigand E, Vieillard S, Madurell F, et al. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts[J]. *Cognition and Emotion*, 2005, 19(8): 1113-1139.
- [16]Biswas D, Lund K, Szocs C. Sounds like a healthy retail atmospheric strategy: Effects of ambient music and background noise on food sales[J]. *Journal of the Academy of Marketing Science*, 2019, 47(1): 37-55.
- [17]Black J W. Relationships among fundamental frequency, vocal sound pressure, and rate of speaking[J]. *Language and Speech*, 1961, 4(4): 196-199.
- [18]Chang H H, Mukherjee A, Chattopadhyay A. More voices persuade: The attentional benefits of voice numerosity[J]. *Journal of Marketing Research*, 2023, 60(4): 687-706.
- [19]Chattopadhyay A, Dahl D W, Ritchie R J B, et al. Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising[J]. *Journal of Consumer Psychology*, 2003, 13(3): 198-204.
- [20]Choi K, Fazekas G, Sandler M, et al. Convolutional recurrent neural networks for music classification[A]. Proceedings of 2017 IEEE international conference on acoustics, speech and signal processing[C]. New Orleans: IEEE, 2017.
- [21]Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science[J]. *Behavioral and Brain Sciences*, 2013, 36(3): 181-204.
- [22]de Benito-Gorron D, Lozano-Diez A, Toledano D T, et al. Exploring convolutional, recurrent, and hybrid deep neural networks for speech and music detection in a large audio dataset[J/OL]. EURASIP Journal on Audio, Speech, and Music Processing, <https://doi.org/10.1186/s13636-019-0152-1>, 2019-12-01.
- [23]de Paulo F J, Simiqueli A A, Tabet C J, et al. The effect of background music on sensory evaluation of craft beer[J]. *Research, Society and Development*, 2022, 11(9): e27611931620.
- [24]Failing M, Theeuwes J. Selection history: How reward modulates selectivity of visual attention[J]. *Psychonomic Bulletin & Review*, 2018, 25(2): 514-538.
- [25]Friston K. The free-energy principle: A unified brain theory?[J]. *Nature Reviews Neuroscience*, 2010, 11(2): 127-138.
- [26]Fu S X, Wu Y, Du Q Z, et al. The secret of voice: How acoustic characteristics affect video creators' performance on Bilibili[J]. *Decision Support Systems*, 2024, 179: 114167.
- [27]Ge J Q, Peng G, Lyu B J, et al. Cross-language differences in the brain network subserving intelligible speech[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(10): 2972-2977.
- [28]Hasan R, Shams R, Rahman M. Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri[J]. *Journal of Business Research*, 2021, 131: 591-597.
- [29]Hofbauer L M, Rodriguez F S. Emotional valence perception in music and subjective arousal: Experimental validation of stimuli[J]. *International Journal of Psychology*, 2023, 58(5): 465-475.
- [30]Hopfield J J. Neural networks and physical systems with emergent collective computational abilities[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1982, 79(8): 2554-2558.
- [31]Huang X, Labroo A A. Cueing morality: The effect of high-pitched music on healthy choice[J]. *Journal of Marketing*, 2020, 84(6): 130-143.
- [32]Ijiga O M, Idoko I P, Enyejo L A, et al. Harmonizing the voices of AI: Exploring generative music models, voice cloning, and voice transfer for creative expression[J]. *World Journal of Advanced Engineering Technology and Sciences*, 2024, 11(1): 372-394.
- [33]Jia X S. A music emotion classification model based on the improved convolutional neural network[J/OL]. *Computational Intelligence and Neuroscience*, <https://doi.org/10.1155/2022/6749622>, 2022-01-01.
- [34]Juslin P N, Västfjäll D. Emotional responses to music: The need to consider underlying mechanisms[J]. *Behavioral and Brain Sciences*, 2008, 31(5): 559-575.
- [35]Kaya E M, Elhilali M. Modelling auditory attention[J]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2017, 372(1714): 20160101.
- [36]Kimble C E, Seidel S D. Vocal signs of confidence[J]. *Journal of Nonverbal Behavior*, 1991, 15(2): 99-105.

- [37]Klein K, Melnyk V, Völckner F. Effects of background music on evaluations of visual images[J]. *Psychology & Marketing*, 2021, 38(12): 2240-2246.
- [38]Lai C J, Chiang C C. Effects of placement point of background music on shopping website[J]. *Work*, 2012, 41(S1): 5419-5421.
- [39]Lang M, Shaw D J, Reddish P, et al. Lost in the rhythm: Effects of rhythm on subsequent interpersonal coordination[J]. *Cognitive Science*, 2016, 40(7): 1797-1815.
- [40]Loui P, Bachorik J P, Li H C, et al. Effects of voice on emotional arousal[J]. *Frontiers in Psychology*, 2013, 4: 675.
- [41]Lowe M L, Loveland K E, Krishna A. A quiet disquiet: Anxiety and risk avoidance due to nonconscious auditory priming[J]. *Journal of Consumer Research*, 2019, 46(1): 159-179.
- [42]Menon V, Levitin D J. The rewards of music listening: Response and physiological connectivity of the mesolimbic system[J]. *NeuroImage*, 2005, 28(1): 175-184.
- [43]Moreno-Lobato A, Di-Clemente E, Hernández-Mogollón J M, et al. How emotions sound. A literature review of music as an emotional tool in tourism marketing[J]. *Tourism Management Perspectives*, 2023, 48: 101154.
- [44]Morreale F, Masu R, De Angeli A, et al. The effect of expertise in evaluating emotions in music[A]. Proceedings of the 3rd international conference on music & emotion[C]. Jyväskylä: University of Jyväskylä, 2013.
- [45]Nestor M, Loredana F, Erik B, et al. The AI index 2023 annual report[R]. California: The Stanford Institute for Human-Centered Artificial Intelligence, 2023.
- [46]Oleszkiewicz A, Pisanski K, Lachowicz-Tabaczek K, et al. Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults[J]. *Psychonomic Bulletin & Review*, 2017, 24(3): 856-862.
- [47]Pardi M B, Vogenstahl J, Dalmay T, et al. A thalamocortical top-down circuit for associative memory[J]. *Science*, 2020, 370(6518): 844-848.
- [48]Plack C J. The sense of hearing[M]. 3rd ed. London: Routledge, 2018.
- [49]Puligadda S, Vanbergen N. The influence of sound logo instruments on brand personality perceptions: An investigation of brand ruggedness and sophistication[J]. *Journal of Business Research*, 2023, 156: 113531.
- [50]Rodero E. Effectiveness, attention, and recall of human and artificial voices in an advertising story. Prosody influence and functions of voices[J]. *Computers in Human Behavior*, 2017, 77: 336-346.
- [51]Rodero E. Do your ads talk too fast to your audio audience? How speech rates of audio commercials influence cognitive and physiological outcomes[J]. *Journal of Advertising Research*, 2020, 60(3): 337-349.
- [52]Roschk H, Loureiro S M C, Breitsohl J. Calibrating 30 years of experimental research: A meta-analysis of the atmospheric effects of music, scent, and color[J]. *Journal of Retailing*, 2017, 93(2): 228-240.
- [53]Scherer K R, Ekman P. Approaches to emotion[M]. New York: Psychology Press, 2014.
- [54]Shang J C, Liu Z H. The “beauty premium” effect of voice attractiveness of long speech sounds in outcome-evaluation event-related potentials in a trust game[J]. *Frontiers in Psychology*, 2022, 13: 1010457.
- [55]Spence C, Keller S. Sonic branding: A narrative review at the intersection of art and science[J]. *Psychology & Marketing*, 2024, 41(7): 1530-1548.
- [56]Venkatesh S, Moffat D, Miranda E R. You only hear once: A YOLO-like algorithm for audio segmentation and sound event detection[J]. *Applied Sciences*, 2022, 12(7): 3293.
- [57]Vuust P, Heggli O A, Friston K J, et al. Music in the brain[J]. *Nature Reviews Neuroscience*, 2022, 23(5): 287-305.
- [58]Wang X, Bendale N, Pan Y J. Beyond text: Marketing strategy in a world turned upside down[J]. *Journal of the Academy of Marketing Science*, 2024a, 52(4): 939-954.
- [59]Wang X, Lu S J, Li X J, et al. Audio mining: The role of vocal tone in persuasion[J]. *Journal of Consumer Research*, 2021, 48(2): 189-211.
- [60]Wang Y, Yang Y, Wang X Q, et al. How do voice characteristics affect tourism interpretation purchases? An empirical study based on voice mining[J]. *Journal of Travel Research*, 2024b, 63(2): 481-495.
- [61]Wei W, Huang Z R, Feng C, et al. Predicting phonological information in language comprehension: Evidence from ERP

- representational similarity analysis and Chinese idioms[J]. *Cerebral Cortex*, 2023, 33(15): 9367-9375.
- [62]Zhou X Y, Yan X, Jiang Y W. Making sense? The sensory-specific nature of virtual influencer effectiveness[J]. *Journal of Marketing*, 2024, 88(4): 84-106.
- [63]Zwicker E, Fastl H. Psychoacoustics: Facts and models[M]. Berlin Heidelberg: Springer, 1990.

“Sound” Touches Hearts: The Marketing Effect, Mechanism, and Automatic Extraction Technique of Audio

Ran Yaxuan, Dong Linkang, Huang Yuting, Xiang Lizi

(School of Business Administration, Zhongnan University of Economics and Law, Wuhan 430073, China)

Summary: As an important communication tool, audio is ubiquitous in marketing. Compared with other marketing communication tools (such as text, picture, smell, etc.), audio has the advantages of wide channels, implicit and deep influence, and unobservable marketing purposes, which has attracted the attention of academia and practice. Understanding the specific role and theoretical mechanism of audio in marketing is the key premise for the effective use of audio, and an important issue to promote audio research. This paper systematically introduces the development of audio research and reveals the keywords of three major themes of audio research, including sensory marketing, multimodal, and AI. Audio can lead to a series of emotional and cognitive responses, and then promote individual purchase and prosocial behaviors to achieve the marketing effect or marketing performance. Neural mechanism, attentional control theory, prediction processing theory, and associative psychology are the key perspectives to understand the marketing effect of audio. This paper summarizes and compares the analysis methods of audio, and sorts out the technical process of automatic extraction and calculation. Finally, this paper puts forward four future topics, including the research on AI audio, Chinese native sound, and AIGC multimodal sound, and the development of automatic extraction technique.

Key words: audio; marketing effect of audio; sensory marketing; automatic extraction technique; unstructured data

(责任编辑: 王舒宁)