

人工智能语音在有声读物中的应用研究

◎ 刘一鸣 高 玥

摘要：本文以人工智能语音技术为基础，探讨有声读物与智能语音融合发展的方法与途径，从而提高有声读物制作的效率与质量。通过文献调研，分析当前有声读物生产与服务中存在的问题和智能语音技术的发展情况，提出相应的建议。目前有声读物存在生产效率低、成本高、版权保护欠缺、服务形式单一等问题，本文基于语音合成和语音识别技术，从有声读物的内容生产、审核与用户服务三方面提出融合发展的建议。

关键词：人工智能 智能语音 有声读物 数字出版

中图分类号：G237.6

DOI:10.19393/j.cnki.cn11-1537/g2.2019.11.008

20世纪50年代诞生，经历三次高潮的人工智能如今正在全球掀起新一轮的产业革命。传统出版产业也搭乘“AI快车”，试水人工智能，当下对选题策划、编校审核、出版发行等各方面实现人工智能对出版流程再造都有了进一步的研究和实践，而在传统出版的基础上发展的有声读物，“声音”部分尚处于初探阶段，智能语音技术作为当下人工智能的重点发展方向，与同属“声音系”的有声读物有诸多契合点，二者的融合在今后将迎来怎样的机遇与发展，值得期待与探讨。本文将以智能语音技术作为主要探讨对象，辅以必要的人工智能的其他技术来对智能语音技术在有声读物中的应用进行研究。

北京：清华大学出版社，2019：13，206。

[14] 黄华.技术、组织与“传递”：麦克卢汉与德布雷的媒介思想和时空观念[J].新闻与传播研究，2017（12）.

[15] 黄旦.听音闻道识媒介——写在“媒介道说”译丛出版之际[J].新闻记者，2019（9）：47.

[16] 喻国明.边缘创新与价值准则：互联网“下半场”的发展关键[J].新闻界，2017（10）.

[17] 王文利，甄烨.区块链技术在数字出版供应链融资中的应用[J].出版发行研究，2018（8）.

[18] 张志强，杨阳.直挂云帆济沧海——中国出版业改革开放40年成就回顾[J].中国出版，2018（21）.

[19] 冉景刚.区块链在数字出版业中的应用研究[J].出

一、有声读物发展热潮下的困局

移动互联网时代的到来和移动终端的发展为有声读物带来了更广阔前景，随着人们的快消费、阅读碎片化，有声读物以其即时即听、陪伴性等优势成为近年来兴起的热门阅读方式，由最初面向小众群体生产转向了大众市场，热度不断上升，各类有声读物平台纷纷涌现，行业竞争激烈，但同时也存在成本高、侵权、平台发展不充分等问题。

1. 人工录播耗费时间与成本

移动终端的普及和互联网的发展使有声读物阅读由

版发行研究，2018（6）.

[20] 聂静.基于区块链的数字出版版权保护[J].出版发行研究，2017（9）.

[21] 井底望天，等.区块链与大数据打造智能经济[M].北京：人民邮电出版社，2016：153—156.

[22] 简书网.你对未来版权UIP了解多少？[EB/OL].[2019-09-09].<https://www.jianshu.com/p/dfdd214efc9e>.

[24] 世代.出版社真的需要区块链吗？[EB/OL].[2019-09-19].<https://www.jianshu.com/p/47350795a283>.

（作者单位：南京大学信息管理学院 南京大学出版研究院）

传统的实体光盘转向专业网络平台，如喜马拉雅、懒人听书。艾瑞咨询将“高开发成本”列入2018年有声读物内容供给缓慢的主要原因之一。相较于其他音频产品，有声读物的制作在生产步骤、开发周期、人力投入方面都更为复杂，相应地需要耗费更高的成本。首先在正式开始制作之前需要获得作品的有声版权授权，在内容制作中，人工录播需要对文本进行分析、修改润色、情绪把握，效率较低，加上后期需要搭配背景音乐、剪辑、校对，无论是在时间、人力还是物力上都会积累大量成本。丁洪玲的调研发现，目前我国聘请专业播音员做有声读物，国家专业级播音员一般为100元左右/分钟（录音员级别不同价位不等），版权费用按国家标准，音频版权费50元/分钟，以此标准计算，一本15万字的图书，制作费用为10多万元^[1]。而且我国的内容付费还处于发展阶段，艾媒咨询公布的关于我国有声书市场的报告中，2018至2019年我国一线、二线及三四线城市中不愿为有声读物付费的用户分别占比71.4%、64.7%和55.1%，有声读物生产商面临着成本高且销售难的问题。

2. 内容审核与版权保护欠缺

作为开放的有声读物平台，UGC成为平台有声读物内容资源的主要来源之一，但UGC模式下参与的用户水平不一，制作出的内容质量也优劣不一，逐渐激烈的内容竞争局面使一些UGC用户无视版权界限随意侵权，或是通过刺激、暴力等内容来博眼球。用户从接受者变成生产者和传播者，平台的主动权逐渐转向被动，加上UGC内容不断增多，仅靠人工进行内容审核困难重重。

在版权保护方面，既要重视对原作者的权利保护，即是否获得了原作品版权授权，也要重视有声读物本身的版权，有声读物制作者的权利也应受到保护。录音录像者的工作不单单依靠设备，在劳动过程中也付出创造性的思维，其智力成果也应得到保护^[2]。当前对有声读物被盗录、修改、二次上传等问题的发现和处理机制还不完善。

3. 平台个性化推荐基准单一

当下的有声读物平台个性化推荐主要是根据用户的检索行为、浏览记录、购买记录、点赞、订阅等行为的记录，通过算法分析进行推送。这种方法能够为用户提供相似题材和类型的有声读物，是以有声读物的文本内容为基准，但有声读物由文本内容和音频组成，对用户在音色、

韵律方面的需求满足较薄弱。单一的推荐基准会形成“信息茧房”，既不利于受众多样化内容的接收，也不利于有声读物吸引和丰富受众群体。

4. 音频使用传播受载体局限

有声读物区别于传统纸质图书，能够适应碎片化阅读趋势，降低了“能够进行文字阅读”这一门槛，无论是儿童还是特殊群体，无论受教育水平如何都能成为有声读物的受众。但有声读物多以手机终端为主要载体，不能满足多样群体的需求，如儿童、盲人对于手机终端的使用存在障碍。单一的运营形式使有声读物的受众群体和传播范围受限，也使得有声读物的陪伴性特点难以得到更深入的发展。

二、智能语音赋能有声读物，唤醒“新声活”

人工智能语音技术近年来得到不断发展和完善，在音频处理与制作方面的独特优势在相关行业得到了充分发挥，对以音频为主要构成要素的有声读物来说，找到与智能语音技术的契合点，充分融合发展，有利于走出困境，迎来新一轮智能驱动的转型升级。

智能语音技术的发展可以追溯到20世纪50年代，主要由语音合成技术和语音识别技术组成，将人工智能的应用锁定于音频领域，赋予机器“听”和“说”的能力，是目前应用较广泛的人工智能技术之一。语音合成是一门跨学科的技术，它涉及声学、语言学、心理学、数字信号处理、人工智能、计算机科学等多个学科技术，是信息处理领域的一项前沿技术^[3]。语音识别技术是以人的语音信号为研究对象，能通过模式识别或深度学习的方法将语音信号转换成文字信息^[4]。语音识别技术目前发展较为成熟，在深度学习推动下，近场语音的识别率可达98%。据艾媒咨询的报告，语音识别被46%的手机网民选为人工智能技术发展水平较高的领域之一。

小到手机上的语音助手、汽车上的语音导航，大到AI合成主播、机器人记者、智能客服，从个人服务到产业运用，智能语音逐渐融入生活各方面，打造出新的音频生态。有声读物与智能语音技术，都以“音频”为主体，紧扣这一交点，在融合发展中更具针对性和专业性，在未来发展中存在着无限的可能。

1. 有声读物内容生产智能化

有声读物的内容包括文本内容与音频内容，将人工

智能技术应用于有声读物的内容生产，通过人机协同在文本转换、内容分析、智能配音等方面提高有声读物的制作效率，丰富内容资源。

(1) 文本转换，内容制作高效化

多语种读物的制作存在翻译费时且人才短缺的问题，对内容生产商来说，难以兼备各语种的翻译人员，而人工智能却具备多语种即时翻译能力，如科大讯飞推出的翻译机能进行33种语言翻译。虽然智能语音翻译不能做到百分百精确，但可采用人工智能协同翻译的形式来作为有声读物制作的智能补充。一方面，对于引进的国外有声读物，存在受众需要翻译后的文本来辅助理解、收听的情况，借助语音识别技术进行初步翻译，在此基础上再进行人工精确校对，相比一开始就采用人工逐句翻译，工作量减少，效率提升。另一方面，将外语图书制作为有声读物，录制者需要对原文本的思想、情感有准确的理解，同样可借助语音识别技术和自然语言理解技术进行翻译，辅助制作者理解文本内容。对于有声读物出口来说，也可采用人机协同翻译来提高效率，由此形成有声读物“出口—进口”“中文—外文”都覆盖的闭环。

通过语音识别技术将音频转化为文字，实现音频内容搜索。江苏电台“大蓝鲸”的AI内容编译系统，可将音频转化为文字再进行二次编辑，最终使得用户在终端上可进行音频内容的搜索、归类、传播。^[5]也可尝试将这种模式运用到有声读物制作，通过语音识别技术将有声读物的音频内容转化为文字，形成内容数据库，对一些高频词、关键词进行标记，每段音频拥有自己的标签后，有声读物制作者在后期加工或修改时，通过关键词、标签的检索就能从录制好的大量音频素材中找到需要的片段；对于有声读物平台的用户来说，即使不知道目标读物的准确信息，如读物的名字、类型或配音者，也能通过对有声读物内容的直接检索来锁定目标，避免了有的内容提供者为了曝光度而给作品贴上许多不相干的标签，导致检索结果混乱的问题。另外，也可通过音频片段对比和匹配来实现检索，将音乐平台的“听歌识曲”功能移植到有声读物平台，打造“听音识书”，通过用户提供的片段识别出原有声读物。音频内容检索功能的实现，一方面扩展了用户检索有声读物内容的途径，另一方面也有利于内容制作者在后期制作中对音频中重要与多余内容的查找与裁剪。

(2) 内容分析，音频匹配精准化

2015年，新华社推出机器人记者“快笔小新”，依托大数据采集与分析，通过“采集清洗”“计算分析”和“模板匹配”三个流程进行新闻稿件编写。机器自动写作实质上是自然语言生成引擎，简单来说需要经过“获取数据—分析数据—提炼观点—生成结构和格式—发布初版”几个流程^[6]，“快笔小新”的写作也是基于模板。而有声读物内容制作也可通过训练人工智能深度学习来辅助创作或直接创作。首先收集整理大量优质的、受到受众欢迎的有声读物进行数据分析，对于内容提供商和平台运营商来说，有声读物的销量和用户流量、评论反馈等都是很好的依据；在提炼观点阶段，人工智能对大量有声读物进行深入学习和理解后，抽取特征，在有声读物的属性（如种类、受众群体的年龄等）、音频属性（如录制者的性别、年龄、音色、韵律等）和受众的反馈（如好评、差评、分享次数、购买数量等）三者之间建立联系，生成“某类有声读物用具有某种特征的声音进行配音能够得到受众喜爱/讨厌”的模板，把握文本与音频间的关系与规律，“定制模板”，在以后有声读物制作过程中，能够根据文本类型调用模板，向制作者提供适配音色、情感层次等建议信息，甚至根据模板独立创作有声读物。

2018年7月发布的第六代微软小冰，进行了有声读物创作的尝试，能够为垂直有声读物领域搭建内容生产线，在20分钟内能够完成50个小时的内容生产，据艾瑞咨询发布的《2018年中国人工智能行业研究报告》，微软小冰创作的童话有声读物质量超过98%的人类创作者，用时仅为同水平人类的1/500，成本低至同水平人类的1/80000。虽然微软小冰还限于童话故事有声读物的创作，但为未来人工智能创作各类型的有声读物提供了很好的借鉴，随着技术的完善和成熟，通过内容分析，定制模板，实现有声读物精准制作，有助于优质内容的生产和效率的提高。

(3) 声音复刻，有声读物多样化

目前的语音合成技术在情感的认知、情绪的表达方面还不成熟，但如教材教辅、专业图书这类对情绪表达要求低、配音多枯燥而重复性大的图书，可采用人工智能进行配音，如微软运用语音识别、语音合成等技术推出的“朗文小英”能够为用户提供教材的朗读和教学。对于外语教材教辅来说，目前语音合成达到的水平已经完全能满足其

有声化的要求。

除了基本的合成语音以外，通过不断扩大语音库，丰富语音类型，能缩小人工智能制作的有声读物与人制作的有声读物间的音频种类多寡的差距。一方面，语音合成技术能够支持多种类的外语以及独具特色的地方方言，如科大讯飞的语音合成服务就提供了河南话、东北话、台湾普通话等12种方言语音，同一本有声读物能够制作各种不同的版本，增强了趣味性，满足了不同类型受众群体的需求。另外，个性化定制也是有声读物多样化的方式之一，对制作有声读物的方法、技巧等了解不多、没有制作经验或没有制作时间的受众，可以根据要求上传声音素材，由系统进行分析、优化、处理等后形成自己的音库，通过语音合成技术实现个性化有声读物定制。2018年1月，央视推出的纪录片《创新中国》通过搜集、选取、处理已故“配音大师”李易生前的声音素材，形成音库，还原了李易的声音，通过对配音文本的语言处理、韵律处理等最终合成声音，完成了整部纪录片的配音，效果可以假乱真，复刻个人的声音已初见成效。通过个性化声音的复刻，一千个人能拥有“一千种配音的哈姆雷特”，有声读物的选择不再固定于内容提供商或平台运营商，而是成为流动的资源，就像德国社会学家鲍曼所说的“液态化”传播，接受者也可能是传播者。每个人都将成为“流动的语音库”，通过语音合成制作自己的有声读物、收听自己制作的有声读物。

2. 有声读物内容审核智能化

以有声读物平台为主的内容审核方，可借助智能语音技术进行有声读物内容的审核，将人力从枯燥且庞杂的工作任务中解放，提高效率、节约成本的同时提升UGC内容的质量，并引入声纹识别技术为有声读物制作者提供维权依据。

（1）语音鉴别，协助内容筛选

建立音频过滤系统来提高有声读物内容审核效率。音频过滤系统依托语音识别下的语音转写技术，基于深度全序列卷积神经网络，将长段音频数据转换成文本数据，并运用自然语言处理技术，将音频的审核文字化，审核效率提高90%以上。^[7]音频审核主要用于视频网站、直播平台的音视频审核，如拥有秒拍、一直播等应用的一下科技，采用百度AI提供的技术，从画面、音频、文字等多方面入手综合进行内容审核，原先需要2分钟的审核任务

缩短到24秒就能完成，达到了96%以上的内容审核模型准确率，人力成本则降低了50%以上。有声读物平台可以基于审核内容和标准，建立敏感词、违规词等内容数据库，通过语音识别技术将音频内容转写为文本，利用关键词识别、语义识别对文本内容进行分析和判断，与数据库进行匹配，及时发现涉及数据库中的不合规内容的有声读物，对音频中一些无意义但涉黄、暴力等的拟声词或声音，利用敏感声音检测技术进行检测，实现对有声读物音频的多角度判断，以此过滤不合规定的内容。依托人工智能进行内容审核，能够一站式准确检测涉黄暴、政治敏感信息的内容，相比人工审核，速度更快且能够不间断工作，节约了大量人力成本，而与视频审核相比，有声读物省去了对图像等其他方面的审核，仅针对音频的审核降低了复杂度，简化了流程，既节省了成本，又提高了效率。当下有声读物平台都在尝试通过培养UGC内容打造竞争优势，孵化音频IP，如喜马拉雅拥有700万主播，1亿多种的音频内容，加上其推出的“万人十亿新声计划”，平台原有且依然在增加的内容加上这类活动生产的内容形成的庞大内容资源，在筛选与审核上单纯靠人工进行的工作量是难以想象的，质量也难以保证，采用智能语音提供技术支持，机器初审，筛选出可疑内容后人工复审、精审，以此提高效率。

（2）声纹识别，助力版权保护

声纹识别技术能够通过提取说话人的声音特征和说话内容信息来进行说话人辨别或确认。目前声纹识别技术主要用于刑侦、用户认证、门禁系统等的辅助。声纹鉴定依据的重要原理是语音同一认定，主要依据语音反映性、个人语音特征稳定性、个人语音特征总体差异性。^[8]不同的人音色、音调、发音习惯等都各不相同，可借助这些特性，建立有声读物制作者的声纹数据库。一方面，尝试与数字水印技术结合，将智能分析提取出的制作者的这些特性作为音频水印的内容嵌入作品，为其上传的作品打上隐形的“声音版权印记”，另一方面，也可用于音频内容的对比和认证，科大讯飞提供的声纹识别技术，能够将说话人声纹信息与库中的已知用户声纹进行1:1比对验证和1:N的检索，以此进行检索和验证。将声纹识别技术运用于有声读物的版权保护，对于音频被盗录、二次上传甚至恶意篡改、剪辑的侵权行为进行识别和检验，为有声读物制作者维护自身权益提供帮助和证据。

3. 有声读物用户服务智能化

聚焦于用户的听觉，以用户的多样需求为导向，将人工智能技术运用到为用户精准定制、个性化服务中，提升用户体验，实现有声读物内容与用户间的智能连接。

(1) 音频分析，读物精准推荐

随着用户需求的多元化，个性化成为重要的服务之一。对于有声读物平台来说，一方面要不断提高以用户行为分析为基准的内容推荐的精准率，基于人工智能的大数据分析，以用户需求为导向，对用户的检索行为、常听类型等进行分析，另一方面可将有声读物的音频特征纳入推荐基准，对用户检索、收藏的有声读物的音频进行分析和归类。美国的内容提供商 Trajectory 通过分词技术从转折情节中抓取关键词，提供特征分析并形成可视化情感走势图，使平台根据读者已读内容的情感曲线推送相似曲线的其他作品。^[9]语音识别技术可对录制者的性别、年龄、音色、音调等属性进行分析，形成画像，加上自然语言处理技术、语义分析、分词技术综合对有声读物内容、情感进行分析，形成以音频和情感为标准的分类，为用户推荐与其喜欢的音色、情感导向相匹配的有声读物。

(2) 智能终端，交互式陪伴体验

开发智能终端产品，也能提升有声读物交互式陪伴体验。智能家居是人工智能的一大发展领域，将有声读物扩展到线下也成为许多有声读物平台的发展趋势。智能语音嵌入终端，能够为不同类型、处于不同场景的用户提供智能点播服务，相较于手机，智能音箱这类终端始终处于开机状态，语音入口具有优势，可以随时被唤醒，复杂空间内的可用度更高，不需要动手便可实现有声读物的检索与播放，并且通过语音识别、自然语言处理等技术能够实现智能机器人与用户的对话互动，提升了陪伴感，在提供听书服务的同时又实现了人机交互体验，对于平台来说则打通了线上资源与线下服务，增加了传播端口。喜马拉雅率先推出小雅 AI 音箱，基于自身积累的资源，将有声读物引入线下。而对于缺少资源的终端制造商来说，可以与有声读物内容提供商合作，实现内容与技术的融合。除了智能音箱，也可尝试将有声读物引入其他智能家居、智能穿戴设备，并根据终端不同的特性和所处的场景来进行专业化内容的提供，如陪伴机器人针对的受众为幼儿群体，其中配置的有声读物资源就集中于儿童读物、幼儿教育一类，减少其他类型的提供，做到内容专、精、覆盖面广。

虽然目前智能家居还处于新生阶段，但有较强发展势头，据艾瑞咨询发布的《2018 年中国智能家居行业研究报告》，我国 2017 智能家电市场规模为 2828 亿元，未来三年复合增长率将超过 20%。智能家居拥有可观的发展前景，随着智能家居消费人群的扩大，有声读物的受众范围也会相应得到扩展。

三、结语

智能语音技术和有声读物都是当下新兴且热门的领域，以“声音”为主要内容的特性使二者的结合具有更可靠的基础和更多发展的可能性。本文仅对有声读物的内容生产、审核和用户服务三方面进行了研究，其他方面与智能语音技术的融合有待探讨，并且智能语音技术只是人工智能的一个分支，不可能完全独立，在与有声读物的融合发展，也需要其他人工智能技术加以辅助。而在有声读物与整个人工智能科学的融合之路上，还存在更多的可能性和更广阔的空间。

参考文献

- [1] 闫伟华, 申玲玲. 我国有声书行业的发展现状与策略研究 [J]. 出版发行研究, 2017 (2).
- [2] 熊宏蕾. 网络环境下有声读物版权利益平衡探讨 [J]. 出版参考, 2019 (2).
- [3] 阿日木扎, 包春梅, 马占新. 语音合成技术及其研究进展 [J]. 内蒙古科技与经济, 2010 (18).
- [4] 尹展鹏. 语音识别技术应用于现有三维绘图软件的探索 [J]. 科技创新与应用, 2019 (12).
- [5] 宫承波, 陈曦. 社交场景用户体验再升级——智媒时代音频传播观察之一 [J]. 新闻论坛, 2018 (3).
- [6] 胡郁, 袁春杰, 王玮. 人工智能技术在传媒领域的应用——以智能语音技术为例 [J]. 新闻与写作, 2016 (11).
- [7] 姜波. 利用智能审核平台对数字内容作品把控的探讨 [J]. 传媒论坛, 2018 (5).
- [8] 白海莉. 情感语音合成技术或对声纹鉴定准确性产生影响 [J]. 科技创新与应用, 2018 (36).
- [9] 徐晨耀. 人工智能技术在出版领域的应用研究 [D]. 北京: 北京印刷学院, 2019.

(作者单位: 湘潭大学公共管理学院)