# 智能工具辅助影视口述影像制作的案例比较研究——以公益短片与科幻电影片段为例

张思怡[1)]
1）中央民族大学，联系邮箱：1913512925@qq.com

**摘　要：**生成式人工智能、智能配音和轻量化剪辑工具正在改变影视口述影像的制作方式。本文以香港社会服务联会公益短片《点先可以方便轮椅人士出街》和《流浪地球2》00:19:40—00:24:40片段为案例，比较慢节奏公益文本与快节奏科幻文本中的工具适配、脚本修订和质量控制。研究发现，智能工具可提升画面记录、脚本初稿、配音和基础对齐效率，但信息筛选、术语一致、伦理表达和音画同步仍需人工把关。口述影像制作宜形成“工具辅助生成—人工语境判断—受众反馈优化”的协同路径。

**关键词：**口述影像，智能工具，人机协同，视听翻译，无障碍传播

## 1　引言

口述影像（audio description）是将影视文本中的视觉信息转化为口头语言的无障碍传播方式，通常插入原片对白、音乐和关键音效的间隙之中，使视障受众能够理解人物动作、场景转换、空间关系、表情姿态和文字标识等信息。它不是对画面的机械复述，而是一种受时间、声音和受众需求共同制约的跨符号翻译活动。

随着生成式人工智能、自动语音合成和移动端剪辑工具的发展，口述影像制作的技术门槛有所降低。文本生成工具可以提供脚本初稿，智能配音可以快速试听不同音色，剪辑软件能够完成基础音轨对齐。这些变化为高校实践项目、公益传播和基层无障碍服务提供了新的可能，但也带来信息臆测、表达失衡和音画不同步等问题。

本文选取一慢一快两类视频材料进行比较：慢节奏材料为香港社会服务联会公益短片《点先可以方便轮椅人士出街》，快节奏材料为《流浪地球2》00:19:40—00:24:40“太空电梯危机”片段。研究关注智能工具在不同文本中的适配边界，说明人机协同如何在提高效率的同时保持口述影像的准确性、可听性和伦理分寸。

## 2　文献综述与问题提出

国内口述影像研究早期多与公共文化服务、无障碍电影和残障权益保障相关。罗杏芬（2019：86-90）以公共图书馆无障碍电影项目为例，呈现脚本撰写、志愿协作和受众需求之间的关系；鄢秀、罗康特（2019：8-12）从理论与实践层面强调口述影像在新时代翻译研究中的价值。这类研究为口述影像进入公共文化服务和翻译研究视野奠定了基础。

从翻译学角度看，口述影像的关键在于视觉符号向语言符号的转换。蒋莉华、

李颖（2020：99-103）指出，口述影像需处理视觉、听觉、语言和非语言符号之间的互补关系。詹成（2022：30-36）从知识翻译学视角讨论口述影像研究范式，詹成、张晗（2023：151-158）进一步将其置于多模态口译教学之中，提示研究者关注“看见—理解—转述”的训练过程。

技术介入后，研究焦点开始转向工具适配、质量评价和人才培养。肖维青、董琳娜（2020：24-30）关于语音合成接受度的研究表明，技术可用并不等于用户可接受；韩鑫雨、肖志清（2026：36-45）则关注生成式人工智能辅助口述影像员培训。既有成果较少比较不同节奏、题材文本中的工具差异，本文因而以公益短片和科幻片段为例，讨论智能工具辅助制作中的修订机制。

## 3 研究设计与实践流程

本文采用探索性案例分析方法，将制作过程概括为画面记录、提示生成、脚本初稿、人工修订、语音合成、音轨对齐和反馈优化七个环节。案例选择遵循三项原则：一是节奏差异明显，便于观察工具在不同信息密度中的表现；二是题材差异明确，能够比较公益议题与科幻动作场面的描述重点；三是片段时长适中，适合在实践项目中形成可记录、可复盘的制作材料。

工具选择以可获得、低成本和易操作为标准。文本生成类工具可选 ChatGPT、豆包、智谱清言等；配音类工具可选剪映 AI 配音、TTSMaker、讯飞听见等；剪辑与对齐工具可选剪映、LiveDescribe 等。分析维度包括信息准确性、语言可听性、时间适配性、受众适切性和伦理可靠性。

表1 实践材料及分析重点

| 案例 | 材料来源与时段 | 视听特征 | 主要考察点 |
| --- | --- | --- | --- |
| 慢节奏公益短片 | 香港社会服务联会《点先可以方便轮椅人士出街》 | 节奏平缓，人物行动和空间关系清晰 | 客观描述、残障议题表达、服务流程呈现 |
| 快节奏科幻片段 | 《流浪地球2》00:19:40—00:24:40片段 | 镜头切换快，动作、警报、爆炸和对白密集 | 信息压缩、术语校准、动作排序、音效避让 |

## 4 慢节奏公益短片的口述影像实践分析

《点先可以方便轮椅人士出街》以轮椅人士日常出行为主题，画面围绕街道、交通、坡道、建筑入口、服务人员和行动过程展开。此类短片节奏较平缓，人物行动和空间关系相对清晰，口述影像的重点不是制造戏剧张力，而是帮助受众理解人物如何与环境发生关系。

在慢节奏材料中，智能工具较容易生成“人物—动作—地点—结果”结构的初稿。例如，在提示词中要求“只描述画面中可见信息，避免主观评价，句子适合口头表达”时，工具通常能够给出可修改的基础文本。这说明工具在画面记录和初稿生成环节具有实际价值。

但公益短片也容易暴露工具的概括化和道德化倾向。工具可能把可见画面直接写成“社会充满关爱”“无障碍设施完善”等评价性句子，或使用“艰难前行”“令人感动”等同情化表达。人工修订应回到可见动作和空间条件，如“轮椅使用者沿坡道进入大楼”“工作人员扶住玻璃门，让出通道”，以保持描述的客观与尊重。

因此，慢节奏公益文本的协同重点在于伦理把关和信息选择。对白、采访或

字幕已经提供的信息不宜重复，口述影像应补充画面中的行动路径、设施状态和空间关系。工具可以提高初稿效率，但最终表达仍需由制作者根据无障碍传播原则进行判断。

## 5 快节奏科幻片段的口述影像实践分析

《流浪地球2》00:19:40—00:24:40片段属于高密度科幻动作场面，包含太空电梯、升降舱、基地设施、飞行器、爆炸、警报和人物反应等多重信息。与公益短片相比，该片段镜头切换快、声音层次复杂，留给口述影像插入的时间更少。

在此类材料中，智能工具生成初稿时常见的问题是信息过满。面对密集画面，工具倾向于罗列人物、设备、火光和运动方向，形成长句或并列句，容易压住对白和关键音效。人工修订的首要任务是建立信息优先级：先说明影响情节理解的动作结果，再补充必要的空间方位，舍弃装饰性描写。

科幻片段还会放大术语处理问题。工具可能混用“太空电梯”“升降舱”“基地”“飞行器”等词语，甚至依据类型片经验扩写画面中并未明确出现的设定。口述影像不宜替电影补充世界观，而应坚持可见、可听、可验证原则，保证术语前后一致。

快节奏材料的核心难点在于叙事压缩和音画同步。较可行的做法是采用短句、断句和分段插入：在对白间隙说明动作结果，在音效强烈处减少解说，保留警报、爆炸和机械声本身的叙事功能。该案例表明，工具能辅助整理视觉元素，却难以独立完成取舍判断。

## 6 两类案例的人机协同差异

两类案例显示，智能工具辅助口述影像不能采用单一模板。公益短片的难点在于社会语境、残障议题表达和服务细节呈现；科幻片段的难点在于动作压缩、术语一致和音画同步。前者要求工具输出更客观、更克制，后者要求输出更短促、更有层级。

从提示词设计看，公益短片适合加入“避免评价性语言”“突出人物行动和环境设施”“尊重轮椅使用者主体性”等约束；科幻片段则适合加入“按动作先后排序”“只保留影响剧情理解的信息”“避开对白和关键音效”等约束。提示词不是固定模板，而是制作者理解文本之后对工具施加的任务边界。

从人才培养看，公益短片适合训练学习者的无障碍意识、伦理表达和社会议题敏感度；科幻片段适合训练注意力分配、多模态信息筛选和口头表达压缩能力。二者结合，可使口述影像实践从单一公益体验转向更系统的翻译能力训练。

表2 两类案例的工具适配与修订重点比较

| 维度 | 公益短片 | 科幻片段 | 修订策略 |
|---|---|---|---|
| 信息类型 | 人物行动、环境设施、服务过程 | 装置、动作、空间方位、危机场面 | 按文本类型设定信息优先级 |
| 工具常见问题 | 概括化、道德化、情绪化表达 | 描述过满、术语混用、动作顺序不清 | 公益片去评价，科幻片做压缩 |
| 时间压力 | 对白间隙较充足，但不可重复原声信息 | 对白和音效密集，插入空间有限 | 前者控制密度，后者使用短句 |

| | | | |
|---|---|---|---|
| 伦理风险 | 同情化叙述或替代残障经验发言 | 过度扩写世界观或臆测技术设定 | 坚持可见、可听、可验证原则 |

## 7 评价机制与质量控制

智能工具辅助口述影像的评价不应只看制作速度，也应看作品是否真正改善视障受众的理解体验。ADLAB 指南（Remael 等，2014）和 Fryer（2019：155-177）关于质量评价的讨论提示，口述影像质量需要同时考虑信息选择、语言表达、声音呈现和受众接受。Fryer（2016）的实践指南也强调，描述者必须在准确、简洁和可听之间取得平衡。

实践中可采用制作日志、脚本修订记录、评分量表和访谈反馈相结合的方式。制作日志记录耗时和返工次数，修订记录保留工具初稿与人工修改的差异，受众反馈用于解释评分背后的原因。版权和伦理同样不可忽视，电影片段用于研究和课堂实践时应控制传播范围，公益短片涉及残障议题时更应避免以工具生成的概括性语言替代真实经验。

## 8 结语

以《点先可以方便轮椅人士出街》和《流浪地球2》片段为案例，可以看到智能工具辅助口述影像制作具有明显的效率价值，也存在不可忽视的人工把关需求。慢节奏公益短片要求制作者处理好客观描述、社会语境和伦理表达之间的关系；快节奏科幻片段要求在动作压缩、术语校准和音画同步之间取得平衡。

较为可行的人机协同路径是“需求拆解—画面记录—提示生成—脚本初稿—人工修订—智能配音—音轨对齐—用户反馈—二次优化”。智能工具承担初稿、配音和基础对齐等辅助任务，人工主体负责信息选择、语境判断、伦理把关和最终质量控制。未来研究可进一步积累脚本语料和视障用户听评数据，检验不同工具组合的实际效果。

## 参考文献：

[1] Remael, A., Reviers, N. & G. Vercauteren. Pictures Painted in Words: ADLAB Audio Description Guidelines[Z]. European Union Lifelong Learning Programme, 2014.

[2] Braun, S. Audiodescription research: State of the art and beyond[J]. Translation Studies in the New Millennium, 2008(6):14-30.

[3] Fryer, L. An Introduction to Audio Description: A Practical Guide[M]. London/New York: Routledge, 2016.

[4] Fryer, L. Quality assessment in audio description: Lessons learned from interpreting[A]. In E. Huertas-Barros, E. Iglesias-Fernandez & S. Vandepitte (eds.). Quality Assurance and Assessment Practices in Translation and Interpreting[C]. Hershey: IGI Global, 2019:155-177.

[5] Jankowska, A., Milc, M. & L. Fryer. Translating audio description scripts: Translation as a new strategy of creating audio description[J]. SKASE Journal of Translation and Interpretation, 2017,10(2):2-16.

[6] Zhan, C. & R. Moratto (eds.). Audio Description and Interpreting Studies:

Interdisciplinary Crossroads[M]. London/New York: Routledge, 2025.
[7] 韩鑫雨, 肖志清. 生成式人工智能辅助的口述影像员培训模式探究[J]. 译苑新谭, 2026(1):36-45.
[8] 蒋莉华, 李颖. 翻译研究新动向：口述影像的多模态分析[J]. 外语教学, 2020(5):99-103.
[9] 罗杏芬. 公共图书馆无障碍电影口述脚本研究——以广东省立中山图书馆“心聆感影”项目为例[J]. 图书馆学研究, 2019(17):86-90.
[10] 全国人民代表大会常务委员会. 中华人民共和国无障碍环境建设法[Z]. 2023.
[11] 肖维青, 董琳娜. 视听翻译新发展：语音合成在口述影像中的应用——基于针对视障人士的接受实验与调查[J]. 东方翻译, 2020(2):24-30.
[12] 鄢秀, 罗康特. 新时代的翻译研究：口述影像理论与实践的重要性[J]. 东方翻译, 2019(4):8-12,24.
[13] 吕洁, 张霄军. 翻译学视角下的口述影像研究[J]. 译苑新谭, 2026(1):24-35.
[14] 詹成. 知识翻译学视阈下的口述影像研究：范式与路径[J]. 当代外语研究, 2022(6):30-36,45.
[15] 詹成, 张晗. 多模态口译教学中的口述影像：研究现状与前景展望[J]. 外国语文, 2023,39(6):151-158.