香港城市大學
City University of Hong Kong

Audio Description in the Era of AI: Challenges and Future Development

Jackie Xiu Yan & Su Lin
City University of Hong Kong

Funding information:
GRF Project: CityU No. 11609621) and DON-RMG Grant 9229125

專業 創新 胸懷全球
Professional · Creative
For The World

Abstract

Audio Description (AD) provides accessibility service to people with visual impairment. AD has received increasing attention from both the public and academia over the past 20 years, as a result, research and practice in this field have developed rapidly. Given the high degree of similarity between the processes of translating between languages and translating visual content into words, it is possible to apply the principles of translation studies (TS) to AD, consequently, AD has been considered a form a translation (audio-visual translation or intersemiotic translation). Just like the industry of translation, AD is facing the challenges brought by the recent development in AI. For example, technological innovations such as ChatGPT, automatic speech recognition, machine translation have dramatically changed the prospect of translation and AD studies. This study, based on a review of AD-related technological development and an experiment comparing AD output generated by human professionals and AI-powered tools, attempts to examine the challenges and possible direction for future development in this field. A sample of video slips is chosen for the production of both human-created and AI-generated AD transcripts, which will be evaluated by a group of 4 researchers consisting of both sighted and visually impaired people. Factors such as choice of details, language use, contextual appropriateness, and overall user satisfaction will be discussed. The results will provide insights into the strengths and limitations of human-generated and AI-based AD, and will shed light on how AI-powered AD may complement or augment human efforts, focusing on opportunities for collaboration and workflow optimization.

Funding information:
The work described in this paper was partially supported by the Hong Kong RGC Grant CityU 11609621 and DON-RMG Grant9229125.

AI development in AD

Accuracy and efficiency of machine-generated AD is growing

- AI can learn better: Self-attention mechanism enables encoder to learn
- Annotations contained in datasets for more accurate and efficient methods
- An automatic mixing system created to keep the AD narrator voice intelligible
- Machine-generated AD sounds more and more natural
- Interaction immediately with users: AI develops AD that that can interact immediately with users while audio describers may not be able to do so

Current Applications of AI in Audio Description

Key Points:

1.Streaming Services

1. **Example:** Netflix, Amazon Prime Video
2. AI-generated audio descriptions for movies and shows, improving accessibility for visually impaired users.
3. Descriptions are integrated into the user interface, allowing for easy selection.

2.Social Media Platforms

1. **Example:** YouTube, Facebook
2. Automatic generation of audio descriptions for videos uploaded by users.
3. Enhances engagement and accessibility for diverse audiences.

3.Media Production Tools

1. **Example:** Descript, Adobe Premiere Pro
2. AI-assisted tools that enable content creators to generate audio descriptions during the editing process.
3. Simplifies the workflow and reduces the time needed for manual descriptions.

4.Real-Time Descriptive Services

1. **Example:** Live sports events, theater performances
2. AI systems provide real-time audio descriptions for live broadcasts, enhancing accessibility for attendees and viewers.

5.Mobile Applications

1. **Example:** Seeing AI, Aira
2. Apps that use AI to provide audio descriptions of the environment, objects, and text in real time.
3. Empower users to navigate their surroundings with greater independence.

6.Automated Content Moderation

1. AI tools analyze video content to ensure compliance with accessibility standards, automatically generating descriptions where needed.

The role of human in AD production

1. AD post-editing involving less technical effort

Scriptwriter's role: In an automatic mixing system created to keep the AD narrator voice intelligible

Currently available machine-generated video descriptions, using even the most advanced neural network models, still lack granularity, accuracy and nuance, when compared to human-generated video descriptions.

Whilst this is a relatively simple task for most human beings, our analysis suggests that training the computer in a way that allows it to detect multiple cues and assimilate this knowledge into a reasoned conclusion remains challenging.

Gender: The need to train AI about identities representation in AD

More appearance information in image descriptions: Issues related to prejudice like racism, ableism, and transphobia

Labels, word ordering, community language preferences, skin colors

The role of human in AD production

2. Bias in AI-generated appearance descriptions: Concerns about AI bias: AI classified people and the resulting inequities.

Moving image sequences: Challenges for machine- generated descriptions

Unpredictable machine-generated descriptions affected by elements not detected by the human eye

poor lexicon in both variety and nuance

restricted repertoire of syntactic structures

striking errors in action identification

The computer vision algorithms often do not select the most salient actions even within individual frames.

The role of human in AD production

3. Emotions in AD: Facial expressions: laughing and grinning are difficult to distinguish in current models.

Scene and people: The need of necessary details: scene attributes (weather, etc.) ; personal attributes (name, gender, race, facial expression, etc.), which were stated as important by both the sighted and VI participants

The role of human in AD production

Challenges for machine- generated descriptions:

Human perception: Life experience

↓

Machine perception: availability of training datasets

香港城市大學
City University of Hong Kong

Thank you!

The research is supported by GRF Project 9043268 (CityU No. 11609621) and DON-RMG Grant9229125

專業 創新 胸懷全球
Professional · Creative
For The World