

Audio Description and Artificial Intelligence: A Preliminary Discussion for the Greek Audience

Dr. Ismini Karantzi, Ionian University, Greece

International Postgraduate Conference in Translation and Interpreting
13-14 November 2025, School of Modern Languages & Cultures, University of Warwick

Introduction

- Given the rapid advancements of **artificial intelligence (AI)** and **natural language processing (NLP)**, chatbots such as **ChatGPT** have attracted millions of users worldwide.
- However, this evolution does not mean that these tools are accurate, but possibly **supportive**.
- AI has several applications (research, business, academia, etc.), **but can it be used for audio description (AD) production?**
- Although text-to-speech technologies have been utilized, **research in the field of AD and AI remains relatively unexplored**.
- This poster makes a **first attempt to approach AD and AI for the Greek blind and visually impaired audience using ChatGPT**.
- In particular, a request was made to ChatGPT for creating an AD script and audio subtitles (AST) in Greek for an **English-language short film** already audio described as part of my PhD to **compare** our result to the AI one.

Methodology

Research questions

- Can ChatGPT create adequate-quality AD and AST?
- What impact has such a result for the Greek adult blind and visually impaired audience (BVIP)?
- How a human-based & an AI version are compared?

Corpus – case study

- Film used:

My Stuffed Granny (English narrative, 10')

Genre: stop-motion animation film,
Comedy/Drama | UK, 2014



With regard to the actual performance of the AD script, McGonigle (2007:43) notes that “the way an audio description is read is as important as what is being said, since the sound of a voice can bring life to a text through intonational meaning, as well as conveying a certain image, intimating with or alienating the audience”.

Human AD & AST

Steps followed:

- ✓ Drafted annotated AD-script (VIP consultant)
- ✓ Translated original dialogue for AST use
- ✓ Assigned voice talents (4 different versions)
- ✓ Recorded, mixed, synchronized, edited files
- ✓ Designed a multisensory approach (included hand-crafted puppets, etc.)
- ✓ Screened the film to the target Greek BVIP
- ✓ Presented a pre-screening touch tour
- ✓ Read questionnaires (including discussion)

Participants:

- ✓ 50 adult BVIP from Greek blind associations

Feedback received:

- Linguistic aspects
 - ✓ Adverbs to be avoided (interpretative)
 - ✓ Adjectives to be vivid, succinct and concrete (prevention of indefinite interpretations)
 - ✓ Syntax to be simple (avoidance of complex structures, passive voice, long sentences)
- Paralinguistic aspects
 - ✓ The target audience preferred the actor-voice version: his intonation, style, tone, theatrical elements, diction, articulation, rhythm and clarity played a crucial role in their presence (communicating also a range of emotions).
 - ✓ Respect to silence was deemed ideal for the scene that granny dies (i.e., quality AD is not a running commentary).
- Multisensory aspects – engagement
 - ✓ The touch tour (i) stimulated the audience imagination, (ii) provided BVIP with further information about the protagonists, (iii) did not overload the user, (iv) added information to the overall experience, since direct touch is the best way to explore an object (active sense), (v) reduced the over-reliance of a single medium to convey information (smell was also embedded).

ChatGPT input

🔊 Ακουστική Περιγραφή – “Η Γεμιστή Γιαγιά μου” (My Stuffed Granny)
(Προτεινόμενη) για τυφλό ή μερικώς βλέπον κοινό
Διάρκεια: περίπου 10 λεπτά
Τόνος φωνής: ζεστός, ουδέτερος, μελαγχολικός αλλά τρυφερός.

[Εισαγωγή]
Ηπια μουσική.
Η κάμερα κινείται ανάμεσα σε παλιά πολυκατοικίες μιας ελληνικής πόλης. Το φως είναι θαμπό, σαν να είναι χωρίς το πρωί.
Μια γκριζομάλλα γυναίκα, η Γιαγιά, κάθεται σε πολυθρόνα στο σαλόνι ενός φτωχικού σπιτιού. Δίπλα της, ένα μικρό κορίτσι με κόκκινα μαλλιά, η Σοφία, κοιτάζει σιωπηλά.

[Σκηνή 1 – Το σπίτι]
Το σπίτι είναι παλιό και σκοτεινό. Η Γιαγιά δείχνει ακίνητη, σχεδόν άψυχη.
Η Σοφία προσπαθεί να της μιλήσει, αλλά η Γιαγιά δεν αντιδρά.
Η μητέρα μπαίνει στο δωμάτιο, αγχωμένη. Κρατά ένα γράμμα — ειδοποίηση σύνταξης.
Η Γιαγιά είναι νεκρή, αλλά η μητέρα δεν το αποκαλύπτει στο παιδί.

[Σκηνή 2 – Η Απόκριση]
Η μητέρα “γεμίζει” τη γιαγιά — ντύνει και στρώει το σώμα της στην πολυθρόνα.
Η κάμερα εστιάζει στο βλέμμα της Σοφίας: απορία και φόβος.
Η μουσική γίνεται ειρηνικά χαρούμενη, θυμίζοντας παλιά κινούμενα σχέδια.
Η γιαγιά φαίνεται “ζωντανή” μόνο χάρη στο τεχνητό στήριγμα.

[Σκηνή 3 – Η Καθμερινότητα]
Οι δύο γυναίκες συνεχίζουν τη ρουτίνα τους.
Η Σοφία παίζει με τη γιαγιά, της δείχνει ζωγραφιές.
Η μητέρα παίρνει τη σύνταξη κάθε μήνα.
Η γιαγιά, άψυχη, στέκει πάντα στο ίδιο σημείο, σαν μακάβριο φυλαχτό.

AI AD & AST

➤ ChatGPT:

- **Pros:**
 - ✓ Offered a draft AD with basic info
 - ✓ Script was written in less than 1 minute
- **Cons:**
 - ✓ Did not recognize what AST are: it offered SDH instead
 - ✓ Did not understand that the old lady was a granny and not little girl’s mother
 - ✓ Did not annotate the AD script
 - ✓ Missed much info in the AD script

Participants (comparative study):

- ✓ 6 adult BVIP from the Patras (GR) blind association (pilot study)

Feedback received:

- ✓ Not adequate or acceptable AD

Conclusions

- AI can be supportive in the process of AD and AST, as well as their voicing.
- Human touch is necessary to ensure quality.
- Text-to-speech technologies can be used with positive feedback; however, it is preferred to use human voices (voice actors) and not robotic ones.
- Multisensory approaches (e.g. touch tours, etc.) can only be implemented by humans and not AI.

• Future direction:

- ✓ To evaluate AI tools in AD production (e.g., ViddyScribe and DescribePro), including their input quality etc.
- ✓ To conduct reception studies to Greek BVIP, comparing human and AI versions on AD, AST and voicing.

Feel free to get in touch
for further information and discussion!
is.karantzi@gmail.com

Indicative references

- Cheema, M., Elahimanesh, S., Martin, S., Fazli, P. & Seifi, H. (2025). DescribePro: Collaborative Audio Description with Human-AI Interaction. Available at: <https://arxiv.org/html/2508.01092v1#S2>.
- Karantzi, I. (2023). Audio describing (to Greek) foreign-language short animation films combined with a multi-sensory approach [PhD]. Available at: <https://www.didaktorika.gr/eadd/handle/10442/54733?locale=en>.
- RNIB, 2025. Exploring AI-generated audio description: can emerging technologies help expand access to broadcast media?. Available at: <https://www.rnib.org.uk/professionals/research-and-data/reports-and-insight/exploring-ai-generated-audio-description-can-emerging-technologies-help-expand-access-to-broadcast-media/>