A multilingual exploratory study of generative AI usage for museum audio description

María Isabel Rivas Ginel (Dublin City University)
Martin Kappus (Zurich University of Applied Sciences)
Laurie Le Roux (Université Bourgogne Europe)
Alexa Lintner (Zurich University of Applied Sciences)
Joss Moorkens (Dublin City University)
Will Noonan (Université Bourgogne Europe)
Lucía Pintado-Gutiérrez (Dublin City University)

DCU Ollscoil Chathair Bhaile Átha Cliath Dublin City University

ADAPT

UNIVERSITÉ BOURGOGNE EUROPE

zhaw

arsad advanced research seminar on audio description

Table of Contents

Technology and AD

Areas of technology use in AD

- Machine translation of AD-scripts (Vercauteren, G., Reviers, N., & Steyaert, K. (2021)).
- Synthetic voices for AD (Walczak, A., & Iturregui-Gallardo, G. (2022)).
- Automated, computer-generated video/image descriptions (Braun, S., & Starr, K. (2022); Bergin, D., & Oppegaard, B. (2024)).
- Image creation from AD (Bergin, D., & Oppegaard, B. (2024)).

Many of these topics and others (as well as their social and ethical implications) have been discussed in various presentations at this conference.

The Project
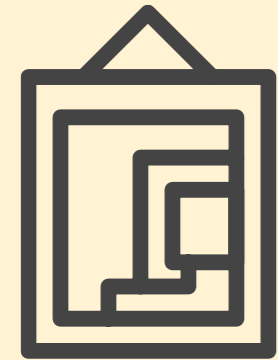
Objectives

Assess the use of generative AI in creating museum audio descriptions and analyse differences between AI, students, and professionals.
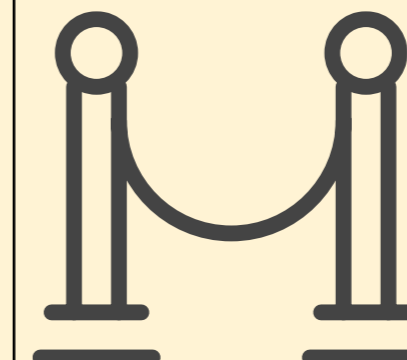
Methods

Compare descriptions of two paintings created by:

Claude.ai

Students

Professionals

Instructions

Between 200 & 300 words per painting.

Seminar/class based on materials by ADLAB PRO

Class content also included in the instructions to Claude.ai

Partners

Dublin City University, Zurich University of Applied Sciences, and Université Bourgogne Europe.
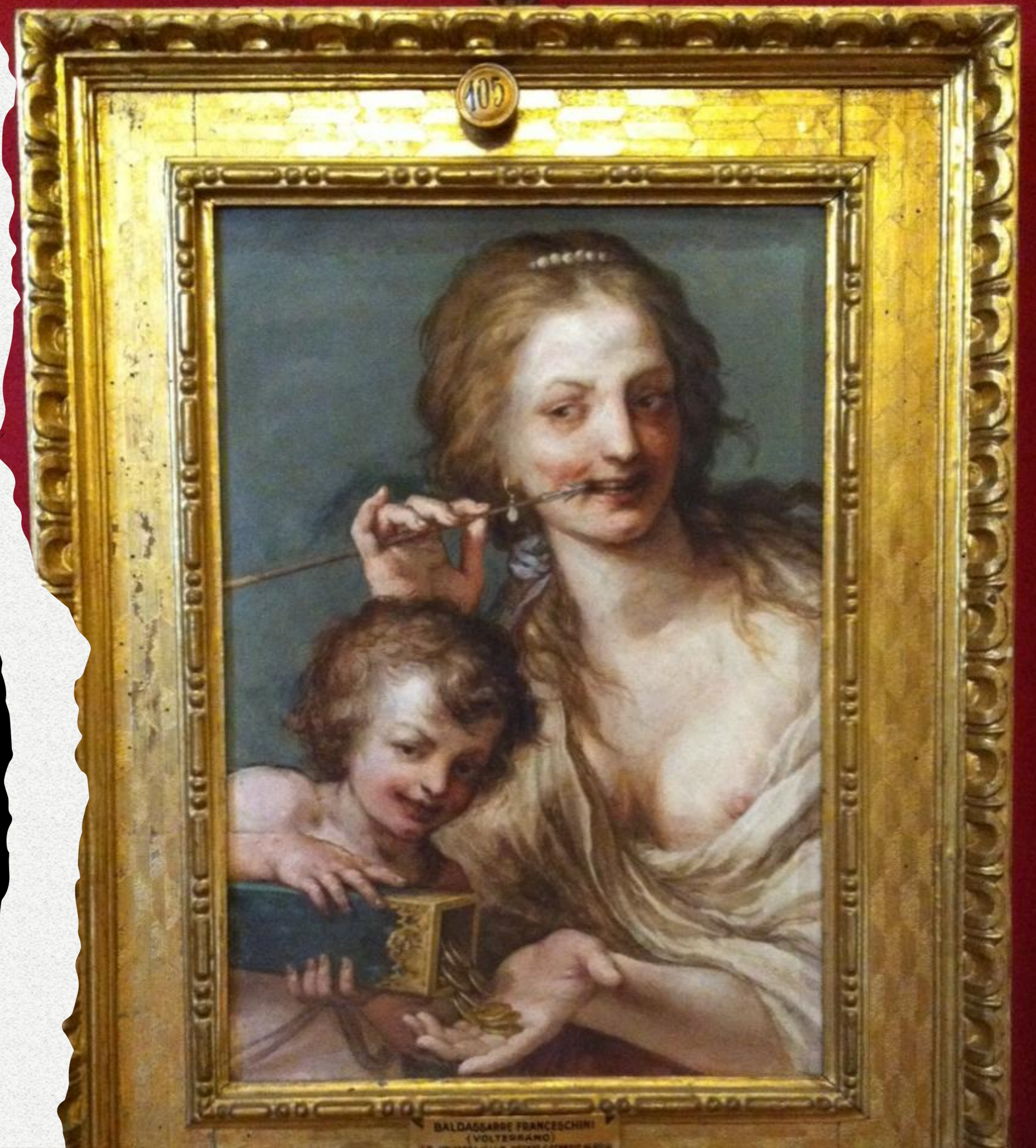
Number of files per language

EN	DE	FR	ES	IT	GR
9	10	14	3	3	1

Fernando Botero,
The Death of Pablo Escobar,
1999

Baldassare Franceschini,
Amore Venale,
17th century

Criteria

Based on previous corpus analysis (Perego, 2019; 2023)

"Units of Creative Potential" (Guerberof-Arenas and Toral, 2022)

Taxonomy of the components of visual communication in art and design (Soler & Luque, 2018)

Criteria

- Characters, Syllables, Tokens, Sentences, n° of words & frequency, Mean sentence length, very long words / multiword units (≥ 15 characters)

- Lexical diversity, Lexical density, Gunning fog index, reading level

- Parts of speech, types of verbs, deictic expressions

- Active/passive voice; Tense, aspect, modality; Third person vs. second person

- (Semi-)technical art terms; Basic colour names; Non-basic colour names

- Background information, structure, visual vs. non-visual information

- Subjectivity

- "Units of Creative Potential" (Guerberof-Arenas and Toral, 2022)

- Taxonomy of the components of visual communication in art and design (Soler & Luque, 2018)

Visual communication	Content	Icon	
		Symbol	
		Opinion	
	Form	Technique	
		Material	
		Style	
		Composition	
		Visual elements	Dot
			Line
			Shape
			Space
			Dimension
			Movement
			Direction
			Texture
			Tone
			Colour

✳English✳

Word Count & Sentence Length: Professional descriptions are longer than student/AI descriptions, relatively long but consistent sentences.

Student	218 / 24.2	233 / 15.5
Professional	279 / 20	285 / 19.2
Claude	246 / 18.9	209 / 17.4

Readability: All Escobar descriptions difficult to read. Professional and student descriptions of *Venale* easier to read, similar readability scores.

Background Information: Professionals provide historical and artistic context. Claude/students provide minimal information and misinterpret details.
   - ➢ Claude: describes "a figure **in the style of** Fernando Botero", potentially misidentifying the artist.
   - ➢ Claude & Student: fail to recognise Cupid. Claude mis-describes arrow as 'reed', student misses extra meaning of mother/child trope.

Structure: Professionals move from context to image. Humans move from background to foreground and general to specific, Claude less consistent.

Visual vs. Non-visual Information: Professionals integrate visual and contextual details, students and Claude focus more on visual elements.
   - ➢ Professional: "…in Botero's signature, cartoonish style with Escobar – 'The King of Cocaine' – portrayed in exaggerated proportions…"
   - ➢ Student: "…his signature Boterismo style, which is characterized by large and exaggerated volumes in the depiction of people and figures."
   - ➢ Claude: "…a figure in the style of Fernando Botero dominates this dramatic composition."

Subjectivity: Professionals combine description with interpretation, using a literary style to build up a sense of what each painting evokes. Students describe in more neutral/detached terms, Claude is less consistent (*Escobar* highly emotive, *Venale* much more academic)

✳ German ✳

Word Count & Sentence Length: Claude's descriptions are shorter than student's and professional. Student descriptions include longer sentences than professional or AI-generated descriptions.

	Escobar	Venale
Student	243 / 15.2	293 / 19.5
Professional	300 / 11.5	299 / 15.6
Claude	192 / 14.6	177/ 15.8

Readability/complexity: Professional and student description are more readable and less complex than Claude's descriptions

	Escobar (LIX / complexity / CFER)	Venale (LIX / complexity / CFER)
Student	42 / low /B2	45.0 / low / C1
Professional	42.9 / low / B2	42.2 / low / C1
Claude	49.9 / medium / B2	47.4 / low / C1

Background Information: Professional and student provide more historical and artistic context than Claude (date, size of paintings)

Structure: Professional follows a clear, structured approach (background to foreground). For both paintings Claude includes an introductory sentence directly addressing the listeners, placing them in front of the painting.

Subjectivity: Professional remains mostly neutral and factual, Claude is descriptive but lacks depth, and student explicitly provides the interpretation of the painting.

✳French✳

Word Count & Sentence Length: Professional descriptions (and sentences!) are long. Claude descriptions are shorter than requested.

Student	265 / 17.7	295 / 18.4
Professional	303 / 21.6	316 / 31.6
Claude	196 / 16.3	166 / 15.1

Readability: Professional descriptions are difficult (university level). Sophisticated vocabulary including infrequent words.
Student and Claude descriptions are easier (secondary level) but not necessarily accurate or stylistically coherent!

Background Information: Professionals and students provide context, professionals give external information (medium, nickname).
Claude gives limited context, often incoherent and/or inaccurate: "*art figuratif volumineux*", "*typique des villes méditerrannéennes*".

Structure: Professionals follow a clear structure: introduction, then foreground (general ➔ specific), then background. Students describe foreground, then background (fairly general description). Claude offers introduction, then (general) description, then concluding interpretation.

Visual vs. Non-visual Information: Professionals provide context, then focus on visual information with occasional interpretation.
Students focus on visual information with little analysis and occasional omissions.
Claude misinterprets details (arrow described as 'reed') but also effects, concluding with "a touching intimacy between the figures depicted".

Subjectivity: Professionals are relatively neutral and descriptive, explaining context and justifying interpretations through reference to visual details.
Student descriptions are neutral, relatively superficial, and clumsily written. Claude tends to use emotive language and unsupported statements.

✳Spanish✳

Word Count & Sentence Length: Student descriptions are longer and more complex than professional and AI-generated ones. Claude often fails to follow instructions.

Student	258 / 32.3	222 / 22.2
Professional	217 / 24.1	212 / 19.3
Claude	191 / 17.4	174 / 19.3

Readability: All Spanish descriptions are difficult, requiring a university-level reading ability.

Background Information: Professionals provide historical and artistic context, while Claude and students often miss key details.
- ➤ Claude: "arte latinoamericano contemporáneo".
- ➤ Claude & Student: fail to recognise Cupid, the arrow and the reference.

Structure: Professionals follow a clear, structured approach; Claude mixes interpretation with description, and students use a more evocative style.

Visual vs. Non-visual Information: Professionals balance visual and contextual details, whereas Claude and students sometimes introduce inaccuracies.
- ➤ Claude: mouth open / hand pressed against the chest / arrow as small branch / box as casket (cofrecillo).
- ➤ Student: describes the arrow as a feather

Subjectivity: Professionals remain neutral and factual, Claude is descriptive but lacks depth, and students use interpretative, literary language.

Key takeaways

English: Professional descriptions combine description and interpretation in a structured (and meaningful?) way. Student descriptions are more superficial and Claude descriptions are less coherent overall.

German: Professional descriptions are most clearly structured and descriptive, leaving interpretation to the listener. Student descriptions explicitly present an interpretation. AI-generated descriptions give partial interpretations for individual aspects.

French: Professional descriptions are the most accurate/informative and pleasant to read, though require greater reading ability. Student descriptions are simpler in style but lack interpretation of key details. Claude descriptions tend to be incoherent and/or inaccurate.

Spanish: Professional descriptions are the most accurate and structured, while student descriptions are emotionally engaging but less precise. Claude's descriptions are visually detailed but lack contextual depth and have errors in the description

Special thanks to Laurie Le Roux

Our amazing first-year Master T2M intern student funded via the "Appels à projets « blanc » de la Fédération des MSH de Bourgogne et de Franche-Comté"

Thank you!

Let's play!

Can you spot the description created by Claude.ai?

https://forms.gle/K7oFeEJrHBUXua2p9

OPTION 1

"The Death of Pablo Escobar" shows the leader of the notorious Medellín drugs cartel being fatally shot. It's an online image of a painting by the Colombian artist, Fernando Botero. Botero is from Medellín – also the hometown of Pablo Escobar.

In portrait orientation, painted in 1999, it's in Botero's signature, cartoonish style, with Escobar - "The King of Cocaine" - portrayed in exaggerated proportions, balancing precariously on the pan-tiled rooftop of a house. It's one of many houses packed closely together that spread towards lumpy moss-green mountains on the horizon, under a cloudy dark grey sky. But it is the only house to have a chimney pumping out white smoke. Escobar, much bigger than the house on which he's standing, takes up most of the image.

He's facing us, holding up a pistol in his right hand, aiming at the sky while his left palm is held up level with his chest as if to fend off a hail of black bullets that stream towards him and past him like a swarm of insects. He wears black trousers and a white shirt. The shirt is untucked and unbuttoned, as if pulled on in haste over his flabby body and he is barefoot, his delicate toenails glinting a surprising pink. His feet are tanned but the skin of his chest, hands and jowly, bestubbled face, is a sickly yellow. His eyes are closed. A red droplet in the middle of his forehead suggests that one bullet has already met its mark, and other wounds fleck his chest and bulging stomach. He has a thatch of brown hair and a moustache and his body twists in a final dance of death.

OPTION 2

Standing on terracotta roof tiles, a figure in the style of Fernando Botero dominates this dramatic composition. The painting uses present action to convey intense movement - the figure arches backward, his white dress shirt falling open to reveal his rounded torso marked by small crimson wounds. His bare feet grip the tiles as his body contorts in response to incoming bullets, which appear as dark shapes suspended in the gray sky. In his upraised right hand, he clutches a pistol, while his left hand presses against his chest in a gesture of pain. His face, with eyes tightly shut and mouth grimacing, expresses profound agony. The artist employs Botero's characteristic volumetric style, giving the figure an exaggerated, sculptural quality.

Below the figure stretches a Mediterranean townscape - densely packed houses with orange-tiled roofs cascade down gentle green hills. The houses feature white walls and small windows, creating a pattern of warm earth tones. The sky, painted in muted grays, adds an ominous atmosphere that contrasts with the warm colors below. The composition creates diagonal movement from the bottom right to the top left, emphasizing the figure's dramatic pose.

The painting combines elements of violence with theatrical presentation, freezing a moment of crisis. Through careful positioning, the artist places viewers directly on the rooftop with the figure, creating an immediate sense of danger and drama. The exaggerated style transforms what could be a purely violent scene into something more complex - both tragic and somehow dreamlike.

References:

ADLAB PRO. (2017). Assessment of current AD training practices. Project report no. 1. Retrieved from www.adlabproject.eu.

Bergin, D., & Oppegaard, B. (2024). Automating Media Accessibility: An Approach for Analyzing Audio Description Across Generative Artificial Intelligence Algorithms. Technical Communication Quarterly, 1–16.

Braun, S. & Starr, K. (2022) Automating audio description. In Taylor, C. & Perego, E. (Eds). The Routledge handbook of audio description. No. 191371. New York: Routledge. 391–406

Guerberof-Arenas, A., & Toral, A. (2022). Creativity in translation: Machine translation as a constraint for literary texts. Translation Spaces, 11(2), 184–212.

Soler, S. & Luque, M. O.(2018).Paintings to my ears: A method of studying subjectivity in audio description for art museums. Linguistica Antverpiensia New Series, 17, 140–156.

References:

Perego, E. (2019). Into the language of museum audio descriptions: a corpus-based study. Perspectives, 27(3), 333-349.

Perego, E. (2023). Audio Description for the Arts: A Linguistic Perspective (1st ed.). Routledge. https://doi.org/10.4324/9781003134237

Vercauteren, G., Reviers, N., & Steyaert, K. (2021). Evaluating the effectiveness of machine translation of audio description: the results of two pilot studies in the English-Dutch language pair= Avaluació de l'efectivitat de la traducció automàtica de l'àudiodescripció: els resultats de dos estudis pilot per a la combinació lingüística anglès-holandès. Revista Tradumàtica: Traduccio i Tecnologies de la Informacio i la Comunicacio, (19), 226-252.

Walczak, A., & Iturregui-Gallardo, G. (2022). Artificial voices. In The Routledge Handbook of Audio Description (pp. 365-376). Routledge.