Andha-Dhun: A First Look at Audio Descriptions in Hindi

Ritabrata Chakraborty*[1,2], Divy Kala*[1], Nisheeth Bhooshan Gupta[3], Ganji Sreeram[3], Pailla Balakrishna Reddy[3], and Makarand Tapaswi[1]

[1]CVIT, IIIT Hyderabad, [2]Manipal University Jaipur, [3]Jio Platforms Ltd.
https://github.com/katha-ai/AndhaDhun-HindiAD
* denotes equal contribution

Abstract. Audio Descriptions (ADs) narrate visual content for Blind and Low Vision (BLV) audiences during gaps in audiovisual media. There is growing momentum around ADs in movies and TV shows, and with mandates from India's Central Board of Film Certification (CBFC), there is a need to expand ADs beyond English. Yet, there is no work that generates ADs for any Indian language. To address this gap, we present the first systematic study of ADs in Hindi, contributing to aspects such as data, generation, and evaluation. We introduce Andha-Dhun, the first dataset of human-authored Hindi ADs collected from 8 full-length movies. We explore two approaches for generating ADs in Hindi: (i) directly from English dense video descriptions, and (ii) translating English ADs into Hindi. We evaluate these approaches using perplexity and LLM-as-a-judge metrics to assess fluency and quality respectively. We also analyze movies that have both English and Hindi human-authored ADs and find that naïve translation introduces artifacts and narrows diversity compared to original Hindi ADs. Direct machine translation fails to adapt cultural references, while human-translated ADs do better but still fall short. Our findings emphasize that the purpose of Hindi ADs is accessibility for Indian BLV audiences, and that this requires adapting content for the audience more than strict fidelity to the source.

Keywords: Audio Descriptions · Media Accessibility

1 Introduction

Audio Descriptions (ADs) are narrations inserted into silent gaps (*e.g.* between dialog) to help Blind and Low Vision (BLV) audiences understand visual content in movies and TV shows. An example AD is shown in Fig. 1. A sighted viewer perceives the different modes (visuals, music, speech, and sound effects) of a movie jointly to create an experience of a *coherent whole*. It is the job of an expert audio describer, to identify and isolate that information which is not accessible to the BLV audience, but is essential in creating that *coherent whole* [3]. Consequently, creating ADs requires expert knowledge and is a time-consuming and expensive process, taking up to 35-40 working hours for a 90-minute movie [14].

Fig. 1. Example of a Hindi audio description showing a clip from *3 Idiots* with the human-authored Hindi AD. AutoAD-Zero, an automatic AD generation framework, provides a dense description of the scene in English, which is then summarized to an English AD by an LLM. We show examples of two methods to get Hindi ADs: (bottom-left) Dense-to-Hindi, obtained by using a Hindi capable LLM directly on the English dense descriptions; and (bottom right) AD-to-Hindi, where AutoAD-Zero's [52] English AD is translated to Hindi using multiple translation models. The example shows an instance where a culture-specific item such as *scooter* is wrongly perceived as a motorcycle, resulting in downstream errors in Hindi generation.

Growing support for ADs has led to international regulatory efforts [44], and this momentum has led to an explosion of interest in automatic AD generation for English [10,21,39,51,52] and European languages [36,38,11,14]. Similarly in India, the Central Board of Film Certification (CBFC) announced that all movies releasing in India will be required to have ADs [43]. However, there is no prior work on ADs for Indian languages. In fact, the ADP ACB Title Directory[1] lists only around 650 titles of ADs in Hindi, compared to over 13.5K for English.

A natural solution to this problem is to translate existing ADs in English (or other languages) into Hindi. However, prior studies on AD translation across European languages show that machine-translated ADs often require human correction and cannot be assumed to substitute for the target-language ADs directly [11,47,33,31]. They also highlight specific issues with machine-translated ADs, including post-editing burden [11], machine-translation errors, timing mismatches [47], and the lack of visually grounded, time-aligned AD data [12].

Skopos theory [48] (from the German word for "purpose") offers insights to understand why. It states that a translation should be guided by its *skopos*; for Hindi AD, this means creating descriptions that fulfill the purpose of *helping Indian BLV audiences understand the movie and its visuals*, rather than strictly

[1] ADP Title Directory: https://adp.acb.org/adp-search is an initiative of the American Council of the Blind, providing a comprehensive directory of expert written ADs for content in multiple languages.

following the original wording. For instance, consider the English AD: *"He scored a touchdown and celebrated with a Gatorade shower."* A naïve translation to Hindi would be *"Usne touchdown score kiya aur Gatorade shower ke saath celebrate kiya."* However, this preserves culture-specific terms that may mean little to a Hindi-speaking BLV audience. In contrast, a skopos-aware translation might read *"Usne winning score banaya aur teammates ne uspe drinks daal kar celebrate kiya"*, fulfilling the purpose of the AD for the target audience.

We study the problem of Hindi ADs on three fronts: data, generation, and evaluation, all of which have not been explored before. We examine the effectiveness of translation models for converting human-authored English ADs into Hindi, and note the introduction of translation artifacts and a reduction in diversity. We show that the Hindi AD landscape is still at an early stage for both generation and translation systems. Furthermore, they often fail to adapt content for Indian BLV audiences, particularly when handling cultural references. Our contributions are as follows:

1. **Data.** We present Andha-Dhun[2], a dataset of 5,870 AD sentences across 8 full-length movies. This is the first corpus for studying ADs in Hindi.
2. **Generation.** We investigate the automatic generation of ADs in a zero-shot setup [52] and explore two translation-based approaches: (i) directly generate ADs from English dense descriptions using Hindi-capable models, and (ii) translate English ADs into Hindi.
3. **Evaluation.** We evaluate the quality of Hindi ADs using intrinsic (perplexity) and extrinsic (LLM-as-a-judge) metrics. We also study artifacts of Culture-Specific Items (CSIs) in machine- and human-translated ADs.

2 Related Work

Accessibility and translation. Early research [41,42] introduced ADs as an accessibility problem, emphasizing comprehension and usability for BLV users to follow audiovisual content. With rising adoption of ADs in movies after the European Parliament accessibility directives (2007) [9], movie-aware and translation-aware work gained interest [11,14]. Handling culture-bound [23,32] elements was a recurring problem for foreign movie ADs. To this end, many works [34,35] focused on using Skopos theory [8], to emphasize the *purpose* of ADs, and keeping in mind the target demographic. The Skopos framework connected low resource machine translation research [18] to AD translations. Across multiple settings, Vercauternen *et al.* [47,11] showed that mistranslations remain a major limitation in simply using translated ADs, hence human post-editing is needed.

Automatic AD generation. With the rise of Multimodal LLMs (MLLMs), automatic AD generation has grown in popularity [21,45,52,14]. Many approaches are training-based [10,20,19,21,30,49], while a growing body of work explores zero-shot methods [6,52,53,54,27,51,39]. Typically, zero-shot methods first use

[2] *Andha* means 'blind' and *dhun* means 'melody' (or 'sound') in Hindi.

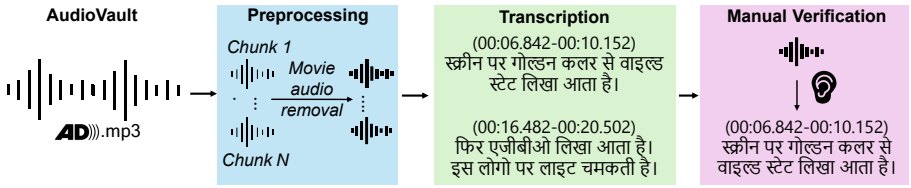

Fig. 2. Andha-Dhun dataset collection pipeline. Given movie ADs in audio format, we perform some preprocessing (chunking and movie audio removal). The cleaned audio chunks are transcribed using Gemini-2.5-Flash to obtain timestamped ADs. A final manual verification ensures high quality transcription and temporal alignment.

an MLLM to produce dense visual descriptions, which are then transformed into ADs by an LLM [51,52]. In recent work, Shot-by-Shot [51] uses shot scale and film-grammar to improve AD generation, while CoherentAD [27] uses a multi-stage approach to ensure ADs are coherent. Across multiple training-based and zero-shot methods, generated ADs consistently show a significant gap from achieving human-authored AD quality [4,26]. All this work is catered primarily towards English language AD generation.

Relevance of our work. We present a first work on AD generation for an Indian language (Hindi). Building on AutoAD-Zero [52], we evaluate translation-based AD generation. We also analyze shortcomings of using translations, both for automatically generated and human-authored English ADs.

3 The Andha-Dhun Dataset

To create the dataset, we source the AD audio of a movie from AudioVault[3], where each file contains the original movie soundtrack and Hindi AD narrations. We split each movie into roughly 30-minute chunks and process them with Resemble Enhance[4] to suppress non-AD sound sources, producing narrator-focused audio for transcription. Each chunk is transcribed with Gemini-2.5-Flash, which returns Hindi transcripts with start and end timestamps and labels each spoken segment as either actor (dialog) or narrator (AD). All outputs are manually verified and corrected to fix transcription errors and inaccurate boundaries. Fig. 2 summarizes this pipeline and the transcription prompt is shown in Sec. A.1.

To align the transcribed ADs with visuals, we source the corresponding movie videos from DVDs or public platforms. We manually synchronize the cleaned Hindi AD audio with the visual stream. Thus, the Andha-Dhun dataset consists of transcribed ADs synchronized with movie visuals. Since existing AD generation methods require a character bank [52,19,21,51] for identifying characters in the movie, we also create character banks for all movies using IMDb[5] cast profiles. Tab. 1 provides some statistics of the dataset.

[3] A non-profit repository for ADs, available at https://audiovault.net.
[4] A tool that improves quality of speech by performing denoising and enhancement, available at https://huggingface.co/ResembleAI/resemble-enhance
[5] IMDb (Internet Movie Database), available at https://www.imdb.com/

Table 1. An overview of the Andha-Dhun dataset consisting of 4 Hindi movies with Hindi ADs and 4 English movies with both Hindi *and* English ADs. The latter facilitate studying effects of translation. AD duration is the narration time in seconds.

Source	Movie	Duration (s)		#ADs	Average words/AD
		Movie	AD		
Hindi	3 Idiots	10,234	2,589	761	18.8
	Airlift	7,385	2,041	686	16.3
	Gehraaiyan	8,880	1,629	606	15.1
	Lage Raho Munna Bhai	8,707	1,428	516	14.6
English	Slumberland	7,237	2,362	895	16.9
	Murder Mystery	5,912	1,197	499	15.0
	Heart of Stone	7,525	3,167	997	17.7
	Extraction 2	7,437	2,489	910	13.2
	Total (8 movies)	63,317	16,902	5,870	16.1

4 Automatic Generation of Hindi ADs

As current English AD generation methods operate on few-second trimmed clips [45,21], we convert our dataset into video-AD pairs containing a short video clip and associated Hindi AD. We use these pairs for AD generation.

4.1 Experimental Setup and Metrics

Zero-shot AD generation. We adopt AutoAD-Zero [52], a zero-shot AD generation framework that uses an off-the-shelf MLLM to generate a dense description of a few-second video clip, which is then distilled into an AD by an LLM. To improve character grounding, the framework augments the visual frames with cast information predicted based on a character bank from IMDb cast profiles[5]. Qwen-2-VL-7B-Instruct [50] is the MLLM and Llama-3-8B is the LLM [17].

AD generation approaches. For generating ADs in Hindi, we explore two approaches (see Fig. 3): (i) **Pred AD-to-Hindi**, where we first generate English ADs using the AutoAD-Zero method and then translate them into Hindi using machine translation models such as IndicTrans2 [13], No Language Left Behind (NLLB-600M) [37], or MADLAD [28], and a commercial LLM-based translation using Gemini-3.1-Pro. (ii) **Pred Dense-to-Hindi** distills the MLLM generated dense-descriptions into Hindi ADs directly by using a Hindi-capable LLM. We use Nemotron-4-Mini-Hindi-4B-Instruct [25] and Gemini-3.1-Pro [16] as the Hindi-capable LLMs. For *Pred Dense-to-Hindi* with Gemini, we provide 5 previous ADs and/or dialogs to establish narrative context for the LLM. The prompt for Dense-to-Hindi is shown in Sec. A.2.

Metrics. We evaluate the quality of generated ADs from two complementary perspectives: extrinsic and intrinsic. As an extrinsic metric, we use *LLM-AD-eval* [21], an LLM-as-a-judge framework [29] that prompts a language model to

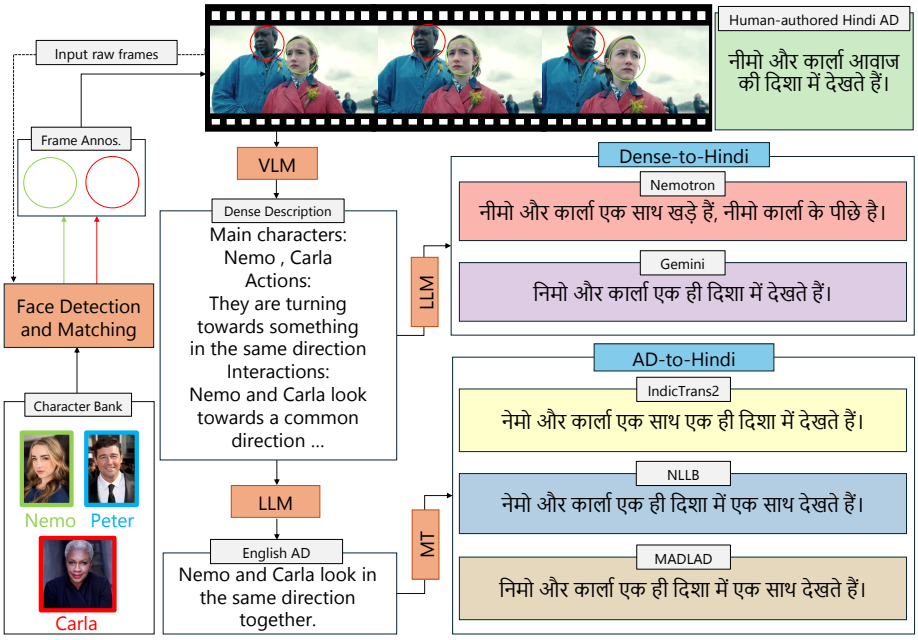

Fig. 3. Zero-shot Hindi AD generation pipeline. (From bottom-left) (i) A character bank of the movie cast is used to perform face recognition and create in-frame annotations for the input video. (ii) The annotated frames are passed to an MLLM to generate a dense description, which is then passed to an LLM to get a summarized English AD. This is the AutoAD-Zero pipeline [52]. (iii) To obtain Hindi ADs, we propose two approaches: (bottom right) *AD-to-Hindi* uses a machine translation model to translate the English AD to Hindi; and (middle right) *Dense-to-Hindi* uses a Hindi-capable LLM to obtain Hindi ADs directly from the English dense descriptions. For reference the human-authored Hindi AD is also shown (top-right).

evaluate the similarity between a predicted and ground-truth AD, assigning a score from 0 (worst match) to 5 (best match). The LLM-AD-Eval prompt is shown in Sec. A.2. As an intrinsic metric, we use *perplexity* [24], which measures how predictable the ADs are for a language model. Perplexity does not require ground-truth and acts as a measure of linguistic diversity [15,22].

The same metrics are also used in assessing human-authored ADs in Sec. 5. Given the recent findings by Kala *et al.* [26], we do not use classical captioning metrics such as CIDEr. Additionally, as Hindi allows flexible word order to convey the same meaning, two sentences can be semantically identical despite having low n-gram overlap [1,5]. Hence, surface-level metrics like CIDEr and BLEU may be unreliable for evaluating Hindi generations.

4.2 Results

Perplexity. We compute perplexity scores using Llama-3.1-8B. Due to a long-tail of extremely rare tokens/words, we remove outliers beyond the 95th per-

Table 2. LLM-AD-Eval scores for predicted ADs. Under both approaches (*Pred AD-to-Hindi* and *Pred Dense-to-Hindi*), we evaluate quality with two different models in LLM-as-a-judge setup: Gemini-3.1-Pro [16] and LLaMA-3-8B-Instruct [17]. **Best results** and second-best results per translation mode (along each row) are highlighted.

	AD-to-Hindi				Dense-to-Hindi	
Judge	IndicTrans2	NLLB	MADLAD	Gemini	Nemotron	Gemini w/ ctxt
Llama	3.07	3.03	2.91	2.99	**3.38**	3.00
Gemini	1.00	0.99	0.84	0.94	**1.12**	1.09

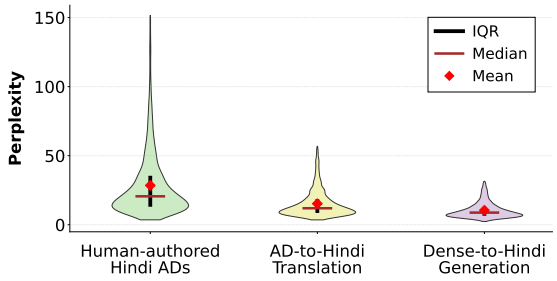

Fig. 4. Perplexity of Hindi ADs. The violin plot shows values up to 95th percentile. AD-to-Hindi translation uses IndicTrans2 to translate ADs, and Dense-to-Hindi uses Nemotron to generate ADs. The base model is Llama-3.1-8B.

centile. In Fig. 4, we observe that perplexity is highest for human-authored Hindi ADs, followed by AD-to-Hindi translation (IndicTrans2). Dense-to-Hindi generations by Nemotron display the lowest perplexity. This suggests that human-authored ADs have most linguistic diversity while capturing the rich visual details, while both AD-to-Hindi and Dense-to-Hindi lack this quality.

LLM-AD-Eval. Tab. 2 reports LLM-AD-Eval scores for the AD generation choices described in Sec. 4.1. Among the AD-to-Hindi translation models, both judges consistently rank IndicTrans2 highest and MADLAD lowest, with NLLB and Gemini falling in between. For the Dense-to-Hindi approach, the smaller Hindi-focused Nemo-tron achieves the best scores under both judges, outperforming the larger multilingual Gemini even with AD/dialog context.

The two judges differ substantially in their absolute scoring scales: the Gemini judge assigns scores in a narrow range around 0.84–1.12, whereas the Llama judge produces scores between 2.91 and 3.38. While absolute scores are different, likely due to varying harshness exhibited by the models, they consistently identify Nemotron as the strongest and MADLAD as the weakest methods.

Interestingly, the Dense-to-Hindi Nemotron surpasses all AD-to-Hindi translation methods, suggesting that directly generating Hindi AD from dense captions can be more effective than translating predicted English ADs. Nevertheless, even the highest scores remain well below the upper end of the 0–5 scale, confirming that current automatic Hindi AD generation methods are still far from human-level (results for translated human-authored English ADs in Sec. 5.2).

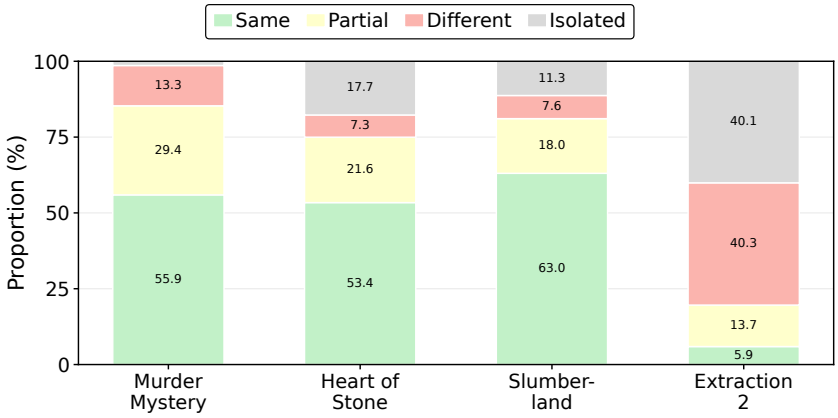

Fig. 5. Human-authored Hindi ADs closely match human-authored English ADs, machine translated to Hindi. We form pairs by temporally aligning human-authored Hindi ADs (*HI-Human*) with Hindi translations of human-authored English AD sentences (*Human AD-to-Hindi*). Each bar shows the fraction of times these paired ADs describe the same details, partially overlapping details, or entirely different details. Isolated or non-aligned cases are also displayed. For the first three movies, we see a strong correspondence with most pairs describing the same or partially overlapping details. On the other hand, *Extraction 2* has many isolated ADs, and even the aligned ones describe different details.

5 Human-Authored English to Hindi ADs

Dual AD tracks. The four English movies in Andha-Dhun are chosen to have human-authored ADs in both languages (*EN-Human*, *HI-Human*). We transcribe the English ADs automatically[6], and observe that the English and Hindi ADs are closely aligned, hinting that the Hindi ADs are likely expert human translations. Next, we create *Human AD-to-Hindi*, a set of Hindi ADs obtained by translating English ADs using automatic methods from Sec. 4.1.

To understand correspondences, we manually align *HI-Human* AD sentences with *EN-Human* based on temporal overlap. We first match entries by their start times across both tracks. When durations do not align, this is often due to over-segmentation during transcription. We merge adjacent entries until the combined duration matches its counterpart and treat the merged span as a single AD unit. This results in 2,350 Hindi and 2,358 English ADs, of which, 2,061 time-aligned pairs are used in further analysis.

Amount of overlapping detail in AD tracks. Given two paired AD sentences from the dual-track subset, we explore how much visual detail they share. We instruct Gemini-3.1-Pro to classify each *HI-Human* and *Human AD-to-Hindi* pair into three categories: (i) same details: same primary action/event even if one

[6] We transcribe the *EN-Human* AD track using WhisperX [2] and classify each sentence as *AD* or *dialog* using Gemini-3.1-Pro, following [26].

Overlap Type	Movie	HI-Human	EN →HI (Human)
Same Details	Slumberland	उसके दाढ़ी वाले पिता, पीटर, एक छोटी सी नाव से खाने का क्रेट उतारते हैं।	उसके दाढ़ी वाले पिता, पीटर, एक छोटी नाव से भोजन का एक टोकरा उतारते हैं।
Same Details	Heart of Stone	दूर से आता एक हेलीकॉप्टर एक पहाड़ के ऊपर स्थित रिज़ॉर्ट की ओर उतरता है।	एक दूर का हेलीकॉप्टर पहाड़ों में से एक की चोटी पर स्थित एक रिसॉर्ट की ओर उतरता है।
Same Details	Murder Mystery	प्लेन उतरता है। स्क्रीन पर शब्द। मालिगा, स्पेन।	जैसे ही विमान उतरता है, शब्द दिखाई देते हैं। मलिका, स्पेन।
Partial Overlap	Slumberland	फिलिप दरवाजा बंद कर देता है। वह हॉल में खड़ा है और नजरें नीचे झुकाए रहता है।	वह हॉल में खड़ा होता है और अपनी नज़रें नीची करता है।
Partial Overlap	Heart of Stone	यंग एक ड्रिंक्स की ट्रे से हमलावर का मुक़ाबला करती है।	यांग एक ड्रिंक्स ट्रे के साथ एक हमलावर से लड़ता है। पार्कर अपने प्रतिद्वंद्वी का गला घोंट देता है।
Partial Overlap	Murder Mystery	कार एक छोटे डॉक के पास आकर रुकती है।	वे एक गोदी के पास रुकते हैं जहाँ एक नौका को बांध दिया जाता है।
Different Details	Slumberland	वो एक कोने में रुकती है और मुस्कुराती है।	तहखाने में, वह एक सिंडरब्लॉक गलियारे के साथ चलती है।
Different Details	Heart of Stone	स्टोन हमलावर के चेहरे पर किक मारती है।	वह उसे एक मेज पर फेंक देता है, फिर उसे एक तरफ फेंक देता है।
Different Details	Murder Mystery	ऑड्री मुस्कुराती है।	डेक पर, ऑड्रे निक की एक तस्वीर खींचती है।

Fig. 6. Qualitative examples of semantic overlap in dual-track subset. We show aligned AD pair examples with same details, partial overlap, and different details. *HI-Human* refers to human-authored Hindi ADs. *EN→HI (Human)* refers to machine translated Hindi ADs from human-authored English ADs (*Human AD-to-Hindi*).

sentence is more detailed or uses synonyms, (ii) partially overlapping details: tracks share some visual content, but one also describes a clearly distinct additional action absent from the other, or (iii) different details: completely unrelated with no meaningful overlap in action, object, or character. The LLM also provides a one-sentence justification for each classification. We also perform manual validation for 50 pairs per movie to confirm that the assignments are reliable. The detailed prompt is provided in Sec. A.3.

Fig. 5 shows the breakdown for each movie and Fig. 6 shows some qualitative examples. Across 3 movies, same and partial correspondences dominate, confirming that *HI-Human* ADs are largely human-edited translations of the corresponding *EN-Human* ADs. This makes them a useful reference to analyze how meaning is preserved and adapted across languages in ADs. *Extraction 2* is an outlier, since its AD tracks describe different salient details (often for fast-paced action scenes) and is hence excluded from subsequent experiments.

5.1 Culture-Specific Items in ADs

Experimental setup. To ensure that the cultural references in translated ADs are genuinely accessible to an Indian BLV audience, we designed an evaluation pipeline grounded in *Skopos* theory. Here, the goal of the translation is to serve its intended purpose—accessibility for an Indian BLV audience. This is, in part, achieved by prioritizing cultural aspects over strict source-text fidelity.

We conduct this experiment in two phases. In the first phase, we identify Culture-Specific Items [7] (*e.g.* "quarterback", "ivy league colleges", "kitchen

Table 3. CSI resolution rates across categories. ↑ is better. Human-authored Hindi ADs (*HI-Human*) consistently outperform machine translated human-authored English ADs (*Human AD-to-Hindi*), with the highest resolution in cases of same detail (N=63). Lower resolution rates in partial (N=39) and different (N=18) cases suggests that humans adapt or avoid CSI-heavy content when direct resolution is difficult.

Method	Aligned	Same	Partial	Different
HI-Human	42.5%	50.8%	43.6%	11.1%
Human AD-to-Hindi	10.0%	7.9%	12.8%	11.1%

island") in the *EN-Human* side of the 4 movie dual track subset, which includes *Extraction 2*. Since the target audience might not understand Culture-Specific Items (CSIs) from the source culture, it is imperative to resolve them while translating. We use Gemini-3.1-Pro[7] to flag CSIs in 2,358 unique *EN-Human* ADs using the prompt given in Fig. 13. We manually verify the flagged samples to obtain 141 ADs (6.0%) with CSIs.

In the second phase, we translate the *EN-Human* ADs into Hindi to produce *Human AD-to-Hindi* ADs and manually evaluate whether these translations resolve the identified CSIs. We use 120 of 141 ADs aligned with *HI-Human* ADs from the dual track for further analysis. We categorize CSI resolution strategies based on Pederson's taxonomy [40]: (1) *retention*: keeping the original term, *e.g. gurney*; (2) *specification*: adding clarification, *e.g. pahiyen wala stretcher*; (3) *direct translation*: literal rendering, *e.g. White House* as *Safed Ghar*; (4) *generalization*: using a broader term, *e.g. Five Guys* as a fast food joint; (5) *substitution*: replacing with a cultural equivalent, *e.g. as wise as an owl* with *Chanakya jaisa samajhdar*; and (6) *omission*: removing the reference altogether. We also manually analyze the corresponding *HI-Human* ADs to assess CSI resolution. Finally, we compare how human authors and machine translations handle CSIs, including the strategies they employ, by manually assigning strategy label(s) to each culture-specific AD.

Results and discussion. We present results for CSI resolution in Tab. 3. Overall, among the aligned ADs, we observe that *HI-Human* ADs (42.5% resolution rate) substantially outperform machine translation (10.0% resolution rate) in resolving CSIs. We study the resolution rate across different levels of overlapping detail and find that when humans preserve the original content (same details), they resolve CSIs at the highest rate (50.8%), indicating deliberate effort to adapt cultural references. In contrast, resolution drops in partial (43.6%) and different (11.1%) cases. This suggests that when CSIs are difficult to adapt, humans may avoid resolving them and instead reframe the AD itself. Machine translation, on the other hand, shows low resolution rates consistently (hovering around 10%) across all categories. This likely stems from the fact that the translations are not *Skopos*-oriented, as they are designed to prioritize literal fidelity over adapting content to the target culture.

[7] Gemini-3.1-Pro gave better results than 2.5-Pro for identifying cultural nuances.

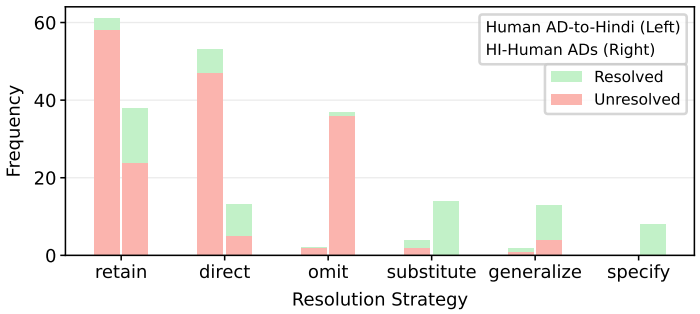

Fig. 7. CSI resolution strategies in *Human AD-to-Hindi* ADs (left) vs. *HI-Human* ADs (right). Bars indicate the strategy used for resolved (green) and unresolved (red) CSIs. Model translations (left) primarily rely on retention and direct translation and feature low resolution rates, while humans (right) use more adaptive strategies (*e.g.* substitute, generalize, specify) and resolve CSIs more effectively.

We discuss the CSI adaptation strategies used by humans and machine translations in Sec. A.3, with examples in Fig. 14. We see that translated ADs (*Human AD-to-Hindi*) overwhelmingly rely on retention and direct translation, leaving most CSIs unresolved. This conflicts with the purpose of the translation, Indian BLV accessibility. In contrast, AD experts (*HI-Human*) also employ retention and direct translation, but they resolve CSIs more frequently, indicating that these strategies can be used more effectively. Translations, by design, almost never use omission. Meanwhile, humans rely heavily on it, even though it often results in unresolved CSIs. This suggests that human AD experts omit details to avoid difficult-to-resolve CSIs or because such CSIs cannot be meaningfully adapted. This is consistent with Toury [46], which identifies omission as a legitimate and frequent strategy. For example, in subtitling, omission is sometimes the only feasible strategy [40]. In general, we find that humans favor adaptive strategies such as substitution, generalization, and specification. These strategies show higher resolution rates because they reinterpret content for the target culture, aligning well with Skopos theory. Examples of CSIs and their resolution strategies are presented in Sec. A.3.

In summary, translations are faithful to the source but unaware of purpose, whereas humans are more purpose-driven and culturally adaptive. These findings suggest that translation systems, if used for creating ADs in Hindi, should adapt culturally grounded ADs to the needs of the target audience.

5.2 Automatic translation of English human-authored ADs

While translating automatically generated ADs produced poor results (Tab. 2), does automatic translation of *EN-Human* ADs fare better? We present results on 3 movies; *Extraction 2* is excluded due to differences in Hindi/English ADs.

LLM-AD-Eval. From Tab. 4, we observe that machine-translated Hindi ADs do preserve the broad content of human-authored English ADs, resulting in scores close to 4 (5 being the highest). However, they still do not match the quality of

Table 4. LLM-AD-Eval scores for human-authored ADs. We evaluate AD quality with two models as LLM-as-a-judge setup: Gemini-3.1-Pro [16] and LLaMA-3-8B-Instruct [17]. We show results for two sets of ADs based on overlapping details: *All events* and *Same+Partial events*. **Best results** and second-best results are highlighted.

Detail	Judge	IndicTrans2	NLLB	MADLAD	Gemini	Gemini w/ ctxt
All	Llama	3.98	3.91	3.61	4.04	**4.06**
	Gemini	3.82	3.12	3.04	4.11	**4.14**
Same + Partial	Llama	4.03	3.96	3.71	4.08	**4.12**
	Gemini	3.99	3.82	3.21	4.24	**4.43**

human-authored Hindi descriptions. Gemini-3.1-Pro with previous AD/Dialog context outperforms all machine translation models. Scores are slightly higher for the *Same+Partial details* setting because it removes cases where the two AD tracks (English and Hindi) focus on different salient details. This also indicates that part of the challenge in the *All details* setting arises from humans choosing to describe different details rather than translation quality. Despite these high scores, recall that Tab. 3 shows that translations resolve only 10% of CSIs compared to 42.5% by human-authored ADs. Thus, while LLM-AD-Eval captures surface-level semantic overlap, it ignores whether the translation fulfills its purpose for the target audience.

Perplexity. We report diversity of the ADs of both tracks (*HI-Human* and *Human AD-to-Hindi*) in Fig. 8. Similar to Sec. 4, we observe that machine translating human-authored English ADs to Hindi results in reduced perplexity. This suggests that translated ADs are more predictable, and are likely less diverse compared to human-authored Hindi ADs.

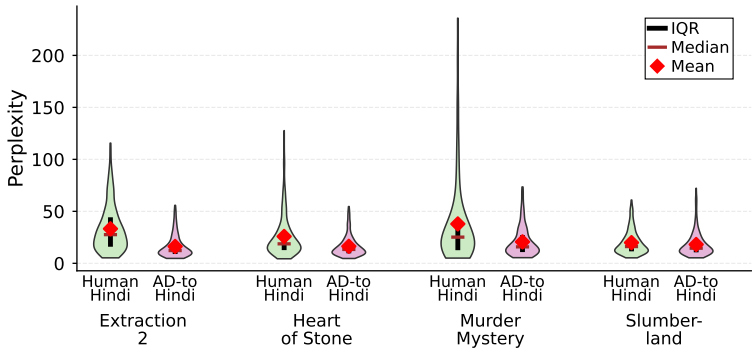

Fig. 8. Perplexity of human-authored Hindi ADs and machine translations of human-authored English ADs. Perplexity of translated English ADs is lower than human-authored Hindi ADs across all movies with dual AD tracks. IndicTrans2 is used to perform translation. The base model is Llama-3.1-8B. The plot shows values up to the 95th percentile.

6 Conclusion

We presented the first systematic study of Hindi ADs, addressing the gap in AD research in the Indian context. Through the Andha-Dhun dataset and evaluation of translation-based approaches, we found that current methods remain significantly below human-authored quality, especially in handling cultural references and preserving linguistic diversity. We find that directly generating Hindi ADs from dense descriptions can be more effective than translating predicted English ADs, indicating a promising direction for future research. Our findings also identified that the true purpose of Hindi ADs, to make Indian BLV audiences understand the movie and its visuals, requires prioritizing adaptation over strict translation. This hints that we need to develop methods that incorporate cultural and audience-aware adaptation.

Acknowledgments. This project was supported by SERB SRG/2023/002544, Adobe Research India, and Amazon Prime Video. We thank Google for sponsoring Gemini API credits.

References

1. Ananthakrishnan, R., Bhattacharyya, P., Sasikumar, M., Shah, R.M.: Some Issues in Automatic Evaluation of English-Hindi MT: More Blues for BLEU. Icon **64** (2007)
2. Bain, M., Huh, J., Han, T., Zisserman, A.: WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. In: Interspeech (2023)
3. Braun, S.: Audio Description Research: State of the Art and Beyond. Translation Studies in the New Millennium **6**, 14–30 (2008)
4. Braun, S., Qian, S., Zou, Y., Orasan, C.: Evaluation of AI-generated audio description for factual TV/media genres. Royal National Institute of Blind People (2025)
5. Chatterjee, N., Balyan, R.: Towards Development of a Suitable Evaluation Metric for English to Hindi Machine Translation. International Journal of Translation **23**(1), 07–26 (2011)
6. Chu, P., Wang, J., Abrantes, A.: LLM-AD: Large language Model based Audio Description System. arXiv preprint arXiv:2405.00983 (2024)
7. Dickins, J.: Language Studies: Stretching the Boundaries, chap. The translation of culturally specific items. Cambridge Scholars Publishing (2012)
8. Du, X.: A Brief Introduction of Skopos Theory. Theory and Practice in Language Studies **2**(10) (2012)
9. European Parliament and Council: Directive 2007/65/EC of the European Parliament and of the Council of 11 December 2007. WIPO Lex (2007)
10. Fang, B., Wu, W., Wu, Q., Song, Y., Chan, A.B.: DistinctAD: Distinctive Audio Description Generation in Contexts. In: CVPR (2025)
11. Fernández-Torné, A., Matamala, A.: Machine Translation in Audio Description? Comparing Creation, Translation and Post-Editing Efforts. SKASE Journal of translation and interpretation **9**(1), 64–87 (2016)
12. Fischer, L., Gao, Y., Lintner, A., Rios, A., Ebling, S.: SwissADT: An Audio Description Translation System for Swiss Languages. In: NAACL-HLT (2025)

13. Gala, J., Chitale, P.A., Raghavan, A., Gumma, V., Doddapaneni, S., Kumar, A., et al.: IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages. arXiv preprint arXiv:2305.16307 (2023)
14. Gao, Y., Fischer, L., Lintner, A., Ebling, S.: Audio Description Generation in the Era of LLMs and VLMs: A Review of Transferable Generative AI Technologies. In: NAACL Findings (2025)
15. Garbacea, C., Carton, S., Yan, S., Mei, Q.: Judge the Judges: A Large-Scale Evaluation Study of Neural Language Models for Online Review Generation. In: EMNLP-IJCNLP (2019)
16. Gemini Team: Gemini: A Family of Highly Capable Multimodal Models. arXiv preprint arXiv:2312.11805 (2025)
17. Grattafiori, A., Dubey, A., Jauhri, A., Team: The Llama 3 Herd of Models. arXiv preprint: arXiv 2407.21783 (2024)
18. Haddow, B., Bawden, R., Miceli Barone, A.V., Helcl, J., Birch, A.: Survey of Low-Resource Machine Translation. Computational Linguistics (2022)
19. Han, T., Bain, M., Nagrani, A., Varol, G., Xie, W., Zisserman, A.: AutoAD II: The Sequel-Who, When, and What in Movie Audio Description. In: CVPR (2023)
20. Han, T., Bain, M., Nagrani, A., Varol, G., Xie, W., Zisserman, A.: AutoAD: Movie Description in Context. In: CVPR (2023)
21. Han, T., Bain, M., Nagrani, A., Varol, G., Xie, W., Zisserman, A.: AutoAD III: The Prequel-Back to the Pixels. In: CVPR (2024)
22. Hashimoto, T.B., Zhang, H., Liang, P.: Unifying Human and Statistical Evaluation for Natural Language Generation. In: NAACL-HLT (2019)
23. Homayouni, G., Khoshsaligheh, M.: Audio Description Technology: Enhancing Communication of Culture-Bound Elements in Films. Journal of Business, Communication & Technology **3**(1) (2024)
24. Jelinek, F., Mercer, R.L., Bahl, L.R., Baker, J.K.: Perplexity—a Measure of the Difficulty of Speech Recognition Tasks. The Journal of the Acoustical Society of America **62**(S1), S63–S63 (1977)
25. Joshi, R., Singla, K., Kamath, A., Kalani, R., Paul, R., Vaidya, U., et al.: Adapting Multilingual LLMs to Low-Resource Languages using Continued Pre-training and Synthetic Corpus. In: Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages (2025)
26. Kala, D., Khandelwal, E., Tapaswi, M.: What You See is What You Ask: Evaluating Audio Descriptions. In: EMNLP (2025)
27. Khandelwal, E., Xie, J., Han, T., Bain, M., Nagrani, A., Zisserman, A., Varol, G., Tapaswi, M.: More than a Moment: Towards Coherent Sequences of Audio Descriptions. arXiv preprint: arXiv 2510.25440 (2025)
28. Kudugunta, S., Caswell, I., Zhang, B., Garcia, X., Xin, D., Kusupati, A., Stella, R., Bapna, A., Firat, O.: MADLAD-400: A Multilingual And Document-Level Large Audited Dataset. In: NeurIPS (2023)
29. Li, D., Jiang, B., Huang, L., Beigi, A., Zhao, C., Tan, Z., Bhattacharjee, A., Jiang, Y., Chen, C., Wu, T., et al.: From Generation to Judgment: Opportunities and Challenges of LLM-as-a-Judge. In: EMNLP (2025)
30. Lin, K.Q., Zhang, P., Gao, D., Xia, X., Chen, J., Gao, Z., et al.: Learning Video Context as Interleaved Multimodal Sequences. In: ECCV (2024)
31. Lüthi, N., Lintner, A., Fischer, L., Kappus, M., Ebling, S.: An Exploratory Analysis of LLM-Based Machine-Translated Audio Description Scripts in the French-German Language Pair. Advanced Research Seminar on Audio Description (2025)

32. Maszerowska, A., Mangiron, C.: Strategies for dealing with cultural references in audio description. In: Audio description, pp. 159–178. John Benjamins (2014)
33. Matamala, A., Boix, C.O.: Accesibility and Multilingualism: An Exploratory Study on the Machine Translation of Audio Descriptions. TRANS: Revista de Traductología (2016)
34. Mazur, I.: A Functional Approach to Audio Description. Journal of Audiovisual Translation **3**(2) (2020)
35. Mazur, I.: Same Film, Different Audio Descriptions: Describing Foreign Films from a Functional Perspective. Universal Access in the Information Society **23**(2) (2024)
36. Mazur, I., Chmiel, A.: Towards Common European Audio Description Guidelines: Results of the Pear Tree Project. Perspectives **20**(1), 5–23 (2012)
37. NLLB Team: No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv preprint arXiv:2207.04672 (2022)
38. Orero, P.: Sampling Audio Description in Europe. In: Media for all, pp. 109–125. Brill (2007)
39. Park, J., Ye, J., Lee, S., Ka, H.W., Han, D.: NarrAD: Automatic generation of audio descriptions for movies with rich narrative context. In: WACV (2025)
40. Pedersen, J.: Subtitling Norms for Television. John Benjamins Pub. Co. (2011)
41. Peli, E., Fine, E., Labianca, A.: Evaluating Visual Information Provided by Audio Description. Journal of Vision Impairment and Blindness **90** (1996)
42. Pettitt, B., Sharpe, K., Cooper, S.: AUDETEL: Enhancing Television for Visually Impaired People. British Journal of Visual Impairment **14** (1996)
43. Press Information Bureau: Central board of film certification (cbfc). Government of India, Press Information Bureau (Feb 2026), release ID: 2233786. Accessed: 2026-04-06
44. Reviers, N.: Audio Description Services in Europe: an Update. The Journal of Specialised Translation pp. 232–247 (2016)
45. Soldan, M., Pardo, A., Alcázar, J.L., Caba, F., Zhao, C., Giancola, S., Ghanem, B.: MAD: A Scalable Dataset for Language Grounding in Videos from Movie Audio Descriptions. In: CVPR (2022)
46. Toury, G.: Descriptive Translation Studies and Beyond. John Benjamins (1995)
47. Vercauteren, G., Reviers, N., Steyaert, K.: Evaluating the Effectiveness of Machine Translation of Audio Description: the Results of two pilot studies in the English-Dutch Language Pair. Tradumàtica tecnologies de la traducció (2021)
48. Vermeer, H.: A Skopos Theory of Translation: (some Arguments for and Against). Reihe Wissenschaft - TEXTconTEXT, Textcontext (1996)
49. Wang, H., Tong, Z., Zheng, K., Shen, Y., Wang, L.: Contextual AD Narration with Interleaved Multimodal Sequence. In: CVPR (2025)
50. Wang, P., Bai, S., Tan, S., Wang, S., Fan, Z., Bai, J., et al.: Qwen2-VL: Enhancing Vision-Language Model's Perception of the World at Any Resolution. arXiv preprint: arXiv 2407.10671 (2024)
51. Xie, J., Han, T., Bain, M., Nagrani, A., Khandelwal, E., Varol, G., Xie, W., Zisserman, A.: Shot-by-Shot: Film-Grammar-Aware Training-Free Audio Description Generation. In: ICCV (2025)
52. Xie, J., Han, T., Bain, M., Nagrani, A., Varol, G., Xie, W., Zisserman, A.: AutoAD-Zero: A Training-Free Framework for Zero-Shot Audio Description. In: ACCV (2024)
53. Ye, X., Chen, J., Li, X., Xin, H., Li, C., Zhou, S., Bu, J.: MMAD: Multi-Modal Movie Audio Description. In: LREC-COLING (2024)
54. Zhang, C., Lin, K., Yang, Z., Wang, J., Li, L., Lin, C.C., et al.: MM-Narrator: Narrating long-form Videos with Multimodal In-Context Learning. In: CVPR (2024)

A Supplementary Material

We describe in detail the prompts for each generation/evaluation pipeline.

A.1 Dataset

Prompt for Andha-Dhun Transcriptions. We use Gemini-2.5-Flash to transcribe Hindi AD audio files into timestamped, speaker-classified annotations. The prompt instructs the model to distinguish between narrator speech (audio descriptions) and actor speech (dialogs), producing structured 4-column output. The full prompt is shown in Fig. 9.

You are given a Hindi audio file that contains a mixture of narration and dialogues.

Your task is to generate structured annotations in the following 4-column format: [start_time] [end_time] [Speaker: Narrator or Actor] [Text]

Instructions:

1. Classify each line of speech as either spoken by a "Narrator" or an "Actor".
2. Maintain continuous timestamps from start to finish (timestamps should never reset).
3. Text must be the spoken line in Hindi, using Devanagari script.
4. Do not include speaker names such as "Farhaan", "Raaju", etc. Just use "Narrator" or "Actor".
5. Ensure each annotation has accurate start and end times in the format minutes+seconds+milliseconds (e.g., 02:00.010 to 02:00.050).
6. The output should follow the style shown below:[8]

```
01:59.98  02:02.66  Narrator:  Usse jhempte hue apni jeb se
phone nikala, number dekha.
02:03.11  02:04.14  Actor:  Hello?
```

Do not include any extra commentary, metadata, character names, or formatting other than what is asked above.

Fig. 9. Prompt used with Gemini-2.5-Flash for transcribing Hindi AD audio into timestamped, speaker-classified annotations. Hindi text is romanized for simplicity.

A.2 Machine-Generated Audio Description

Dense-to-Hindi Generation Prompt. For *Dense-to-Hindi* AD generation, we use the following system prompt with both Nemotron-Mini-Hindi-4B-Instruct and Gemini-3.1-Pro. The prompt instructs the model to convert English dense visual descriptions into natural, narrator-style Hindi ADs in Devanagari script. The full prompt is shown in Fig. 10.

You are a professional writer for film audio description (AD) in Hindi.

Your task is to convert an English dense visual description of a movie clip into natural, fluent, narrator-style Hindi audio description in Devanagari script.

Important requirements:

1. This is for FILM AUDIO DESCRIPTION, not subtitles or dialogue.
2. Preserve the visible meaning from the input and do not hallucinate details.
3. Keep the AD concise, natural, and suitable for spoken narration.
4. Focus on the most attractive / salient characters and their actions.
5. For characters, use their first names when available, and remove titles such as 'Mr.' and 'Dr.'.
6. If names are not available, use natural pronouns where possible; avoid generic phrases like 'a man' unless truly necessary.
7. For actions, avoid mentioning the camera, and do not focus on talking or position-related actions such as sitting and standing unless essential.
8. Do not mention characters' mood unless absolutely necessary from the visible description.
9. Prefer natural spoken Hindi, not stiff or overly literal Hindi.
10. If a term does not have a good Hindi equivalent in this film context, keep it as a natural English-origin word written in Devanagari.
11. Adjust output length according to the clip duration.
12. Return valid JSON only in exactly this format: {"summarised_AD": "<YOUR OUTPUT>"}

Fig. 10. System prompt used for *Dense-to-Hindi* AD generation with Nemotron and Gemini.

LLM-AD-Eval Prompt. We use the LLM-AD-Eval framework to evaluate the semantic match between predicted and ground-truth ADs. The system prompt, shown in Fig. 11, instructs the judge model to focus on visual elements and assign a score from 0 (worst) to 5 (best).

You are an intelligent chatbot designed for evaluating the quality of generative outputs for movie audio descriptions. Your task is to compare the predicted audio descriptions with the correct audio descriptions and determine its level of match, considering mainly the visual elements like actions, objects and interactions. Here's how you can accomplish the task:

Instructions:

- Check if the predicted audio description covers the main visual events from the movie, especially focusing on the verbs and nouns.
- Evaluate whether the predicted audio description includes specific details rather than just generic points. It should provide comprehensive information that is tied to specific elements of the video.
- Consider synonyms or paraphrases as valid matches. Consider pronouns like 'he' or 'she' as valid matches with character names. Consider different character names as valid matches.
- Provide a single evaluation score that reflects the level of match of the prediction, considering the visual elements like actions, objects and interactions.

Fig. 11. System prompt for LLM-AD-Eval, following the same evaluation setup as [52]. The LLM judge assigns a score from 0 (worst) to 5 (best) reflecting the semantic match between a predicted AD and the ground-truth AD.

A.3 Human-Authored Audio Description Analysis

Detail Overlap Prompt. To classify the detail overlap between *HI-Human* and *Human AD-to-Hindi* AD pairs, we use the prompt shown in Fig. 12.

CSI Detection and Examples. To identify Culture-Specific Items (CSIs) in English ADs, we use the prompt shown in Fig. 13. Examples of identified CSIs and their resolution strategies across human-authored and machine-translated Hindi ADs are provided in Fig. 14.

You are an expert bilingual annotator for Hindi movie audio description (AD) tracks. Audio descriptions narrate on-screen visual events for visually impaired viewers.

You will receive two inputs per row:

`HI_HUMAN` : human-authored ground-truth Hindi AD
`HUMAN_AD_TO_HINDI`: machine-translated Hindi AD from human-authored English AD

TASK: detail_overlap

Decide whether the two Hindi sentences describe the SAME visual details.

Focus ONLY on visual content: actions, objects, characters, spatial relations. Ignore style, level of detail, and transliteration variants — "*Taaylar*" and "*Taailar*" are the same character Tyler; "*Ejeebeeoh*" and "*Egbo*" are the same production house; "*vah*" standing in for a named character is not a mismatch.

Labels:

- `SAME_DETAILS` — same primary action/event even if one sentence is more detailed or uses synonyms/paraphrases.
- `PARTIAL_OVERLAP` — share some visual content but one sentence also describes a clearly distinct additional action absent from the other.
- `DIFFERENT_DETAILS` — completely unrelated details with no meaningful overlap in action, object, or character.

OUTPUT FORMAT

Return ONLY a JSON object with exactly these two keys — no extra text:

```
{
"detail_overlap":   "<SAME_DETAILS | PARTIAL_OVERLAP |
DIFFERENT_DETAILS>",
"overlap_reason":  "<one concise sentence>"
}
```

Fig. 12. Prompt used with Gemini-3.1-Pro for classifying detail overlap between *HI-Human* and *Human AD-to-Hindi* AD pairs.

You are an expert Audio Description (AD) translator and cultural consultant. We are translating English Audio Descriptions into Hindi for an Indian Blind and Low Vision (BLV) audience. We apply Skopos theory, meaning the translation's primary purpose is to be functionally effective and culturally accessible to the target audience, rather than a literal word-for-word translation.

Your task is to analyze a batch of English Audio Descriptions and determine if each contains any Culture-Specific Items (CSIs) that might not be easily understood by an Indian BLV audience.

Examples of CSIs between English and Hindi contexts:

- "He scored a touchdown in the final quarter." (American football reference)
- "They celebrated Thanksgiving with turkey and pumpkin pie." (American holiday and food)
- "He's applying to Ivy League colleges." (US education system)
- "They watched the Queen's Speech on Christmas Day." (British cultural event)
- "They had fish and chips by the seaside." (British food)

Here is the batch of English ADs to evaluate, provided as a JSON array of strings: {batch_json}

Evaluate each AD and return a JSON array of objects. Each object must have:

1. "ad_text": The exact English AD text from the input.
2. "is_csi": "yes" if the AD contains a cultural element likely unfamiliar to the target audience, or "no" if it is culturally neutral or globally understood.
3. "reason": A brief explanation of your decision.

Respond ONLY with a valid JSON array of objects.

Fig. 13. Prompt for detecting Culture-Specific Items (CSIs) in English to Hindi AD translation.

EN-Human	CSI Reason	Human AD-to-Hindi		HI-Human	
		AD	Resol.	AD	Resol.
They pass a steaming furnace on their left, then stop at a chute in the wall covered by a metal plate.	The word 'chute' is specific Western architectural terminology that will require careful functional translation for clarity.	वे अपनी बाईं ओर एक भाप वाली भट्टी से गुजरते हैं, फिर एक धातु की प्लेट से ढकी दीवार में एक च्यूट पर रुकते हैं।	no, retain	टाइलर ने अपनी गन तान रखी है।	no, omit
A goateed man descends the stairs.	A 'goatee' is a Western descriptive term for a specific beard style; requires localization (like French beard).	एक बकरी वाला आदमी सीढ़ियों से उतरता है।	no, direct	गोपी तल से एक सफेद दाढ़ी वाला आदमी आता है।	yes, general
Nick and Audrey push over a bookshelf, knocking successive bookshelves over like dominoes.	The 'domino effect' is an idiom based on the Western game of dominoes. The game is not historically widespread in India, so a descriptive translation might be more effective.	निक और ऑड्रे एक बुकशेल्फ़ को धक्का देते हैं, लगातार बुकशेल्फ़ को डोमिनोज़ की तरह धक्का देते हैं।	no, retain	स्टैंड हिलता है और अगले स्टैंड पर जाकर गिरता है। इस तरह पूरे स्टैंड्स की श्रृंखला ढह जाती है। ऑड्री और निक भागते हैं।	yes, specify
She opens the front door. A middle-aged black woman with cropped salt and pepper hair approaches, followed by a pair of white men.	The idiom 'salt and pepper hair' is a distinct English metaphor for graying hair. A literal translation would be confusing in Hindi, where expressions like 'khichdi baal' or 'safed aur kaale baal' are used instead.	वह सामने का दरवाजा खोलती है। कटे हुए नमक और काली मिर्च के बालों वाली एक अधेड़ उम्र की काली महिला आती है, जिसके बाद सफेद पुरुषों की एक जोड़ी आती है।	no, direct	वो मुख्य दरवाजा खोलती है। एक अधेड़, छोटे सफेद बालों वाली अश्वेत महिला और कुछ गोरे अनुचर उसकी ओर बढ़ते हैं।	yes, substitute
Shafts of light beam on a stone slab filled with hieroglyphics and the words Vincent Films.	Hieroglyphics refers to ancient Egyptian writing, often taught in Western curricula, requiring translation to 'chitralipi' or ancient scripts in Hindi.	चित्रलिपि और विंसेंट फिल्म्स शब्दों से भरे एक पत्थर के स्लैब पर प्रकाश किरण के शाफ्ट।	yes, direct	प्राचीन शिलालेख पर चमकती धूल उड़ती है। स्वर्णिम पत्तियों से बनी आकृति उभरती है।	yes, general
A hologram of a blimp-like craft materializes in the large, dark space. Jack stares open mouth.	Blimps or Zeppelins are not commonly seen or referenced in India. Explaining its shape requires describing it as a large airship or balloon.	एक झपकी जैसी शिल्प का एक होलोग्राम बड़े, अंधेरे स्थान में होता है। जैक मुँह खोलकर देखता है।	no, direct	एक अंडाकार आकृति का होलोग्राम हवा में उभरता है। जैक उसे खुले मुंह निहारता है।	yes, substitute

Fig. 14. Examples of Culture-specific Items and their resolutions. Each row presents the source English AD, the identified CSI with explanation, the machine-translated Hindi AD from English AD, and the human-authored Hindi AD, along with resolution outcomes and strategies. The figure shows how different strategies are applied in practice to either resolve CSIs or leave them unresolved.