Evaluating ASR Systems Accuracy Using Word Error Rate Across Indian and Standard US Dialects

Oishani Bandopadhyay, supervised by Professor Will Styler

1. Abstract

Automatic Speech Recognition (ASR) systems are increasingly integrated into everyday technology, yet disparities in transcription accuracy for speakers of non-standard English dialects are apparent to users. This study investigates the extent of these disparities by evaluating the transcription error rates of three major ASR systems - Google Cloud Speech-to-Text, OpenAI's Whisper, and Microsoft Azure Cognitive Services - on speech from Indian English dialect speakers and Standard US English speakers. Audio data were collected from participants with diverse linguistic backgrounds using prompts of varying complexity. Word error rates were analyzed across dialect groups and models. Statistical tests revealed significantly higher error rates for Indian speakers across models, with Whisper exhibiting the greatest disparity and Azure the least. Linear mixed models showed that specific categories of prompts were significant predictors of elevated error rates, even when controlling for prompt type and speaker-level variation. Prompts including Indian food names, place names, or phrases common to Indian English lexicon but uncommon outside it had significantly higher WER rates. These findings highlight the limitations of current ASR systems in accommodating dialectal variation and underscore the need for more inclusive speech technology.

2. Introduction

While ASR has improved greatly for speakers of Standard US English, there remains a disparity in ASR performance across non-standard dialects of English, such as speakers with an Indian accent or dialect.

English is fairly widely spoken in India, being the primary language of instruction in a couple of large centralized academic education board exams. Many people grow up speaking English, making them L1 speakers of the language rather than L2 speakers, or begin schoolwork in English from a young age (4-6 years old). Indian English has distinct features in vowel quality, syllable timing, and segmental pronunciation, probably influenced by British English and features of regional languages such as Hindi, Tamil, Bengali, and Telugu, among others. These differences are hypothesized to result in a higher likelihood of transcription errors when Indian English speech is processed by ASR systems trained predominantly on Standardized US English data. Our hypothesis states that Indian English speakers will have significantly higher WERs (Word Error Rates) in their ASR transcriptions, with some influence of prompt type and language background. Prompt types are categorized based on kinds of expected variation across dialects, with some standard sentences, some improbable sentences, and some sentences with words from the Indian English lexicon and phrasing common to Indian English.

Prior work has demonstrated differences in WERs across Indian and Standard US dialects, with transcription errors consistently higher for Indian dialects. WERs are calculated as the percentages of error words over all words in a set, such as a sentence. By breaking down these errors into specific prompt categories designed to capture more variation in errors, capturing language backgrounds of speakers, and testing across Google, Whisper, and Azure models, this paper aims to capture the nuanced difference in error rates by these models pertaining to Indian and Non-Indian dialects.

3. Background

Standardized US English and Indian English differ substantially in phonetic features such as vowel quality, and vowel length. Indian English, for example, exhibits unique patterns in these domains, influenced by speakers' regional languages and sociolinguistic backgrounds (Dodd et al., 2023). These phonetic differences cause a wide gap between US and Indian English, making accurate transcriptions difficult for ASR systems primarily trained on US English data.

Recent research with ASR evaluations using diverse datasets show consistently higher WERs for Indian English, with the accuracy difference attributed to accent variation and a lack of Indian English training data (Javed et al., 2023). Open-source and commercial ASR models show a substantial increase in WER when tested on Indian-accented speech compared to Standardized US English benchmarks (Javed et al., 2023).

Efforts to address these disparities include Indian English-specific pronunciation lexicons and phoneme dictionaries, which reduce WER for Indian English speakers (Jain et al, 2021). This improves the mapping of sound sequences to original words said in Indian dialects. For example, using an Indian English Common Phone Set (IE-CPS) instead of a standard US English lexicon reduced the WER by 3.95% (Jain et al., 2021).

Building on this existing research, the present study investigates the impact of dialectal variation on ASR error rates by directly comparing the performance of three speech-to-text systems: Google Cloud Speech-to-Text, OpenAI's Whisper, and Microsoft Azure Cognitive Services, on speech from speakers of Indian and US English dialects. Using this comparative analysis, the same set of prompts given to all speakers allows us to evaluate differences between groups and account for various differences in WER. Our speech dataset contains variation in terms of speaker ages, gender, language background, naturalness of speaking, and prompt types that we expect different accuracies for. By accounting for speaker language background and varying prompt complexity, we aim to more precisely quantify the extent and sources of error rates across dialects and models, and to identify which linguistic backgrounds are most predictive of ASR performance disparities in this paper. Our creation of a unique dataset of prompts and collection of relatively natural speech data in various conditions through speakers recording themselves results in a comprehensive evaluation.

4. Methods

4.1 Subject Demographics

Subjects were individuals with a language background from India or individuals with a Southern California English accent. Participants varied in age and were grouped based on their language backgrounds and dialects. Multiple participants were recruited for each dialect tested: Tamil, Telugu, Bengali, Hindi, and Southern California English.

4.2 Experimental Design

Each subject completed an online questionnaire involving audio recordings hosted on the online FindingFive platform. Participants read and agreed to a consent form, adapted from IRB-approved form #131094 (UCSD Phonetics Lab Audio Recording Consent Form), and modified for online administration. We submitted our own protocol to the IRB: #810448 Testing Voice Assistants' Failure Rates with Different Dialects and Accents (Indian Accents in Particular) after designing the experiment.

Participants were asked to read a series of sentences aloud, each of which served as a prompt for an ASR system. Prompts were recorded individually.
Initial instructions asked participants to speak naturally as if speaking to someone with a similar language background, prompting natural speech. By allowing participants to record themselves by clicking a button, we gave them the opportunity to read the prompt and start recording when they were ready, improving the quality of recordings especially for sentences with unfamiliar words or incorrect grammatical structures. They could also re-record themselves if they felt they had made a mistake.

4.3 Language Background

Participants also completed demographic questions, including checkboxes for languages they spoke fluently and their place of origin at the state level (e.g., California, Karnataka, Tamil Nadu, West Bengal, Andhra Pradesh, Telangana, Madhya Pradesh, etc.). From the language background questions, we focused on Question 3: ‘Which languages do you think have an influence on your accent or dialect?’ for our analysis. This simplified our process of quantifying language background and dividing speakers into Indian dialects and non-Indian dialects. Speakers with Non-Indian dialects were generally US English speakers, so this question was effective to distinguish between speakers who may be regionally situated in the US but have influences of an Indian accent or dialect.

Error variation across groups was calculated using this question encoded as a binary variable with Indian and non-Indian dialects. Subsequently, average error rates were calculated by each speaker, aggregated across all of the prompts they said. An approach accounting for all the language background questions together was significantly more complex, since variation among speaker language backgrounds was quite high across the other questions. Since several speakers mentioned multiple languages in various parts of their responses, isolating the effect of a singular language on their dialect was difficult. However, responses to these other language background questions were also stored and could be analyzed further.

4.3 Prompt Design

Speech prompts read by participants were divided into four main categories:

1. Generic commands with relatively low complexity (ExpPrompt): These were predictable prompts expected to yield high transcription accuracy, or low WER. Some included moderate complexity, such as multi-word contractions (“She wouldn’t’ve known how it ain’t all rosy”), brand names with nonstandard spellings (“In-n-Out”), or unexpected sentence structures.

2. Difficult commands with noticeable dialectal differences (DialPrompt): These prompts included phrases likely to elicit pronunciation patterns that varied by dialect. For example, vowel rounding by Bengali speakers, where front vowels such as /a/ are pronounced with more rounding, closer to /ɑ/ or /ɒ/, could create lexical ambiguities that complicated accurate transcription. The distinction between /w/ and /v/ that is often less pronounced in Indian English was addressed, as some speakers use them interchangeably. The alveolar /t/ instead of the alveolar tap /ɾ/ in words such as ‘water’ (often /wɒːtəɹ/ in Indian English and /wɑːɾəɹ/ in Standard US dialects) was included. Voicing for labials can also be different due to lenition in a more pronounced way for Indian English speakers, such as ‘p’ sounding more like ‘b’ for some speakers.

3. Generally difficult prompts (HardPrompt): These sentences had inherently low predictability due to randomized or less semantically coherent word combinations. Some other complexities in the prompts included unpredictable final words and niche vocabulary. High lexical ambiguity or syntactic irregularity would make them difficult to transcribe for models regardless of dialect.

4. Region-specific words (LocPrompt): These included names of Indian cities, foods (e.g., “biryani”), and English words frequently used in the Indian English lexicon. Phrases that are colloquially used in Indian English dialects were also included here. Another example included city names, such as small town names in the US and small town names in India, as well as bigger city names in both, and city names that have a colonial naming convention and an original Indian name. These prompts were designed to test how well speech recognition systems handled local and culturally specific pronunciations.

Participants read each prompt aloud, generating audio files that were subsequently played back to cloud-based speech-to-text systems. Each system’s transcription was compared to the original target sentence. Each FindingFive session’s data was stored with the audio files of every prompt read out by every participant, and their demographic and language background data.

4.4 Data Processing

4.4.1 Transcription Pipeline

After we downloaded the raw audio data, we prepared the files for transcription by converting them to a speech-to-text API-compatible format. The source audio files, provided in either ogg or mp4 formats depending on the version of FindingFive used at the time, were converted to wav with the ffmpeg library through a Python script. This was applied to all the files of the input directory. To ensure compatibility across APIs, stereo audio recorded from earphones was converted to mono, which is a requirement for some transcription services.

We used three transcription services: Google Cloud Speech-to-Text, OpenAI’s Whisper, and Microsoft Azure Cognitive Services. Each ASR service was called in a function that returned the transcript as a string. We used the Google Speech client with linear PCM and US English as the language, passing audio from speech.RecognitionAudio to the API call. We used OpenAI’s Whisper ‘base’ model with the built-in transcribe() function, choosing this model for less compute load after some sample comparisons with ‘large’. For Microsoft Azure, we used the latest endpoint of Cognitive Services v1 for speech recognition, using the audio file content as a data parameter. Error handling with a try/except block caught empty audio files. Each of the transcript strings were cleaned and stored as a txt file in the same directory as the original audio. The Python pipeline was optimized to check for existing transcriptions by filename before calling a transcription function. Results from each transcription were also written as a new row into a CSV file for analysis. Participant metadata was later merged into this CSV using Pandas.

4.4.2 WER Measurement

We applied a word-level error counting function to assess transcription quality. This function involved turning the prompts and transcriptions into lists and comparing strings within the lists to account for insertions, deletions, and substitutions to some extent (adding an error for any words that do not overlap correctly across both lists). We compared every model's transcript with the original prompt sentence using this method. We tokenized and stripped punctuation from both sentences prior to converting them to lowercase. Lists were then made to be equal in length for proper alignment and indexing. Words in the prompt but not in the transcript, and vice versa, were counted. Total error count was the maximum number of words unmatched in either direction. This approach tallied substitutions, deletions, and insertions within one measure. The function yielded a list of missing words on each side as well as the total word-level error count.

4.4.3 Language Background Collection

Participant data were exported from session result files containing answers to the language background questionnaire. Responses to our specific question of interest (LangBgQ3 - Which languages do you think have an influence on your accent or dialect?) were extracted to encode column-wise. This made the encoding of language background for each speaker more consistent and precise, chosen as the best representation of language background from our questionnaire.

A binary one-hot encoding model was used to encode every subject's language origin, with markers present or not for languages such as Hindi, Bengali, Tamil, and Telugu. This accounted for multilingual speakers, so a speaker could have 1's in multiple language columns, or 0's in multiple language columns if they do not speak any non-English languages. The binary column noted "English-only" where subjects indicated English only without any additional language.

4.4.4 Merging Language Background and Transcriptions

Transcription outputs were matched with participant IDs truncated for anonymity and consistency, and original prompts. The word error counter was applied on this merged dataset, facilitating per-sentence and per-speaker analysis. An additional column was added to denote if a participant had an Indian dialect (Hindi, Bengali, Tamil, or Telugu). This final merged dataset as a CSV was used for group-level statistical analysis and visualization in subsequent sections.

5. Analysis

Analyzing this dataset, we saw that transcriptions of Indian dialect speakers had higher median error rates, with Whisper having the highest error rates overall, and even higher rates for Indian dialect speakers. Azure had the lowest error rates overall, and a smaller difference across Indian and Non-Indian dialect speakers, but there remained a visible gap, as demonstrated in Figure 1.

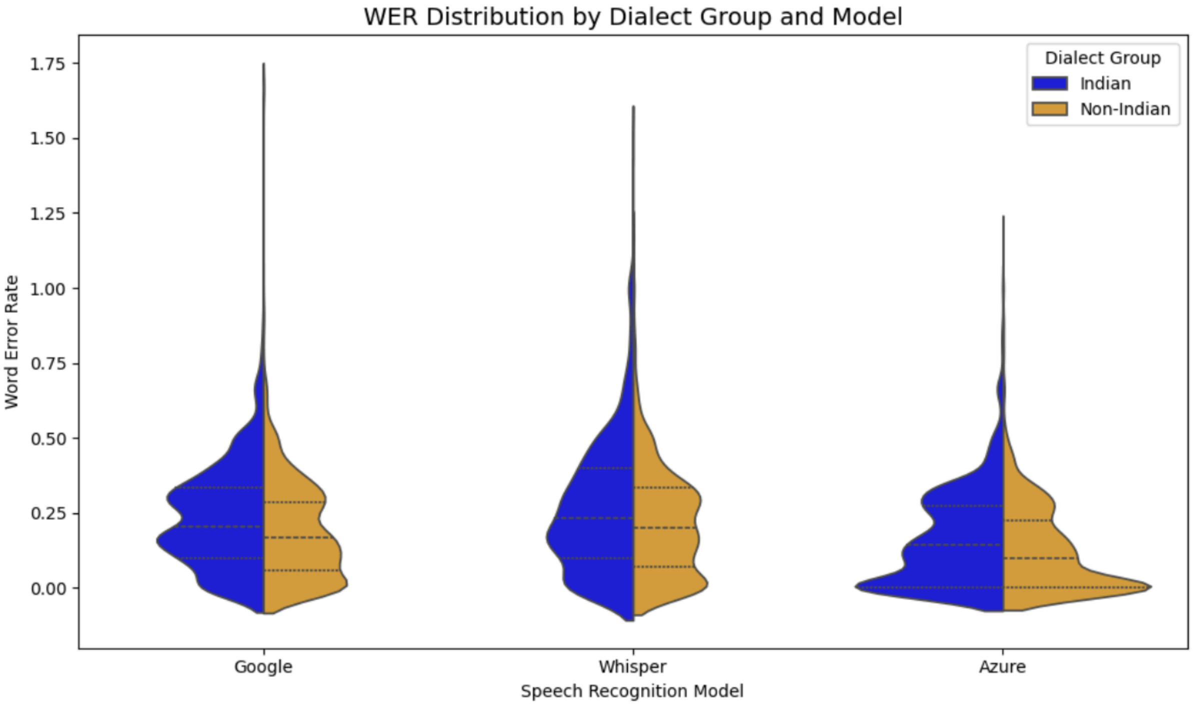

Figure 1. Distribution of speaker-level error rates across models grouped by dialect

Using a split violin plot in Figure 2, we looked further into the word error rates broken down across the models and for each prompt type.

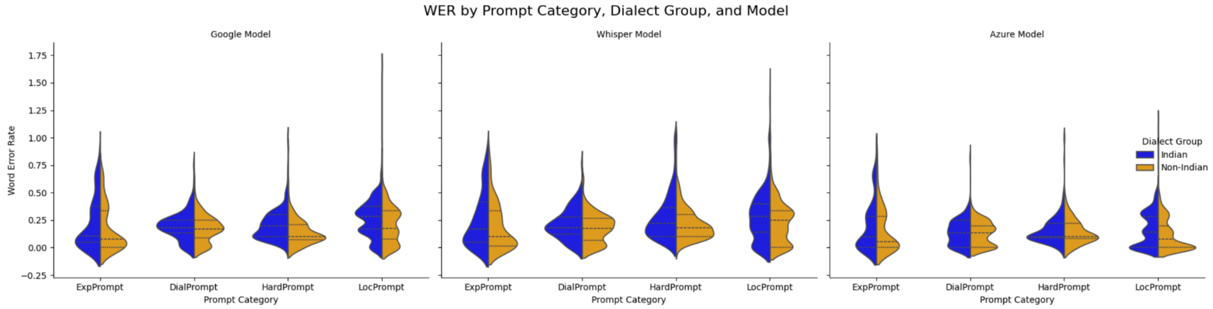

Figure 2. Violin plots of speaker error rates across prompt categories and dialect groups for each model, with speaker-level word error rates for Indian and Non-Indian speakers

From Figure 2, we can see that transcription error rates for Indian dialect speakers are generally higher than Non-Indian speakers. This gap is visible across all the models and the prompt types as well. The prompt types do have an effect, with LocPrompt (region-specific words, Indian English lexicon) generally having wider variability across speakers in terms of error rates, both for Indian and Non-Indian dialects. Some differences across prompt types are more visible for certain models, such as ExpPrompt (simplest prompts) having more variability in error rates in

Azure and Google. Overall, we see Whisper having the highest error rates, followed by Google, with Azure having noticeably lower error rates, and a smaller gap between dialects.

The general trends across prompt types are clearly seen in Figure 3, which uses a strip plot to show each speaker's error rate. This visualisation shows variation within groups, and central tendencies for models. The gap between dialects remains clearly visible, with Indian dialect error rates consistently higher, and Whisper having the highest gap between them, with the highest overall errors as well. Azure has the lowest overall error rate, and also tighter group clustering.

```
<Figure size 1400x600 with 0 Axes>
```

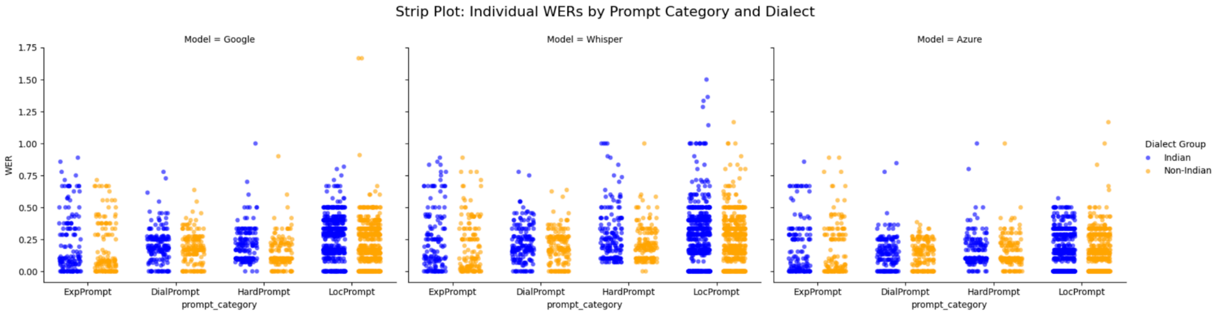

Figure 3. Speaker-level transcription word error rates by model and dialect groups

6. Results

We used a linear mixed model (LMM) to evaluate the effects of dialect group, ASR model, and prompt category on word error rate (WER) as the outcome variable, keeping a random intercept for each speaker (participant ID).

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.145	0.015	9.695	0	0.116	0.175
Dialect[T.Non-Indian]	-0.015	0.021	-0.718	0.473	-0.057	0.027
Model[T.Google]	0.057	0.018	3.181	0.001	0.022	0.092
Model[T.Whisper]	0.072	0.018	4.052	0	0.037	0.107
prompt_category[T.ExpPrompt]	0.046	0.019	2.465	0.014	0.009	0.083
prompt_category[T.HardPrompt]	0.016	0.018	0.864	0.387	-0.02	0.051
prompt_category[T.LocPrompt]	0.019	0.014	1.314	0.189	-0.009	0.047
Dialect[T.Non-Indian]:Model[T.Google]	-0.011	0.026	-0.448	0.654	-0.061	0.039
Dialect[T.Non-Indian]:Model[T.Whisper]	-0.023	0.026	-0.884	0.376	-0.073	0.027
Dialect[T.Non-Indian]:prompt_category[T.ExpPrompt]	-0.008	0.027	-0.313	0.755	-0.061	0.044
Dialect[T.Non-Indian]:prompt_category[T.H	0.002	0.026	0.09	0.928	-0.049	0.053

ardPrompt]						
Dialect[T.Non-Indian]:prompt_category[T.LocPrompt]	-0.031	0.021	-1.529	0.126	-0.072	0.009
Model[T.Google]:prompt_category[T.ExpPrompt]	-0.024	0.026	-0.915	0.36	-0.076	0.028
Model[T.Whisper]:prompt_category[T.ExpPrompt]	-0.016	0.026	-0.614	0.539	-0.068	0.036
Model[T.Google]:prompt_category[T.HardPrompt]	-0.012	0.026	-0.468	0.64	-0.063	0.038
Model[T.Whisper]:prompt_category[T.HardPrompt]	0.049	0.026	1.883	0.06	-0.002	0.099
Model[T.Google]:prompt_category[T.LocPrompt]	0.044	0.02	2.177	0.03	0.004	0.084
Model[T.Whisper]:prompt_category[T.LocPrompt]	0.046	0.02	2.233	0.026	0.006	0.086
Dialect[T.Non-Indian]:Model[T.Google]:prompt_category[T.ExpPrompt]	-0.015	0.038	-0.403	0.687	-0.089	0.059
Dialect[T.Non-Indian]:Model[T.Whisper]:prompt_category[T.ExpPrompt]	-0.006	0.038	-0.154	0.878	-0.08	0.068
Dialect[T.Non-Indian]:Model[T.Google]:prompt_category[T.HardPrompt]	-0.039	0.037	-1.045	0.296	-0.111	0.034
Dialect[T.Non-Indian]:Model[T.Whisper]:prompt_category[T.HardPrompt]	-0.033	0.037	-0.886	0.376	-0.105	0.04
Dialect[T.Non-Indian]:Model[T.Google]:prompt_category[T.LocPrompt]	0.003	0.029	0.115	0.909	-0.054	0.06
Dialect[T.Non-Indian]:Model[T.Whisper]:prompt_category[T.LocPrompt]	0.02	0.029	0.694	0.488	-0.037	0.077
Group Var	0.001	0.002				

Figure 4. Mixed Linear Model Regression Results Across Models

The model revealed significant main effects for model type and prompt category. Specifically, compared to Azure, both Whisper (β = 0.072, p < .001) and Google (β = 0.057, p = .001) exhibited significantly higher WERs, showing poorer performance. ExpPrompt (simplest prompts) yielded significantly higher WERs (β = 0.046, p = .014), while LocPrompt (region-specific words, Indian English lexicon) also showed higher error rates (β = 0.019, p = .189), though the effects were not statistically significant. HardPrompt (generally challenging across dialects) did not differ significantly from the reference prompt category, DialPrompt (prompts to capture dialectal variation).

However, the main effect of dialect group was not significant (p = .473), suggesting no consistent WER difference between Indian and non-Indian speakers across all models and prompt types. The two- or three-way interaction terms involving dialect did not reach significance, which leads us to think that the relative model and prompt performances were stable across dialect groups.

To further probe dialectal effects, we fit a separate LMM focusing exclusively on each of the models. Beginning with the Google transcription error data, dialect group and prompt type were kept as fixed effects and participant ID was a random intercept. The model revealed a significant effect of Indian dialect status (β = 0.050, p < .001), indicating higher word error rates for Indian English speakers. Among the prompts, the LocPrompt category was also significantly associated with increased error rates (β = 0.049, p < .001). No significant effects were observed for the ExpPrompt or HardPrompt categories.

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.163	0.012	13.808	0	0.14	0.186
prompt_category[T.ExpPrompt]	0.01	0.013	0.824	0.41	-0.014	0.035
prompt_category[T.HardPrompt]	-0.014	0.012	-1.111	0.267	-0.038	0.011
prompt_category[T.LocPrompt]	0.049	0.01	5.023	0	0.03	0.069
is_indian_dialect	0.05	0.012	4.022	0	0.025	0.074
Group Var	0.001	0.002				

Figure 5. Mixed Linear Model Regression for Google Transcription Results

Similarly, an LMM was fit to the Whisper transcription error data. The model identified a significant effect of dialect group, with Indian English speakers exhibiting higher word error rates than non-Indian speakers (β = 0.052, p < .01). As with the Google model, the LocPrompt category also showed a significant positive effect (β = 0.059, p < .001), suggesting increased transcription difficulty for the local Indian English lexicon word prompts. The HardPrompt category also had a significant positive effect (β = 0.049, p = .001), pointing towards the model struggling with improbable prompts. No significant effect was observed for ExpPrompt.

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.173	0.016	10.652	0	0.141	0.205
prompt_category[T.ExpPrompt]	0.023	0.015	1.492	0.136	-0.007	0.053
prompt_category[T.HardPrompt]	0.049	0.015	3.313	0.001	0.02	0.078

prompt_category[T.LocPrompt]	0.059	0.012	5.028	0	0.036	0.082
is_indian_dialect	0.052	0.018	2.788	0.005	0.015	0.088
Group Var	0.002	0.004				

Figure 6. Mixed Linear Model Regression for Whisper Transcription Results

The Azure model's LMM results showed a significant effect of Indian dialect status on word error rate (β = 0.032, p = .001), demonstrating dialect-based performance disparities. Interestingly, ExpPrompt was a significant predictor of higher error rates (β = 0.042, p < .001). No other prompt categories showed significant associations.

	Coef.	**Std.Err.**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	0.121	0.01	11.954	0	0.101	0.141
prompt_category[T.ExpPrompt]	0.042	0.012	3.594	0	0.019	0.065
prompt_category[T.HardPrompt]	0.017	0.011	1.499	0.134	-0.005	0.04
prompt_category[T.LocPrompt]	0.003	0.009	0.348	0.728	-0.015	0.021
is_indian_dialect	0.032	0.01	3.198	0.001	0.012	0.051
Group Var	0	0.001				

Figure 6. Mixed Linear Model Regression for Azure Transcription Results

This consistency across models shows that the accuracy using WER for Indian English dialects compared to Standardized US English is generally lower, with more errors, especially for certain types of prompts. The significance of prompt type demonstrates that there continue to be certain gaps in ASR in terms of words, names, and places that are from India. While the main LMM did not detect overall dialect-based disparities across all models, model-specific analyses did reveal meaningful dialect effects.

7. Discussion

Generally, accuracy for Non-Indian dialects seems to be higher, although as seen in the overall LMM, when accounting for other factors such as language background and prompt type, they are not statistically significant. This could be due to the non-independence in the predictor variables, making it difficult to appropriately predict the effects of Indian dialects independent of individual language backgrounds. Trying to isolate each language as a variable for language background and compare them led to multicollinearity issues. Therefore, we tried to isolate other factors and maintained the one-hot binary encoding of speakers having an Indian dialect or not.

The gap between the performances of the models themselves is likely due to training data and an emphasis on Indian data for a wide base of Indian users. Whisper and Google ASR's poor performances with the LocPrompt category using region-specific words and the Indian English lexicon is probably due to the absence of wider vocabulary including Indian geography, food, and place names. Azure's better performance might be the result of Microsoft's work in expanding into India by adding training data in local language words and Indian English speech.

8. Conclusion

This study reveals some disparities in ASR accuracy across dialect groups, with higher WERs for Indian English speech than Standard US English. While Whisper and Google both had substantial gaps in performance across dialects, Azure had relatively lower overall error rates, and a smaller gap between Indian and Non-Indian speakers. Statistical testing confirmed these differences as significant.

The LMM results reflect certain prompts being significant predictors for errors, but due to the interactions between models and language backgrounds, the independent effect of speaker dialect on word error rate seems to be non-significant in most cases. These findings suggest that while dialectal disparities are real and measurable, they may depend on language-specific phonological features and are specific to models and language backgrounds of speakers.

The results and previous work in the area emphasize the need for more representative training data and dialect-aware modeling in ASR systems. Differences in dialect, especially words and phrasing common to the Indian English lexicon, should be accounted for when training these ASR systems. The gaps in accuracy between models and prompt types shows that there can be several improvements made in terms of expanding training data. As speech-based interfaces continue to grow in popularity and utility, accounting for dialectal bias could improve equitable access and usability of ASR systems for global English speakers.

9. Future Directions

This research could be expanded in several ways to investigate WER variations in ASR across phonetic differences in specific prompts, and more sociolinguistic factors from the language background questionnaire. By using this experimental design to collect more speech data, more representative samples of specific language speakers would be helpful to evaluate nuanced effects. The collected speech data could be analyzed phonetically using computational tools, and perceptual differences to prompt types, speech from different dialects, and between speakers with varying language backgrounds could also lead to further analysis. Finally, expanding beyond WER to account for CER (character error rates) and phoneme-level errors would be useful for a more detailed and thorough discussion.

10. References

Dodd, N., Cohn, M., & Zellou, G. (2023). Comparing alignment toward American, British, and Indian English text-to-speech (TTS) voices: influence of social attitudes and talker guise. Frontiers in Computer Science, 5. https://doi.org/10.3389/fcomp.2023.1204211

Javed, T., Joshi, S., Nagarajan, V., Sundaresan, S., Nawale, J., Raman, A., Bhogale, K., Kumar, P., & Khapra, M. M. (2023). Svarah: Evaluating English ASR systems on Indian Accents. Interspeech 2022, 5087–5091. https://doi.org/10.21437/interspeech.2023-2588

Shelly Jain, Aditya Yadavalli, Ganesh Mirishkar, Chiranjeevi Yarra, and Anil Kumar Vuppala. (2021). IE-CPS Lexicon: An Automatic Speech Recognition Oriented Indian-English Pronunciation Dictionary. In Proceedings of the 18th International Conference on Natural Language Processing (ICON), pages 195–204, National Institute of Technology Silchar, Silchar, India. NLP Association of India (NLPAI).

11. Appendix

Relevant code, stimuli, and questionnaire are also available at the following link:
github.com/oishani-b/asr_indian_dialects
Language Background Questions:

1. Which languages, if any, do you speak besides English?
2. Which languages were spoken around you where you grew up?
3. Which languages do you think have an influence on your accent or dialect?
4. If you have been told you have an accent associated with a region, select the region:
5. If you have been surrounded by this language/these languages for extended periods of time, select the language/s
6. Where did you grow up or spend a lot of time in your life? Feel free to list places if you have mutliple. You can use City, State, Country to format the place/places.
7. If you selected Other, please name the other language/languages, otherwise you may leave this blank:

Options for Questions 1, 2, 3, 5:

1. Bengali
2. Hindi
3. Kannada
4. Tamil
5. Telugu
6. None besides English
7. Other (you may specify in the next slide)

Options for Question 4:

1. North India
2. South India
3. East/Northeast India
4. West/Northwest India
5. None of these
6. Other (you may specify in the next slide)

Questions 6, 7 are open-ended and user can optionally input a text response.

Prompts that participants will be asked to read:

1. Finding the derivative of x-dash y cos z should not be this difficult.
2. Take it alright where it belongs you decide.
3. Bro sure with all my work I've made it to the brochure.
4. He got to the plot to plant bland bread plenty late and had to ladle the blend prior.
5. Henry heard raging hot heads rockily hover round his risky head, hoping regulation homes remain.
6. Bro, I don't even know what she said dude she's going on and on as if I did something so bad bro
7. She gives it a one of ten when often it deserves more.
8. Did you order seventy parts apart from the prayer part?
9. I don't know why you'd go fart then that forgetful one wouldn't hear a word
10. You'll adhere to going where mars and venus were visibly singing.
11. They abhor the abolition absconders that speak of aborting arbitrary overlooking.
12. Why would veins have caught on to the caroling war jobholders.
13. Wrap the wrap in aluminium foil wrap and tell that fellow where to go da.
14. Turn on the living room, bedroom and kitchen lights
15. The beige hue on the waters of the loch impressed all, including the French queen, before she heard that symphony again, just as young Arthur wanted
16. She wouldn't've known how it ain't all rosy
17. There aren't enough people with S-tier rizz
18. Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.
19. You haven't even been to the In-n-Out in the Outback Steakhouse neighbourhood?
20. It's not all fun-n-games with y'all subbing for 'em
21. The complex houses married and single soldiers and their families

22. Histories make men wise; poets, witty; the mathematics, subtle; natural philosophy, deep; moral, grave; logic and rhetoric, able to contend
23. The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak. They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other. Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him; and at last the North Wind gave up the attempt. Then the Sun shined out warmly, and immediately the traveler took off his cloak. And so the North Wind was obliged to confess that the Sun was the stronger of the two.
24. Could you pare a pair of pears for me?
25. Everyone expected instant metanoia when Meta launched and met a mere man to mete out meters of metamorphosing moves.
26. He took a wok from the chinese restaurant at the end of his walk.
27. Rose rose to put rose roes on her rows of roses.
28. She flew out to feel flu-flouting flounder few knew to flu flout.
29. She came back from the sweltering heat of the dessert.
30. soldiers married and single their the houses complex families and
31. Every night he can't help but see iridescent laser beams ephemerally shining.
32. Though it's a bastion of wonder, the bougie bungalow uses effete schemes.
33. population windy trout head birthday brim caught exalting grudge bundt
34. Her dress was made of the most beautiful, delicate laze.
35. What's the best papdi chat to have with friends in Mysore Palace?
36. That was a dabba movie, chumma you said it's amazing n something n all
37. They knew where Snigdha would be on a Friday evening.
38. He is leaving for Srirangam tomorrow.
39. I enjoy eating luchi and aloor dum.
40. Guru, help me out boss, I'm running out of money
41. She is going to see Charminar tomorrow.
42. They knew where Rituparno would be on a Friday evening.
43. I enjoy eating poori with chana masala.
44. What da, why you keep laughing?
45. They knew where Nagaraju would be on a Friday evening.
46. She ate up the naan, confrontation awaited her
47. She is leaving for Gorakhpur tomorrow.
48. I enjoy eating roti with daal tadka.
49. They knew where Anirudh would be on a Friday evening.
50. Those two were dancing full-on at the wedding.
51. He is leaving for Malleshwaram tomorrow.
52. I enjoy eating maddur vada and chutney.
53. They knew where Saathvik would be on a Friday evening.

54. Macha what are you doing, come off to the movie on Friday no
55. He is leaving for Kolkata tomorrow.
56. She is leaving for Calcutta tomorrow.
57. They knew where Manjunatha would be on a Friday evening.
58. He is leaving for Bengaluru tomorrow.
59. She is leaving for Bangalore tomorrow.
60. He is leaving for Bombay tomorrow.
61. She is leaving for Mumbai tomorrow.
62. He is leaving for New York tomorrow.
63. She is leaving for Los Angeles tomorrow.
64. He is leaving for Chicago tomorrow.
65. She is leaving for Houston tomorrow.
66. He is leaving for Temecula tomorrow.
67. She is leaving for Savannah tomorrow.
68. I ate the kolumb and naan described as San Diego's best
69. He is leaving for Newton tomorrow.
70. She is leaving for Brentwood tomorrow.
71. He is leaving through Golden Gate bridge tomorrow.
72. Lakshmi and Palash have been sceneing so much, they're definitely dating
73. They knew where Uma would be on a Friday evening.
74. I enjoy eating sambar with dosa.
75. They knew where Shweta would be on a Friday evening.
76. She is leaving through Howrah Bridge tomorrow.