


Bodhi


Speech Recognition Benchmarking Report 2024


Executive Summary

The aim of this report was to evaluate the speech recognition offerings, particularly those in streaming, provided by Navana.ai's Bodhi, Google, Whisper, and Deepgram. We conducted benchmarking using the [FLEURS dataset](#) released by Google. The report provides insights into accuracy and latency benchmarks, showcasing top-of-the-market performance in both aspects.

Core advantages of Bodhi

 **Highly customised models** that perform well on downstream tasks.

 Ability to **fine tune** based on regular **custom vocabulary** additions.

 Bodhi API supports coverage of **11+ Indic languages** and **40+ dialects**.

Available features & languages

Languages	Navana.ai Bodhi	Google	Deepgram	Whisper
Customised speech models	✓	✗	✓	✗
Supports real-time streaming recognition	✓	✓	✓	✗
Real-time streaming latency	450 - 550 ms	1000+ ms	< 300 ms	N/A
Fine-tuning with custom vocabulary	✓	✓	✓	✗
Supported Indic languages	Supports 10+ Indic languages covering 40+ dialects	Supports 10 Indic languages	Supports 1 Indic language: Hindi	Supports 5 Indic languages

Accuracy Benchmarking

Accuracy benchmark test performed on FLEURS test with CER (Character Error Rate) and WER (Word Error Rate) as the evaluation metric without any normalisation.

FLEURS - CER

Languages	Navana.ai Bodhi	Google V2 (Chirp/USM)	Deepgram	Whisper
Hindi	9.9	9.2	22.3	13.2
Kannada	11.8	9.8	N/A	28.1
Bengali	12.9	N/A	N/A	N/A
Marathi	15.1	21.0	N/A	27.0
Telugu	14.5	10.7	N/A	N/A
Tamil	19.0	23.4	N/A	18.0
Average	14.7	14.8	22.3	21.6

NON STREAMING SPEECH RECOGNITION

Languages	Navana.ai Bodhi	Google V1 (Enhanced)	Deepgram	Whisper
Hindi	10.4	9.3	25.3	N/A
Kannada	12.2	14.3	N/A	N/A
Bengali	13.7	8.9	N/A	N/A
Marathi	15.4	11.7	N/A	N/A
Telugu	15.3	18.5	N/A	N/A
Tamil	19.2	15.8	N/A	N/A
Average	14.4	13.1	25.3	N/A

STREAMING SPEECH RECOGNITION

FLEURS - WER

Languages	Navana.ai Bodhi	Google V2 (Chirp/USM)	Deepgram	Whisper Large V2	Whisper Large V2 on paper (normalised)
Hindi	20.6	17.6	34.1	35.8	21.9
Kannada	38.3	27.2	N/A	75.4	37.0
Bengali	37.3	N/A	N/A	N/A	104.1
Marathi	37.9	34.4	N/A	86.2	38.3
Telugu	37.3	27.9	N/A	N/A	99.0
Tamil	45.0	41.7	N/A	59.8	17.5
Average	36.1	29.8	34.1	64.3	53.0

NON STREAMING SPEECH RECOGNITION

Languages	Navana.ai Bodhi	Google V1 (Enhanced)	Deepgram	Whisper Large V2	Whisper Large V2 on paper (normalised)
Hindi	21.2	19.9	36.4	N/A	N/A
Kannada	38.6	35.8	N/A	N/A	N/A
Bengali	37.8	28.3	N/A	N/A	N/A
Marathi	38.2	33.7	N/A	N/A	N/A
Telugu	37.7	38.8	N/A	N/A	N/A
Tamil	44.6	40.6	N/A	N/A	N/A
Average	36.4	33	36.4	N/A	N/A

STREAMING SPEECH RECOGNITION