

LAB NOTES · 23 OCTOBER 2026

Speech at home

Ten speech recognizers on a base M5 MacBook, scored on 89 hours of open audio in 10 languages. Two more finish tomorrow.

Results so far, one machine, one run per model · Sources on the last slide

Space, Page Down or arrow keys to move. Or use the links at the bottom.



A small model wins English. Whisper still covers the most languages.

English: Granite Speech 470M

3.24% WER on all 7,996 English clips. Also the fastest: an hour of audio in about 20 seconds. Parakeet v2 is next at 3.67%.¹

Swedish, Danish, Norwegian: Whisper

Turbo: 11.05% Swedish, 13.61% Danish, 9.07% Norwegian. Parakeet v3 and Qwen3-ASR sit at 18 to 21% in Swedish and Danish.¹

Hindi: Qwen3-ASR

13.68% WER, less than half of Whisper's 28.72% (large-v3) and 32.05% (turbo).¹

Tamil and Telugu: nothing usable yet

Only Whisper ran. Best is large-v3: 54.85% WER in Tamil, 76.09% in Telugu. Kannada was dropped.¹

German, Spanish, French: close race

All models land between 3% and 7%. Cohere is lowest, on a 3-hour subset. On the full set, Granite and Whisper turbo are close.¹

All of it fits at home

Every finished model peaked under 4 GB of memory. Parakeet v2 used the least, 1.2 GB.¹

Lower WER is better. * in the tables means a 3-hour subset per language, not directly comparable with the full set. Haku EG2 and Moonshine v2 finish tomorrow.

THE SETUP

One base M5 MacBook, ten models, one run each

Machine	MacBook with a base Apple M5, 32 GB RAM, macOS 26.5.2.
Full set	All 40,958 clips (about 89 hours) for Whisper large-v3-turbo. All clips in their languages for Parakeet v2 and v3, Bark 3.0, Bark 3.0 Plus, Granite Speech 470M and Phonon-2.
3-hour subset *	Whisper large-v3, Cohere Transcribe (4-bit) and Qwen3-ASR 1.7B (8-bit) ran on about 3 hours per language, to save time. Same sources, fewer clips.
Runtime	Apple MLX for everything finished: mlx-whisper, parakeet-mlx and mlx-audio. Haku EG2 runs on the CPU with PyTorch.
When	Full run started 22 Oct 2026 at 19:27 Stockholm time. Phonon-2 finished 23 Oct at 21:45. Paused overnight at 22:02.

Source: our test logs and results files.¹

ENGLISH · 7,996 CLIPS, 17.2 HOURS

Granite leads English. Every full-set model is within 1.7 points.

MODEL	WER %	CER %	RTF	PEAK RAM
Granite Speech 470M	3.24	1.68	0.006	2.0 GB
Parakeet v2	3.67	1.83	0.025	1.2 GB
Parakeet v3	4.00	1.99	0.028	1.3 GB
Phonon-2	4.55	2.22	0.007	3.6 GB
Whisper large-v3-turbo	4.55	2.43	0.083	2.1 GB
Bark 3.0 Plus	4.72	2.36	0.023	1.8 GB
Bark 3.0	4.95	2.44	0.031	1.8 GB
Qwen3-ASR *	4.69	2.67	0.104	3.2 GB
Cohere Transcribe *	4.78	2.84	0.030	2.1 GB
Whisper large-v3 *	5.86	3.54	0.156	3.8 GB

* 3-hour subset, 1,239 clips. Lower is better in every column. RTF is processing time over audio length, so 0.025 means an hour of audio in about 90 seconds. Peak RAM is the peak memory of the process. Bark 3.0's model card claims 4.94% English WER; we measured 4.95% on a different mix of test sets.¹²

GERMAN, SPANISH, FRENCH · WER %

In German, Spanish and French, every model lands between 3% and 7%

MODEL	GERMAN	SPANISH	FRENCH
Granite Speech 470M	6.01	3.63	5.46
Whisper large-v3-turbo	6.14	3.52	5.87
Parakeet v3	6.58	6.59	6.37
Cohere Transcribe *	5.21	3.04	4.52
Qwen3-ASR *	5.72	3.59	5.05
Whisper large-v3 *	5.93	3.56	5.60

Full set: German 1,566 clips, Spanish 1,676, French 1,444 (FLEURS plus MLS). * 3-hour subset, about 740 to 850 clips per language. Cohere has the lowest numbers in all three, but on the subset. Parakeet v3's Spanish is pulled down by MLS (8.92%); on FLEURS it scores 3.64%. Bold is the lowest in each column, subset included.¹

For the Nordic languages, use Whisper

MODEL	SWEDISH	DANISH	NORWEGIAN
Whisper large-v3-turbo	11.05	13.61	9.07
Parakeet v3	17.86	18.74	65.77
Whisper large-v3 *	8.92	12.69	8.19
Qwen3-ASR *	19.48	21.22	—

Swedish without NST

NST is a training split, so models may have seen it. On FLEURS alone: Whisper large-v3 7.58%, turbo 8.99%, Parakeet v3 15.54%, Qwen3-ASR 18.88%. The order doesn't change.¹

Parakeet v3 and Norwegian

Norwegian is not on Parakeet v3's list of supported languages, which explains the 65.77%.³ Danish and Norwegian are FLEURS only: 930 and 357 clips.

Full set: Swedish 12,885 clips. Danish and Norwegian are the same clips in both columns, because they are smaller than 3 hours. Qwen3-ASR didn't run Norwegian. * 3-hour subset.

HINDI, TAMIL, TELUGU · WER % (CER %)

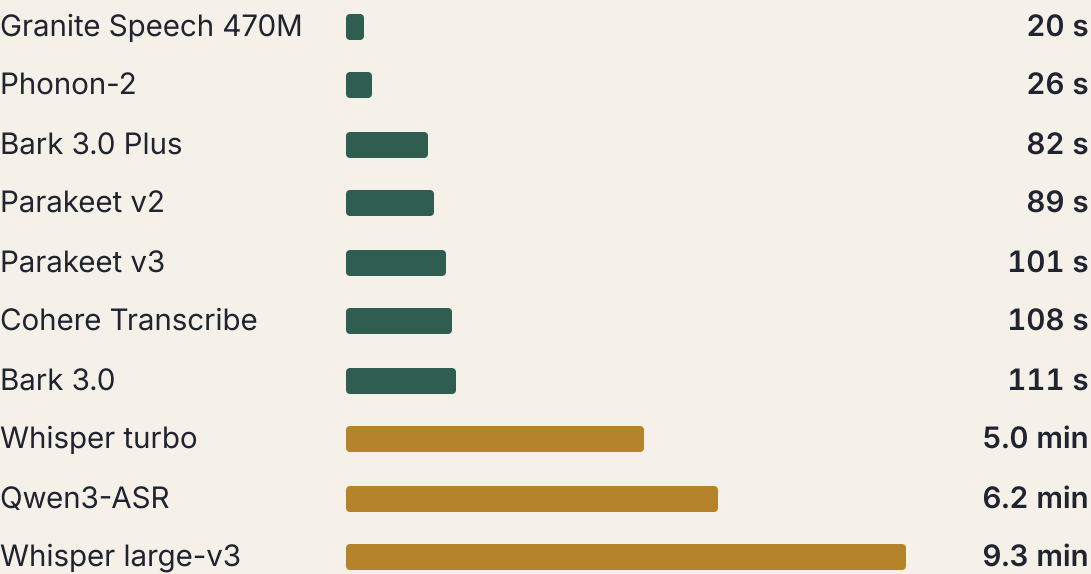
Qwen3-ASR halves Whisper's errors in Hindi. Tamil and Telugu are not usable yet.

MODEL	HINDI	TAMIL	TELUGU
Whisper large-v3-turbo	32.05 (11.25)	74.95 (30.26)	81.39 (24.27)
Qwen3-ASR *	13.68 (5.28)	—	—
Whisper large-v3 *	28.72 (10.71)	54.85 (15.78)	76.09 (25.33)

Full set: Hindi 8,945 clips, Tamil 2,767, Telugu 2,392 (Kathbath plus FLEURS). * 3-hour subset. WER is harsh on these languages because words are long; CER (in brackets) is the fairer read. Whisper turbo's Tamil is worst on FLEURS: 86.90% WER and 57.09% CER, against 69.89% and 18.97% on Kathbath. Kannada was dropped from the run, and IndicConformer, the Indic-specific model we planned, was not run.¹

Granite and Phonon-2 are at least 3x faster than the rest

Time to transcribe one hour of English



Memory and disk

- Peak memory: Parakeet v2 1.2 GB, Parakeet v3 1.3 GB, Bark 1.8 GB, Granite 2.0 GB, Cohere 2.1 GB, Whisper turbo 2.1 GB, Qwen3-ASR 3.2 GB, Phonon-2 3.6 GB, Whisper large-v3 3.8 GB.
- Whisper turbo is 1.5 GB on disk, large-v3 2.9 GB.
- Whisper slows down most in Tamil: turbo's RTF runs from 0.055 (German) to 0.20 (Tamil), large-v3's from 0.14 (German) to 0.81 (Tamil).
- Hindi RTF: Whisper turbo 0.12, Qwen3-ASR 0.28, Whisper large-v3 0.41.

Speed from English runs, after warm-up. One run per model on one machine, so treat small speed gaps as noise. Qwen3-ASR and Whisper large-v3 speeds are from the subset.¹

WHAT IT MEANS

What I'd run at home, based on these numbers

English	Granite Speech 470M. Lowest error, fastest, 2 GB of memory. Parakeet v2 if memory is tight.
German, Spanish, French	Granite or Whisper turbo. The gaps are small. Cohere Transcribe is worth a full run.
Swedish, Danish, Norwegian	Whisper turbo. Within about 2 points of large-v3, about twice as fast, half the size on disk.
Hindi	Qwen3-ASR. Slower than Whisper turbo, but less than half its error rate.
Tamil and Telugu	Nothing good locally yet. Whisper large-v3 is the least bad.

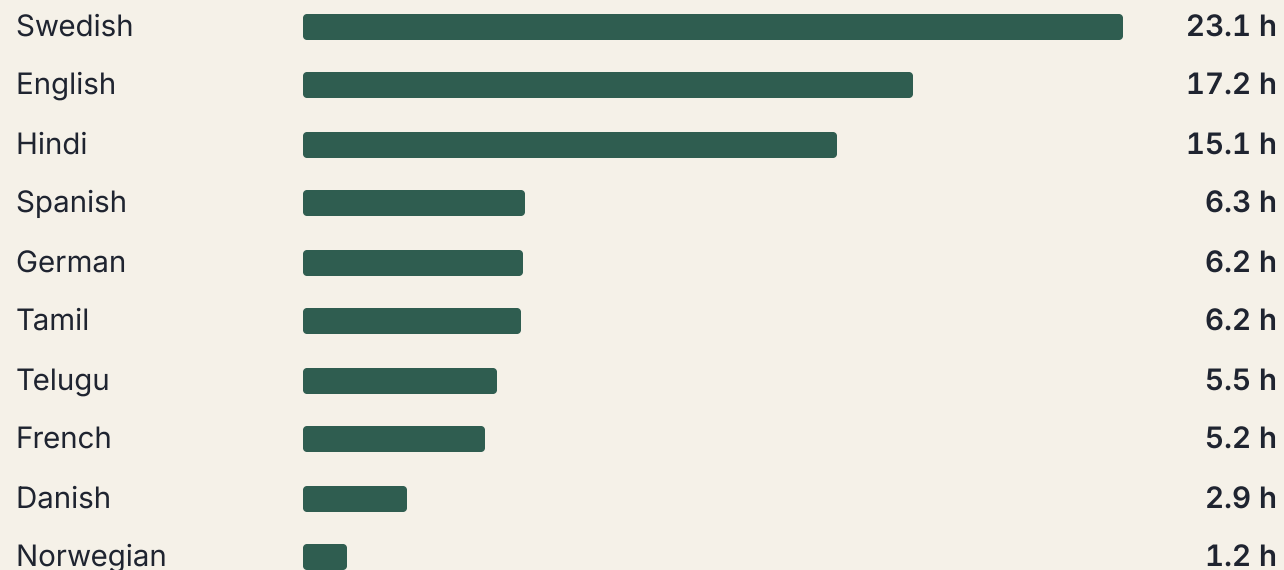
One machine, one run, open read-speech test sets. Your own audio (meetings, kids, accents, noise) will likely score worse.

THE EVAL SET

89 hours of open audio in 10 languages

88.9 hours
40,958 clips

16 kHz FLAC. Only open licenses: CC-BY-4.0 (FLEURS, MLS, LibriSpeech) and CC0 (Kathbath, NST, VoxPopuli). We built 106.5 hours in 11 languages, then dropped Kannada (17.6 hours).



Swedish: NST 19.8 h plus FLEURS. Hindi, Tamil, Telugu: Kathbath plus FLEURS. English: LibriSpeech 10.8 h, VoxPopuli 4.7 h, FLEURS 1.8 h. Spanish, German, French: MLS plus FLEURS. Danish, Norwegian: FLEURS.¹

How the scores are counted, and where they can mislead

Scoring

- WER and CER with jiwer, pooled over all clips in a language.
- English text through Whisper's English normalizer.
- Other languages: lowercase and strip punctuation. Indic vowel signs are kept.
- RTF is processing time divided by audio length, measured after warm-up.
- Long audio is cut into chunks (60 s for Granite, 120 s for the rest).

Caveats

- **Subset vs full set.** Starred models saw about 3 hours per language. Same sources, but not directly comparable.
- **NST is a training split.** Swedish NST scores may be flattering.
- **Read speech.** Audiobooks, parliament and read sentences. Real conversation is harder.
- **Skipped:** Common Voice and IndicVoices (behind logins), Kannada, IndicConformer.
- **One run each** on one Mac. Small gaps in WER or speed are noise.

Source: our test harness and results files.¹

STILL RUNNING

Two English models are pending. Results tomorrow.

Haku EG2

PENDING

476 of 7,996 English clips done. Partial WER 14.85% on the first 402 scored clips. Not final. It runs on the CPU, at RTF of about 0.21, slower than every finished English model.

Moonshine v2 medium

PENDING

English. Starts after Haku EG2. No numbers yet.

Paused

The run is paused overnight. It resumes about 08:00 Stockholm time on 24 Oct, then both models get scored and added here.

Everything else in this deck is final for this run.

SOURCES · CHECKED 23 OCT 2026

Sources

1. Our own results files, logs and harness, 22 to 23 Oct 2026 (not published)
2. [Hugging Face: ConwayResearch/Underdog-Bark-3.0 \(model card\)](#).
3. [Hugging Face: nvidia/parakeet-tdt-0.6b-v3 \(model card, supported languages\)](#)
4. Models as run: [ibm-granite/granite-speech-5.0-470m-turboctc](#), [mlx-community/parakeet-tdt-0.6b-v2](#), [mlx-community/parakeet-tdt-0.6b-v3](#), [mlx-community/Phonon-2](#), [mlx-community/whisper-large-v3-turbo](#), [mlx-community/whisper-large-v3-mlx](#), [ConwayResearch/Underdog-Bark-3.0-Plus](#), [lyzgeorge/cohere-transcribe-03-2026-mlx-4bit](#), [mlx-community/Qwen3-ASR-1.7B-8bit](#)

Vendor claim, not checked like for like: Bark 3.0's 4.94% English WER. Every other number is our own measurement from one run on one Mac.