

LAB M1.07 REPORT

Dina Bosma-Buczynska | February 6, 2026

Differences between prompted and unprompted transcriptions:

The unprompted transcription struggled with technical terms like 'rendezvous' and 'kangaroo walk,' sometimes misspelling them or missing exact terminology. With prompts that included context about Apollo missions and spacecraft vocabulary, accuracy improved significantly. Whisper correctly identified specialized terms and handled overlapping dialogue better. Prompts helped Whisper make better guesses when audio isn't perfectly clear.

Benefits of chunking for long audio:

Chunking is necessary because Whisper has a 25MB file size limit. Beyond this requirement, chunking provides practical benefits:

- Easier error handling: if one chunk fails, you don't re-transcribe everything.
- Progress tracking: see each chunk complete instead of waiting for one large file.
- Accurate timestamps: by adding cumulative time from previous chunks, the final transcript maintains correct timestamps matching the original recording.

Challenges faced:

Main challenges included:

- Installation issues: librosa and llvmlite had compilation errors. I learned these were optional for visualization only.
- API response types: Initially tried dictionary access (segment['start']) but learned segments are objects requiring dot notation (segment.start).
- Homebrew permissions: FFmpeg installation required fixing directory permissions using sudo commands.

Recommendations for improving accuracy:

- **Use prompts with relevant terminology:** Include product names, technical jargon, and key terms specific to the subject matter.
- **Ensure good audio quality:** Clear recordings with minimal background noise transcribe significantly better.
- **Review output:** Even with prompts, manually verify important terms and names.

My key takeaways: prompts significantly improve accuracy, chunking enables handling longer recordings, and proper setup (despite technical challenges) produces professional results. Understanding how to effectively use Whisper, with context-aware prompts, strategic chunking, and quality audio, enables quite reliable speech-to-text transcription.