

## Spec sheet

# Open-source speech corpora vs. Ocular Hi-Fi

Attribute	Ocular Hi-Fi Commercial	LibriTTS Open source	VCTK Open source	Common Voice Open source	LJSpeech Open source
Audio format	<b>Sample rate</b> Hz of analog capture preserved end-to-end <b>48 kHz</b>	24 kHz	48 kHz (downmixed to 16 kHz in most distros)	32–48 kHz source → 16 kHz MP3	22.05 kHz
	<b>Bit depth</b> Per-sample dynamic range <b>24-bit</b>	16-bit	16-bit	MP3 (lossy)	16-bit
	<b>Channels</b> Per-speaker channel isolation when the format calls for it <b>Stereo / per-speaker L / R on conversational sessions</b>	Mono	Mono	Mono	Mono
Capture	<b>Capture modalities</b> Recording formats supported at the Hi-Fi tier <b>Full-duplex conversation, monologue ASR, narrated TTS, multilingual reads</b>	Solo read audiobooks	Solo read prompts	Solo read prompts (crowdsourced)	Solo read (non-fiction)

Attribute	<b>Ocular Hi-Fi</b> Commercial	<b>LibriTTS</b> Open source	<b>VCTK</b> Open source	<b>Common Voice</b> Open source	<b>LJSpeech</b> Open source	
	<b>Overlap &amp; turn-taking</b> Backchannels, interruptions, dual-talk on conversational sessions	<b>Preserved verbatim (full-duplex sessions)</b>	None	None	None	None
Coverage	<b>Speaker metadata</b> Demographic + acoustic context per recording	<b>Age, gender, region, dialect, native language, environment</b>	Gender + speaker ID	Age, gender, accent	Self-reported age / gender / accent (optional)	Single speaker
	<b>Languages</b> Catalogue breadth	<b>14 (and growing)</b>	English only	English only	100+ (long-tail thin)	English only
Licensing	<b>Licensing</b> Commercial usability	<b>Commercial — clean consent at scene level</b>	CC BY 4.0 (research-friendly, audiobook attribution required)	ODC-BY 1.0	CC0	Public domain
	<b>Provenance</b> Source + consent trail	<b>Paid contributors, signed consent, scene-level audit trail</b>	LibriVox volunteers reading public-domain books	Newspaper sentences read by paid actors	Anonymous web volunteers	Single LibriVox volunteer (Linda Johnson)

*Ocular Hi-Fi values shown in bold are the per-row reference. Open-source values cited from the LibriTTS paper, the VCTK README, Common Voice corpus stats, and the LJSpeech project page.*