Label	Dataset Name	Purpose / Use Case	Link	Total Duration	Number of Speakers	Gender Distribution	Accent/Language	Sampling Rate	Environment
1	LibriSpeech	Large-scale read English speech, commonly used for training Automatic Speech Recognition (ASR) systems.	https://www.openslr.org/12	Approx. 1000 hours	2484	Balanced (51.65% male and 48.35% female)	English	16kHz	"Clean" and "Other"
2	RAVDESS	Audiovisual emotional speech dataset, commonly used for Emotion Recognition research.	https://datasets.activeeon.ai/ml/datasets/ravdess-dataset	Approx. 25 minutes (pure speech)	24	Balanced (12 Male, 12 Female)	North American English	48 kHz	Sound Studio (Controlled)
3	SAVEE	Audiovisual expressed emotion dataset, commonly used for Emotion Recognition.	http://kahlan.eps.surrey.ac.uk/savee	Approx. 10-15 minutes (480 segments)	4	Male-only (4 Male)	British English	44.1 kHz	Visual Media Lab
4	EMO-DB	Berlin emotional speech database, commonly used for small-scale, high-quality Emotion Recognition training.	http://emodb.bilderbar.info/download	Approx. 3 minutes (535 segments)	10	Balanced (5 Male, 5 Female)	German	16 kHz	Anechoic Chamber
5	VoxCeleb-1 & 2	Large-scale face and voice recognition dataset, mainly used for speaker identification/verification.	https://www.robots.ox.ac.uk/~vgg/data/voxceleb/index.html#about	Approx. 2000 hours	7000+ (VoxCeleb2)	Imbalanced (61% Male, 39% Female)	Broad English/Multilingual	16 kHz	YouTube/Natural Environment
6	CMU-MOSEI	Large-scale multimodal emotion dataset, commonly used for multimodal sentiment analysis.	http://multicomp.cs.cmu.edu/resources/cmu-mosei-dataset	Approx. 23 hours (dialogue)	1000+	Imbalanced	Broad English (American English)	16 kHz	YouTube/Natural Environment
7	MUSAN	Multi-purpose background, music, and noise dataset, commonly used for audio data augmentation and	https://www.openslr.org/17	Approx. 109 hours	N/A (Non-speaker)	N/A	N/A (Noise/Music)	16 kHz	Various Sources (Background/Music)
8	RIR	Room Impulse Response dataset, used for model training and audio enhancement.	https://www.openslr.org/28	N/A	N/A (Non-speaker)	N/A	N/A	16 kHz	Various Environments (Simulated)
9	CN-Celeb	Chinese speaker recognition database, focused on speaker identification/verification tasks.	https://cnceleb.org	271.72 hours	997	N/A	Mandarin/Chinese	16 kHz	Web/Natural Environment
10	CommonVoice	Large-scale multilingual public speech recognition database, used for ASR training.	https://commonvoice.mozilla.org/en/datasets	Over 24,000 hours	300,000+	Imbalanced (Crowdsourced)	Multilingual/Broad Spectrum	16 kHz	Crowdsourced/Natural Environment
11	IEMOCAP	Interactive emotional motion capture database, used for dialogue and emotion recognition.	https://sail.usc.edu/iemocap/iemocap_info.htm	Approx. 12 hours (audio/video)	10	Balanced (5 Male, 5 Female)	American English	16 kHz	Motion Capture Lab
12	Emo-DB	Berlin emotional speech database, same as label 4, used for emotion	http://emodb.bilderbar.info/index_1280.html	Approx. 3 minutes (535 segments)	10	Balanced (5 Male, 5 Female)	German	16 kHz	Anechoic Chamber
13	Toronto Emotional Speech Database	Emotional speech database, focused on emotion recognition research.	https://tspace.library.utoronto.ca/handle/1807/24487	Approx. 7 hours (movie clips)	20 (Actors)	Balanced (10 Male, 10 Female)	Broad English	16 kHz	Movie Clips/Natural
14	LIRIS-ACCEDE	Audiovisual content annotation and emotion database, used for audiovisual emotion analysis.	http://liris-accede.ec-lyon.fr/database.php	Approx. 98 hours (movie clips)	N/A	N/A	Broad/Multilingual	N/A	Movie Clips/Natural
15	AESDD	Greek emotional speech database, used for emotion recognition research.	http://mcl.ece.uth.gr/research/speech/speech-emotion-recognition	Approx. 700 segments (short)	4 (Actors)	Balanced (2 Male, 2 Female)	Greek	44.1 kHz	Sound Studio

18	VoxForge	Public speech recognition corpus, used for training ASR systems.	http://www.voxforge.org	N/A (Crowdsourced)	N/A (Crowdsourced)	Imbalanced (Crowdsourced)	Multilingual (Crowdsourced)	16 kHz	Crowdsourced/Various Environments
19	TED-LIUM3	Large English speech recognition corpus, extracted from TED talks,	https://www.openslr.org/30	Approx. 450 hours	Approx. 2300	Imbalanced	Broad English (American English)	16 kHz	Lecture Venue
20	AISHELL-1	Chinese speech recognition dataset, used for Chinese ASR.	https://www.openslr.org/33	Approx. 178 hours	400	Balanced (186 male, 214 female)	Mandarin, different accent areas in China	16 kHz (downsampled	Quiet Indoor
21	AISHELL-WakeUp-1	Chinese wake-up word database, used for wake-up word recognition.	https://www.aishelltech.com/wakeup_data	1561.12 hours	254	N/A	Mandarin	Six 16kHz, 16bit and one 44.1kHz, 16bit	Real Home Environment
22	CMU-MOSEI	Large-scale multimodal emotion dataset, same as label 6, used for multimodal sentiment analysis.	http://multicomp.cs.cmu.edu/resources/cmu-mosei-dataset	Approx. 23 hours (dialogue)	1000+	Balanced	Broad English (American English)	16 kHz	YouTube/Natural Environment
23	VoxBlink2	Face liveness detection and speaker verification dataset.	https://voxblink2.github.io	N/A	N/A	N/A	N/A	N/A	N/A
24	Emotional Voices Database (EmoV-DB)	Emotional speech database, used for emotion recognition and speech synthesis.	https://github.com/mmerlart/EmoV-DB	Approx. 3 hours	4 (Professional)	Balanced (2 Male, 2 Female)	English	24 kHz	Sound Studio
25	DEMOS	Italian emotional speech database, used for Italian emotion recognition.	https://zenodo.org/record/2544029	Approx. 2 hours	4 (Professional	Balanced (2 Male, 2 Female)	Italian	48 kHz	Sound Studio
26	Multilingual LibriSpeech (MLS)	Large-scale multilingual speech recognition dataset, used for training multilingual ASR.	https://www.openslr.org/94	Approx. 44,500 hours	N/A	N/A	10 European Languages	16 kHz	Read Audiobooks