Table 1: Speech corpora

Name	# of samples	Languages	License
BibleTTS [1]	113,348	Akuapem, Asante, Ewe, Hausa, Lingala, Yoruba	CC BY-SA
Buckeye [2]	255	English-US	Noncommercial uses
EUROM1 [3]	6,623	Danish, Dutch, English, French, German, Norwegian, Swedish	Research uses
HiltonMoser2022 [4]	785	English	CC BY-NC-SA 4.0
LibriSpeech [5]	27,144	English	CC BY 4.0
MediaSpeech [6]	10,022	Arabic, French, Spanish, Turkish	CC BY 4.0
MozillaCommonVoice (16.1)[1] [7]	491,982	Abkhaz, Arabic, Armenian, Bashkir, Basque, Belarusian, Bengali, Breton, Bulgarian, Cantonese, Catalan, Central Kurdish, Chuvash, Czech, Danish, Dhivehi, Dutch, English, Esperanto, Estonian, Finnish, French, Frisian, Galician, Georgian, German, Greek, Guarani, Hakha Chin, Hindi, Hungarian, Igbo, Indonesian, Irish, Italian, Japanese, Kabyle, Kazakh, Kinyarwanda, Kurmanji Kurdish, Kyrgyz, Latgalian, Latvian, Lithuanian, Luganda, Malayalam, Maltese, Mandarin-China, Mandarin-Taiwan, Meadow Mari, Mongolian, Occitan, Odia, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Spanish, Swahili, Swedish, Taiwanese, Tamil, Tatar, Thai, Turkish, Ukrainian, Urdu, Uyghur, Uzbek, Vietnamese, Welsh, Yoruba	CC BY-SA 4.0
Primewords Chinese Corpus Set 1 [8]	47,166	Mandarin-China	CC BY-NC-ND 4.0
Room Reader [9]	146	English	Noncommercial uses
Clarity Speech [10]	555	English-UK	CC BY-SA 4.0
TAT-Vol2[2] [11]	793	Taiwanese	Noncommercial uses
THCHS-30 [12]	13,327	Mandarin-China	Apache License v.2.0
TIMIT [13]	479	English-US	Research uses
TTS-Javanese [14]	3,068	Javanese	CC BY-SA 4.0
Zeroth-Korean [15]	22,255	Korean	CC BY 4.0

[1] Including a language if the number of voices is greater than 1000. Only verified samples were included. If the number of files is greater than 10,000, randomly sample 10,000 files.

[2] Only the free-access samples were included.

[3] This set only includes the first 30 seconds of each audio.

Table 2: Music corpora

Name	# of samples	License
Albouy2020[16]	123	Noncommercial uses
FMA (large)^3 [17]	106,257	CC BY 4.0
The Garland Encyclopedia of World Music [18]	296	Fair use
HiltonMoser2022 [4]	803	CC BY-NC-SA 4.0
IRMAS [19]	2,838	CC BY-NC-SA 3.0
ISMIR04 [20]	2,185	CC BY-NC-ND 2.5
MagnaTagATune [21]	25,859	CC BY-NC-SA 3.0
MTG-Jamendo [22]	55,699	CC BY 3.0 ES
NHS2 [23]	1,007	CC BY-NC-SA 4.0

References

[1] J. Meyer, D. Adelani, E. Casanova, A. Öktem, D. Whitenack, J. Weber, S. Kabongo Kabenamualu, E. Salesky *et al.*, "BibleTTS: a large, high-fidelity, multilingual, and uniquely African speech corpus," in *Proc. Interspeech 2022*, 2022, pp. 2383–2387.

[2] M. A. Pitt, K. Johnson, E. Hume, S. Kiesling, and W. Raymond, "The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability," *Speech Communication*, vol. 45, no. 1, pp. 89–95, Jan. 2005.

[3] D. Chan, A. Fourcin, D. Gibbon, B. Granstrom, M. Huckyale, G. Kokkinakis, K. Kvale, L. Lamel, B. Lindberg, A. Moreno, J. Mouropoulos, F. Senia, I. Trancoso, C. i. Veld, and J. Zeiliger, "EUROM- A Spoken Language Resource for the EU," in *Proceedings of the 4th European Conference on Speech Communication and Speech Technology*, vol. 1, Madrid, Spain, Sep. 1995, pp. 867–870.

[4] C. B. Hilton, C. J. Moser, M. Bertolo, H. Lee-Rubin, D. Amir, C. M. Bainbridge, and others., "Acoustic regularities in infant-directed speech and song across cultures," *Nature Human Behaviour*, vol. 6, no. 11, pp. 1545–1556, Nov. 2022.

[5] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *IEEE ICASSP*, Apr. 2015, pp. 5206–5210, iSSN: 2379-190X.

[6] R. Kolobov, O. Okhapkina, O. Omelchishina, A. Platunov, R. Bedyakin, V. Moshkin, D. Menshikov, and N. Mikhaylovskiy, "MediaSpeech: Multilanguage ASR Benchmark and Dataset," Mar. 2021, arXiv:2103.16193.

[7] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber, "Common Voice: A Massively-Multilingual Speech Corpus," Mar. 2020, arXiv:1912.06670.

[8] L. Primewords Information Technology Co., "Primewords Chinese Corpus Set 1," 2018, data retrieved from https://www.openslr.org/47/. [Online]. Available: https://www.primewords.cn

[9] J. Reverdy, S. O'Connor Russell, L. Duquenne, D. Garaialde, B. R. Cowan, and N. Harte, "RoomReader: A Multimodal Corpus of Online Multiparty Conversational Interactions," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference.* Marseille, France: European Language Resources Association, Jun. 2022, pp. 2517–2527.

[10] S. Graetzer, M. A. Akeroyd, J. Barker, T. J. Cox, J. F. Culling, G. Naylor, E. Porter, and R. Viveros-Muñoz, "Dataset of British English speech recordings for psychoacoustics and speech processing research: The clarity speech corpus," *Data in Brief*, vol. 41, p. 107951, Apr. 2022.

[11] Y.-F. Liao, C.-Y. Chang, H.-K. Tiun, H.-L. Su, H.-L. Khoo, J. S. Tsay, L.-K. Tan, P. Kang, T.-g. Thiann, U.-G. Iunn, J.-H. Yang, and C.-N. Liang, "Formosa Speech Recognition Challenge 2020 and Taiwanese Across Taiwan Corpus," in *O-COCOSDA*, Nov. 2020, pp. 65–70.

[12] D. Wang and X. Zhang, "THCHS-30 : A Free Chinese Speech Corpus," 2015, arXiv:1512.01882.

[13] J. S. Garofolo, L. F. Lamel, W. M. Fisher, D. S. Pallett, N. L. Dahlgren, V. Zue, and J. G. Fiscus, “TIMIT acoustic-phonetic continuous speech corpus,” *Linguistic Data Consortium*, 1993.

[14] K. Sodimana, K. Pipatsrisawat, L. Ha, M. Jansche, O. Kjartansson, P. D. Silva, and S. Sarin, “A Step-by-Step Process for Building TTS Voices Using Open Source Data and Framework for Bangla, Javanese, Khmer, Nepali, Sinhala, and Sundanese,” in *Proc. The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages*, 2018, pp. 66–70.

[15] L. Jo and W. Lee, “Korean Open-source Speech Corpus for Speech Recognition,” https://www.openslr.org/40/.

[16] P. Albouy, L. Benjamin, B. Morillon, and R. J. Zatorre, “Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody,” *Science*, vol. 367, no. 6481, pp. 1043–1047, 2020.

[17] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, “FMA: A dataset for music analysis,” Sep. 2017, arXiv:1612.01840 [cs].

[18] R. M. Stone, Ed., *The Garland Encyclopedia of World Music: The World’s Music: General Perspectives and Reference Tools.* New York: Routledge, Aug. 2017.

[19] J. J. Bosch, J. Janer, F. Fuhrmann, and P. Herrera, “A comparison of sound segregation techniques for predominant instrument recognition in musical audio signals,” in *13th ISMIR 2012*, 2012, pp. 559–564.

[20] P. Cano Vila, E. Gómez Gutiérrez, F. Gouyon, P. Herrera Boyer, M. Koppenberger, B. S. Ong, X. Serra, S. Streich, and N. Wack, “ISMIR 2004 audio description contest,” 2006.

[21] E. Law, K. West, and M. Mandel, “Evaluation of algorithms using games: the case of music tagging,” in *10th ISMIR 2009*, 2009.

[22] D. Bogdanov, M. Won, P. Tovstogan, A. Porter, and X. Serra, “The MTG-Jamendo dataset for automatic music tagging,” in *Machine Learning for Music Discovery Workshop, ICML 2019*, 2019.

[23] M. Bertolo, M. Snarskis, M. Singh, and S. Mehr, “Cross-cultural music corpus: The Expanded Natural History of Song Discography,” Aug. 2023, zenodo.org/records/8378337.