

We are pleased to announce the release of Speech and Language Corpus free of charge for access and fee-based for usage. Corporuses are indispensable for research and development of communication AI. Our Speech and Language Corpus is the collection of data (or edited works) we gathered through our originally planned projects and copyrighted by Time Hill Inc. (and related authors) . You can use the product by signing a Non-Exclusive Copyright Agreement (for commercial or research use only) . The three main features are as follows.

- 1) Size of collection ---- one of the largest in the world
- 2) Multilingual and multiregional----coverage of more than 100 cities worldwide
- 3) Natural, spontaneous speech----voices recorded in various scenes, which considered difficult to collect

It is our hope that you will be able to use the data for a wide range of purposes not only for AI learning and evaluation for speech recognition and machine translation but also for the areas including linguistics, phonetics, pragmatics, semantics, lexicology, cognitive science, psychology, sociology, cultural anthropology, dialect research, language study and as a material for speech therapists and translators and a creative source for novels and movies.