Product specification of Data-Baker's TTS data

SUPPORT NON-COMMERCIAL ONLY

【Speech corpus of the standard Mandarin female】

Product introduction

Speech synthesis is a technique that produces artificial speech by mechanical and electronic methods. Text-to-Speech service converts written text to natural-sounding speech to provide speech-synthesis capabilities for applications. Text-to-Speech service is a part of speech synthesis.

Text-to-Speech is one of the key technologies for realizing human-machine voice communication. It gives the computer the ability to speak like a human. Text-to-Speech functionality allows our characters to speak any text dynamically. Compared with Automatic Speech Recognition, Text-to-Speech is more mature and has a wider range of applications.

With the rapid development of the artificial intelligence industry, the TTS has also been used more widely. In addition to requiring the speech synthesis effect to be clear enough and understandable, people are increasingly demanding the naturalness, rhythm and sound quality of TTS speech synthesis. One of the key factors in determining the TTS synthesis effect is the quality of the speech corpus.

【Speech corpus of the standard Mandarin female】Speech corpus is a female who pronounces standard Mandarin. The sound recording environment is a professional recording studio. The sound recording uses professional recording software. Recording environment and equipment remain unchanged throughout the recording. The signal-to-noise ratio of the recording environment is not less than 35dB, Mono recording, using 48KHz 16-bit sampling frequency, PCM WAV format. The recording corpus is pure Mandarin text corpus. The corpus is designed to cover the syllables and phonemes as much as possible within the limited corpus data. After recording, the Speech corpus will be proofread; the Rhythm and the Phoneme boundary will be manually edited. The above operations are performed according to the synthesized speech labeling standard.

Data application scenario

Text-to-Speech functionality can be incorporated into custom application. Text-to-Speech is also available to developers building their own applications, and APIs are available to integrate the module with third-party applications. Speech corpus can be used in the following areas:

- ✧ Scientific research, which can be used for speech synthesis system model training;
- ✧ Daily life, PS navigation voice broadcast
- ✧ Smart technology products
- ✧ Education
- ✧ Recreation

And so on.

Data product details

Technical Parameters

Data specification	
Data content	Speech corpus of the standard Mandarin female
Recording corpus	Comprehensive corpus. Cover the phoneme syllable as much as possible.
Effective duration	About 12 hours
Average number of words in a sentence	About 16 words per sentence
Language type	Standard Mandarin
Speaker	Chinese female;Between 20 to 30 years old

Recording environment	✧ Sound collection environment is a professional studio environment ✧ The recording studio meets the professional sound library recording standards ✧ Recording environment and equipment remain unchanged throughout the recording. ✧ The signal-to-noise ratio of the recording environment is not less than 35dB.
Recording tool	Professional recording equipment and professional recording software
Sampling format	Uncompressed PCM WAV format, sampling rate is 48KHz, 16bit.
label content	Proofreading of Chinese syllables and letters proofreading. Rhythm level annotation. Chinese initial consonant and vowel boundary segmentation
Label format	The text is labeled as .txt; Syllable phoneme boundary segmentation file as .interval
Quality Standard	✧ The audio file is in 48k 16bit wav format. The pronunciation of the person's voice will not change greatly, the volume is basically the same, and the pronunciation speed is consistent. Audio file has no clipping. ✧ The accuracy of the word of the marked document is not less than 99.5%; ✧ The proportion of phoneme boundary errors greater than 10ms will be less than 1%; the syllable boundary accuracy is greater than 98%
Storage method	FTP storage
file format	Audio file: WAV; Text annotation file: TXT

	Boundary dimension file: INTERVAL
Data desensitization	
Data sensitive item	none
Applications	
Field of application	Scientific research, smart home, life, education, entertainment and other fields
ownership of copyright	
copyright holders	Databaker（Beijing）Technology Co.,Ltd.

Data hierarchy

Data directory tree

Data directory tree

```
│DB-1 speech corpus of the Mandarin female
││├─CH （Chinese data folder）
│││Wave
││││*.wav （Audio file)
│││Prosody Labeling
││││*.txt （Label text file)
│││Phone Labeling
││││*.interval （Phoneme boundary label file)
```

File naming rules

content	Detailed Description	format
Audio file	Recording file in sentences	Wav 48k 16bit
Label text file	Text phonetic, rhythm-labeled text file	TXT file

Phoneme boundary label file	Chinese initial consonant and vowel boundary labeling results	Interval file

Example：000001.wav

000001-012000.txt

000001.interval

synthesized speech labeling standard

- ✧ **Label format (Chinese)**
- ✧ The text format is *.txt, one line of text, one line of phonetic symbols. The first line of the text is the sentence number, and the sentence number is composed of half-width Arabic numerals, separated by Tab, followed by the text content; the phonetic line begins with the Tab key followed by the phonetic symbol. Words are separated by "/". Phonemes are separated by spaces. The syllables are separated by ".".
- ✧ Chinese tune:marked with 1 to 5, 1 to 4 corresponds to level tone; rising tone; falling-rising tone; falling tone, and 5 to a light voice.
- ✧ rhythm-labeled text：
  - ■ Chinese prosodic structure tagging includes prosodic words (# 1), prosodic phrases (# 2), intonation phrases (# 3) and sentence end (# 4)

Example：

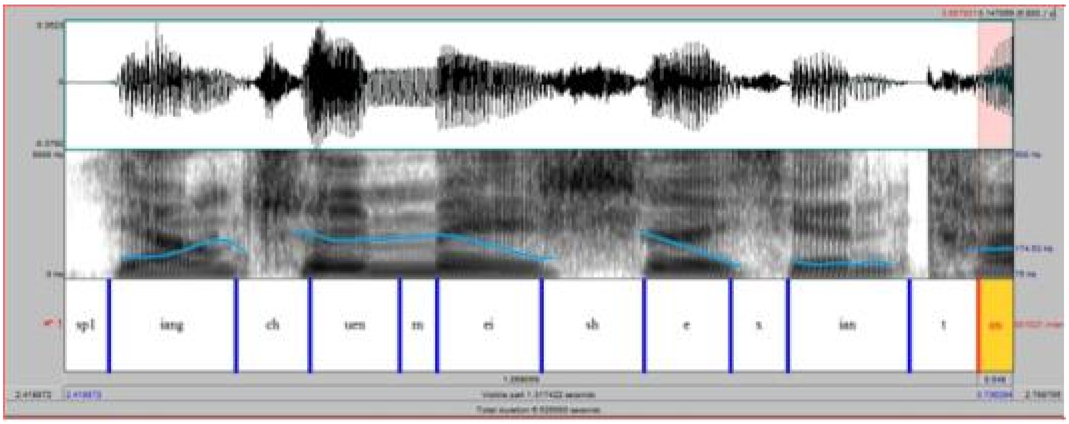
100001 该公司#1 当时#1 表示#3，将于#1 本周一#2 公布#1 正式#1 消息#4。

gai1 gong1 si1 dang1 shi2 biao3 shi4 jiang1 yu2 ben3 zhou1 yi1 gong1 bu4 zheng4 shi4 xiao1 xi4

Phoneme boundary label file：

- ■ Consonant boundary segmentation: Chinese segmentation to vowel, annotated format for interval file format

Example：


sp1
iang
ch
uen
m
ei
sh
e
s
ian
t