

1. Number of utterances : 52487
2. Duration in total : 360000 seconds
3. Number of speakers : 253
4. Randomly sampling 100 audio files, the 95% confidence interval for the mean correct rate is [98.79%, 98.85%]
5. Corpus composition :
 Modern articles > 95%, Ancient articles and poems < 5%
 Number of characters in total: 1085315
 Number of unique characters: 6047
6. Portrait of speakers :

Ages :

Under 17	0.0282
18-24	0.1456
25-29	0.3068
30-39	0.4194
40-49	0.0563
Above 50	0.0437

Gender :

Male	0.6573
Female	0.3291
Unknown	0.0136

Province Distribution :

Guangdong	0.1126
Beijing	0.1029
Jiangsu	0.1029
Zhejiang	0.0786
Anhui	0.0745
Hubei	0.0694
Shandong	0.0553
Fujian	0.0527
Henan	0.0482
Tianjin	0.0476
Jiangxi	0.0428
Sichuan	0.0296
Liaoning	0.0257
Shanghai	0.0220
Hunan	0.0204
Shaanxi	0.0194
Jilin	0.0165
Shanxi	0.0155
Guangxi	0.0155

Others	0.0479
--------	--------

Platforms :

Android	0.6214
iPhone	0.3660
Others	0.0126

(The portrait data is from console of Tencent Mini apps, timeslot 20180328-20180403)