

1. Number of utterances : 48823
2. Duration in total : 360007.8 seconds
3. Number of speakers : 481
4. Randomly sampling 100 audio files, the 95% confidence interval for the mean correct rate is [98.84%, 98.89%]
5. Corpus composition :
 Modern articles > 95%, Ancient articles and poems < 5%
 Number of characters in total: 1008571
 Number of unique characters: 5980
6. Portrait of speakers :

Ages :

Under 17	0.0065
18-24	0.1065
25-29	0.2741
30-39	0.4981
40-49	0.0713
Above 50	0.0380

Gender :

Male	0.6954
Female	0.2990
Unknown	0.0055

Province Distribution :

Beijing	0.1593
Shanghai	0.1028
Jiangsu	0.1009
Guangdong	0.0917
Zhejiang	0.0713
Sichuan	0.0593
Shandong	0.0454
Henan	0.0380
Fujian	0.0315
Hubei	0.0300
Anhui	0.0269
Hunan	0.0231
Liaoning	0.0222
Jiangxi	0.0213
Shaanxi	0.0194
Shanxi	0.0186
Heilongjiang	0.0148
Chongqing	0.0130
Guangxi	0.0130

Others	0.0975
--------	--------

Platforms :

Android	0.5954
iPhone	0.4009
Others	0.0037

(The portrait data is from console of Tencent Mini apps, timeslot 20180322-20180328)