

1. Number of utterances : 52100
2. Duration in total : 360000.7 seconds
3. Number of speakers : 268
4. Randomly sampling 100 audio files, the 95% confidence interval for the mean correct rate is [99.30%, 99.32%]
5. Corpus composition :
 Modern articles > 95%, Ancient articles and poems < 5%
 Number of characters in total: 1074816
 Number of unique characters: 6023
6. Portrait of speakers :

Ages :

Under 17	0.0197
18-24	0.1464
25-29	0.3261
30-39	0.3562
40-49	0.0768
Above 50	0.0748

Gender :

Male	0.4933
Female	0.4787
Unknown	0.0280

Province Distribution :

Guangdong	0.1049
Shandong	0.0914
Shanghai	0.0717
Jiangsu	0.0706
Beijing	0.0613
Hebei	0.0561
Henan	0.0550
Jiangxi	0.0508
Sichuan	0.0467
Anhui	0.0447
Zhejiang	0.0363
Liaoning	0.0312
Hubei	0.0301
Hunan	0.0280
Fujian	0.0239
Tianjin	0.0239
Shanxi	0.0218
Shaanxi	0.0208
Guangxi	0.0208

Others	0.11
--------	------

Platforms :

Android	0.6324
iPhone	0.3437
Others	0.0239

(The portrait data is from console of Tencent Mini apps, timeslot 20180403-20180409)