

1. Number of utterances : 53259
2. Duration in total : 360001.3 seconds
3. Number of speakers : 298
4. Randomly sampling 100 audio files, the 95% confidence interval for the mean correct rate is [98.41%, 98.44%]
5. Corpus composition :
 Modern articles > 95%, Ancient articles and poems < 5%
 Number of characters in total: 1099181
 Number of unique characters: 6076
6. Portrait of speakers :

Ages :

Under 17	0.0218
18-24	0.1736
25-29	0.3091
30-39	0.3609
40-49	0.0736
Above 50	0.0610

Gender :

Male	0.4609
Female	0.5191
Unknown	0.0200

Province Distribution :

Guangdong	0.1091
Shandong	0.0918
Jiangsu	0.0882
Zhejiang	0.0727
Beijing	0.0545
Hebei	0.0545
Jiangxi	0.0482
Shanghai	0.0455
Henan	0.0455
Liaoning	0.0436
Anhui	0.0345
Fujian	0.0309
Hubei	0.0291
Hunan	0.0255
Sichuan	0.0245
Tianjin	0.0227
Shanxi	0.0218
Shaanxi	0.0200
Heilongjiang	0.0191

Others	0.1183
--------	--------

Platforms :

Android	0.6545
iPhone	0.3291
Others	0.0164

(The portrait data is from console of Tencent Mini apps, timeslot 20180410-20180416)