RASC863 - A Chinese Speech Corpus with Four Regional Accents

Aijun Li, Zhigang Yin, Tianqing Wang, Qiang Fang, Fang Hu
The Institute of Linguistics, Chinese Academy of Social Sciences
Yueliang Qian,
Institute of Computing Technology, Chinese Academy of Sciences
liaj@cass.org.cn; ylqian@863-306.org.cn

Abstract

This paper introduces RASC863 (Regional Accented Speech Corpus funded by National 863 Project), a Chinese speech corpus with 4 regional accents of Shanghai (Wu), Guangzhou (Yue), Chongqing (Southwestern Mandarin) and Xiamen (Min) respectively. The corpus consists of spontaneous speech, read speech and selected dialectical words. For the spontaneous speech, each speaker was asked to select a topic himself or from our prepared topic sheet with a variety of 160 topics and then to give a 4-5 minute spontaneous speech on the topic. Besides, each speaker was asked to answer 15 questions spontaneously. The read speech consists of 2200 phonetically balanced sentences selected automatically, and 460 sentences frequently used in daily life. For each dialectal region, we prepared those words that are frequently used in daily life and are different from Standard Chinese, and each speaker was asked to read 15 dialectal words. 800 speakers (200 from each region; balanced in terms of the age, sex, and educational background) were recruited in the project.

1. Introduction

There are 10 dialect families in China, viz. Guan (Mandarin), Jin, Wu, Hui, Xiang, Gan, Kejia (Hakka), Yue (Cantonese), Min, and Ping. It is well known that Chinese dialects differ greatly from each other and are not mutually intelligible. Thus it is quite naturally that Putonghua (Standard Chinese, hereafter SC), which is phonetically based on Beijing Mandarin, has been chosen as the communicative spoken language between people from different dialectal regions. However, people with different dialectal backgrounds typically speak SC with a certain degree of accent due to the influence of their mother tongue dialect. And this kind of influence could be phonetic, lexical, and/or syntactical.

As standardized in PSC (Putonghua Shuiping Ceshi, i.e. Test of Spoken Chinese) issued by the State Language Committee of the Education Ministry of China, the accent of mandarin is categorized into three levels (A, B and C) and each level can be divided into two degrees (1 and 2) (refer to http://www.china-language.gov.cn/ for details). Accent A-1 is the best and speakers with A-2 accent are acceptable for working as a broadcaster in local broadcasting stations. Accent C-2 is the worst and speakers with C-2 accent sometimes couldn't be understood by others from a different dialectal region. In this study, we referred to the criterion in PSC as our working principle of the accent evaluation.

In the recent years, with the development of the ASR techniques, collecting accented spontaneous speech corpora becomes an urgent demand in the field of speech technology, as well as in the field of phonetic sciences. Funded by the National 863 High-Tech Project, we collected a speech corpus with four representative regional accents, namely Chongqing, Shanghai, Guangzhou and Xiamen. The annotation for the whole corpus is still in progress. Up to now, the Chinese Character transcription, as well as the paralinguistic and non-linguistic labeling, has been made for both the spontaneous and read speech. In addition to these, phonetic annotation has been made for the read speech for up to 80 speakers. This paper will firstly give an introduction to RACS863 and then the pronunciation dictionaries will be listed for these four regions.

2. RASC863 production

2.1 Meta data – speaker information

Totally 800 speakers were recruited (200 in each region). Tables 1, 2, 3, and 4 describe the distribution of speakers' age, sex, educational background, and accent category. The speaker's accent category was recorded directly from his/her report if he/she had got a recent PSC test score through the local government's PSC test center; otherwise, it was judged by trained phoneticians referring to the PSC evaluation criterion.

2.2 Corpus design

The prompt sheet for each speaker is shown in Table 5.

The spontaneous speech is composed of two types of content. First, the speaker was asked to give a 4-5 minute talk focusing on one particular topic. And the speaker could propose the speech topic himself/herself or select one from our

* 本文在 ICSLT-O-COCOSDA 上发表

prepared topic list with a variety of 160 different topics. Second, the speaker was asked to answer 15 questions spontaneously. Table 6 shows the 15 eliciting questions.

Table 1：Speaker information of Chongqing

	Category	Male	Female
Age	16-25 (y)	52	55
	26-50 (y)	39	35
	Older than 50 (y)	9	10
Education	Junior high school	4	5
	Senior high school	23	40
	University	73	55
Accent	A-1	0	1
	A-2	11	2
	B-1	28	24
	B-2	53	62
	C-1	7	9
	C-2	1	2
Total		100	100

Table 2：Speaker information of Guangzhou

	Category	Male	Female
Age	16-25 (y)	47	55
	26-50 (y)	48	39
	Older than 50 (y)	5	6
Education	Junior high school	3	1
	Senior high school	11	16
	University	86	83
Accent	A-1	1	2
	A-2	8	
	B-1	29	12
	B-2	49	63
	C-1	4	15
	C-2	9	8
Total		100	100

Table 3：Speaker information of Shanghai

	Category	Male	Female
Age	16-25 (y)	45	46
	26-50 (y)	46	43
	Older than 50 (y)	9	11
Education	Junior high school	10	9
	Senior high school	37	42
	University	53	49
Accent	A-1	0	0
	A-2	0	0
	B-1	46	34
	B-2	41	57
	C-1	13	7
	C-2	0	2
Total		100	100

Table 5: Prompt sheet for each speaker

Items	Speech style	Content
0	Spontaneous	4 to 5 minutes
1-15	Spontaneous	15 question answers
16-388	Read	23 common sentences
36-50	Read	15 dialectal words
51-165	Read	110 phonetically balanced sentences (<30 syllables each)

Table 4: Speaker information of Xiamen

	Category	Male	Female
Age	16-25 (y)	45	46
	26-50 (y)	46	43
	Older than 50 (y)	9	11
Education	Junior high school	10	9
	Senior high school	37	42
	University	53	49
Accent	A-1	0	0
	A-2	0	0
	B-1	42	43
	B-2	51	45
	C-1	6	11
	C-2	1	1
Total		100	100

Table 6: 15 questions

a0001.wav ?你叫什么名字？(What is your name?)
a0002.wav ?出生年月日是什么？ (When is your birthday?)
a0003.wav ?你的联系电话？(What is your telephone no?)
a0004.wav ?工作、学习单位？ (Where do you work or study?)
a0005.wav ?教育程度？ (What is your educational background?)
a0006.wav ?父亲和母亲分别是哪里人？ (Where are your parents from?)
a0007.wav ?说出一个好朋友的名字。(Would you tell the name of one of your good friends?)
a0008.wav ?说出一个网址。(Would you tell a website?)
a0009.wav ?有什么个人爱好？(What is your hobby?)
a0010.wav ?最喜欢吃什么菜？ (What kind of food do you like?)
a0011.wav ?家里有私人汽车吗？准备买什么样的车？ (Do you have a car? What car do you want to buy?)
a0012.wav ?你经常到哪家餐厅吃饭？ (Which restaurant do you usually go?)
a0013.wav ?说出一个手机号码。(Would you speak a mobile phone no?)
a0014.wav ?说出一部电影或电视剧的名字。 (Would you tell one of the film or TV play?)
a0015.wav ?今年多大了？ (How old are you?)

The read speech consists of 2200 phonetically balanced sentences and 460 sentences frequently used in daily life. In addition to these, for each dialectal region, we also prepared those words that are frequently used in daily life and are different from Standard Chinese, and each speaker was asked to read 15 dialectal words as well. 2200 phonetically rich sentences are automatically selected from the newspaper or the Internet on-line talk shows. The sentences cover all Chinese syllables, intersyllabic diphones, and 84% intersyllabic triphones (cf.: 89% in the original text corpus). The sentences were divided into 20 sets

for each region, and each speaker was assigned to read one set of 110 sentences. 460 daily used sentences were also divided into 20 sets and each speaker was assigned to read on set of 23 sentences. See sentences in table 7 for example.

1 天气怎么样？ (What's the weather?)
2 我可以打对方付款的电话吗？ (May I make a collect call?)
3 你们有计算器服务吗？(Do you have a calculator?)
4 轮椅能进吗？(Can a wheelchair come in?)
5 我留个话好吗？ (Would you please take a message for me?)
6 请把信带到邮局去。(Take the letter to the Post Office please.)
7 我得去银行。(I have to go to the bank.)
8 在这里游泳安全吗？(Is it safe to swim here?)
9 你有空房间吗？(Do you rooms?)
10 我要订个房间。(I want to make a room reservation.)
11 住一个晚上多少钱？(How much for one night?)
12 我看看房间行吗？(Could you show me your room?)
13 甚么时候退房？(When will you check out?)

Table 7: Examples of commonly used sentences

Table 8: Meta data for each sound file

```
LHD: SAMPA
DBN: RASC863
SES: Xmf001
CMT: *** Speech Label Information
SRC: A0001.a.wav,A0001.b.wav
DIR: ..\Xmf001
CCD: 1
BEG: 0 //labled sequence begin position
 END: 3.133375 //labled sequence end position
REP: office
RED: 12-23-2003
RET: 17:20:28
CMT: *** Speech Data Coding ***
SAM: 16000 //SampleRate
SNB: 2 signed //BytesPerSample
SBF: lohi //Sample byte order
SSB: 16 //BitsPerSample
QNT: PCM
NCH: 2
CMT: *** Speaker Information ***
SCD: 001 SEX: female
AGE: 31-40
ACC: Xiamen
CMT: *** Recording Conditions ***
SNQ:
MIP: close,medium
MIT: SENNHEISER_ME64,797_CR722
SCC: ENV=office
DBA: 38
CMT: *** Label File Body ***
LBD: A0001.textgrid
LBR: 0,3.133375
 //Prompt text的begin end    gain    min max
LBO: 0,3.133375
//Prompt text的begin end    gain    min max
ELF:
```

2.3 Recording

Speech data were recorded directly into a laptop computer via an external USB M-audio sound card. And a Sennheiser earphone and a CR 722 capacitor microphone (20-20000Hz) were used simultaneously to acquire the audio signal. For each recording session, the acoustic environment and background noise (in db) was recorded.

The software Cooledit Pro 2.0 was used in recording the 4-5 minute spontaneous speech, and YYSRecorder was used in recording other speeches. All speech data were sampled in 16KHz and 16 bit.

A meta data file, as exemplified in table 8 for instance, was generated for each sound file.

2.4 Annotation

Chinese Characters, as well as paralinguistic and no-linguistic events, were transcribed for both the spontaneous and read speech.

The read speech was phonetically annotated for 80 speakers, 20 from each dialectal region. The speech software Praat was employed for the phonetic annotation. Syllables are segmented on the SY tier with canonical syllabic labels. Phonemes of initials and finals (+ tone) are annotated on the SY tier with the time alignment. What should be pointed out is that on the SY tier, we not only annotated sound alternations, such as weakening (voiceless sounds become voiced), insertion, and deletion, due to the effect of co-articulation, but also annotated the phonemic change due to the influence of the speaker's dialect. An example of the phonetic annotation is shown in Figure 1, where # denotes the sound change caused by the dialectal influence.

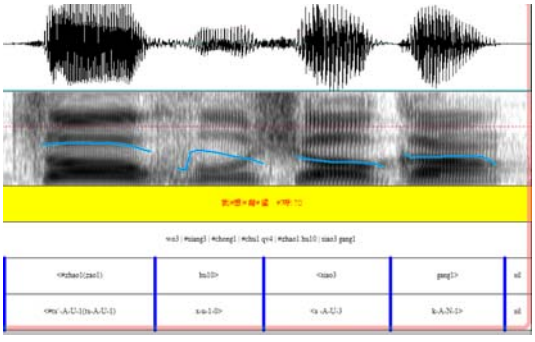


Fig 1. Mis-pronunciation caused by accent is annotated

2.5 Pronunciation and Lexical dictionaries

Based on the extraction from the spontaneous speech of 200 speakers, pronunciation and lexical dictionaries were produced for each region: (1) Canonical Pinyin with tones, as shown in Table 9; (2) Canonical Pinyin without tones (not exemplified here due to the limited space); (3)

Initial and final lists, as shown in Table 10; (4) Lexical words, as shown in table 11.

It is interesting to note that the distributions of tone-included syllables, tone-excluded syllables, and initials/finals are similar across the four dialectal regions. Roughly speaking, about top 200 tone-included syllables cover 80% of the all data; top 100 tone-excluded syllables cover 80% of all; and top 50 initials/finals cover 80% of all. Figures 2 and 3 are two examples.

However, we find that some lexical words with high frequency in the monologue speech of RASC863 are different from those in the conversational speech of CADCC, a conversation or dialogue spontaneous corpus, and those in the children monologue and dialogue speech of CASSCHILD, a child speech corpus (5 to 6 years old). The comparative study on the lexical frequency difference is in process and will be reported elsewhere.

2.6 Corpus distribution

National 863 Information Technology Office is responsible for the distribution of the database, which is about 80 GB. The contact telephone number is +86-10-68339172 and the email address COMPUTER@HTRDC.COM.

3. Ongoing work

Data collecting of 6 more other representative dialectal regions is scheduled in the coming year. Please feel free to contact the authors if you need more information about our speech corpus.

4. References

[1] ZU, Y. Scientific issues on continuous speech corpus design. *Phonetic Study Report of CASS*, 1998.

[2] Schiel, F. & Draxler, C. Production and validation of speech corpora, *Bastard Verlag Munchen, Erstausgabe*, 2003.

[4] Li, A. Chinese Prosody and Prosodic Labeling of Spontaneous Speech, *Speech Prosody 2002*, Aix-en-Provence.

[5] Chen, X., Li, A., et. al. Application of SAMPA-C in SC, *ICSLP2000*, Beijing.

[6] Li, R. (ed.), *Dialectal Dictionaries of Shanghai, GuangZhou, Guizhou and Xiamen*, Institute of Linguistics, CASS.

[7] Li, A. & Wang, X. A contrastive investigation of Standard Mandarin and Accented Mandarin, *Eurospeech2003*.

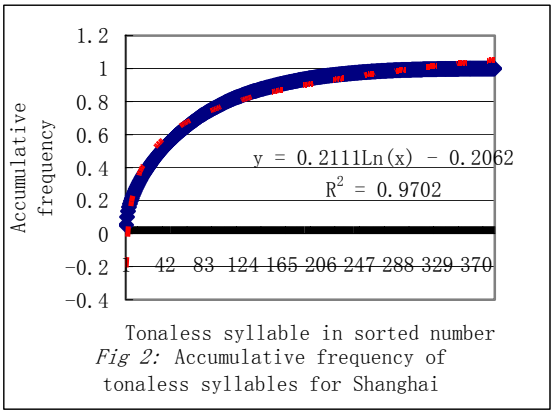
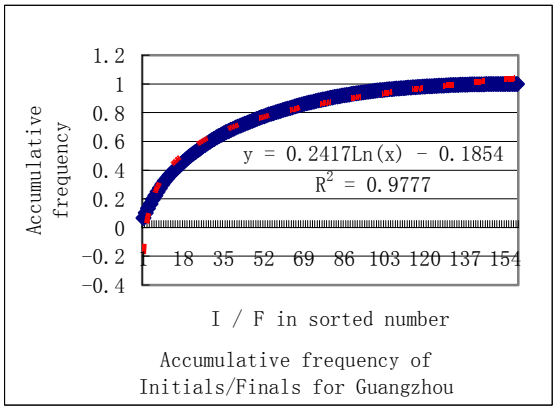


Fig 2: Accumulative frequency of tonaless syllables for Shanghai



Accumulative frequency of Initials/Finals for Guangzhou

Table 9: Top 15 Pinyin+tone in RASC863[i]

GZ	N0.	%	∑%	SH	N0.	%	∑%
de0	8910	0.051352691	0.051352691	de0	9438	0.052255	0.052255
shi4	5464	0.031491706	0.082844397	shi4	6012	0.033286	0.085541
wo3	4268	0.024598573	0.10744297	wo3	4587	0.025397	0.110938
yi1	4088	0.023561145	0.131004115	yi1	3872	0.021438	0.132375
ta1	3077	0.017734257	0.148738372	ta1	2959	0.016383	0.148758
jiu4	2919	0.016823626	0.165561998	jiu4	2525	0.01398	0.162738
you3	2539	0.0146335	0.180195498	you3	2459	0.013615	0.176353
na4	2495	0.014379906	0.194575404	zai4	2378	0.013166	0.189519
zai4	2221	0.01280071	0.207376114	ge4	2350	0.013011	0.20253
bu4	2121	0.012224361	0.219600475	ne0	2244	0.012424	0.214954
ge4	2033	0.011717174	0.231317649	zhe4	2165	0.011987	0.226941
shi2	1962	0.011307966	0.242625615	bu4	2054	0.011372	0.238314

hen3	1829	0.010541422	0.253167037	liao3	2050	0.01135	0.249664
liao3	1807	0.010414625	0.263581663	na4	1956	0.01083	0.260493
men0	1701	0.009803696	0.273385358	men0	1948	0.010785	0.271279
XM	**N0.**	**%**	**∑%**	**CQ**	**N0.**	**%**	**∑%**
de0	6243	0.04791	0.04791	de0	8374	0.052345	0.052345
shi4	4065	0.031195	0.079105	wo3	5005	0.031286	0.08363
wo3	3184	0.024434	0.103539	shi4	4867	0.030423	0.114053
yi1	3012	0.023114	0.126654	yi1	3467	0.021672	0.135725
na4	2449	0.018794	0.145448	jiu4	2658	0.016615	0.15234
you3	2049	0.015724	0.161172	ta1	2416	0.015102	0.167442
ta1	2047	0.015709	0.176881	zai4	2323	0.014521	0.181963
jiu4	1925	0.014773	0.191654	na4	2139	0.013371	0.195333
zai4	1792	0.013752	0.205406	you3	2090	0.013064	0.208397
bu4	1619	0.012424	0.21783	ge4	1941	0.012133	0.22053
ge4	1517	0.011642	0.229472	shi2	1915	0.01197	0.232501
shi2	1464	0.011235	0.240707	men0	1897	0.011858	0.244359
zhe4	1343	0.010306	0.251013	zhe4	1869	0.011683	0.256041
liao3	1289	0.009892	0.260905	bu4	1819	0.01137	0.267412
hou4	1281	0.009831	0.270735	liao3	1815	0.011345	0.278757

Table10: Top 15 initials and finals in RASC863

GZ	**N0.**	**%**	**∑%**	**SH**	**N0.**	**%**	**∑%**
e	20984	0.066297	0.066297	e	22750	0.068779742	0.068779742
a	11130	0.035164	0.101461	j+i	10622	0.032113337	0.100893078
j+i	10169	0.032128	0.133589	iii	10610	0.032077057	0.132970136
iii	10103	0.031919	0.165509	a	10213	0.030876813	0.163846949
d+e	9260	0.029256	0.194765	d+e	9818	0.029682616	0.193529565
i	8998	0.028428	0.223193	i	9269	0.028022832	0.221552397
sh+iii	8181	0.025847	0.24904	sh+iii	8773	0.026523282	0.248075679
ai	8061	0.025468	0.274508	en	8421	0.025459086	0.273534765
en	7997	0.025266	0.299774	ai	7924	0.023956513	0.297491278
x+i	7385	0.023332	0.323106	u	7345	0.022206031	0.319697309
u	6534	0.020644	0.34375	x+i	7200	0.021767654	0.341464963
yi	5915	0.018688	0.362438	uo	6263	0.018934836	0.360399799
an	5493	0.017355	0.379792	yi	6078	0.018375528	0.378775328
uo	5240	0.016555	0.396348	ian	5843	0.017665056	0.396440384
ao	5195	0.016413	0.412761	ao	5440	0.016446672	0.412887056
XM	**N0.**	**%**	**∑%**	**CQ**	**N0.**	**%**	**∑%**
e	14934	0.062646559	0.062646559	e	19103	0.065250732	0.065250732
a	8725	0.036600457	0.099247016	a	9744	0.033282894	0.098533626
j+i	7560	0.031713405	0.130960421	j+i	9691	0.033101861	0.131635487
iii	7303	0.030635317	0.161595738	iii	9188	0.031383747	0.163019234
en	6507	0.027296181	0.188891919	d+e	8686	0.02966905	0.192688284
d+e	6488	0.027216478	0.216108396	en	8010	0.027360015	0.220048298
i	6408	0.026880886	0.242989282	i	7938	0.027114082	0.24716238
ai	6191	0.025970594	0.268959876	sh+iii	7413	0.025320823	0.272483203

sh+iii	5968	0.025035132	0.293995008	ai	7128	0.024347339	0.296830542
x+i	5855	0.024561109	0.318556117	x+i	6462	0.022072461	0.318903003
u	5301	0.022237137	0.340793255	u	6235	0.02129709	0.340200094
uo	4570	0.019170669	0.359963924	uo	5622	0.019203246	0.35940334
yi	4391	0.018419783	0.378383707	yi	5130	0.017522706	0.376926046
ao	4355	0.018268767	0.396652474	wo	5024	0.017160638	0.394086684
ian	4143	0.017379449	0.414031923	ian	4780	0.0163272	0.410413884

Table11. Top 15 words in RASC863

GZ	**N0.**	**%**	**∑%**	**SH**	**N0.**	**%**	**∑%**
的	8125	0.067493	0.067493	的	8676	0.070147	0.070147
我	3256	0.027047	0.09454	我	3213	0.025977	0.096124
是	2216	0.018408	0.112948	是	2293	0.018539	0.114663
就	1929	0.016024	0.128972	呢	2244	0.018143	0.132806
了	1702	0.014138	0.14311	了	1945	0.015726	0.148532
在	1561	0.012967	0.156077	在	1465	0.011845	0.160376
他	1402	0.011646	0.167723	也	1387	0.011214	0.171591
呃	1326	0.011015	0.178738	我们	1351	0.010923	0.182514
也	1300	0.010799	0.189537	就	1350	0.010915	0.193428
呢	1294	0.010749	0.200286	他	1346	0.010883	0.204311
有	1244	0.010334	0.210619	有	1231	0.009953	0.214264
啊	1069	0.00888	0.219499	说	1140	0.009217	0.223481
很	1035	0.008598	0.228097	就是	1069	0.008643	0.232124
嗯	1032	0.008573	0.23667	很	963	0.007786	0.23991
我们	987	0.008199	0.244868	恩	950	0.007681	0.247591
XM	**N0.**	**%**	**∑%**	**CQ**	**N0.**	**%**	**∑%**
的	5651	0.062963788	0.062964	的	7545	0.068409	0.068409
我	2294	0.025559889	0.088524	我	3737	0.033882	0.102291
是	1552	0.017292479	0.105816	是	1734	0.015722	0.118013
了	1197	0.013337047	0.119153	就	1732	0.015704	0.133717
就	1190	0.013259053	0.132412	了	1731	0.015695	0.149411
在	1188	0.013236769	0.145649	在	1520	0.013781	0.163193
有	1016	0.011320334	0.156969	恩	1371	0.012431	0.175623
那个	1010	0.011253482	0.168223	我们	1240	0.011243	0.186866
说	943	0.010506964	0.17873	说	1193	0.010817	0.197683
他	943	0.010506964	0.189237	呢	1093	0.00991	0.207593
也	935	0.010417827	0.199655	也	1047	0.009493	0.217085
呃	906	0.010094708	0.209749	有	983	0.008913	0.225998
我们	885	0.009860724	0.21961	他	966	0.008758	0.234757
恩	869	0.009682451	0.229292	那个	905	0.008205	0.242962
那	814	0.009069638	0.238362	就是	834	0.007562	0.250524

[i] GZ: Guangzhou; SH: Shanghai; XM: Xiamen; CQ: Chongqing.