

# 基于连续话语语料库的 语音音段的初步统计分析\*

陈肖霞 祖漪清

**摘要** 本文是在连续话语语音切分和标记的基础上,对连续话语的语音单元进行的初步分析。主要包括以下几个方面:1.三位发音人声韵母出现频率的分别统计;2.声母时长的统计及发音人之间差异性的检验;3.韵母时长的分别统计。

## 1. 前言

大型语音语料库的建立,为进行大规模语音研究提供了条件,使语音研究不单单停留在定性的阶段,而可以进行定量研究,然而,对大量语料的处理又是这个阶段一个亟待解决的问题。无论是进行音段的研究还是超音段的研究,都需要首先解决语音的切分和标记问题。只有在上述问题解决之后,才可能得出大量的数据,进行进一步的研究。国外著名的 TIMIT 语料库,就是在标记的基础上,对英语作了大量的研究,将过去定性的实验研究扩展到较自然的语言环境中,为语音研究开辟了新路(Keating 1992,1994);言语工程方面,更是利用大量的语料,寻求语音规律,最后用算法加以实现,德国的 KIEL 语言研究也利用大型语料库进行标记,对其中的语音现象及其出现语境进行切分和标记,对语音现象利用语音学和音系学的结合进行解释(Kohler 1994)。这些说明利用语料库进行研究,有重要的价值。对认识语音的本质,解释语音现象的出现也是非常重要的。

本文的研究,是在对一个大的语音语料库(祖漪清 1997)进行了切分和标记的基础上,将语音单元进行的最初步的统计,为下面进一步的研究作准备。关于语音的切分和标记,另有专文介绍,本文对此不再多谈。本文主要谈与此次研究有关的切分。语音切分和标记包括五部分:声母段,过渡段,韵母段,闭塞段,停顿段,所有标记的部分都可以直接得到,可以对不同的音段做统计,从连续话语中得出一些定量的数据,目前做了以下几方面的工作:(1)声、韵母出现频率的统计;(2)发音人之间声母时长值及其差异性检验;(3)韵母时长的统计。

## 2. 研究方法

本文使用的语料是选自 1993、1994 年《人民日报》,由 3 个发音人说出的 1560 句自然话语,覆盖普通话中绝大多数三音子组合。说话速度为每秒 3、4 个音节。录音直接进入计算机,采样率为 16K。利用 KAY 4300B CSL 语音分析软件,对这些语料进行频谱分析,结合宽带语图、波形及听辨,在波形上,从时间点上切分上述五段,并加以标记。根据切分的样点,得出数据,见图一。对声、韵母出现频率进行了统计,对声、韵母时长得出均值和标准差。对不同发音人进行方差分析,考察发音人之间的差异。下面以图一为例对切分进行简单的解释。

---

\* 本研究得到国家“863”高科技研究基金的资助。

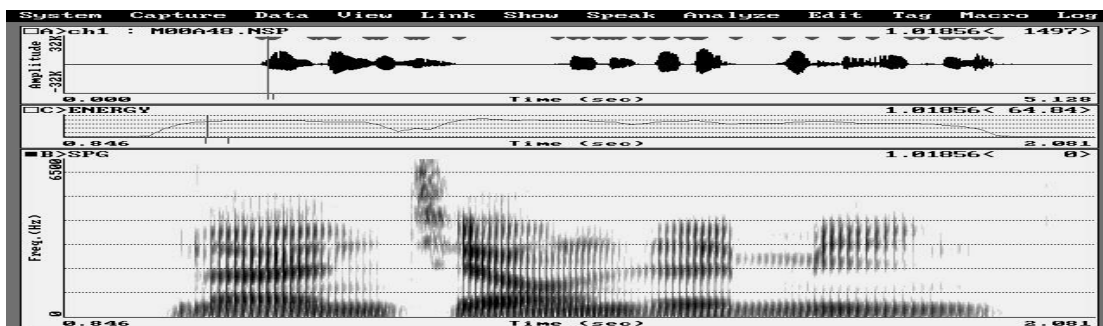


图1 南 nan2 疆 jiang1 人 ren2 民 min2

图1是“南疆人民”，在上面的波形中，倒三角是波形切分标记的时间点，其中带有说明。中间是振幅，最下面是宽带语图。根据宽带语图、波形和振幅，可将各段切出。声母的切分根据声韵之间频谱的明显变化，可分为两种情况，声母为清音和浊音。一般来说，第一种情况下，声母与后面大韵母过渡有比较清楚的分界，浊声母虽然与韵母连接密切，但在高频部分却与韵母有明显大差别，也是比较好区分的；过渡段根据元音共振峰转折点确定起止点的过渡，在时长上差别较大；元音段则是过渡段之后的部分。但在连续话语中，音节间的切分也是要注意的，尤其是后音节为浊声母或零声母时，如“人民”中对声母的切分，从频谱可以看出，浊擦音由于带有噪音，与前面音节韵母的共振峰相连接，没有明显的分界，但在高频部分其能量很弱，所以可以将这个信息作为分界的标志。

### 3. 研究结果

#### 3.1 本文的统计结果同声学手册的比较

《声学手册》(马大猷等 1983)中对声、韵母的出现频率做过统计，得出了一个语音分布自然平衡结果。因此，本文希望将计算结果与《声学手册》中关于声、韵母出现频率的结果做比较，看数值相差大不大，并以此考察一个人念的520句所包含的声韵母是否已经足够，已达到自然平衡，这是下面计算声、韵母的时长并做比较的前提。表1给出了三位发音人各自实际发音的语料中声母出现频率的值，零声母因归入韵母未做统计。声母中的浊化音单独考虑，不在计算中。

表1 声母出现的频率(%)

人\声	b	p	m	f	d	t	n	l	g	k	h
m00	5.7	1.8	3.7	3.6	7.7	4.0	2.2	6.5	6.6	2.4	5.8
m01	4.4	1.8	3.6	3.7	9.3	4.2	2.4	7.0	6.0	1.8	5.3
m45	4.0	1.9	4.6	4.3	5.2	3.9	3.0	6.5	5.5	2.1	6.4
手册	5.15	0.98	3.74	2.45	12.00	3.53	2.53	5.69	5.50	1.83	4.42
人\声	j	q	x	zh	ch	sh	r	z	c	s	
m00	10.0	4.3	5.8	7.1	3.9	8.6	2.5	4.1	1.8	1.9	
m01	9.2	3.8	6.3	7.7	3.8	9.0	2.5	3.8	1.7	2.5	
m45	9.4	3.8	6.8	7.6	4.1	9.8	2.2	4.5	1.8	2.6	
手册	6.98	3.11	4.86	7.18	2.75	7.66	1.94	3.01	1.15	1.08	

\*这里的“手册”指《声学手册》中的声、韵母出现频率。

本文的计算是从实际发音来考虑的。从本文大统计数值看, 本文与《声学手册》中大结果比较接近。零声母未参加计算, 对声母的浊化未考虑, 会影响到结果, 这是造成数值不同的两个原因。但从结果看, 很接近, 只有 d 和 j 差别大些。三位发音人之间差别不大。

表 2 韵母出现频率 (%)

	a	o	e	ɪ	u	v	-ɪ2	er	ai	ei	ao	ou
m00	3.56	0.48	9.51	9.02	6.30	2.90	4.70	0.77	4.31	1.89	2.84	1.86
m01	4.16	0.46	9.72	8.77	6.27	2.17	5.42	0.64	4.24	1.18	2.61	1.90
m45	3.46	0.52	9.27	9.07	6.82	2.19	5.01	0.68	3.53	1.41	2.85	1.94
手册	3.89	0.54	12.38	8.80	7.11	1.80	6.41	0.28	2.83	1.28	3.10	1.88
	ia	ie	iao	iu	ua	uo	uai	ui	ve	an	en	ang
m00	1.53	2.14	2.04	2.30	0.55	3.63	0.52	3.30	0.77	3.54	2.50	3.02
m01	1.12	2.10	1.96	2.23	0.52	3.66	0.32	2.95	0.90	3.19	3.00	3.02
m45	1.16	2.08	1.90	2.27	0.63	3.69	0.44	3.56	0.71	3.66	2.97	2.86
手册	1.09	2.42	2.06	2.60	0.44	4.40	0.32	2.75	1.01	3.41	3.62	2.87
	eng	ong	ian	in	iang	ing	iong	uan	un	uang	van	vn
m00	2.78	3.94	3.93	2.60	2.14	3.65	0.33	2.00	0.89	0.77	1.18	0.45
m01	3.25	4.60	4.46	2.49	1.84	3.64	0.30	1.79	0.90	0.70	1.21	0.59
m45	3.39	4.48	4.50	2.96	1.74	3.25	0.31	1.60	1.02	0.63	1.22	0.72
手册	3.09	4.18	4.10	1.95	1.80	3.05	0.42	1.24	0.89	0.65	0.85	0.52

\*v, ve, van, vn 等表示韵母为“鱼”韵或以此为介音的韵母。

由于 ueng 出现的很少, 本文没有选, i1 的出现相对较少, 也没选, i2 为舌尖后卷舌元音, 本文据此与手册的 i 做对比。三位发音人的语料与手册比较, 从韵母出现频率看, 只有 e 差别大些, 其它韵母出现频率很接近。因此, 语料基本达到了自然平衡。

### 3.2 三位发音人之间声母时长值及其 F 检验

本文表三给出了三位发音人的各声母的时长值。过去对时长的测量主要在单音节(吴宗济、曹剑芬 1979)、两音节(齐士铃、张家录 1982)或引导句中(冯隆 1985)进行的。真正从连续话语中做大量的统计是没有的, 这是本文与以前做的研究的不同之处。这样做只是希望考察在相对自然的连续话语中, 声母时长的真正变化到底是怎样的, 与过去在这方面的研究差别大不大。表三给出了三位发音人声母时长的数值, 三位发音人的时长略有不同, 第一位发音人时长最长, 第三位最短, 但每位发音人内部自成系统。

表 3 声母的时长(ms)

人\声	b	p	m	f	d	t	n	l	g	k	h
m00	19	71	62	102	23	78	53	42	31	80	105
m01	16	78	56	93	15	78	54	40	26	79	92
m45	13	56	53	76	14	62	44	33	22	65	79
qi	11.7	77.7	65.4	90.1	7.9	95.7	65.0	60.6	18.0	75.0	88.2

人\声	j	q	x	zh	ch	sh	r	z	c	s	
m00	67	134	133	46	98	133	44	70	134	138	
m01	45	113	110	34	92	112	46	46	113	120	
m45	51	95	97	29	71	93	42	42	90	98	
qi	42.5	114.0	105.1	20.2	95.4	118.7	54.7	59.3	111.0	120.4	

\*本文的三位发音人分别是 m00、m01、m45, qi 是指“汉语普通话辅音音长分析”的作者齐士钤、张家录得出的数据。

从结果看,本文的三位发音人每个人都自成系统,但发音方法与时长的关系没变,三个人声母的长度依次下降,第三个人的最短,有些音甚至比齐文的还短,如 q 和 x。声母越长,差别越大,依次为:擦音、送气塞擦音、送气塞音、不送气塞音。本文在测量上与齐文不同,本文将过渡单独切出,声母的时长不包括韵母中的过渡。但从本文与齐文所做的两音节声母时长的比较看,在绝对时长值上,本文的结果与齐文的结果只在舌尖和双唇不送气塞音中略有不同,时长的差别不大。这说明声母时长在连续话语中和在孤立两音节情况下差别不大,即声母时长受语境的影响不大。另外,本文的三位发音人的声母浊化分别占声母总数的 6%, 5% 和 13%。第三位发音人的声母时长最短,浊化的也最多。

下面要考虑的是三位发音人之间每一个声母的时长是否有显著差别,这个问题也是下面进一步研究而必需作的。如果发音人之间差异较大,则说明个体的差异是一个必需考虑的因素,而不能将三位发音人的结果作为整体考虑。本文对三位发音人的声母时长两两分别做 F 检验,考察不同人之间是否存在显著差异。结果见表 4。

表 4 声母时长的 F 检验值( $p>0.01$ )

b	F(310,201)=2.38	F(310,240)=1.14	F(240,201)=2.08
p	F(95,96)=1.50	F(97,95)=1.41	F(97,96)=2.13
m	F(200,231)=1.95	F(200,194)=1.58	F(194,231)=1.24
f	F(192,216)=1.28	F(202,192)=1.40	F(202,216)=1.80
d	F(415,263)=2.37	F(415,505)=2.54	F(263,505)=1.07
t	F(218,199)=2.58	F(218,226)=1.36	F(226,199)=1.89
n	F(121,150)=1.32	F(121,128)=1.40	F(150,128)=1.06
l	F(349,332)=1.44	F(349,381)=1.07	F(381,332)=1.35
g	F(355,277)=2.22	F(355,325)=1.70	F(325,277)=1.30
k	F(99,132)=1.19	F(107,132)=1.04	F(99,107)=1.13
h	F(315,327)=1.70	F(289,315)=1.06	F(289,327)=1.82
j	F(542,477)=1.70	F(542,498)=1.84	F(477,498)=1.08
q	F(232,193)=2.13	F(232,207)=1.03	F(207,193)=2.08
x	F(316,347)=1.95	F(316,343)=1.21	F(343,347)=1.61
z	F(219,227)=2.93	F(219,206)=2.40	F(206,227)=1.22
c	F(96,93)=3.56	F(96,90)=2.07	F(90,93)=1.71
s	F(101,130)=1.59	F(136,101)=1.27	F(136,130)=2.00
zh	F(381,386)=1.78	F(381,417)=1.54	F(417,386)=1.15

ch	F(213,206)=1.71	F(213,206)=1.84	F(206,206)=2.05
sh	F(465,498)=1.99	F(484,465)=1.08	F(484,498)=2.13
r	F(136,114)=1.58	F(134,136)=1.22	F(134,114)=1.92

表 4 对三位发音人发的每个声母时长的方差做两两 F 检验,考察发音人之间有无差异。从 F 值看,在显著性水平为 0.01 时,三位发音人之间没有明显差异。

### 3.3 对韵母时长的统计

本文给出韵母时长统计值,区分了四声。见表五。

表五 韵母的时长 (ms)

	a	o	e	ɪ	u	v	-i2	er	ai	ei	ao	ou
m00 1	110	120	97	107	86	98	73		106	98	116	107
2	115	127	106	111	101	169	73	201	116	98	137	100
3	121		90	106	123	106	77	163	110	125	105	111
4	116	94	84	109	92	106	81	177	111	109	130	109
m01 1	119	127	81	116	103	98	92		117	121	135	134
2	121	145	110	118	108	178	82	201	138	101	147	120
3	129		116	113	117	124	89	175	122	138	137	119
4	128	114	96	114	97	122	90	202	137	117	129	109
m45 1	105	95	88	91	79	91	82		112	96	116	116
2	111	104	98	105	88	154	81	168	109	102	108	100
3	103		89	98	91	107	87	140	99	110	99	102
4	108	102	79	101	93	96	89	151	128	104	128	104
	ia	ie	iao	iu	ua	uo	uai	ui	ve	an	en	ang
m00 1	163	107	147	160	175	122		121	155	152	136	143
2	149	130	156	176	159	124		196	111	160	144	136
3	200	124	128	169	213	135	134	130	107	138	132	134
4	154	145	164	107	123	113	155	132	162	142	128	132
m01 1	157	119	154	175	152	117	160	159	170	153	140	147
2	168	135	153	166	162	119	146	182	123	154	141	152
3	125	139	167	171	222	141		161	115	142	142	147
4	158	153	168	149	129	120	182	144	138	154	133	132
m45 1	135	95	141	139	137	127		114	184	127	131	121
2	178	98	116	146	106	104	100	162	105	129	119	132
3	134	114	115	142	137	123		114		120	114	119
4	152	130	150	127	124	103	171	128	146	140	132	119
	eng	ong	ian	in	iang	ing	iong	uan	un	uang	van	vn
m00 1	128	137	142	135	162	137	227	166	146	146	137	150
2	130	130	182	152	191	143	167	177	190	233	191	140
3	137	138	178	154	170	148	193	176	143	190	186	
4	129	123	146	129	160	125	201	169	147	170	206	156

m01	1	137	139	149	139	180	143	112	159	146	162	148	138
	2	130	139	176	146	193	148	137	171	158	225	197	148
	3	157	144	155	155	181	142		161	196	197	180	
	4	140	134	152	148	157	137	191	179	160	180	171	172
m45	1	111	113	122	113	139	111	143	131	119	133	109	107
	2	108	107	152	133	144	117		134	154	164	184	136
	3	112	116	137	123	132	117	200	157	150	143	144	
	4	113	110	129	100	147	130	180	135	141	182	213	153

从结果看,第三位发音人的韵母时长是最短的,而每个人都自成系统。每位发音人每个韵母的时长在四声中的差别不大,尤其在上声时没有表现出明显的长,这是与在单音节中很不相同的。从韵母的长度看,单元音韵母略短于复合元音韵母或带有鼻韵尾的韵母,其差值大多在 100ms 之内。凡是与 a 音相拼的韵母一般都比较长,而与 e 音相拼的则较短。

#### 4. 结论

本文给出了作为语音音段单元的声、韵母出现率和时长的部分统计结果。1 对出现率的统计是考察每个单元的覆盖情况是否足够,对下一步的深入研究,是否有意义。从结果看,三位发音人的语音覆盖情况与过去《声学手册》所给的结果近似,对这些语料进行研究还是有意义的。2 对声、韵母时长的测量结果分别给出平均值,并与过去在这方面作的研究进行了比较,目前的工作与过去在这方面的研究不同之处有两点,其一是语料多,其二是采用的比较自然的连续话语。声母时长的结果看,本文声母的时长与发音方法的关系更密切,这与曹剑芬、齐士钤等、冯隆的结果相同,即发音方法对声母时长的影响更大;韵母的时长,发音人之间差别不是太大,每一韵母的四声之间长度变化不是很有规律,并不是上声最长。韵母的长度与韵母的单复以及舌位大高低有一定关系,不如声母稳定,这可能与语境的变化有关。

另外,目前在使用语料中主要考虑了有明显静止段的停顿,其它形式的停顿没有考虑,也没有仔细区分声韵之间的互补。这些对声、韵时长也有一定的影响,这是以后要进一步研究的。

#### 参考文献

- 冯隆 (1985) 北京话语流中声韵调的时长,《北京语音实验录》,北京大学出版社。
- 马大猷、沈豪等(主编)(1983)《声学手册》,科学出版社。
- 齐士钤、张家录(1982)汉语普通话辅音音长分析,《声学学报》第2期。
- 吴宗济、曹剑芬(1979)普通话辅音声学特征的几个问题,《第二届全国声学学术会议论文摘要》。
- Klaus J. Kohler. (1994) Complementary phonology: a theoretical frame for labelling an acoustic database of dialogues.
- Keating, P., B. Blankenship, D. Byrd, E. Flemming, and Y. Todaka (1992) Phonetic analyses of the TIMIT corpus of American English, UCLA Working Papers in Phonetics 81.

- Keating, P., Peggy MacEachern, Aaron Shryock, and Sylvia Dominguez (1994) A manual for phonetic transcription segmentation and labeling of words in spontaneous speech. UCLA Working Papers in Phonetics 88.
- Keating, P., Dani Byrd, Edward Flemming, and Yuichi Todaka. (1994) Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication* 14, 131-142.
- Zu, Yiqing (1997) Sentence design for speech synthesis and speech recognition database by phonetic rules, EuroSpeech' 97, Vol. 2, 743-746.

## **An Statistical Analysis of Segments Based on Continuous Speech Database**

Chen Xiaoxia , Zu Yiqing

This report is a preliminary statistical analysis for segments in continuous speech of Standard Chinese. The segmentation and labelling of the continuous speech have been done, upon which the current report is based. The purpose is to give some results from three aspects. Firstly, we have calculated the occurrence frequency of each initial (shengmu) and final (yunmu), and the results are presented. Secondly, acoustic durations of initials and finals are obtained after the segmentation, and statistical analyses for initials are done to see whether there are significant differences among three speakers. Finally, a comparison is made between our results and the results from previous works.

The present study differs from the previous ones in that it is based on the continuous speech database rather than on isolated contexts or carrier sentences.

The speech data in this study, which consist of 1560 sentences spoken by three male speakers, cover all triphones in Standard Chinese. All initials and finals in the data have been segmented out for use. According to the segmentation, statistical analyses were given, with the following results. (1) There are no significant differences between our results and the results of *Manual of Acoustics* in occurrence frequency of all initials and finals. (2) The initial duration is mainly related to the manner of articulation; it varies with different speakers, but the difference is not significant. (3) The distributions of final durations are irregular in four tones, and we leave it open for further research.