第 24 卷第 3 期　　声 学 学 报　　Vol. 24, No. 3
1999 年 5 月　　ACTA ACUSTICA　　May, 1999

汉语连续语音数据库的语料设计*

祖漪清

(中国社会科学院语言研究所　北京　100732)

1997 年 10 月 27 日收到
1998 年 3 月 27 日定稿

摘要　质量优良的语音识别系统或语音合成系统需要高质量的、在语音学和语言学知识指导下设计的科学合理简洁有效的连续语音数据库的支持. 在目前阶段, 汉语语音数据库应限制在朗读言语 (read speech) 的音段方面. 为了描写语流中的音变现象, 考虑如下语音单元: (1) 不计声调的音节 (401 个). (2) 音节间的双音子 415 个. (3) 音节间的三音子 3035 个, 这是根据 37 个基本音子, 利用音节间共振峰过渡的研究结果, 按规则规纳的结果. (4) 所有音节间过渡段的韵母 - 声母结构, 采用和同三音子相同的归并方法, 共 781 个. 为了增加不同的韵律结构, 并考虑语音识别系统的后处理, 语料还包括汉语的 17 类基本句型. 选用 1993 、 1994 两年的“人民日报”、“百家报刊精选”及若干电视剧本、词典词库作为语料库的原始语料, 从中选出 2185 个句子和 388 个短语作为朗读语料, 它们覆盖了 99.8% 个无调音节, 100% 的双音子, 99.6% 的三音子, 以及 17 类句型.

PACS 数: 43.70

The text design for continuous speech database of standard Chinese

ZU Yiqing

(*Institute of Linguistics Cinese Academy of Social Sciences* Beijing 100732)

Received Oct. 27, 1997
Revised Mar. 27, 1998

Abstract　Well developed continuous speech recognition systems need a higher quality, scientific designed, succinct and valid continuous speech database. At the first stage the database should be mainly limited in read speech. To describe very complex variances in continuous speech, we propose the following speech units: (1) 401 syllables witout tone. (2) 415 inter-syllabic diphones. (3) 3035 inter-syllabic triphones. (4) 781 inter-syllabicfinal-initial structures. We also give 17 sentence patterns to include the prosodic phenomena. Using automatic method 2185 sentences and 388 phrases are collected by above phonetic rules from a big corpus - recent years “Peple's Daily” and so on, as the read text of continuous speech recognition database in Standard Chinese. This set of sentences covers 99.8% syllables without tone, 100% inter-syllabic diphoes, 99.6% inter-syllabic triphones and 100% sentence patterns.

* 国家 863 高科技计划资助项目 (课题号: 863-306-03-09-1)

引言

汉语语音识别和语音合成已走出孤立音节和孤立词，进入了大词汇的连续语音阶段。质量优良的语音识别及合成系统需要在语音学和语言学知识指导下设计的科学合理简洁有效的连续语音数据库的支持。

目前国际上关于连续语音语料库的言语类型可分为三类，第一类是朗读言语 (read speech)，第二类是流畅言语 (fluent speech)，第三类是自由言语 (spontaneous speech)。第二类和第三类的区别在于言语的内容是否有所准备 (planning)。这三类的语音学问题都包括音段 (segmental) 和韵律 (prosodic) 两个方面。

国外的语音数据库无论在音段方面还是在韵律方面都较为先进，例如 80 年代就制成的英语语音数据库 TIMIT 数据库[1,2]，在连续语音的层面上考虑了声学语音学规律，并用自动方法为语音数据做了音段标音[3]，90 年代又作了手工的语音学音段标音[4]。在韵律研究方面，研制了用于语音数据库韵律标音的 TOBI 韵律标音系统[5]。随着言语工程技术的发展，许多研究机构都为各自的研究目的和各自的语言制作了专门的数据库。例如，日本 ATR 的语音数据库为处理自由口语的语音数据库。它不仅考虑了语音学现象，而且考虑了对韵律结构起作用的语言学现象[6]。Bell 实验室用于文语转换的语音数据库在考虑音段的同时也考虑了韵律方面的内容[7]。AT&T 研制了用于口语对话系统的语音数据库[8]。为了进行电话语音识别，许多国家都在电话线上采集语音数据[9]。

汉语语音数据库在海内外已有几套版本，为语音识别制作语音数据的库有清华大学电子工程系、清华大学计算机系、中国科学院自动化所、香港大学计算机系等[10−13]。用于语音研究与合成的数据库有中国科学院声学所、中国社会科学院语言所等。所有这些汉语的数据库在覆盖汉语连续语句的语音特性方面都没有作出全面的设计。以服务于语音识别的语音数据库为例，文本设计时所考虑的语音现象只能在音节和词组中加以控制，连续语句多从大的语料库中随机选取，不能驾驭。而训练一个连续语音识别系统，最好的训练数据是连续语音。因此我们受 “863” 专家组的委托，负责连续语音数据库的语料文本设计，中国科学技术大学电子工程系负责录音及管理。我们希望通过这一工作将语音研究的成果较好地服务于言语工程。

连续语流中极为复杂的语音现象 —— 音变，为言语工程带来了许多困难，我们认为在目前阶段首先考虑音段中的语境音变是比较合适的。根据目前国内语音研究的现状，连续语音数据库文本设计首先针对朗读言语 (read speech) 的音段方面。

1　连续语音的声学特征

1.1　连续语音的音变问题

连续语音的音变是指语音性质离开其孤立存在时的变化，在音段方面主要表现为语境 (context) 音变 (语音单位在不同语境下发生的音变) 和非语境音变 (语速、语气、句子结构、不同说话人引起的音变)。超音段方面也有这两方面的音变，它们是指基频、时长、能量等因素在不同环境下发生的变化以及音段和超音段之间发生的相互影响。

1.2　连续语音和共振峰过渡

连续语流是由一个个音节级联而成，每个音节是由一个个更小的语音单元构成的。为了研究连续语流中的语音现象，我们需要确定普通话语音的基本成分。音位是特定语言具有区别意义的类别，每个音位的语音体现是多种多样的，称音位变体，石锋将对应音位变体的声学表现定义为音子[14]。本文将音子定为连续语音的最小音段，在不同的语境下，音子可以描写连续语音的音

变. 在连续语流中，基本音子通常是以变化的形式出现. 连续语音同孤立字或连接词的根本区别在于：不仅在音节内，而且在音节间以及词组间，各音子之间也存在着相互影响. 尽管汉语普通话是由一个一个音节连接而成，但在连续语流中，音节的声学表现与孤立音节的情形十分不同，它受到左右音段的影响，偏离了本来位置，这就是所谓音段方面的音变.

图 1 是连续语音“腊月初六”的语图，它们是连续变化的，不存在明显的稳定段，这一事实充分证实了“一个音位可以由语音信号中前后接续的几个音段来实现[15]”，也就是说一个语音学层面的音段对应着声学层面的多个音段，其中的原因就是存在着过渡音段. 在声学层面上描写语流中的音变现象以及音段间的过渡在脱离语境的情况下仅用音子是不够的.

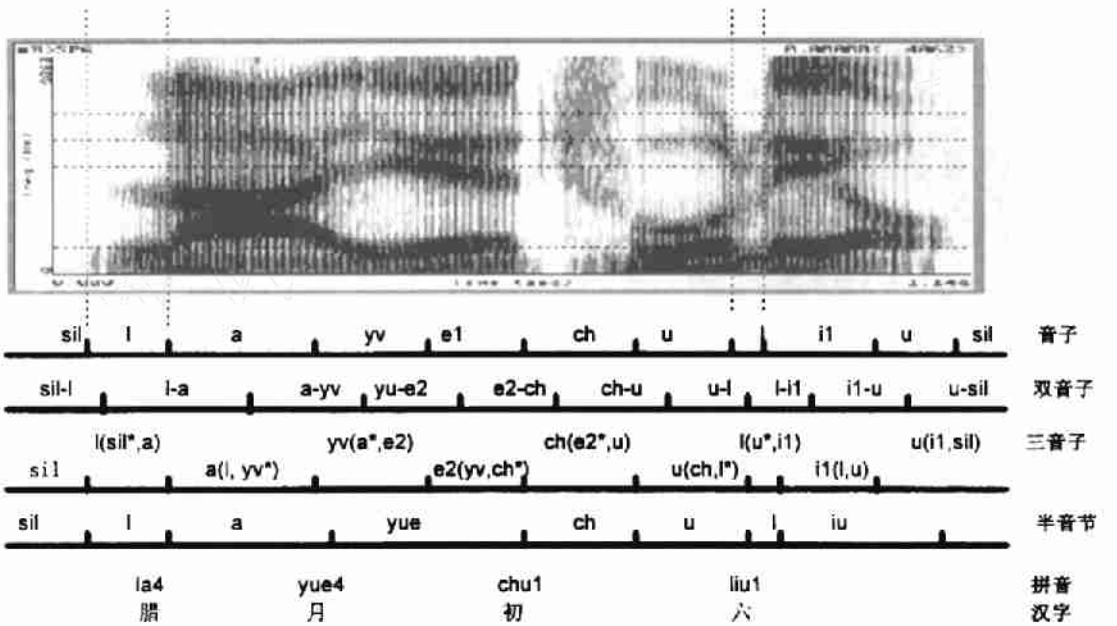

图 1 语音“腊月初六”的语图及其对应的不同的语音单元

在国家“七五”、“八五”期间，中国社会科学院语言所受“863”等方面资助，对全部单音节和两音节间的共振峰过渡进行了全面的研究，特别是音节间的共振峰过渡为本研究提供了重要参考. 利用普通话两音节结构表[16]，陈肖霞研究了两音节组 c1v1c2v2 中唇、唇齿音 (b, p, m, f)、舌尖音 (d, t, n, l)、舌根音 (g, k, h) 三个部位的协同发音情况，v2 为 a, i, u 三个极端部位，v1 为 22 个韵母，结果表明，唇音组明显存在 v2 通过 c2 影响 v1 的现象[17]. 颜景助研究了两音节结构 c1v1c2v2 中 c2 为零声母的情况[18]，这时的共振峰过渡是十分明显的. 孙国华研究了 c1v1c2v2 中 c2 为 /z, c, s, zh, ch, sh, j, q, x, r/，以及 v1 后有鼻尾 n, ng 的情况，给出了许多规律性的结论[19]. 综合他们的研究结果，可规纳为如下规则，在两音节结构 c1v1/n1-c2v2 中，(1) c2 为零声母 (元音) 或浊声母 (m, n, l, r) 时，v1 向 c2 的共振峰过渡是十分明显的. (2) c2 为塞音或塞擦音时，由于存在着闭塞段，受 v2 的影响相对较小. (3) 鼻尾 n, ng 和其它韵相比，n1 向 c2 的过渡有着特殊性.

许毅[20] 指出：音联是语流中各语音单元之间的连接和分界，并归纳了四个不同等级的音联：(1) 闭音联 —— 音节内各音子间的音联. (2) 音节音联 —— 音节间的音联. (3) 节奏音联 —— 词组间的音联. (4) 停顿音联 —— 语流中的短暂停顿. 两音节间的音联属第二类 —— 音节音联，将上述的 c1v1/n1-c2v2 结构放在连续语流中，将受到语句的韵律结构影响，而且还有节奏音联和停顿音联尚待研究. 语流中情况极为复杂，不同的语速会使不同层次的音联发生转化.

1.3　连续语流中的停顿和韵律结构

连续言语 (continuous speech) 中的停顿 (pause) 是指两个音节间的空白段 (silence)，当停顿的时间长到一定程度，两边的音段才相互不发生关联。事实上，一个话语 (utterance) 包含着不同的韵律结构 (prosodic structrue)。韵律结构与句法、语法结构有着一定的对应关系，但又不是完全一一对应，韵律结构规定了上述的几种音联。韵律结构由大至小有如下几种：

韵律结构		句法结构
	大	
语调短语 (intonational phrase)	|	句子 (utterance)
音系短语 (phonological phrase)	|	短语 (phrase)
韵律词 (prosodic word)	↓	词 (word)
音步 (foot)		语素 (morpheme)
	小	

韵律结构的具体分类有所不同，它们反应了连续语音的节奏。韵律结构之间在感知上存在着边界 (break)，这种边界绝非仅仅是停顿，它对应的声学表现可能是停顿、音高曲线的变化 (pitch movement / F0 reset) 以及边界之前的音段延长 (pre-lengthening / final lengthening)[21,22]。停顿对应的声学表现是无声段 (silence)。一些研究指出，语法结构引起的停顿发生在句子之间或较复杂的句子内部的语块之间[23]，平均句长为十个词左右。这表明停顿发生在较大的韵律结构间。李爱军[24] 也对新闻广播进行了停顿的研究，结果表明，在句子中被听为有边界的位置，只有 30% 左右存在着真正的停顿。这说明听感上的停顿不一定是无声段，还可能是其它声学征兆。因此即使在句子中，词间、音段间都存在着过渡。我们可以得出这样的结论：一个不是很长的连续话语内部，几乎所有音节间都存在着声学上的过渡。

1.4　汉语普通话的基本音子

我们提出 37 个音子作为普通话的基本音子，表 1 给出这些基本音子及其存在环境，其中 sil 为无声段 (silence 的缩写)。

由表 1 可以看出，所有辅音以及多数元音都和汉语拼音字母对应的音位一致，只有 /a,i,o,e/ 给出多个变体，它们及其它音子的变化可通过左、右语境来表现。

众所周知，衣服的衣 (yi) 中的高元音 /i/，与辅音 z、c、s 后接的舌尖前元音，以及辅音 zh、ch、sh 后接的舌尖后元音是不同的，我们分别记作 i1、i2、i3。低元音 a 在不同语境下是十分不同的，在一般情况下为通常的 /a/，在韵母 ai、an 中记为 a2，在韵母 ang、ao 中记为 a3。同样，歌、河中的韵母记为 e1；韵母 ei、ie、yve 中的 e 记为 e2；韵母 en、eng 中的 e 记作 e3。韵母 uo、ou 中的 o 分别记作 o1、o2。在音节内，给出不同的语境，就可基本区别 a、i、o、e 的不同变体。但当考虑到它们对前面的辅音的影响时，就有必要将它们分成几个音子，例如，ga、gai 与 gao 中的 a 是不同的，其中声母 g 向韵母的过渡也是明显不同的，我们用 a1、a2、a3 就可以描述这种区别。图 2 为辅音 g 与 a、o、e、u 的不同变体相拼的情况，从过渡段的音轨走向可看出区别性，元音 u 的不同变体的区别性不如 a、o、e 明显，在此我们只设一个音子。

关于普通话基本音子的个数应当是多少，存在着不同看法，争议多在于元音.

表 1 基本音子、对应的国际音标（IPA）及其出现的例子

音子	IPA	例子	音子	IPA	例子
a1	a	b*a*, *a*	m	m	*m*a, *m*i
a2	ɛ	*an*, *ai*	n	n	*n*a, *n*i
a3	ɑ	*ang*, *ao*	ng	ŋ	u*ng*, q*ing*
b	p	*b*a, *b*u	o1	ɔ	w*o*, p*o*
c	ts^h	*c*i, *c*a	o2	o	*ou*
ch	$tʂ^h$	*ch*a, *ch*u	p	p^h	*p*o, *p*ang
d	t	*d*a, *d*i	q	$tɕ^h$	*q*i, *q*u
e1	ɤ	h*e*, g*e*	r	ʐ	*r*i, *r*ong
e2	e	*ei*, y*e*, yu*e*, i*an*	s	s	*s*i, *s*un
e3	ə	*en*, *eng*, u*en*	sh	ʂ	*sh*i, *sh*ui
er	er	*er*	t	t^h	*t*e, *t*i
f	f	*f*a, *f*ei	u	u	*w*an, b*u*
g	k	*g*e, *g*ei	x	ɕ	*x*ia, *x*i
h	x	*h*e, *h*a	yv	y	*u*, *yu*e
i1	i	b*i*, x*ia*	z	ts	*z*i, *z*uo
i2	ɿ	z*i*, c*i*, s*i*	zh	tʂ	*zh*u, *zh*ong
i3	ɻ	zh*i*, ch*i*, sh*i*	sil	无声段	
j	tɕ	*j*i, *j*in			
k	k^h	*k*e, *k*a			
l	l	*l*a, *l*ang			

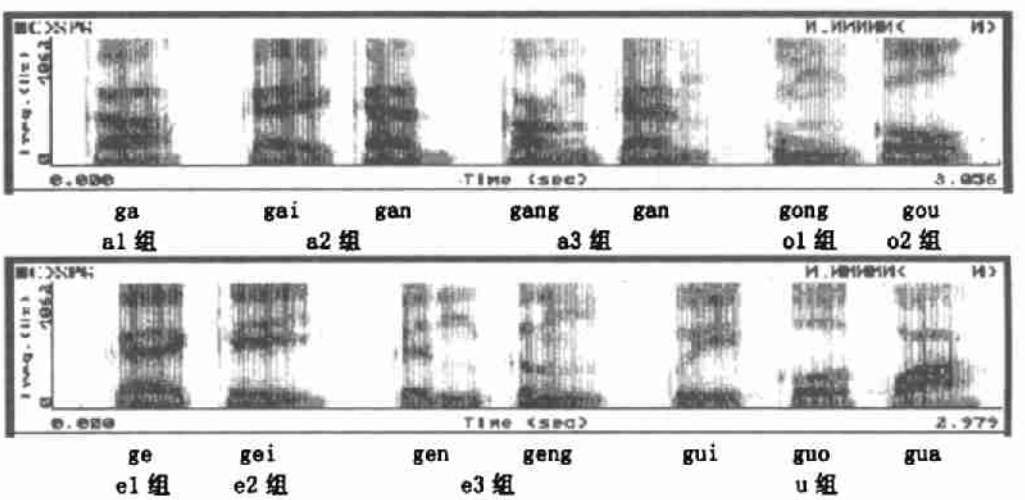

图 2 辅音 g 后接音子 a1 、 a2 、 a3 、 o1 、 o2 、 e1 、 e2 、 e3 、 u 的语图

1.5 双音子和三音子

音子在连续语流中很难以稳定形式存在，它受左、右音子的影响，同时又会影响相邻音子，因此，描述连续语音，用音子是不够的. 在语音学层面，以基本音子为基础，考虑两个相邻音子，就可形成一个双音子 (diphone)；同时考虑左、右相邻音子，就可形成三音子 (triphone)，左、右不同语境可扩大到声母、韵母，并由此构成半音节 (semi-syllable)，对应到声学层面，双音子

描写的是两个相邻音子的稳定段之间、包括了过渡段的音段；三音子描写的是一个音子的稳定段及向左、右两边音子的过渡部分；半音节对应着一个音节的声母段或韵母段。

图 1 中“腊月初六”的语图下面详细地给出了对应的文字、拼音、音子、双音子、三音子以及韵－声母的标注序列 (图中符号“*”表示音节边界)，从中可比较不同语音单位之间的关系。以图 1 中的两个“l”为例，如用三音子描写，前一个“l”为“l(sil*, a)”，后一个“l”为“l(u*,i1)”，其中 * 代表音节末尾音子或音节起始音子，即一个音节边界。

要系统地描写汉语普通话连续语音，我们应当给出基本音子存在的所有语境。汉语的音位配列规则给出了音节内音子的所有语境。至于音节间，原则上说任何音节都可以同所有其它音节搭配，如果不考虑声调的区别，普通话约有 400 个音节，就会有 160,000 种音节搭配。这样的数目无论对于语料设计，还是语音合成、识别模型的建立都是太庞大了，而且也不一定必要。如果从基本音子出发，音节间的双音子有 400 多个，三音子则有近万个。音节间的韵母－声母的搭配，有 38 韵 33 声 (含零声母) 一千多个，工程上容易实施。

语音单元的选择是语音识别和语音合成建模时要考虑的重要问题。采用的语音单元可以是：音子、多音子、半音节、音节、词，甚至句子。简单构成的基元 (如音子)，较容易得到足够的训练数据，但模型的准确性方面却不能令人满意。较复杂的语音单元增加了模型的准确性，同时模型的复杂性也随之增加了。实际上语音单元的选择就是要在模型的准确性和语料库的数据共享之间作出合理的折衷，当然，这种选择离不开语音学知识的指导。

双音子数目为数百个，在有限的语料中较容易达到穷尽，用它描写连续语音序列，虽比音子更能反应音变，但仍显粗糙。三音子较详尽地描写了连续语音的音变和过渡，著名的 SPHINIX 系统就是在语音学知识的指导下选用了三音子，使英语语音识别获得了明显的提高[25]。许多英语识别系统都采用三音子为识别单元。然而，数目庞大的三音子对系统来说是沉重的负担。一些语音合成系统的语音单元也采用了部分三音子加双音子[26]。根据我们的统计结果，由 37 个基本音子构成的三音子的数目有近万个，大多数理论上可能存在的三音子在自然语言中很少出现，因此，如果采用三音子作为识别基元，应当对其进行合理的归类。根据汉语的特性，声母、韵母及其之间的过渡，或称半音节也是值得考虑的语音单位，它同三音子一样，较好地描述了语境的搭配。半音节组合结构约有两千多个。

2　语音识别语料库所考虑的语音现象

2.1　音段方面的语音覆盖

2.1.1 不含声调区别的音节：除去 m、lo、yo、hm、nou、nie、tei、kei、rua 口语中罕见的音节，共有 401 个。它覆盖了音节内的所有语音现象 —— 包括了所有的音子，元音、辅音，声母、韵母，以及音节内的双音子、三音子，因此，后面考虑其它语音现象时，只需对音节间进行处理。

2.1.2 音节间的双音子：可作韵尾的音子 (有 12 个) 和音节起始音子 (33 个) 的搭配，共有 415 个。

2.1.3 音节间的三音子：如果考虑上述的 37 个基本音子，共有 8000 多个三音子，数目相当大，并且很多三音子在实际语言中极为罕见。对于语音识别，所考虑的发音方法和发音部位应尽量全面。利用目前关于音联的研究结果，可按以下规则对它们进行归类和精简。

(1) 单音子音节构成的三音子

单音子音节有：阿，一，无，语，恶，欧，二等 (其中“一、无”前有半元音，在此仍将它们归为单音子音节)，与它们有关的三音子是它们与前后的不同音节搭配形成的，如：a(?*, ?*)，

i(?*, ?*) …… 共有 3000 多个. 由于它们在连续语流中保持着相对的稳定段, 可将这样的三音子看成两个双音子, 如: a(i*, n*)→ i-a, a-n .

(2) 中心音子为韵尾的三音子 (右边为音节边界: ? (?, ?*))

将第二音节的声母按发音部位归类, 因为前一音节的韵尾向相同部位的声母过渡时, 变化基本相同 (鼻尾例外), 图 3 给出"起步 一批 积木 欺负"四个词组的语图, 这四个词组的第一个音节的韵尾都是"i", 第二个音节的声母分别是"b, p, m, f", 属同一发音部位, 由图可见这四种情况下的"i"的过渡情况相差无几. 因此, 前一音节的韵尾向后一音节的三十多个韵头的过渡可简化成向六个部位过渡:

		双唇 {b,p,m,f}			{b,p} + m,f
		舌尖 {d,t,n,l}			{d,t} + n,l
韵尾	→	舌根 {g,k,h}	鼻尾 (n,ng)	→	{g,k} + h
		齿间 {z,c,s}			{z,c} + s
		齿龈 {zh,ch,sh,r}			{zh,ch} + sh,r
		舌面 {j,q,x}			{j,q} + x

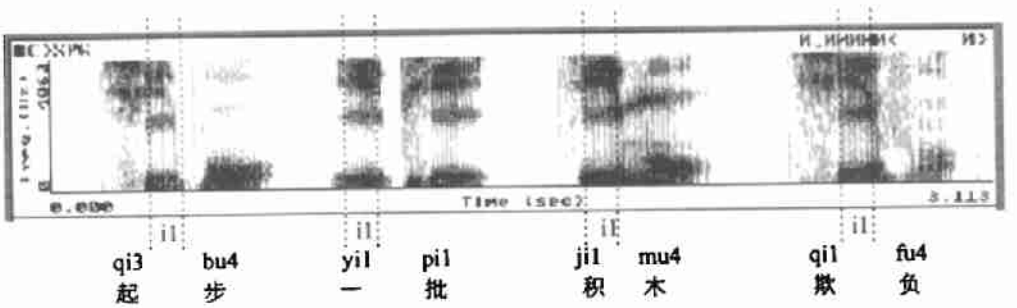

图 3 语音"起步 一批 积木 欺负"的语图

(3) 中心音子为声母的三音子 (左边为音节边界: ? (?*, ?))

对所有零声母情况必需考虑所有搭配的三音子. 所有的塞音和塞擦音前都有一闭塞段, 可看作与孤立字发音的情形相同. 图 4(a) 是单音节"把"和词组"一把"的对比语图, 图 4(b) 是单音节"大"和词组"北大"的对比语图. "把、大"的声母"b,d" (声母段很短, 过渡段表现了其特征) 在单音节和词组情况下的声学表现无大差别.

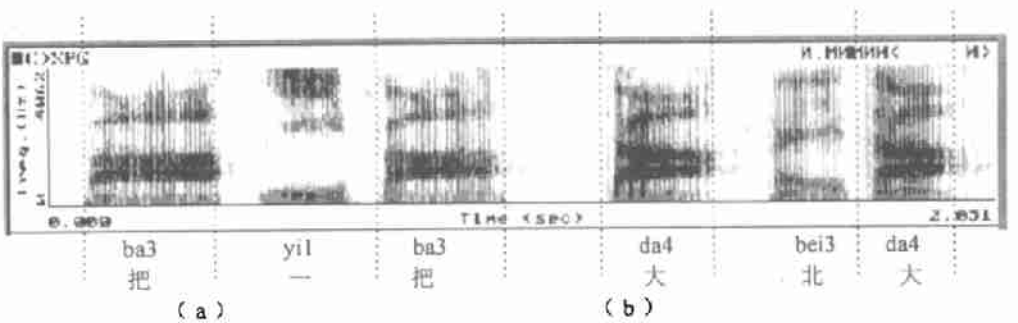

图 4 单音节"把、大"和词组"一把、北大"的对比语图

(4) 生僻搭配不计

与音节： eng 、 o 、 ei 、 yo 、 lo 、 hm 、 tei 、 nou 、 kei 、 rua 等生僻音节形成的三音子可以不考虑，它们多为自由口语中使用，在朗读口语中不出现.

经上述精简，三音子的数目为 3035 个，它们基本覆盖了主要音变现象.

2.1.4 音节间的韵母 - 声母 (韵 - 声) 搭配

38 个韵母与 32 个声母 (包括零声母) 搭配，并采用与三音子同样的简化方法，共有 781 个韵 - 声搭配.

以上考虑的几种语音平衡基本满足了语音识别各种建模方式的需要.

2.2 韵律方面的语音现象

如 1.3 节所述，连续语音中存在着不同的韵律结构，这些结构和句法结构有一定的关系，但又不是一一对应. 韵律结构对于提高语音合成系统的自然度、进行语音识别系统的后处理是十分重要的. 为了使语料包括不同的韵律结构，语料设计还应当包括汉语的不同句型[27] . 句型可概括为 18 大类，见表 2 .

表 2　普通话基本句型表

1 名词谓语句	11 “被” 字句
2 形容词谓语句	12 存在句
3 主谓谓语句	13 连动句
4 主 || 动词	14 兼语句
5 主 || 动词 + 宾语	15 疑问句
6 主 || 动词 + 直接宾语 + 间接宾语	16 独词句
7 主 || 动词 + 补语	17 动词 + 从句
8 “有” 字句	形容词 + 从句
9 “是” 字句	18 特殊句式 (感叹句，祈使句)
10 “把” 字句	

3　连续语句的选取方法

将 1993 、 1994 两年的 “人民日报” 等作为语料库的原始语料，从中选出连续语句作为朗读语料，每个句子的长度限制在 20 个音节以内，这个句长限制了复杂的韵律结构，基本保证了句内基本不出现停顿，使问题相对简化. 语音数据库是以一个个话语 (utterence) 为单位管理的，每个话语有对应的汉字、拼音和波形. 在我们的数据库中，一个话语对应的是一个句子或一个短语.

在前述语音学规则条件的规定下，是不能随机挑选句子的，而是由一套程序，加上人工干预自动实现的， Greedy 算法[28] 就是语料库文本设计的经典方法. 在汉语语料设计中，也多在一定范围内采用了自动方法实现[10−12] . 图 5 为本研究使用的自动选句流程图，分句的原则是凡遇到 “？” 、 “！” 、 “，” 、 “ .” 、 “；” 、 “：” 等标点符号均自动断为一句. 分词的目的是将汉字文本自动转换为拼音文本. 人工干预的作用是消除自动分句、自动分词、拼音自动转换当中的误差，排除内容上不合适的句子.

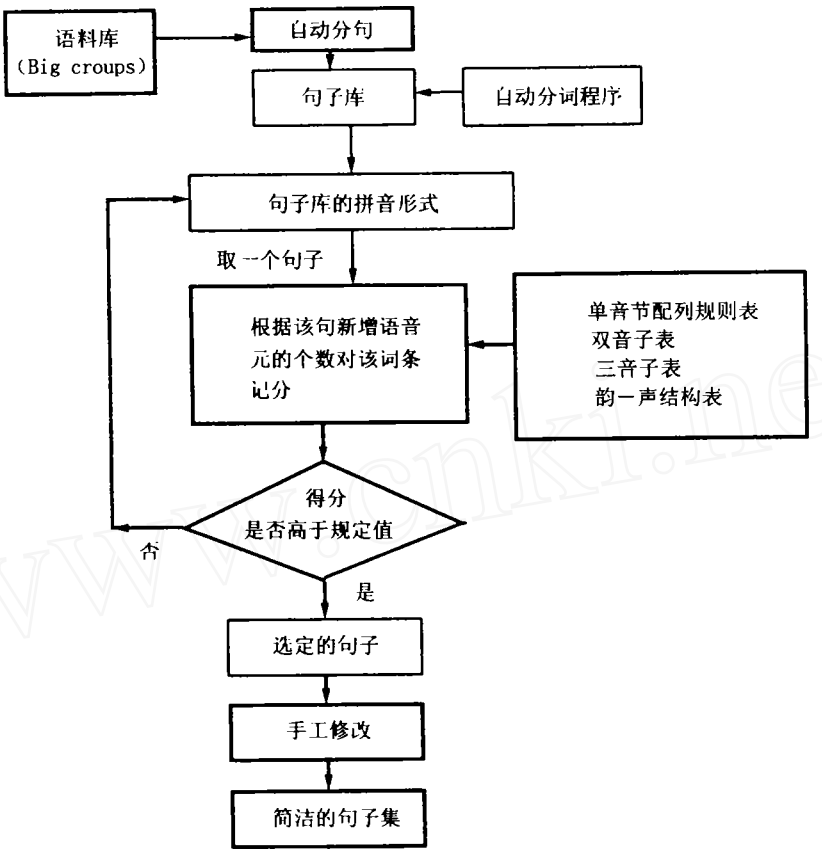

图 5 连续语句自动选取流程图

4 结果及讨论

4.1 语料选取结果

图 6 给出语料的来源，表 3 和图 7 表示语料的统计结果。

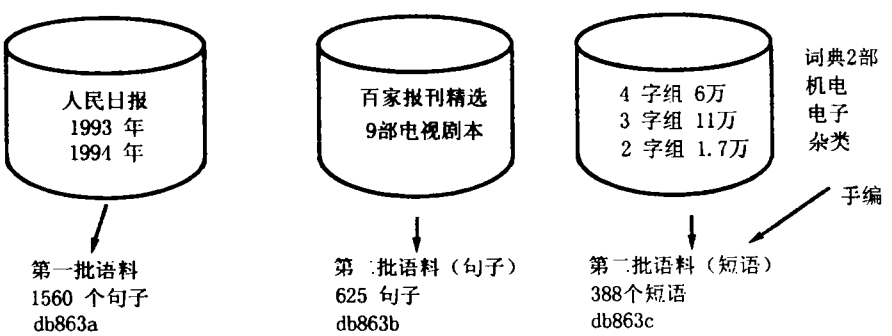

图 6 语料原始来源及选取情况

表 3　连续语音识别数据库语料的统计结果

	语音单位	db863a		db863b		db863c	
		个数	覆盖率	个数	覆盖率	个数	覆盖率
音节	401	396	98.7%	397	99%	400	99.9%
音节间双音子	415	415	100%	415	100%	415	100%
音节间三音子	3035	2128	70%	2644	87%	3023	99.6%
韵－声结构	781	668	85.5%	724	92.7%	781	100%
句型	17	17	100%	17	100%	17	100%

图 7　表 3 结果的直方图表示

4.2　关于未出现的语音现象

由表 3 可见，双音子和韵－声结构是完整的。401 个无调音节有一个未出现，该音节为“这”的另一读音“zhei”，文本中对“这”的拼音全部注为“zhe4”，它的读法因说话人的习惯而不同。三音子有 12 个未出现，它们是与句尾字“呢”不出现在句尾时有关的三音子、“么”出现在句首的三音子、以及与“zhei(这)”有关的三音子。它们在实际语流中确为罕见。

4.3　关于原始语料库

根据我们的统计，“人民日报”内容多为政论性的文章，其中的语音信息量不够大。它虽含盖了很容易获得的全部无调音节和音节间双音子，但只覆盖了 70% 的音节间三音子和 85.2% 的音节间韵－声结构 (表 3 中的 db863a)。“百家报刊精选”包括十种报刊，内容广泛，语音信息量较大，在对它进行搜索之后，音节间三音子和韵－声结构的覆盖率分别达到 87% 和 92.7%(表 3 中的 db863b)。这时其它剧本、词库已不能提供太多的内容了，专业词库也不能提供比大量篇章更多的语音现象，通过手编才达到最后结果 (表 3 中 db863c)。

4.4　语音数据库文本设计与录制的界面

如果将设计的文本直接交予发音人，发音人并不能完全理解文本设计的意图，这里主要是指语气、情感、句型、轻重等对韵律结构影响很大的因素，文本设计与发音人的沟通不仅涉及语言学、语音学的问题，而且还有认知科学和心理学方面。这一问题有待于研究。

4.5　语音数据库语料设计存在的问题和今后的任务

如前所述，本研究进行的语料设计对象是朗读言语的语料，着重考虑的是音段方面的语音学问题，至于超音段现象，仅通过对句型的考虑，使其自然达到平衡。这样做的原因在于我们对于音段的研究相对成熟，而连续语句的韵律结构尚未揭示清楚。通过观察、分析反馈的语音数据，韵律结构、韵律对音段的影响、语速对韵律结构的影响以及广义的语调模型等方面都大有文章可做。即使在音段方面，也仍有许多工作要做，例如连续语句的音变规律、韵律结构对音段的影响等。因此，我们参与进行的语音数据库不仅服务于语音合成、识别系统，而且为连续语音的研究

提供了极好的材料。

致谢

本文在音段方面的工作是在中国社会科学院语言研究所关于音节间共振峰过渡的研究成果基础上进行的，特别是在语音单元的确定方面得到了陈肖霞同志、林茂灿教授的具体帮助。关于语音识别数据库语料的设计思想受益于陈础坚教授指导下的、在香港大学计算机系的工作实践。中国科学院声学所吕士楠教授为本文提出了许多宝贵意见。

参 考 文 献

1 Fisher W M, Doddington G R. The DARPA speech recognition research database: specification and status. Proceeding of the DARPA Speech Recognition Workshop, Palo Alto, CA, 1986: 93—99

2 Zue V W, Cyphers D S, Kassel R H *et all.* The development: design and analysis of the acoustic-phonetic corpus. Proceedings of ICASSP-86, Tokyo, Japan, 1986: 8—11

3 Stephanie Seneff and Victor Zue W. Transcription and alignment of the TIMIT database. The Second Symposium on Advanced Man-Machine Interface through Spoken Language, Oahu, Hawaaii, 1988: 20—22

4 Pat Keating, Blankenship B, Byrd D *et all.* Phonetic analyses of the TIMIT corpurs of American English. Proceedings ICSLP92, 1992 ; **1**: 823—826

5 Kim Silverman, Mary Beckman, John Pitrelli *et al.* TOBI, A standard for labeling English prosody. Proceedings of the International Conference on Spoken Language Processing, 1992 ; **2**: 867—870

6 Tsuyoshi Morimoto, Noriyoshi Uratani, Toshiyuki Takezawa *et al.* A speech and languge database for speech translation research. Proceedings of the International Conference on Spoken Language Processing, 1994, **4**: 1791—1794

7 Jan P H, van Santen, Buchsbaum A L. Methods for optimal text selection. Proceedings of Eurospeech '97, 1997, **2**: 557—561

8 Chih-mei Lin, Shrikanth Narayanan, Russell Ritennour. Database Management and analysis for spoken dialog system: Methodology and Tools. Proceedings of Eurospeech '97, 1997, **5**: 2199—2202

9 Ikuo KUDO, Takao NAKAMA, Nozomi ARAI *et al.* The data collection of voice across Japan (VAJ) project. Proceedings of the International Conference on Spoken Language Processing, 1994, **4**: 1799—1802

10 Wang H -M, Chang Y -C, Lee L -S. An algorithm for automatically selecting phonetically balenced sentences from a large corpus for training and testing a speech recognition system. Proc.Int. Conf. Computer Proc. Oriental Lang. (Korea), 1994: 507—510

11 孙甲松，王作英，王 侠等. 连续语音训练词表的构造. 第二届中国计算机智能接口与智能应用学术会议论文集， 1995: 116—121

12 曲 菲，黄泰翼，张希军. 汉语综合语音库语料设计. 第四届全国人机语音通讯学术会议论文集， 1996: 337—341

13 Li Wenxian, Zu Yiqing, Chan C. A Chinese speech database (Putonghua Corpus). Proceedings of SST-94, 1994: 834

14 石 锋. 语音学探微. 北京：北京大学出版社， 1990

15 G 方特， J 高奋. 言语科学与言语技术. 张家骏等译，商务印书馆， 1994

16 曹剑芬. 两音节音联字表. 语言文字应用. 1997; **1**: 60—68

17 陈肖霞. 汉语普通话两音节 CVCV 间 C2 为三个发音部位的逆向协同发音声学研究. 中国语文， 1997; **4**: 54—63

18 颜景助. 前音节为元音尾和后音节为零声母的普通话双音节的音节间共振峰过渡研究. 语音研究报告，1994-1995: 41—53

19 孙国华. 普通话两音节中 V1-Z 间的共振峰过渡. 第三届语音学研讨会论文集， 1996: 108—110

20 许 毅. 普通话音联的声学语音学特性. 中国语文， 1986; **5**

21 Fant G, Kruckenberg A. Preliminaries to the study of Swedish prose reading and reading style. STL–QPSR 2, 1989

22 Eleonora Blaauw. The contribution of prosodic boundary marks to the perceptual difference between read and spontaneous speech. Speech Conmmunication 14, 1994: 359–375

23 郭锦浮. 汉语句子长度、语速与结构停顿. 计算机时代的汉语和汉字研究学术研讨会论文摘要, 1995: 17

24 李爱军. 普通话新闻广播话语中的停顿. 中国声学学会 1997 年青年学术会议论文集, 1997: 262—266

25 Lee K F. Automatic speech recognition: the development of the SPHINX system. Kluwer Academic Publishers, 1989

26 Peri Bhaskararao. Subphonemic segment inventories for concatenative speech synthesis. Fundamentals of Speech Synthesis and Speech Recognition, Edited by ERIC Keller, University of Lausanne, Switzerland, JOHN WILEY & SONS, 1994: 69—85

27 罗振声, 郑碧霞. 汉语句型自动分析和分布统计算法与策略的研究. 中文信息学报, 8(2): 1—13

28 Cormen T H, Leiserson C E, Riverst R L. Introducton to algorithms, The MIT Press, Cambridge, Massachussetts, 1990