

# 连续语音数据库设计的科学性问题\*

祖漪清

**摘要** 本文描述了连续语音数据库的基本研制过程, 提出在文本设计等方面使用的语言学/语音学知识, 介绍了选择文本时满足语音单元覆盖的算法, 并给出句长分布、语音单元的覆盖等结果。本文还指出, 在篇章的断句和连续语音的韵律结构方面还存在许多问题, 特别是汉语的句法与韵律的界面方面的研究十分薄弱。本文还介绍了基于语音数据库的语音研究, 并给出在 863 语音数据库上做的一些初步结果。

## Issues in the Scientific Design of the Continuous Speech Database

Zu, Yiqing

The design of the continuous speech database suitable for training and testing Chinese speech recognizers has been reported. This paper outlines the general approach and processing of content of database. At first we select a big corpus and all paragraphs in the corpus are segmented automatically into a sentence list. Then the sentences in the list are collected by counting the appearances of speech units according to a set of phonetic rules. Finally, using well known Greedy algorithm we can get a favorable sentences set.

At present stage, considerable attention has been given to segmental variations in reading style. To describe very complex variances in continuous speech, we propose the following speech units inventories: (1) 401 syllables without tone; (2) 415 inter-syllabic diphones; (3) 3035 inter-syllabic triphones; (4) 781 inter-syllabic final-initial structures. We also give 17 sentence patterns to include the prosodic phenomena. With automatic method 2393 sentences and 388 phrases are collected by above phonetic rules from a big corpus -- recent years' "people's Daily" and so on, as the scripts of speech recognition database. This set of sentence covers 99.8% syllables without tone, 100% inter-syllabic diphones, 99.6% inter-syllabic triphones and 100% sentence patterns.

200 speakers' speech data have been collected (by Science and Technology University of China). Among them 1560 sentences of 3 speakers had been labeled manually by Chen Xiaoxia. We provide the statistics on duration of consonants, vowel and *ev* transitions. There are two aspects which influence the quality of continuous speech database. The first one is segmentation of sentences in corpus. Since every utterance is limited within 20 syllables while the sentences of corpus includes more than 20 syllables, we must segment the long one into several utterances. Where are boundaries? We can only segment the corpus automatically only by the punctuation marks “。”, “;”, “:”, “!”, “?”, “:”. The sentence which includes more than 20 syllable must be segmented manually. To solve this problem we must understand where is the boundaries of prosodic structures and the relationship between syntactic structures and prosodic structures. The second problem is about speech units. As mentioned above, we use syllables, inter-syllabic diphones, inter-syllabic triphones and final-initial structure as speech units to control the coverage. None of those four sets of speech involved prosodic information which includes tones, stress and prosodic structures.

This paper will also report following results: (1) the distribution of sentence length; (2) how to deal with segmentation of paragraphs; (3) the coverage of sentences sets created by different collecting methods. (4) introduce some phonetic researches based on speech database.

### 1. 引言

---

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

我们与中国科学技术大学合作已基本完成了国家 863 计划支持的汉语连续语音识别数据库的设计和建立。我们负责语料设计和音段标音,关于语料设计方面的工作我们已经作了总结(祖漪清、李爱军 1997, Zu 1997)。目前阶段的连续语音数据库录制的是朗读言语(read speech),主要考虑音段中的语境音变。该语音数据库的规模和质量在科学、合理、简洁、有效等方面,均比过去有较大进步。通过建立语音数据库的工作,我们认为在语音数据库设计的科学性方面还存在问题有待探讨:在语料设计的过程中,篇章的切分和断句还不能全部实现自动化,只能在自动粗分后再手工细分;在语音现象的归纳方面,我们使用的语音学规则基本局限在音段方面,连续语句的韵律现象,会同句法结构存在一定关系,我们试图通过对不同句型的考虑,使其达到自然覆盖。我们对于音段的研究相对来说较为成熟,而连续语句中的韵律结构等超音段规律以及它们与句法结构等因素的关系均未揭示清楚。这些问题都说明:我们对连续语音的规律尚未揭示清楚。

欧洲对连续语音的研究早在八十年代以前就开始了,瑞典著名言语工程学家 Fant 对连续的朗读语句的重音、节律、短语间的停顿以及韵律边界前的延长都做过定量研究(Fant & Kruckenberg 1989),近年又提出了停顿时长的量子效应,以及重音音节、非重音音节和音位音段的量子效应(Fant & Kruckenberg 1996)。法国语音学家 Mario Rossi 在句法结构和韵律结构方面做了深入的研究,并因此获得 1997 年欧洲言语通讯与技术学会颁发的奖章,他定义了连续语句中句法结构的线性构建和句法成分的层级,提出了用语用-韵律模型、句法-韵律模型和节律模型来确定韵律结构(Rossi 1997)。在这一领域中,汉语研究在连续变调方面给出了研究结果(Shih 1986, Hong 1987, 沈炯, 1994),但总体来说还十分薄弱。

## 2. 863 连续语音数据库制作的基本情况

图 1 是语音数据库建立过程的框图。首先要选择一个大的语料库,863 连续语音识别语音数据库使用的是 1993、1994 年《人民日报》、《百家报刊精选》等。根据语音数据库的目的和要求可选择内容、文体风格不同的语料库。《人民日报》是我国发行量最大的报刊,其电子文本最容易获得,其内容多为新闻报导和论述文章,涉及政治、经济、文化、体育等各个方面,有一定的信息量,文体比较正规。在这个大语料库的范围内,按照语言学/语音学原则挑选出满足一定覆盖量的句子集,作为语音的朗读文本。发音人按照文本朗读,就产生了语音数据。语音数据库加工的主要内容是对语音的波形数据进行标注,包括音段标注和韵律标注。863 连续语音数据库只做了音段标注,即在波形上对构成句子的所有音段标出起始点和结束点的时间位置。图 1 虚线框部分由中国社会科学院语言研究所完成。

语音数据库的制作看似简单,但整个建库过程十分繁琐,其中涉及到言语的理解和产生两个方面的问题。要真正实现科学性,需要语言学、语音学、心理学、声学及言语技术等综合知识与技术。

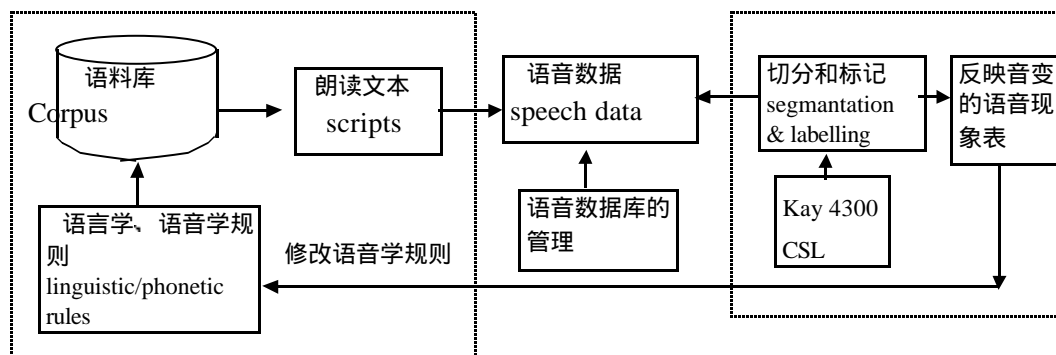


图 1 863 连续语音识别数据库的建立过程

## 3. 语料库朗读文本的设计

863 连续语句文本的设计经历如下步骤。

### 3.1 断句

文本处理的第一步是断句，目的是为文本处理、朗读以及语音数据的处理带来方便。最为合理的是以句号以及内容上完整的问号、感叹号为标志断句，因为在此处必然对应话语中的停顿。863 语音识别数据库的连续语句的断句点是在如下标点符号处：

句号：“。”  
逗号：“，”  
分号：“；”  
感叹号：“！”  
问号：“？”  
冒号：“：”

朗读时，由这些标点符号断开的句子后一般会有明显停顿，因此在一定意义上这样断句是合理的。为了让发音人不感到困难，原则上把句长限制在 20 个音节以内。以逗号断句情况较为复杂，它对应的情况可能是句法结构较完整的句子，也可能只是一个成分，不能作为独立的朗读句。以 1994 年《人民日报》第四季度文本为例，在逗号处断句与否，句长分布情况不同，如果在逗号处断句，句长(音节数)集中在 5-20 个音节内；如果逗号处不断句，句子的音节数在 30-50 内最多。见图 2。

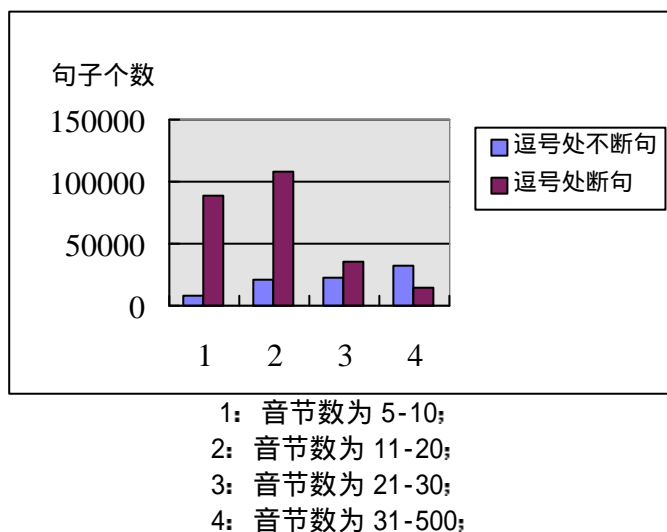


图 2 1994 年第四季度《人民日报》文本  
小于 50 个音节的句长分布

### 3.2 拼音转写

为了确定、统计连续语流中的语音现象，必须将文字转写成拼音，由于汉语有许多多音字、许多复杂的用法以及连读变调，要使拼音转换准确，需要有一个很好的词库用于词的切分，以得到准确的拼音文本。

### 3.3 由拼音生成语音单元

对应音位变体的声学表现定义为音子，我们将它看做连续语音的基本成分，并设它们是连续语音中的最小音段。普通话的基本音子有 37 个：

a1 ( ba,wa), a2 (an,ai), a3(ang,ao), b, c, ch, d, e1(ge), e2(ei,ye), e3(en,eng),  
er, f, g, h, i1 (bi,xia), i2 (zi,ci,si), i3 (zhi,chi,shi), j, k, l, m, n, ng, o1( wo,po),  
o2 (ou), p, q, r, s, sh, t, u, x, yv, z, zh, sil(无声段)

通过基本音子可以方便地描述音节内和音节间的语音现象。有了一个如表 1 形式的音节构成表就可以将一个连续话语转写成语音单元串，我们可以任意定义语音单元，如音子、双音子、三音子以及半音节等。

表 1 音节构成表

音节	由音子生成音节	由声母、韵母生成音节
.....	.....	.....
chu	ch-u	ch-u
chuai	ch-u-a2-i1	ch-uai
chuan	ch-u-a2-n	ch-uau
chuang	ch-u-a3-ng	ch-uang
.....	.....	.....

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，偏离了它孤立存在时的位置。在声学层面上描写语流中的音变现象以及音段间的过渡仅用音子是不够的。在 863 语音连续识别数据库语料设计时，我们考虑了如表 2 所示的几种语音单元的覆盖，其中三音子和韵母-声母结构的数目是根据发音方法和发音部位的变化规律归并而得到的。连续语音中另一个重要的语音现象是韵律结构，一个话语包含着不同的韵律结构，它们与句法、语法结构有着一定的对应关系，但又不是完全一一对应，如表 3 所示。韵律结构的具体分类有所不同，它们反映了连续语音的节奏以及更深层次的内含。韵律结构之间在感知上存在着边界(break)，这种边界对应的声学表现可能是停顿(pause)、音高曲线的变化(pitch movement)，以及边界之前的音段延长(pre-lengthening 或 final lengthening)(Blaauw 1994)。

如上所述，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，863 语音识别和语音合成语料设计中参考清华大学中文系的汉语句型(罗振声、郑碧霞 1994)考虑了 18 大类不同句型。见表 4，其中的例句取自语音合成数据库的独白部分。

表 2 863 语音数据库的语音单位

语音单位	数目
不含声调区别的音节	401
音节间的双音子	415
音节间的三音子	3035
音节间的韵母-声母	781

表 3 韵律结构及句法结构

韵律结构	句法结构
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)
音节 (syllable)	语素 (morpheme)

表 4 普通话基本句型表(句型分类由李智强提供)

句型	例子
1 名词谓语句	全长二百五十一公里，
2 形容词谓语句	这姑娘很大方。
3 主谓谓语句	日本经济力量雄厚，
4 主  动词	一个较为完善的粮食市场体系已经在我国初步形成。
5 主  动词+宾语	我们不鼓励也不支持核扩散。
6 主  动词+直接宾语+间接宾语	可以方便地告诉计算机任何你想做的事情，
7 主  动词+补语	解决卖粮难的办法在于开动脑筋“想”。
8 "有"字句	一部十层电梯就有八九千个零件！
9 "是"字句	中国是奥林匹克运动最积极的支持者。
10 "把"字句	社会上常把它与计划经济体制联系在一起。
11 "被"字句	她们的父母于5年前先后被病魔夺去了生命。
12 存在句	大厅里坐着年青人。
13 连动句	一位老汉存了两千元钱准备给儿子娶媳妇。
14 兼语句	我请她介绍一下客店的情况。
15 疑问句	过去我不照样跟他们一样吗？
16 独词句	“银胆”，即日常用的白色保温瓶胆。
17 动词+从句	张颖忘不了她吃了多少苦。
18 形容词+从句	
杂类(感叹句,祈使句)	多么突兀的秀色啊！

### 3.3 连续语句的选取方法

经过断句和拼音转写，原始语料库已转换成具有拼音标注的句子库，这时在前述语音学规则的指导下，通过算法可实现自动选取句子作为朗读文本。Greedy 算法(Cormen, Leiserson & Riverst 1990, van Santen 1997)就是用于从大的句子库自动选取最佳覆盖的句子子集的算法。图 2 为 Greedy 算法的框图，图中的循环次数可以是 N，这样得到的新的句子集是按照语音单元表覆盖个数排序的，也可以根据需要选择较小的循环次数。通过 Greedy 算法可以得到满足最小冗余度、最大覆盖面的精练的语料。如果采用 Greedy 算法，对句子库有一定要求，每个句子在意义上应当完整，但句子的长度又不好控制，按 3.2 节给出的标点符号断句很难达到要求。另外，对较大的句子库采用 Greedy 算法，运算时间相当长。由于这些困难的存在，863 连续语音识别数据库句子的选择采用了图 3 所示的方法，并通过人工干预粗选一个数目稍大的句子集，然后再使用 Greedy 算法，可取得较好的效果。表 5 给出用不同方法选择的 500 个句子的语音分布情况。

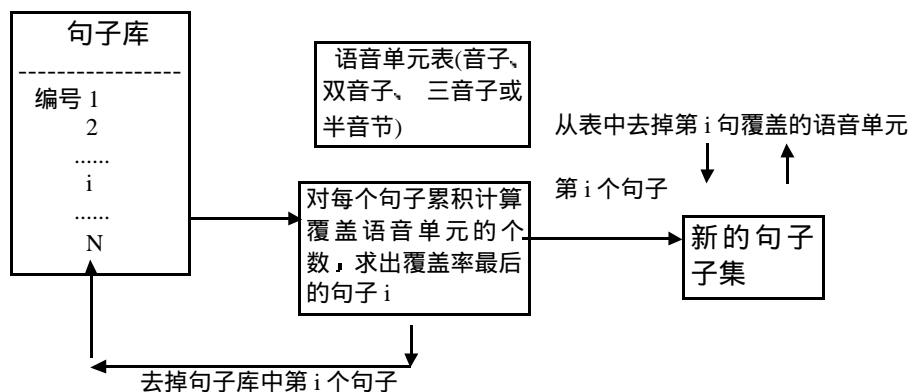


图 2 基本 Greedy 算法

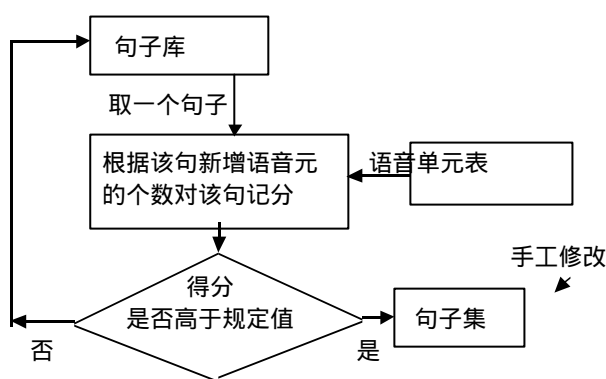


图 3 863 连续语句自动选取流程图

表 5 用不同方法选取 500 个句子的语音覆盖情况

	随机选句	图 3 方法	Greedy 算法	总个数
音节	311	361	382	401
三音子	1167	1458	1986	3035
韵-声结构	485	522	645	781

### 3.4 从朗读文本到语音数据

如何让发音人按照设计要求朗读文本是有待研究的问题，音段方面的语音现象相对来说比较好办，主要困难在于韵律方面。发音人往往在运用韵律结构时不一定满足设计要求。设计人与发音人不能在同一环节发生交流，因此也无法运用语言理解中的合作原则(Horn 1992)。这种不一致就某种意义来说也是不可避免的。但是，可以通过对发音人的选择尽可能克服，例如对发音人的文化程度要有一定要求、录音前熟悉文本等等，同时，录制人对文本的理解对语音数据的质量也起着十分重要的作用。

### 3.5 语音数据库的加工

没有进行标注的语料俗称生语料，如果要对语料库中的语料进行研究，就需要进行词性标注、句法分析，这样便成为加工后的语料库。与此类同，对语音数据库也应当进行加工，一般

是在音段和韵律两个层面上进行。中国社会科学院陈肖霞对 863 语音识别数据库的 1560 个句子在声、韵及其过渡这个层次上做了手工音段标音(陈肖霞 1997); 李智强将对语音合成数据库做手工韵律标音(李智强 1997)。

音段的手工标音, 是对照语音信号的语谱图、结合听辨进行标音。标注的最佳形式是采用国际音标(IPA)进行音位(phoneme)标音, 汉语拼音是语音的正字标音符号, 它对应的国际音标有 30 多个, 如果要描写音段音变, 这 30 多个是远远不够的。因此非训练有素的语音学家难以胜任这一工作, TIMIT 数据库请的是著名语音学家 Keating 进行手工标音(Keating 1992)就足以证明这点。虽然手工标音速度较慢、对标音人的要求较高, 对较大规模的语音数据, 如语音识别训练数据, 做手工标注不够现实, 但由于其准确度较高, 仍十分必要标注一部分, 提供标准。大规模的语音数据应当采用集内语音识别的方法做自动标音。

经过标注的语音数据库对语音研究是极有意义的, 它提供了有别于实验室语言的大量语言事实, 可通过统计的方法总结、修改语音学规则。

### 3.6 基于数据库的语音研究

语音学研究使用的语音材料有两种, 一种是编制的实验室语料, 另一种来自语音数据库。的确, 在许多情况下, 使用实验室语料是十分方便, 可以主动地控制各种条件, 得到可靠的结论。但是, 语音研究的对象应当是自然的言语, 任何生硬的言语不应当是我们主要的研究对象。原则上说, 如果语音库的规模足够大, 满足各种要求的实验材料都可从中找到。我们需要用科学的手段驾驭千变万化的语音数据。基于语音数据库建立语音现象的概率模型是目标之一。

我们已经得到 863 语音数据库 1560 句的标注数据(陈肖霞 1998), 并进行整理, 得出声母、韵母, 以及过渡段时长的数据(均值及其方差), 这些数据可以作为参照值, 用以确定韵律边界、计算语速等。近期我们将得出有关结论。通过标音数据, 我们还可总结出所有在声母、韵母这一层次上的各种音段音变规律表。

## 4. 连续语音数据库有待改进的基本问题

在 863 连续语音数据库的文本设计过程中, 我们的着重点在音段的语音声学特性方面, 我们对于音段的研究、归纳工作比较详细、全面, 就这一层面而言, 对目前的言语工程系统应当是充分的。作者认为存在以下两个方面问题。

### 4.1 篇章的断句

朗读句是最简单形式的话语, 它缺少情景(situation)、场景(scene)、讲话目的(purpose)和交往(interacton)。对于文本的设计者和发音人来说, 涉及到言语的理解和产生两方面的问题。由一个个话语构成的篇章, 意义是动态变化的, 每个句子不仅有意义, 也含有一定的信息, 人们是一步步地处理篇章中的句子, 每个新句子的处理都依赖于上一句, 并更新旧信息, 作为下文的依据( Blakemore 1988)。863 朗读文本以孤立话语的形式呈现给发音人的, 它们在整体上没有统一的话题, 从语用学的角度讲, 发音人无法在更完整层次的语义描写上把句子的语境跟理解关联起来。

为了尽量减少上述矛盾, 在断句方面应多做一些工作。当一个句子较长, 即音节数较多, 需要分成几个子句, 这种断句与句法切分不完全相同(吴竞存, 梁伯枢 1992), 既要考虑意义, 又要考虑韵律边界。汉语有两个音节为一个音步的倾向, 同时音步又受到句法结构的制约(冯胜利 1997)。另外, 从语言产生的角度来看, 断点应放在语言计划单位的边界上, 心理语言学通过对口吃的研究发现, 说话人在实义词之前口吃和停顿多于在功能词之前( Tanenhaus 1988), 也就是说, 人们试图将具有语义的实词说完整。仅此规律仍不足以解决自动断句问题。下例长句给出了句法上的切分, 两个花括号为第一层的两个直接成份:

{诗歌趣味} {[以其日深年久的积韵] [(从一切可能的方面对小说施加了) (巨大的压力)]}。

如果考虑从韵律结构上将其断为两句, 呈如下形式:

{诗歌趣味以其日深年久的积韵}{从一切可能的方面对小说施加了巨大的压力}。

断点不是在两个直接成份之间，即句法和韵律两方面发生了不一致。目前还无法得到一套成系统的断句规则。

#### 4.2 语音单位

如上所述，我们选取的语音单元只在连续语音的音段方面进行了考虑，在 863 语音合成数据库中我们仅在词一级结合音段考虑了声调的搭配。以下为二字、三字和四字词的声调模式个数：

二字词： $N_2 = 4 \times 4 \times 2(\text{清/浊}) = 32$ ；

三字词： $N_3 = 4 \times 4 \times 2(\text{清/浊}) \times 4 \times 2(\text{清/浊}) = 256$ ；

四字词： $N_4 = 4 \times 4 \times 2(\text{清/浊}) \times 4 \times 2(\text{清/浊}) \times 4 \times 2(\text{清/浊}) = 2048$ ；

图 4 为选词的自动算法框图，该方法在实质上还是将声调和音段割裂开考虑的。许毅认为在两音节中，由于言语产生方面的原因，前一音节的声调将扩展到第二音节，第二音节的声调特性附着在音节的后半部(Xu 1997)；美国 IBM 公司的汉语语音识别的模型将音节一分为二，后半携带声调(Chen 1997)。这些为我们提供了很好的参考。

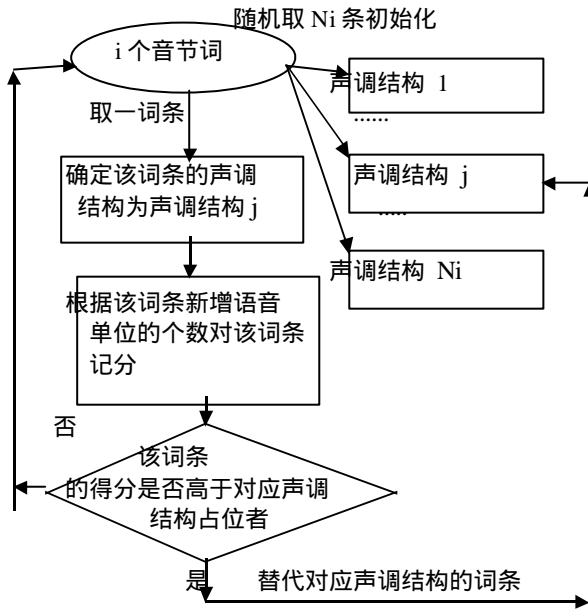


图 4 从原始词库自动选词流程图

我们认为在一个话语中所有的音节之间在声学上都会发生关联，但是由于韵律结构的不同，当句中出现停顿、音高曲线的变化，以及边界之前音段延长等韵律现象时，用表 2 中纯粹的音段语音单元来描述连续语句是不够的，我们应当将韵律结构对音段的影响考虑进去。我们不妨构造如下形式的语音单元：

$$P = p(p_1, p_2, p_3, \dots, p_n)$$

$p$ ：音段语音单元，例如 a, an, ao, ang, e, er, ..... 等 38 个韵母及声母，或其它语音单元；

$p_1$ ：声调；当  $p$  为韵母， $p_1$  = 包括轻声在内的 5 个声调 1, 2, 3, 4, 5；

当  $p$  为声母， $p_1 = 0$ ；

$p_2$ ：在语调结构中的位置； $p_2 = 0$ ：在结构中； $p_2 = 1$ ：在结构末；

$p_3$ ：是否重音， $p_3 = 0$ ：非重音； $p_3 = 1$ ：重音；或分若干重度级；

.....

当  $p$  选择声母、韵母，共有  $21 + 38 = 59$  个， $n$  取 3，就有约  $59 * 6 * 2 * 2 = 1416$  个，如果考虑



音节间的搭配，就要比表 2 中的 781 个韵-声结构多得多。因此在语音单位的自动处理方面还需下一番功夫。

### 4.3 讨论

在语音数据库的语料设计方面，应考虑的是语音现象的覆盖。为此将设计反映这些语音现象的语音单元。语音单元的选择直接反映了我们对语音本质的认识。初步的语料设计只考虑孤立音段，如音位或声母/韵母等数量较有限的语音单元。若考虑音变现象，语音单元就会复杂许多，例如我们现在所选用的三音子或韵-声结构。如果考虑连续语音的韵律现象，就要考虑更高层次的语言学制约。

本文提出的篇章断句和音段-韵律语音单元两方面的问题就是期望揭示更复杂的连续语音现象。我们需要深入研究：音段在不同的韵律层次中是如何变化的；韵律结构的声学表现究竟如何；句重音在句法结构中起的是什么作用；句法结构是怎样影响韵律结构的。这些问题对语音数据库设计的科学性具有十分重要的意义。

## 参考文献

- 陈肖霞 (1998) “基于连续话语语料库的语音音段的初步统计分析”，本刊。
- 冯胜利 (1997) 《汉语的韵律、词法与句法》，北京大学出版社。
- 罗振声、郑碧霞 (1994) “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》，第 8 卷，第二期，pp.1-13。
- 李智强 (1997) “普通话韵律标音的初步研究”，第三届中国计算机智能接口与智能应用学术会议论文集，pp. 169-173。
- 沈炯 (1992) “北京话连读的调型组合和节奏形式”，《中国语文》，第四期，pp.274-281,1994。
- 吴竞存，梁伯枢，《现代汉语句法结构与分析》，语文出版社。
- 祖漪清，李爱军 (1997) “语音识别和语音合成语料库的设计”，第三届中国计算机智能接口与智能应用学术会议论文集，pp.174-179。
- Blaauw, Eleonora (1994) “The Contribution of Prosodic Boundary Marks to the Perceptual Difference between Read and Spontaneous Speech”，*Speech Communication* 14, pp.359-375, 1994.
- Blakemore, Diane (1988) *The Organization of Discourse*, 林书武译，“话语的结构”，《国外语言学》，第四期，1992。
- Cormen, T. H., C.E. Leiserson, and R. L. Rivest (1990) *Introduction to Algorithms*, The MIT Press, Cambridge, Massachusetts.
- Chen, C.J., R.A. Copnath, M.D. Monkowski, M.A. Pcheny, and K. Shen (1997) *New Method in Continuous Mandarin Speech Recognition*, *Proceedings of 5th European Conference of Speech Communication and Technology*, vol.1, vol.3, pp.1543-1546.
- Fant, Gunnar & Anita Kruckenberg (1989) “Preliminaries to the Study of Swedish Prose Reading and Reading Style”，*STL-QPSR* 2.
- Fant, Gunnar & Anita Kruckenberg (1996), “Quantal Theory of Speech Timing”，*TMH-QPSR* 2, pp.41-44.
- Hong, Tony T.N. (1987) *Syntactic and Semantic Aspects of Chinese Tone Sandhi*, University of California at San Diego.
- Horn, Laurence R. (1988) *Pragmatic Theory*, 沈家煊译，“语用学理论”，《国外语言学》，第二期，1991。
- Keating, Pat, B. Blankenship, D. Byrd, E. Flemming & Y. Todaka (1992) “Phonetic Analyses of the TIMIT Corpus of American English”，*Proceedings ICSLP92*, vol.1, pp.823-826.
- Rossi, Mario (1997) “Is Syntactic Structure Prosodically Retrievable”，*Proceedings of 5th European Conference of Speech Communication and Technology*, vol.1, pp. KN 1-8.
- Shih Chi-Lin (1986) *The Prosodic Domain of Tone Sandhi*, Ph.D Thesis, University of California at San Diego.
- Tanenhaus, Michael K. (1988) *Psycholinguistics: An Overview*, 桂诗春译，“心理语言学概

述”， 《国外语言学》， 第一期， 1991.

van Santen, Jan P.H. (1997) “Methods for Optimal Text Selection”, Proceedings of 5th European Conference of Speech Communication and Technology, vol.2, pp. 553-556.

Xu Yi (1997) What can tone studies tell us about intonation? Intonation: Theory, Models and Applications, Proceedings of an ESCA Workshop, pp.337-340.

Zu Yiqing (1997) Sentence Design for speech synthesis and speech recognition, Proceedings of 5th European Conference of Speech Communication and Technology, vol.2., pp.743-746.