

# 汉语文语转换系统的语音规则和声学参数

张 家 骅

(中国科学院声学研究所)

1988年4月12日收到

**摘要** 利用瑞典皇家理工学院(KTH)语言通讯和音乐声学系 Carlson 和 Granström 设计的多语种文语转换系统,进行了汉语文语转换研究。对汉语文语转换用的基本语音规则和声学参数做了分析、设计和调试,特别对轻声、儿化和变调进行了研究。

## Acoustic parameters and phonological rules of a text-to-speech system for Chinese

ZHANG Jialu

(Institute of Acoustics Academia Sinica)

Received April, 12, 1988

**Abstract** A multi-language text-to-speech system developed by Carlson and Granström was used to do text-to-speech conversion for Chinese. Both phonological rules and acoustic parameters are refined. The light tone and tone modification rules and er-coloured final rules are introduced. It is shown that to shorten duration is the most important feature for light tone syllables, and the principle of that the main vowel in the final should be pronounced with retroflex vowel /er/ simultaneously can be used for synthesizing er-coloured finals.

## 引 言

文语转换系统,是一种高级的语言合成系统。它不但具有广泛的应用价值,而且对语言研究也甚为有用。八十年代初瑞典皇家理工学院(KTH) Fant 教授领导的语言通讯和音乐声学系研制了一种多语种文语转换系统<sup>[1,2]</sup>,见图1。这种文语转换系统,完全是靠语音规则系统来工作的,而不是借助于语音词典。它是一个十分灵活的系统。在系统中存入某种语言的所有音位的声学参数和语音区别特征,再制定出该语言的语音规则系统,这样文字材料输入便可以得到口语输出了。

汉语是声调语言,它具有一个复杂的声调系统。轻声和儿化也是两类特殊的语音变化,它不仅影响音质,而且有时具有辨意作用。要想建立一个具有良好自然度的汉语文语转换系统,就必须对轻声和儿化以及声调和语调的关系进行研究。本文便是利用 KTH 文语转换系统对这些问题进行的初步研究。

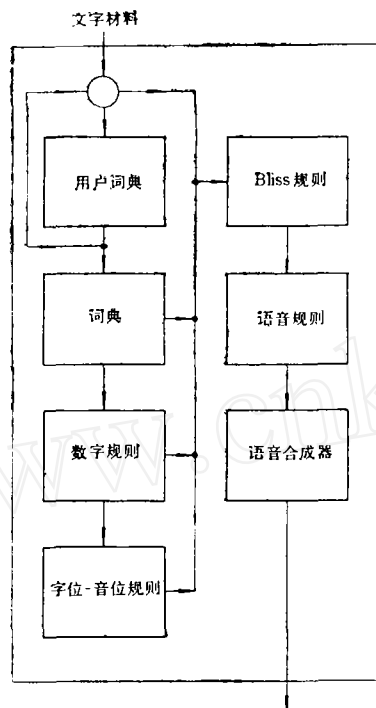


图1 KTH文语转换系统

## 一、汉语音节结构

与印欧语言不同,汉语\*具有自身的语音学特点。首先,汉语音节构词时要具有声调(普通话有四种声调,有的方言可更多)。其次,正因为利用声调作为一种构词手段,所以汉语无调音节数目相当少,约400个左右(加上声调则有1200个左右)。第三,汉语音节结构简明而有严格的规律,见图2。一个汉语音节可分成三部分:声母、韵母和声调。声母都是辅音,总数为21个。声母也可以没有,称之为零声母。韵母是由三个元素组成的,包括介母、主要元音和韵尾。只有三个元音可作介母,它们是*/i, u, ü/*。介母也可以没有。主要元音有*/a, o, e, i, u, ü, ê, er/*等十个。其中拼音符号*/i/*既代表元音*[i]*,又代表元音*[ɿ]*(*z, c, s*之后)和元音*[ɯ]*(*zh, ch, sh, r*之后)。只有两个元音*/i, u/*和两个鼻辅音*/n, ng/*可以作韵尾(韵母 *ao = au,*

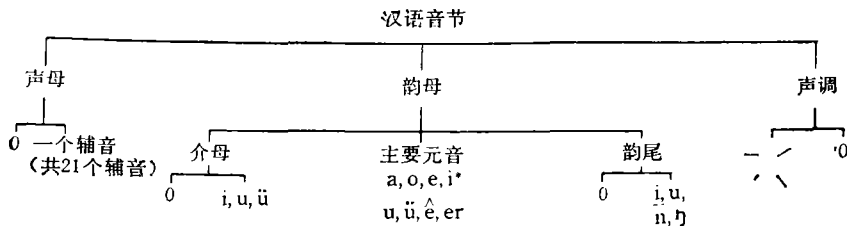


图2 汉语音节结构

\* 汉语在本文中均指普通话。

iao = iau)。一个音节也可以没有韵尾。但是,任何一个音节都不能没有主要元音。因此,汉语的第四个特点就是元音所占的比重较大。

除有四种能区别词义的声调之外,在多音词中还有一种轻声调。有时轻声也产生辨意作用。所以,也可以认为,在连续语言中有五种不同的声调,即阴、阳、上、去、轻。

## 二、语音规则系统

KTH 文语转换系统中所用的语音规则系统,是 Carlson 和 Granström 用特殊的高级程序语言设计的。这种程序语言和语音学家使用的音系学规则的表达方法差不多。这些规则的基本结构是

$$X \rightarrow Y / A \& B$$

其中,结构符号 & 表示,当 X 出现在语境 A 和 B 之间, Y 表示 X 产生的结构变化。→ 和 / 是定义符,表示结构变化及变化的条件。表达式主要使用区别特征,也可以使用语音符号和参数。

### 1. 字位-音位规则

汉字基本上是表意文字。把汉字直接转换成音位是困难的。可是,汉语拼音对于文语转换是非常方便的。因为,汉语拼音系统基本上是按照一音一符的原则设计的。只有声母/zh, ch, sh/和韵尾/ng/是双字母,其中没有同音符号。只有符号/i/既代表元音[i]也代表汉语两个特别元音[ɿ]和[ʊ]。声学语音学测量结果表明,汉语拼音方案把[i, ɿ, ʊ]当成一个音位是合理的,因为舌尖元音[ɿ, ʊ]在 $F_1$ - $F_2$ 平面上是靠近前高元音[i]的。而传统音韵学把舌尖元音放在开口呼,从音系学的角度看,是简明的。但不能说舌尖元音在音值上靠近低元音。在做文-语转换时,必须对两个舌尖元音分别单独给与定义,为此需要写出字位-音位规则如下:

$$I \rightarrow I_1 / \langle \text{CONS}, +\text{ANT}, -\text{COR} \rangle$$

$$I \rightarrow I_2 / \langle \text{CONS}, -\text{ANT}, -\text{COR} \rangle$$

式中 I 代表元音[i],  $I_1$  代表舌尖前元音[ɿ],  $I_2$  代表舌尖后元音[ʊ], CONS 代表辅音, ANT 代表区别特征舌尖前的, COR 代表舌面前的。所以,上面两条规则表示:如果汉语拼音符号 i 出现在舌尖辅音(z, c, s)之后时,则 i 便成为舌尖前元音[ɿ];如果 i 出现在舌尖后辅音(zh, ch, sh, r)之后,则 i 便成为舌尖后元音[ʊ]。

尽管汉语拼音的正字法还不规范,但是从字位到音位的转换,利用汉语拼音符号还是很方便的。

现在我们已经建立了汉字机内码与汉语拼音的关系。因此,任何一种汉字编码输入方式,都可以实现字位音位转换,从而可作为文语转换系统的输入。有些汉字是多音字,其正确读音要在单词一级来决定。由于数目不算太多,大约 240 个左右。完全可以利用图 1 中的用户词典把它们预先存贮起来,根据上下文决定读音。例如,“处理”(chǔlǐ)和“住处”(zhùchù);“当地”(dāngdì)和“适当”(shìdàng)等。

调号要加标记。本系统采用附加字母办法。以 1、2、3、4 分别代表第一、二、三、四声,紧随在主要元音之后。没有调号的音节则为轻声音节。

共使用了 14 条字位-音位转换规则。

### 2. 变调规则

在普通话中,两个第三声(上声)音节相连时,前一个上声要变成阳平(第二声)。对这种现象要给出变调规则

$$\langle \text{VOC}, \text{TONE} = 3 \rangle \rightarrow \langle \text{VOC}, \text{TONE} = 2 \rangle / \& \\ \langle \text{VOC}, \text{TONE} = 3 \rangle$$

其中 VOC 表示元音性的, TONE = 3 即第三声, TONE = 2 表示第二声。

除此之外,在语流之中,也还有一些声调变化。比如,两个第四声(去声)音节相连,前一个音节变成半去;上声跟另外一个非上声相连则变成半上。这些变化都不是从一个调位变成另一个调位,也就是不属于音系学的变化。我们把它放在修调规则之中,即声调弱化规则之中。

### 3. 轻声规则

在普通话之中,除四种声调之外,还有一种特别的声调叫轻声。它是在单词中出现的,也有辨意作用。在词典中轻声的出现率约有 8.6%,是一个不容忽视的特征。例如,“兄弟”一词如果后一个音节重读,则代表哥哥弟弟,要是后一个音节读成轻声则只代表弟弟而不包括哥哥。此外,如“大意”和“大意”;“合计”和“合计”等。实际上,一个词是否读轻声,只由词意或者是读音习惯来决定,没有什么固定的规则。大部分轻声词,虽然不读轻声也不至于造成词意上的错误,但是,听起来却很不舒服。

轻声音节的调值与前一个音节的调型有关。在五度制标调系统中,可表述如下:

1. 阴平调(55)之后,轻声调值为 2,例如,“他的”;
2. 阳平调(35)之后,轻声调值为 3,例如,“谁的”;
3. 上声调(214)之后,轻声调值为 4,例如,“我的”;
4. 去声调(51)之后,轻声调值为 1,例如,“坏的”。

实际上,我们采取了一种简化的规则。那就是在上声之后轻声调值取高调,在其它三种声调之后轻声取低调。写成规则是

$$\langle \text{TONE} = 0 \rangle \rightarrow \langle \text{TONE} = \text{H} \rangle / \langle \text{TONE} = 3 \rangle \\ \langle \text{TONE} = 0 \rangle \rightarrow \langle \text{TONE} = \text{L} \rangle / \langle \text{TONE} = 1 \rangle \\ \langle \text{TONE} = 0 \rangle \rightarrow \langle \text{TONE} = \text{L} \rangle / \langle \text{TONE} = 2 \rangle \\ \langle \text{TONE} = 0 \rangle \rightarrow \langle \text{TONE} = \text{L} \rangle / \langle \text{TONE} = 4 \rangle$$

其中 H 代表高调, L 代表低调,具体取值可以根据发音人事先设定,也可以根据单词在句中所处的位置,由语调生成模型来产生。TONE = 0 代表轻声调。

### 4. 声调弱化规则

在普通话中除了上述变调规则之外,还需要有修调规则或声调弱化规则,来处理语流中声调的协同发音问题。

1. 上声调在除上声以外任何其它声调之前,它都变成半上,即从降升调变成降平调。在 KTH 文语转换系统中,可写成规则为

$$\langle \text{VOC}, \text{TONE} = 3 \rangle \rightarrow \langle \text{LF}_0 = 105 \rangle / \& \langle \text{VOC}, \text{TONE} = 1 \rangle$$

这里表示一个上声在阴平调之前,则上声调的终点  $\text{LF}_0$  只是低调值 105Hz (对本试验的发音人),而不再上升。这里可以看出, KTH 的规则系统中,允许符号和参数并用,非常方便。对于阳平和去声在上声之后时,还得另写两条规则。

2. 在单词中一个去声跟另一个去声相连时,则前一个去声变成半去,即调值由高降

至中而不到低。其规则写作

$$\langle \text{VOC}, \text{TONE} = 4 \rangle \rightarrow \langle \text{LF}_0 = 155 \rangle / \& \langle \text{VOC}, \text{TONE} = 4 \rangle$$

这里去声的终点基频  $\text{LF}_0 = 155\text{Hz}$ , 也是对一个特定的男发音人所取的值。

3. 三音词变调: 三音词的变调问题, 由于三音词的结构和语义, 显得比较复杂一些。三音词的构成不外三种情况: a. 三个音节独立平等, 如“文工团”“迪斯科”; b. 一个双音词加一个单音词或音节, 如“游泳池”“葡萄酒”, c. 一个单音词或音节加上一个双音词, 如“老闺女”“半成品”。但是, 一般三音词都是中间一个音节读得轻。因此, 受弱化规则影响较大。另外, 前述变调、修调规则对三音词同样适用。这里需要特别提出的是, 三音词中间一个音节要是阳平则都变成阴平, 写成规则

$$\langle \text{VOC}, \text{TONE} = 2 \rangle \rightarrow \langle \text{VOC}, \text{TONE} = 1 \rangle / \langle \text{VOC}, \text{STRESS} \rangle \& \langle \text{VOC}, \text{STRESS} \rangle$$

其中 STRESS 表示重读音节, 即具有某种声调。这是很自然的, 因为, 阳平音节在连续言语中时长最长。它出现在三音词当中时, 轻读必须缩短时长, 故而从高升简化为高平。

4. 还有几个词读什么调是与上下文有关的, 主要有“一, 七, 八, 不”等。例如, “一二一”的调是“阴平, 去声、阴平”而“一天”要读成“去声、阳平”, 可是“一夜”又变成了“阳平、去声”了。这种异读调也可以和异读字同样处理。把它们存入用户词典。

### 5. 儿化韵规则

普通话, 特别是北京话, 常有儿化韵出现。儿化一般用于指小, 个别也有词意上的差别, 如: “大鼓”是指一种打击乐器; “大鼓儿”则是曲艺的一种。几乎所有的韵母都可以儿化。所以一个文语转换系统必须具有产生儿化韵的功能。

从言语合成的角度出发, 采用赵元任<sup>[3]</sup>提出的儿化韵发音的基本原则——卷舌韵尾与主要元音可共时发音, 是最方便、合理的了, 根据王理嘉、贺宁基<sup>[4]</sup>关于儿化韵的声学分析结果, 制定了下述语音规则:

1. 开尾韵母: 儿化只决定于主要元音而与介母无关。

a. 主要元音为/a、o、e、u/: 由于这些元音的发音部位与r相近, 有共时发音的基础, 所以, 从这些元音的起点便开始儿化。

鉴于韵母/ie、üe/只是独用或仅与舌面音/j、q、x/相拼, 因此, 其儿化韵与[ər]是相近的。比如, “小街儿”和“小鸡儿”在很多人听来是分不清的, 所以, 并不把[ər]单作一类。

b. 主要元音为/i、ü/: 由于这些元音发音部位靠前, 从前元音到[ǝ]动程较大, 不能共时发音。儿化时必须在主要元音之后引入一个元音[ə], 然后把[ə]当成主要元音, 从起始点就开始儿化。原有的主要元音/i、ü/当作介音。这样处理, 在音系学规则上要简单明了。

c. 主要元音为[ɿ、ɥ]: 由于这些元音发音部位不仅靠前, 而且对于声母/z、c、s; zh、ch、sh、r/有一定的依附性。这两个元音从不单独出现。在儿化时它们便被元音[ə]所取代。

2. 带/i、n/韵尾的韵母: 儿化时韵尾脱落, 以后再按上述1. 项开尾韵母的规则执行。

3. 带/u、ng/韵尾的韵母: 由于/u、ng/发音部位靠后, 有与[ǝ]共时发音的基础, 所以儿化时部分消失, 并对[ǝ]产生影响。实际上, 可以简化为:

a. /u/韵尾: 前面只有元音/a、o/, 所以儿化时从元音/a、o/开始直到韵尾。

b. /ng/韵尾: 前面可有元音/a、o、e/, 儿化时先去掉韵尾/ng/, 按开尾韵母儿化, 再使韵母鼻化。

例如: 婴儿 (yingr) 的程序是:  $yingr \Rightarrow yiengr \rightarrow yier \rightarrow yier^{\sim}$

### 三、声学参数

KTH 文语转换系是一个多语种的文语转换系统。每一语种的所有基本语音都必须用区别特征和声学参数来定义,以便于在语音规则系统中应用。每一个语音最多可包括 32 个区别特征和变数。

这里主要介绍有关儿化韵和轻声的声学参数。

**儿化韵:** 汉语的儿化韵与美国英语的 r-化元音不同。儿化韵的卷舌是从主要元音的起点开始的,而且元音[ɤ]尾既短且弱,有点像英语 art 一词中的 [ar]。一般来说,卷舌韵与声母和介音无关。

北京话中究竟有多少个不同的儿化韵,至今尚无定论。从语音学的角度来看,为了能简化文语转换系统用的语音规则系统,可以把儿化韵分成下述四大类。

1. [ar]: 这一类包括以/a/为主要元音的九个韵母 [a, ai, an, ia, ian, ua, uai, uan, yan]。这些韵母儿化比较容易,其语音规则已如上述。在声学参数上,最重要的是,从元音[a]一开始第三共振峰就向下降低直至韵尾,约从 2500Hz 降至 1900Hz。同时,  $F_1$ ,  $F_2$  也向中央偏低区域收拢。

2. [ər]: 这一类包括韵母 [ɪ, ʌ, i, in, ei, en, uei, uən, y, yn]。这类儿化韵根据语音规则,它们在儿化时或是以元音 [ə] 替代原来的韵母 [ɪ, ʌ], 或是新增元音 [ə] 在原韵母之后,如 [i, in, y, yn]。总之,儿化的来源比较复杂,起点比较分散。这类韵母儿化后,  $F_1$ ,  $F_2$  向元音图的中上方收拢,  $F_3$  常可降至 1600—1700Hz。

3. [ur]: 这类包括韵母 [r, u, o, uo, au, iau, ou, iou]。它们都有可与 r 共时发音的部分。  $F_1$ ,  $F_2$  较低,儿化时  $F_1$ ,  $F_2$  变化不大,但  $F_3$  降得较低,约达 1500Hz 或更低。这一类中有一个词“媳妇”,在儿化时值得注意。它不能直接儿化成 [çifur] 而应当念成 [çifər]。这可能是因为“媳妇”一词应读轻声,元音 [u] 要先弱化为 [ə], 以后再儿化的关系。

4. [ār, ēr]: 这一类包括韵母 [aŋ, ieŋ, uaŋ, əŋ, iŋ, uəŋ, uŋ, yuŋ]。这类韵母由于韵尾 ŋ 可与 r 共时发音,所以儿化时要在主要元音的第一共振峰以下增加鼻音共振峰同时使  $F_1$ ,  $F_2$  的带宽加宽。当然还必须使  $F_3$  从主要元音一开始便逐渐降低。

**轻声:** 一般认为,轻声音节的声强较低音长较短,实验结果<sup>[9]</sup>表明,轻声音节的声强不一定比该音节不轻读时的低,但音长总是短的。我们在言语合成中也证明了这一点。所以,最简单的办法便是缩短音节的时长来产生轻声音节。经过多次试验表明,取正常音长的 60% 较为合适。第一步,采取音节中各音素线性压缩的办法。第二步准备研究非线性压缩、扩张(重读时)问题。比如,不送气塞音 (b, d, g) 可不压缩,介音、韵尾和主要元音的压缩(扩张)比例可有所不同等。

轻声音节的调值只采用了高、低两种。

### 四、语调规则

语调特征是提高文语转换系统的可懂度和自然度的重要参数。特别是对于汉语,在语流

中同时有两种调的变化,声调和语调,交织在一起,情况更加复杂,作用也更加重要。

语调属于韵律特征,对语调的描写,特别是定量的描写,对各种语言都是很不够的。汉语就更加缺乏。在书面语言中,只有有限的标点符号可用来标志语气和停顿。至于重音和着重点,以及句间的关系,如是否复句,则没有任何标志。目前各种语言还都不能进行自动句法分析,由机器产生语调标志。必要的语调标志都是手工附加的。

我们采用 Lund 语调生成模型<sup>[7]</sup>来处理语调,特别是声调和语调关系问题,见图 3。首先就某一发音人确定其调域。调域有四条标志线。中间两条实线是正常声调的上下限,上下两条虚线是调域的上下边界,即当重读或强调时,音高所达到的上下限,见图 3a。第二步根据句型在调域中定出语调走向和声调上下限,称为声调格子,见图 3b。由于这是一句没有疑问助词“吗”的疑句,所以句尾上升。句子主语谓语之间常有一个枢轴点,使语调有一点转折。第三步是把各个音节的调值填入格子。第四步则是根据上下文按变调、修调规则确定各音节在语流中的调值。如遇有着重点则将局部声调上下限扩张,可达到上下虚线;同时还要将相邻部分作必要的压缩(这是韵律特征的对比规则或称突出规则)。第五步就是对基频曲线进行必要的内插和平滑。最后就生成出来所需的语调曲线了。

这里还要提到的是,在修调、变调以及对比规则的运用中,除了对基频  $f_0$  产生变化外,还要影响到时长。这方面的研究还很不够,正在开始阶段<sup>[7]</sup>。

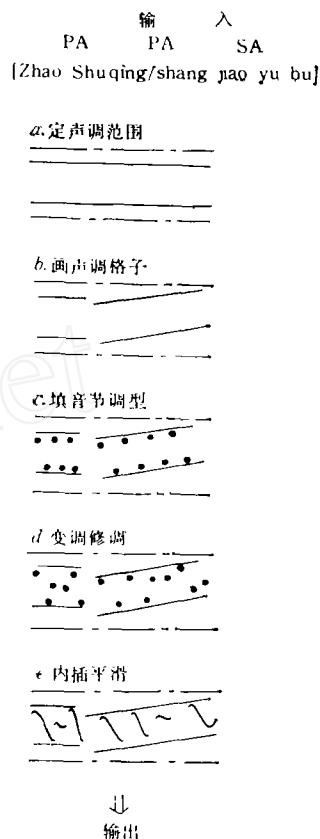


图 3 声调和语调生成流程图

## 五、结 语

利用 KTH 文语转换系统进行汉语文语转换的研究表明:

1. 采用汉语拼音作为输入方式,可以方便地实现汉语文语转换。
2. 儿化韵在合成时可以分成四类,用语音学规则处理比较容易,所合成的语音具有足够的可懂度和自然度。
3. 轻声音节可以只用缩短音长的办法来产生。

**致谢** 作者对 Fant 教授、Carlson 和 Granström 博士在工作中给予的帮助和支持表示感谢。

图 3 中的试验句是沈炯副教授设计的,在此向他致谢。

### 参 考 文 献

- [1] Carlson, R. and Granstrom, B. "A text-to-speech system based entirely on rules", Conference Record, 1976, *IEEE-ICASSP*, Philadelphia, USA, 1976.
- [2] Carlson, R., Granström, B. and Hunicutt, S., A multi-language text-to-speech module, *STL/QPSR*, (1981), No. 4.
- [3] Chao, Yuen Ren, "*A grammar of spoken Chinese*", (University of California Press, 1968).
- [4] 王理嘉、贺宁基, 北京话儿化韵的听辨实验和声学分析, "语言学论丛"第十辑(商务印书馆, 北京, 1983), 38—60.
- [5] 林茂灿、顾景助, 北京话轻声的声学性质, 方言, (1980), No. 3.
- [6] Gårding, E., Zhang, J. and Svantesson, J. "A generative model for Tone and intonation in standard Chinese," *Working Paper*, 25(1983), Lund University, 1983.
- [7] Gårding, E., Zhang, J., "Tempo and prosodic pattern in Chinese and Swedish", *Working Paper* 31(1987), Lund University, 1987.

www.cnki.net