

# 广东话文本语音转换系统的设计与实现

茅于杭 张国贞

语言文字信息处理研究中心  
清华大学自动化系, 北京 100084  
E-mail:myh-dau@mail.tsinghua.edu.cn

**摘要:** 汉语的文本语音转换有广泛的应用, 用音素或音节为单位的合成方法各有利弊, 广东话的语音合成与普通话的合成方法基本类似, 但广东话有较多的音节数和较多的的音调种类, 广东话的多音字和变调字也很多, 所以广东话的合成处理技术比较复杂。用广东话语音合成系统作成的盲人用计算机软件已于香港地区的盲人界, 具有良好的效果。

**关键字:** 语音合成, 文本语音转换, 普通话, 广东话, 方言

## Design and Implementation of a Cantonese Text-to-Speech System

Mao Yuhang Zhang Guo Zhen

Speech and Language information Processing Research Center  
Tsinghua University  
E-mail:myh-dau@mail.tsinghua.edu.cn

**Abstract:** Chinese Speech synthesis used in TTS has been used widely in practice. There are mainly two strategies of synthesis. One is based on phone and the other is based on syllable. Both have their own benefits and shortcomings. Fundamentally, the method used to synthesize Cantonese is similar to that of mandarin. Nevertheless there are more syllables and tones in Cantonese. Besides, tone inflection in Cantonese is quite complicate. This makes the synthesis of Cantonese much difficulty. A practical Cantonese speech synthesis system has been implemented. The system has been used to build special software for the blind. The visually impaired people in Hong Kong and Canton districts are being benefited from it.

**Key words:** speech synthesis, text-to-speech, mandarin, Cantonese, dialect.

### 1, 概述

我国地域辽阔, 方言种类繁多。如今, 在普通话的语音合成已广泛进入实用之后, 地方语的语音合成就提上日程。以广州地方话为代表的广东话流行的地区很广, 除了广州地区外, 在香港、澳门以及海外的许多华侨群体中都流行广东话。分类上讲, 广东话仍属于以音节为单位带调语音, 和普通话有许多相同性。在实现普通话的语音文本转换中所采用的许多技术是可以直接应用到广东话的语音文本转换中的, 但考虑到广东话的特殊性, 又必须采取一些不同的处理方法。

## 2, 汉语语音合成及文本语音转换的基本方法

汉语语音合成有广泛的应用领域, 汉语的文本语音转换系统涉及汉字文本的词切分处理、多音字判别, 三声变调, 一、七、八变调处理等。性能较好的系统还应包括句型分析和语调处理。广东话的语音合成也有同样的问题。

语音合成基本元素若采用音素, 则包括声母、韵母音素, 区分韵母音素中的四声又可以用独立的音素, 也可以在同一音素上加调制的办法来实现。为得到较好的语音效果, 对不同的声、韵母组合要用不同的声韵母音素, 另外, 若考虑到复韵母音节中的声韵母之间的过渡音, 也可单独设定过渡音素。这样实现的系统就相当复杂。

若以音节为基本单元也有两种方案可选, 一是以不带调的音节为基本单元, 采用声调调制的方法获得不同声调, 另一办法是不采用调制而是采取每个音节的不同声调都有独立的音节的办法。显然, 采用音素的方案可以节省数据量, 但是合成语音时要采用较复杂的步骤, 而采用独立音节的方案要求的数据量大, 但合成步骤简单。

在十多年以前, 由于受 DOS 操作系统可用内存的限制, 汉语语音合成系统常采用以音素为单元的方案, 当时我们所实现的系统采用线性预测编码技术, 全部语音数据占存储量不足 50K, 在当时 DOS 操作系统中, 可以连同汉字系统一起都放在内存中, 总共不到 300KB。虽然语音质量不很好, 也能满足一般使用, 在此基础上, 我们制作了第一代的盲人计算机软件。

进入九十年代以后, 微机采用 WINDOWS 操作系统, 系统的存储量已不受限制, 我们开始设计以音节为基本单元的汉语语音合成系统, 这就省去了由音素合成音节的步骤, 简化了算法。汉语普通话的实际有用音节数为 1200 多个 (区分四声), 有许多音节并非四个声调都有, 考虑到变调因素并简化设计, 我们记录了所有的音节的五个声调 (每个音节除四声外都加上轻声), 总共有 2035 个音节。

文本语音转换系统中一般应具备调整语速的功能。在参数合成的系统中语速的调整只要改变某一两个参数, 比较简单。在波形拼接的系统中, 语速的调整只能靠切去部分音节波形来实现。众所周知, 汉语的每一音节大都是由声母和韵母两部分组成, 前部的声母波形变化激烈, 是不稳定区, 后部的韵母部分是准周期性的, 波形有重复性, 添加重复周期就可以延长读音, 使语速变慢, 切除部分重复的周期, 就可以使语速加快。这样, 就可以实现语速的调整。用此方法除在最高速度时感到音质下降外, 基本能满足使用要求。

较完备的语音合成系统应提供多种语声的选择, 如果基本语声是男声, 则还应有女声、童声等等。用参数合成的系统中, 可以用一组语音数据, 借改变某几个参数的方法来获得不同的语声。而在波形合成的系统中, 只能靠设置多套语音数据来提供多语声的选择。典型的方案是系统应提供男、女两种语声, 数据若不进行压缩, 男女声两组数据总共占硬盘空间约 18MB。

在此基础上开发的普通话汉语语音系统于 1997 年用在 WIN98 盲人计算机软件系统中, 供国内广大盲人用户使用。

为适应广东和香港地区的使用, 受香港失明人协进会的委托和资助, 我们开始制作广东话的语音合成系统。

同属于藏汉语系的多种中国方言, 都是以一字一音节发音为特征, 以音节串构成连续语音。绝大多数音节是由一声母和一韵母相拼组成, 韵母还应带有不同声调。在实际读音中可能出现的变调, 所以音节数很多, 语调丰富悦耳。

广东话共有声母 19 个, 韵母 52 个, 九种音调, 比普通话复杂得多。和普通话一样, 广东话中并非每个声母都能和韵母相拼形成音节, 也并非所有音节都有九个声调, 实际

上，广东话中带调音节总共是 1600 个左右。所以在广东话语音合成系统中我们采用以音节为单元的方法，采录 1600 个音节。

下面作具体的介绍。

### 3, 广东话的语音特征

广东话中所用的声母共十九个，用字母表示如下：

b d dz f g gw h j(y) k kw l m n ng p s t ts w

由上可见，广东话的声母和普通话的声母有很多不同，如 dz、gw、kw、ng、ts 等声母发音较特殊，但像 b、p、m、f 等声母的发音和普通话的声母是很相似的。

广东话的韵母有五十二个，罗列如下：

a	ae	ai	ak	am	an	ang	ao	ap	at	au		
e	ei	eik	eing	ek	em	en	eng	ep	et			
i	ik	im	in	ing	ip	it	iu					
ng												
o	oe	oek	oen	oeng	oet	oey	oi	ok	on	ong	ot	ou
u	ui	uk	un	ung	ut							
y	yn	yt										

韵母种类多是广东话的一大特点，普通话中没有闭鼻韵母，广东话中则有如 am em im 等韵母。凡是以 k、t、p 结尾的韵母实际上是三种入声的韵母，发音尤其特殊。

广东话的声调又很特殊。普通话中只有阴平、阳平、上声、去声四种主要声调，加上轻声，可认为有五种声调，而广东话中除了阴平、阴上、阴去、阳平、阳上、阳去六种基本音调外（通常标作 1、2、3、4、5、6 六个声调），还有阴入、中入和阳入三种入声调（标作 7、8、9）。这是三个带调而发音短促的音节，在普通话中是没有对应的声调的，入声类似于日本话中的促音。

### 4, 广东话的音节和多音字问题

在汉字的文本语音转换系统中，多音字的判读是一个大问题。首先对所有汉字都设定一个读音，凡是多音字也设定一个默认的读音。为了判断在文本中出现多音字的读音，一个最简单的，然而不是彻底的解决办法就是设定多音字的词库。按最大匹配法对句子进行词切分，在此过程中只要查到词表中有的词，就可以得到正确的读音，例如词表中有词条“自行车”的标音为“zi4 xing2 che1”和“银行”的标音为“yin2 hang2”，就不会把行字读错。这样，我们设计的多音词库有词条一万余条，其中包括一些相当特殊的词条。例如“还”字的默认读音为“hai2”，但为了在“千里江陵一日还”的诗句中要使其中的“还”字能正确读出，在多音词库中包括了“日还”这样的词条，其读音记为“ri4 huan2”，这样处理方法可解决大部分类似多音词。但是还会有一些词是不能用此方法解决的，例如对于词条“一行”中的“行”字，如果此词条出现在“一行人等”中，应读作“xing2”，如出现在“上一行”、“下一行”等场合，则应读作“hang2”。这类词不易在词库中穷举，靠上下文搜索相关词句来判定其读音实现起来也是十分复杂

的。所以更加智能化的解决方案实现起来是很困难的。

广东话的多音词和普通话有许多是同样的，也有许多是不同的，所以对于广东话的文本语音转换系统必须设立另一词库，如从普通话的多音词库借用过来，先要删除一些在广东话中并非多音的词条，再增加一些在普通话中不是而广东话中是的多音的词条，这是非常细致而琐碎的工作。

## 5, 变调的处理

在文本语音转换系统中，如不注意变调的处理，就会读出不好听的语音来，例如普通话中出现连续两个单字的读音为三声时，前一个就应改读二声。例如“厂长”一词必须读成“chang2 zhang3”，若读成“chang3 zhang3”就会令人觉得很别扭。。在我们的文本语音转换系统中实现这种变调是靠统一的规则。但又有例外，例如在普通话中有一些人称词如“姐姐、婶婶、奶奶、姥姥”等虽然都是两个三声相连，却不能按上述的三声变调来处理。这些词都应将第一字读成3声，第二字读成轻声。其它人称词如“爸爸、妈妈、哥哥、弟弟”等也都要将第二个字读轻声。在广东话中，类似的情况也很多，例如单个“妈”字的读音本应是和普通话一样读作“ma1”，但在称呼语“妈妈”一词中就应读成“ma4 ma5”。

广东话的变调现象还远不止这些。还可以举一些例子如：

“姐”在“姐妹”中读“ze2 mui6”但在“家姐”中读“ga1 ze2-1”

“兰”在“兰花”中读“lan4 fa1”，但在“荷兰”中读“ho4 lan4-1”

“爷”在“老爷”中读“lou5 je4”，但在“伯爷”中读“bak8 je4”

“尾”在“头尾”中读“teu4 mei5”，但在“收尾”中读“seu1 mei5-1”

“晚”在“夜晚”中读“je6 man5”，但在“听晚”中读“ting1 man5-1”（明天）

“毛”在“眉毛”中读“mei4 mou4”，但在“黄毛”中读“wong4 mou4-1”

“人”在“几个人”中读“gei2 go3 jen4”，但在“一个人”中读“jet7 go3 jen4-1”

“儿”在“儿歌”中读“ji4 go1”，但在“乞儿”中读“het7 ji4-1”

“带”在“温带”中读“wen1 dai3”，但在“皮带”中读“pei4 dai3-2”

如此等等，据《广州话入门》（文献1）书中叙述，这样的变调字词不下数千个，可见广东话的语音变调现象是十分复杂的，但我们仍可也用同样方法处理，这就是尽可能在多音词词库中罗列齐全，用最大匹配法在切词过程中将非默认读音的字找出正确读音（此类变调的词库尚在整理之中）。

在此基础上已开发出的广东话语音系统，应香港失明人协进会的要求开发出了广东话的盲人计算机软件，鉴于海外大多实用的是大五码汉字系统，所以该系统可运行在大五码 WONDOWS 上，系统包括盲用编辑器，此编辑器为盲人提供了文本朗读、汉字输入与编辑的功能，还有盲用网络浏览器，允许盲人独立上网浏览网页，还有屏幕朗读者，允许盲人听读 WINDOWS 桌面上的和菜单条中的信息。该系统已于 1998 年开始用在香港、和海外粤语华人的盲人群体中，很受欢迎。

## 6, 结束语

我国虽以普通话为主要流通语言，但是方言种类十分丰富，象广东话、上海话、四川话等地方语就各有数千万乃至上亿人口将其作为生活用语。开发这些方言的语音合成系统就很有必要。本文所介绍的广东话合成系统因其用于盲人计算机上而凸显其实用性。象苏州话、湖南话等等地方语也因各有其特色而受人喜爱，以这类地方语作为日常交流用语的人数也不下数百万。当人们听机器说自己的家乡语时会感到亲切和欣慰，保留地方语音并

使之规范化、标准化也是保留和发扬我国丰富多采的多民族、多文化的重要任务之一，所以地方语的语音合成技术是大有用武之地的，加强这方面的研究和开发是有意义的。

### 参考文献

- (1) 广州话入门，陈宝如编著，广东人民出版社，1997年
- (2) 香港粤语词典，郑定欧编纂，江苏教育出版社，1997年