

## 语言与通信信号处理学会报告摘要

### 普通话字调与语调 在语音处理上的问题

中国社会科学院语言研究所

吴宗济

#### 一、问题的提出

近年来的国际上语音处理工程已不满足于语音的音色处理，而注意到连续的韵律特征的处理，其中又以声调特征和其变化为主要研究对象。不过印欧语系的语调概念用在汉语的声调处理是不够理想的。汉语普通话的字调和语调(本文通称声调)是区别意义(词义和语义)的重要信息之一，它的声学参量在信号处理上与辅音元音同其重要。汉语声调并不象早年人们所认识的那样单纯，它在双字连续时以及在语句中都有变化。再加上语法、语态及速度等等因素更产生数不清的调形。如何分析出它的有效参量，掌握它的变化规律，而定出一套可供语音处理上应用的规则来，是当前的迫切需要解决的问题。

## 二、语音处理上所需的声调声学参量

语音处理上所需的声调参量，总的说采是频率数据，包括基频的上下限 $f_0$ ，拱度 ( contour ) 的时变模式，个人的平均调域，以及振幅的分布，长度的变化等等。由于说话人的性别、年龄、习惯、环境各有不同，就给处理上带来许多困难。特别是在声码通讯上，分析 ( 声调提取 )、合成 ( 高真度的声调合成 ) 与识别 ( 声调模式的提取和匹配 ) 等方面的要求不同，所取参量也有相当差别。本文只对这类参量作一简单叙述。由于自己不是通信专业，有些看法可能不合要求，还希望得到指正！

## 三、普通话的单字调和双字调

(1) 单字调型 普通话单音节 ( 单字 ) 四个声调单读时的基本模式，已为人们熟知。调型通常归纳为：阴平—高平调 ( 55 )，阳平—高升调 ( 35 )，上声—低降升调 ( 214 )，去声—高降调 ( 51 )。但实际的读音与此有些出入。

(2) 双字调型 普通话双音节 ( 二字组 ) 共有十六个组合，构成十五个双调模式 ( 其中上声与上声组合和阳平与上声组合调型相同 )。一般前字调受后字调的影响，在尾音与后部相呼应或连接而成为平滑的过渡调型。后字声母是清辅音 ( 无规则音 ) 时，调型在两音节间断开；是浊辅音或零声母 ( 周期性乐音 ) 时，调型连接。十六个组合的双调模式如下图：

后 前	阴平	阳平	上声	去声
阴平				
阳平				
上声				
去声				

(图中虚线表示调型的断开或连接)

普通话的单字调型和双字调型是构成三字以上组合以及语句调型的基本单元。它在不同语句的语调中保持其一定的稳定性。

(3) 三字以上组合的调型 普通话三音节以上的调型都是单字调型和双字调型这两个基本单元的错综组合而成的。调型的变化与各单字或词在语法上的组合松紧有关。

(4) 弯头与降尾 每个音节的头或尾，或每一语句首字的开头或末字的收尾，在声调提取中往往在所给数据上或图形纪录上，出现一小段不属于基本调型的拱度，其趋势和调型相反。(例如高降调开头有微升的一小段，中升调开头有微降的一小段等等)这一小段调值在听辨上是不显著的，但和声调的自然度有关。

(5) 轻声 普通话的轻声字不能单独存在，一定有其前接字，按语言学的性质可分为(1)歧义轻声，(2)同义轻声，(3)助词轻声和

(4)过渡轻声。举例如下:

歧义轻声: 孙子~~不~~孙·子

同义轻声: 老虎—老·虎

助词轻声: 好·的(语助), 出·来(助动)

过渡轻声: 不·是我, 顷·和园

轻声的调型一般是短而微降。调高大致可分为高、中、低三级, 视前接字调的不同而异。在阴平、阳平后面为中调, 在上声后为高调, 在去声后为低调。举例如下:

酸·的

甜·的

苦·的

辣·的

#### 四、普通话的语调

普通话语调(句调)的调型, 表面看来非常复杂, 其中每个字的调型已和原来的调型不同。汉语语调和单字调的关系既不是代数相加, 也不是削足适履(例如象西方有些文献所提出的, 单字调被语调所约束, 单字调嵌入定型的语调中而变调等等)。普通话语句的调型除了有时服从普遍性规则(如: 陈述句语调尾部趋降, 疑问句尾部趋升之类)外, 并没有固定的语调模式, 而是以上述的单字调型、双字调型作为建筑材料(基本材料), 不过它受语法制约, 并在语态、加速等影响下而使调<sub>调</sub>和调域有一定的变动。这些因素分述如下:

(1) 语法结构的制约 单字或二字组与其它单字或多字组合时

会根据语法结构之不同而改变原来组合规律。如：

好 ㄨ + 马 ㄨ → 好马 ㄨ ㄨ

但 好马 + 跑 ㄨ → ㄨ ㄨ ㄨ ㄨ

又如：

老 + 厂长 → ㄨ ㄨ ㄨ ㄨ (厂长是老的)

但 老厂 + 长 → ㄨ ㄨ ㄨ ㄨ (厂是老的)

(2) 语态环境的影响 不同语态(或称语气)的语句有陈述句(平叙)、加强句(惊叹、祈使、命令等)、疑问句等分别,它对基本单元的调型在平均调 $\bar{\omega}$ 或调域上有一定影响,但不改变拱度的趋势(读得太快时除外)陈述句中的语调总趋势是逐渐微降,称为倾势(declination)。疑问句在句中如没有疑问词(如是不是,有没有,吗等),字句和陈述句基本相同时,则句尾调形可能略升或变形,但其它字调拱度一般不变。如有了疑问词,则调型常和陈述句没有分别。加强句一般是主要词的调 $\bar{\omega}$ 上限抬高或平均调值升高,或调域扩展。

(3) 说话加快的影响 人们日常口语的速度,个人都有其习惯。但如果需要加快说出时,则字调来不及达到其应有调域而被压缩,甚至失去本调而被“拉平”,在句中的次要字眼尤其如此。(但加强的加速语句除外)。

#### 四、声调处理中的问题

语音信息处理多数应用于合成与识别方面。合成要求输出信号的保真，识别要求输入参量的匹配（包括模式匹配与音位提取）声调的合成与识别，似比辅音、元音的简单，但因声调在连续语言中，参量的变化很大，处理起来也有一定的困难。

声调处理按需要不同，出现的问题也不同。分述于下：

(1) 声调的提取 提取声调的方法不一（如自相关、倒频谱、线性预测等），主要都不外是把准周期性的波提出峰值，按其时间间隔而转换成频率数据。常见的问题是：在连续语言中，某些波峰能量过低而得不出数据，或在某些元音的一个周期内提取了两个峰，以致声调忽然跳跃（一般可比实际调值高一个倍频程）等。

(2) 声调的合成 合成的声调在听感上的自然度常与声调的特征是否能正确提取有关。声调特征的参量输入得过于简化，或取得的数据有误差都导致听感上的失真。各种情况略如下述：

1、语句的起讫听来生硬，这是由于未考虑弯头降尾的数量，以及未顾到振幅大小在声调上的分布规律，例如阴平和去声多半是先重后轻，阳平是先轻后重，而上声是重轻重（双峰型）

2、音节间的连接不够平滑，形成跳跃音感。这是没有考虑到过渡调型的连断关系，及其平滑走势。

3、声调听起来呆板 这是未考虑到幅度在语句中的分

情况，以致未脱“机器音”。

4、各个声调的长度分配得不合理。

5、男女声调的声带颤动周期开闭相位不同，一般是女声带开闭的时长相等，而男声带开长闭短。

(8) 声调的识别 声调识别在汉语语音处理中有特别意义，因它能区别同音字，及帮助区分音节。常见的问题是：

1、调域不能匹配 说话人因性别、嗓音的不同，各人的平均调域和限 $\bar{\omega}$ 亦异。处理的带宽有限度，难以复盖。有许多实验证明，把男的平均调域增到1.8倍，听来就是女调，在识别时一般用自动换挡的装置来解决这个问题。

2、调值超出平均限 $\bar{\omega}$  口语中有时有个别字调会出现过高域过低现象，影响了识别的准确度。

3、两音节间由于后字是独辅音（清辅音因连续的同化作用而浊化也属此类）而使调型不能断开，给音位识别造成误差（有时凭调型拱度的抑扬现象（flip-flop或blocking）可以适当补救。

近年来的语音识别经验已经证明：“从声学平面角度来判断语音，必须具备韵律特征的知识”。而声调是其中最主要的特征。单字调型既有种种变化，因此把双字调型作为一个单元来处理（国际上已有些趋势），会简化许多程序而减少差错。

(参考文献略)

1982年10月