

# 表述汉语重音的属性参数研究

王玮<sup>1</sup> 蔡莲红<sup>1</sup> 周同春<sup>2</sup>

<sup>1</sup>清华大学计算机科学技术系 北京 100084

<sup>2</sup>北京师范大学中文系 北京 100875

## 摘要

汉语语句中的重音研究是韵律特征研究中的重要内容。一般认为汉语语流中重音特征主要是通过音高、音长的变化加以体现的，但是在重音确定时表明仅仅选用这两个属性参数对于其分析并不完善，重音的确定还到其他很多属性参数的影响。本文以男声实际语料库中的完整语句作为研究对象，提出一种采用重读音节的三个属性参数作为其表述重音的描述方法，并基于优化的神经网络模型对该属性集进行训练和分类。

## 引言

长期以来，汉语语句重音的确定一直是汉语韵律研究的主要部分。汉语重音的定义如下：在语流中，各音节的响亮程度并不完全相等，有的音节听起来比其它音节响亮，这就是重音。

汉语的重音可以划分成三级：正常重音、强重音和弱重音。正常重音在短语或词中按一般平叙句的发音习惯所表现的轻重；强重音被强调、加重语气读出；弱重音按发音习惯的轻读，一般出现在词后缀、宾语代词、动词或名词重叠部分。这种划分是以词为考查对象的。人们在口语交流中，常把在表情达意方面较重要的词读得重些，其余的词读得轻些。

重音又可以划分成：语法重音，逻辑重音，心理重音。语法重音是由句子语法结构特点决定的，它在表达上，具有提示突出话语中某些语法成分的作用。逻辑重音根据句子在上下文语境中的逻辑联系来确定的语句重音。这种逻辑重音，大都具有对比、强调的含义。心理重音反映说话人在口语交际中的重心，即说话人所要强调的内容。

一般认为汉语的重音的特征描述主要是体现在重读音节的音高和音长上，即重读音节的音高增大，音长变长，但是大量的实际测试中发现仅仅采用这两个属性作为重读音节的有效描述远远不够，音节在语句中的重读与否还受到其他很多因素的制约。本文以一个实际男声言语数据库作为研究对象，提出应该考虑重读音节的五个重要属性，并基于优化的神经网络模型对该属性集进行了训练和分类，开放性的测试结果说明本文选择的重音属性能够比较完善的反映出音节重读时的特征。

## 1 汉语重音的相关研究

汉语里对词层面进行重音研究的语料，一般都是带有正常重音的词组，是一种相对稳定状态下的研究。汉语里词层面上对重音的主要研究结果有：对各种声调组合构成的双字词的声学分析表明，普通话两字组中的轻重音分布主要是后字较重，重读音节的时长均长于非重读音节，其基频表现也更接近单说时的声调模式，但大多数非重读前字的强度大于后字。因此，普通话两字组的正常重音在声学上的表现是具有较长的长度和较完整的音高模式，而不是具有较大的强度（林茂灿等，1984）。对三字组和四字组韵律特征的实验还发现，词组中，从首字到末字基频音域位置逐渐下降，其中末字音域最宽，音域下限下降较为明显，末字具有较完整的音高模式，其时长可能较大，但强度不一定大（杨顺安，1992）。

沈炯(1994)对语势重音的研究中指出“在听辨语势重音时,音长的作用并不明显,音高的作用却很重要”,和“语势重音的音理是声调音域高音线上移,它使声调音域向上扩张,声调音域低音线与语势重音没有明显的关系。”

高明明(1993)对普通话语句中强调重音的声学表现做了深入研究,实验中所用的语料是包含了三个音节的较短语句,她的结论是:1.基频升高是强调重音的重要声学表现。基频升高的方式与声调的音高特征和曲拱特征密切相关。具体的说,阴平的基频整个地上升,阳平主要是高音点上升,有时伴随着低音点下降,去声主要是高音点上升;上声读成半上低音点下降,读成全上时调尾上升。2.强调重音的时长普遍加长。就重读阳平和去声来说,音高升高,时长加长;阴平重读时时长与音高的变化存在着互补关系,基频较高时,时长增加的幅度相对减少,反之,时长增加幅度加大;对于上声重读时基频变化幅度不大,时长有明显的加长。3.音节强调重读时对其强度没有明显的影响。

许洁萍(许洁萍,1999)在对一广播数据库进行了重音标注后,研究得到:“语句重音对音高和音长这两个韵律特征参数都有重要影响,说明语句重音主要是通过音高和音长这两个特征参数的变化体现出来。”“在一般情况下,音高和音长是彼此相互独立的”,“在轻读时,会降低音高,同时读得短一些,因此在轻读时,音高和音长表现为正相关”,“大多数情况下,语句重读会使词的音高上限增高,时长加长但在句头基频很高的情况下,时长加长的作用明显减弱,语句重音往往不会引起时长的加长;当在停顿前和由上声为主组成的词中,提高音高受到制约,重读是词时时长增加较多。”

对于汉语语句重音的研究,由于分析的语音材料和采用的研究方法不同,得到的研究成果不完全一致,但是总体来说,汉语重音的声学表现主要是基频和音长的变化。

## 2 重音的特征描述属性

一般的汉语普通话中存在两类重音:词重音和语句重音。词重音出现在词中,是由于词的含义不同,重读音节的位置也不相同。语句重音出现在句子中,是由于句子的语法结构、语义逻辑、及心理、情感的表达需要而产生的语句中词的重读。在一句话中可以有一个或几个语句重音。

由于目前较为先进的语音处理方法是句子一级为处理对象进行的,因此单个词重音的研究在实际应用中的受限范围较大,本文主要考虑语句中语句重音,即从连续语流中抽取重读音节,并根据其自身的属性特征进行分析。

本文实验数据是一个男声朗读的500句包含不同语气的数据。首先由两个听音员的对语句中各自认为是重音的音节或词进行标注,如“骆驼生活在沙漠里。”这句话,听音员一致标出“骆驼”为重读音节,我们就抽取相应的特征属性进行描述。对于重音的判别个体的差异是比较大的,为了研究的需要,这里我们只研究两个听音员都认为是重音的音节的属性特征。

一般认为重音属性选用的是相应音节的音高和有效音长,这两种属性描述对于无声调语言而言,就已经够了,众所周知,汉语是一种声调化的语言,因此仅仅采用这两个属性进行重音分类正确率就不是很好。本文通过增加重读音节的描述属性来提高其分类的正确性,这里我们选用了重读音节的音高、时长以及音域信息作为重音的有效信息描述。这里的时长指的是整个重读音节的时长,其包含有停顿信息,由于语句中停顿会对整个音节的基频值和时长值产生较大的影响,因此我们在研究重音时,语句中的停顿信息应该加以考虑,而不能仅仅孤立地处理有效时长。本文加入音域信息主要是为了更加有效地描述相应的重音信息。音域信息的计算是采用音高的最大值减去最小值并按照吴宗济先生提出的十二半音进行对数变换得到的,这样可以更好地体现汉语这种声调语言的特性。

因此,本文采用重读音节的基频均值、音节时长、音域等三个信息属性作为后续神经网络的输入训练集,神经网络的输出是重音分类的结果。

## 3 优化神经网络模型进行语句重音的分类

为了检验我们提出的采用重读音节的三个属性值描述重音的合理性,我们采用人工神经

网络进行模拟。

人工神经网络最大的优势在于分类和识别，神经网络是受到动物神经系统的启发，利用大量简单处理单元互联而构成复杂系统，能够解决一些复杂模式识别与行为控制问题。一个神经网络包含大量的简单处理单元，这些单元之间通过激励连接和抑制连接形成一定的拓扑结构，并进行着相互影响。所有神经元一起，接受训练算法的控制，建立起单元之间的连接权，连接权值决定了网络对任意单元的计算响应，所以连接权存储了网络的长时间记忆和知识。

BP 网络是神经网络中最简单、最实用的一种网络，它的学习算法是 BP 学习算法，但是 BP 网络主要缺点是收敛速度慢、容易陷入局部极优等问题，目前 BP 算法的改进主要有两种途径，一种是启发式学习方法，另一种则是对经典算法进行优化。

这里本文采用的是一种的优化学习算法 动量自适应学习率法。

动量—自适应学习率法能够降低网络对于误差曲面局部细节的敏感性，有效地抑制网络陷入局部极小。由于标准的 BP 算法实质上是一种简单的最速下降静态寻优算法，在修正权值  $W(k)$  时，只是按照  $k$  时刻的负梯度方式进行修正，而没有考虑以前积累的经验，即以前时刻的梯度方向，从而容易使学习过程发生震荡，收敛缓慢。我们采用以下权值修正公式 (1) 加以改进：

$$W(k+1) = W(k) + \alpha[(1-\eta)D(k) + \eta D(k-1)] \quad (1)$$

这里  $W(k)$  既可表示单个权值，也可表示权值向量。  $D(k) = \frac{-\partial E}{\partial W(k)}$  为  $k$  时刻的负梯度，

$D(k-1)$  为  $k-1$  时刻的负梯度， $\alpha$  为学习率， $\alpha > 0$ ， $\eta$  为动量因子， $0 \leq \eta < 1$ 。这种方法所加入的动量项相当于阻尼项，它减小了学习过程的震荡趋势，从而改善了收敛性。

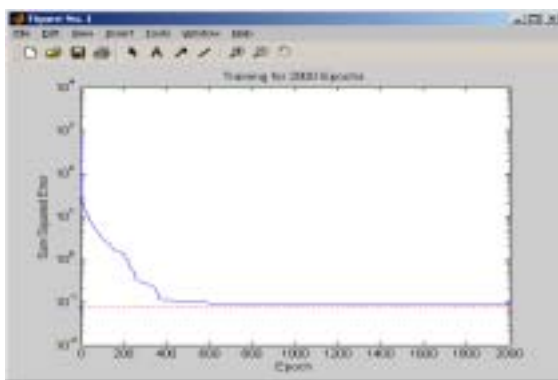
采用自适应调整学习率有利于缩短学习时间，BP 算法收敛速度慢的一个重要原因是学习率选择不当，学习率选得太小，收敛速度太慢，学习率选得太大，则有可能修正过大导致震荡甚至训练完全发散，因此采用了自适应调整学习率的改进算法。

$$W(k+1) = W(k) + \alpha(k)D(k) \quad (2)$$

$$\alpha(k) = 2^{\lambda} \alpha(k-1) \quad (3)$$

$$\lambda = \text{sign}[D(k)D(k-1)] \quad (4)$$

当连续两次迭代其梯度方向相同时，表明下降太慢这时可以使步长加倍，当连续两次迭代其梯度方向相反时，表明下降过大，这时可以调整步长减半。



下面给出基于动量—自适应学习率法的网络误差曲线：

从训练结果可以看出，优化的神经网络模型能够较好地输入的属性集进行学习，能够达到误差设定的目标。

表 1 给出网络采用不同属性的分类结果：

正确率	重读音节	非重读音节
二个属性	71%	80%
三个属性	80%	89%

从表 1 的结果可以看出由于本文增加了音域信息，对重音分类的准确度有了一定程度的提高，所以增加音域信息对于确定重读音节是有积极作用的。

## 4 结束语

本文基于实际言语数据库进行重读音节的分析，提出了采用三个属性值作为重读音节的描述方法，并取自连续语句中的听感重音用优化的神经网络模型进行训练学习，通过和经典的

二值属性描述的比较说明了增加属性的合理性和有效性。

## 参 考 文 献

- <sup>[1]</sup> 许洁萍, 1999, “基于言语数据库的汉语韵律特征研究”, 博士论文, 中国科学院声学研究所
- <sup>[2]</sup> 吕沐华, 2001, “汉语韵律标音系统的设计与实现”, 硕士论文, 清华大学
- <sup>[3]</sup> 郭锦桴, 1993, “汉语声调语调阐要与探索”, 北京语言文化大学出版社
- <sup>[4]</sup> 周俏峰, 蔡莲红, 1996, “汉语重音及在 TTS 系统中的模拟”, 微型计算机, 第 4 期
- <sup>[5]</sup> 蒙以正, 1999, “MATLAB 5. X 应用与技巧”, 科学出版社