

一个反映音节轻重的特征量——声门振动次数

殷治纲

摘要:“轻重”问题是语言韵律研究的重要内容之一。该领域研究的一个重要内容是分析各声学因素对轻重感知的影响。本文将围绕该主题,试图从常见声学因素中找出最能反映轻重感知的特征量。

以往研究认为音高和时长是反映音节轻重感的最重要声学因素。但由于汉语作为声调语言的特殊性和韵律轻重感知的复杂性,音高或时长特征在反映轻重时各有自己的局限性,需要综合考虑。

结合音高和时长的特性,本研究提出了一个可更好反映轻重感知的新特征量——声门振动次数。它是某音节发音时声带振动的次数,也即该音节元音段的声波脉冲数(Pulse)。在数学上,声波脉冲数表现为音高(pitch)和时长(duration)的积分($Pulse = \int f(pitch) dt$),所以它很好的整合了音高和时长在反映轻重时的优势,并补偿了各自劣势。统计证明音节声门振动次数与音节轻重感知结果的相关性很好,高于音高、时长以及音强等其他简单声学量与轻重感知的相关性,可在一定程度上作为反映音节轻重感的特征量。

关键词: 声门振动次数, 轻重, 特征量

1. 序言

语音韵律研究的一个重要内容是对“轻重”问题的探讨。根据所研究语音单元的不同,“轻重”研究又可以细化为音节重音、词重音、短语重音乃至句重音等不同领域的问题。

研究轻重问题首先要解决的问题是:什么是轻重?以及轻重和哪些语音因素有关。

以往研究认为,轻重是对语音单元凸现(prominence)程度的主观心理感知,这种主观的心理感知会受客观语音因素的影响[1]。具体说来,音高、时长、音强、声调、音质(声、韵母类型)、韵律位置等各种因素都可能对轻重感知产生影响,但在诸多影响因素中,某些语音因素对轻重感知的影响力强一些,而另一些则可能比较弱。[2]

在此,我们要研究的问题是希望从所有与轻重感知有关的简单语音因素中,找出一个与之相关性最好的声学/生理参数来作为表征轻重程度的特征量。该特征量虽然不可能作为轻重感知的标准,但至少在很大程度上可以作为衡量语音轻重的参考。

2. 反映音节轻重的主要语音因素

2.1 时长和音高

在诸多语音因素中,音高和时长被认为是与轻重感知关系最密切的两个因素。

具体说来,重音所在语音单元的主要声学表现是音高上限的抬高(上声是压低)和音域(音高范围)的加宽[3][4][5][6][7];在时长方面的表现主要是时长的加长[8][9][10]。

另外,以上结论也可以从重音单元处的发音强化角度来解释——发音强化会延长时长,并凸显承载音节原有的音高特征,使“高者越高,低者越低”。由于汉语普通话中阴平(HH)、阳平(LH)、去声(HL)三个声调中都有“高”(H)特征,所以负载这些声调的音节重读时都会使“高”(H)特征更凸显,即进一步抬高音高上限,从而使得调域加宽。普通话中唯有上声(LL)的目标值是“低”(L)特征,所以上声音节重读的结果是使音高下限更低,但这也使调域变宽了。

那么,音高和时长谁对轻重感知影响更大呢?目前该问题仍存在不同意见。有学者认为时长是影响轻重的首要因素。[11][12]。但也有很多学者认为音高对轻重感知的影响更大,如有研究[13]发现,双音节韵律词重音差与高音点的差值相关性强于时长差值。曹文[14]研究了焦点重音的表现形式,发现韵律信息中最重要的是高音点落级差别,其次是时长。蔡莲红[15]提出了拟合重音听感的线性公式 $L = 1.5F(\text{音高}) + 0.95(\text{时长}) + 0.65R(\text{调域})$,从各因素系数上暗示音高的重要性更高。此外,也有很多研究观察到了重音处音高和时长之间呈现的互补关系或交换关系。[1][16][17]但大多数认为这种互补关系中音高影响占优。

以上对于音高和时长在表现重音时重要性顺序的争议,其实从一个角度说明了它们反映重音时各有局限,即有时重音表现为二者同时凸显,但也可能有时只表现为其中一个特征凸显。这种补偿(或选择)关系意味着它们中任一个在逻辑上都不能单独作为表现重音的充要特征。

2.2 时长和音高作为轻重参考量的局限性

我们前面认为音高特征（又可细分为音高上限和调域等特征量）和时长特征在表现重音时各有局限。这些局限性主要包括：

- 1) 选择音高上限作观测量无法准确描述上声（3声）重音的表现，因为上声的目标音高是“低”，重读只会强化其“低”特征，使音高下降，而不是像其他三个声调一样抬高音高；
- 2) 音高上限作观测量的另一个局限是，较大韵律单元中的音高下倾现象使得前面音节往往比后面音节更有音高优势，但前面的音节未必就一定更重（还要看局部音高对比）。
- 3) 选择调域作为观测量无法准确描述阴平（1声）重音，因为阴平调域本来就较窄。
- 4) 选择时长作为观测量也不全面。曹剑芬（2005）指出音段延长分为重音凸现延长和韵律边界延长等多种情况。对于后者来说，韵律边界处可能存在末尾延长效应，但延长的末音节未必就一定更重过其他音节。

总的说来，音高和时长的各单独特征量都不能反映所有普通话音节的重音情况，只有找到一个可同时涵盖音高和时长的新指标，才能更全面地反映各种重音情况。

2.3 一个新的轻重特征量——声门振动次数

经过分析，我们找到的可同时涵盖音高和时长要素的生理/声学量是声门振动次数。声门振动次数也就是声带振动的次数，它从性质上是个生理数据。但由于声门振动与声波脉冲有严格对应关系，所以一个音节的声门振动次数在数值上也等同于该音节浊音段（一般是元音段）声波脉冲次数。

从下图某音节的音高（基频）曲线图可以看出声门振动次数、音高和韵母时长三者间关系。

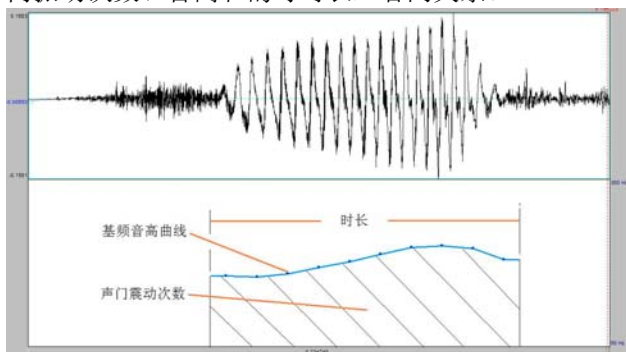


图 1. 声门振动次数、音高和时长的关系图

图中声门振动次数对应于蓝色基频曲线包络（图中斜线部分）面积，声门振动次数 Pulse number 的计算公式为： $P = \int f(\text{pitch}) dt$

公式中 $f(\text{pitch})$ 为基频音高曲线函数， P 为声门振

动次数， t 是时间。所以声门振动次数等于基频音高曲线在时长范围内的积分。

声门振动次数和音节的音高与时长都有关系，并有以下特点。

- 1) 音高（基频）的抬高或降低，可以增加或减少声门振动频率（单位时间内声门振动次数），从而改变其数量。
- 2) 时长的延长或缩短，同样可以可以增加或减少声门振动次数。
- 3) 声门振动次数既是一个生理量，又对应于声学量（声波脉冲数），便于测量。
- 4) 声门振动次数的单位为“次”，无量纲，可以统一音高和时长两种特征。
- 5) 声门振动次数的局限性在于它和音高曲线一样，一般对应音节的浊声段，所以用来描述音节元音部分比较适合，而无法描述噪音段（一般是辅音段）部分。不过梅晓（2010）证明音节时长变化主要表现为元音的变化，大部分辅音的时长并不随音节时长发生太大变化，所以轻重引起的元音变化基本可以代表音节整体变化情况。

3. 数据分析与比较

在上面理论分析中，我们认为声门振动次数在反映音节轻重时同时具备了音高和时长两类因素的优势，并可以通过二者之间的补偿关系克服各自局限，理应与音节的重音级别具有更好的相关性。接下来，我们就来看一下实际数据是否如此。

3.1 实验数据介绍

我们统计分析的语料是中国社会科学院语言研究所语音研究室录制的 ASCCD (Annotated Speech Corpus of Chinese Discourse) 朗读语篇语料库。该语音语料库内容包括各种文体的 18 篇文章(约 1 万音节)，语音数据由 10 位北京地区标准普通话发音人录制而成。本统计使用了其中第 1 位女发音人 (F001) 和第 1 位男发音人 (M001) 的数据。两名发音人发音风格差别较大，F001 发音流畅自然，M001 发音铿锵激昂，便于展开对比研究。

声音文件采用 16KHZ 采样、16 位数据精度、双声道 WAV 格式存储，库容量约 1.5GB。所有语料采用人工标注方法进行了语音学标注，标注信息包括了（正则）拼音、声韵母、韵律边界、重音级别等。其中重音级别标注了 0、1、2、3 共四级重音，分别对应轻、正常、较重和很重四级听感。

下面，我们对比分析一下音节重音级别和基频音高上限(pitchMax)、调域（即音高范围 pitchRange）、音节时长(Duration)、音强(Intensity)以及声门振动次数(Pulse)之间的相关系数，看看哪个变量和重音级别的相关性最好。

以上所有音节数据均未作规整(归一)处理,为原始数据。以下统计结果中所有带**数据表示在 0.01 水平(双侧)上显著相关。

3.2 声门振动次数和重音相关性

我们将发音人 F001 和 M001 的音节重音级别和声门振动次数作了相关性分析,得到相关系数分别为 0.498 和 0.553(在 0.01 水平显著相关)。

表1: 重音级别和声门振动次数的相关性

声门振动次数和重音相关性 (F001)			
		重音级别	声门振动次数
重音级别	Pearson 相关性	1	.498**
	显著性(双侧)		.000
	平方与叉积的和	5768.532	44183.042
	协方差	.658	5.043
	N	8762	8762

声门振动次数和重音相关性 (M001)			
		重音级别	声门振动次数
重音级别	Pearson 相关性	1	.553**
	显著性(双侧)		.000
	平方与叉积的和	6429.746	35544.757
	协方差	.734	4.059
	N	8759	8759

3.3 音高上限和重音的相关性

接下来看一下音高上限和重音的相关性。此处的音高上限就是指每个音节的最高基频值。

下表显示,在所有声调(包括所有四个声调以及轻声)的音节参与统计时,F001 和 M001 的音高上限和重音级别的相关性分别为 0.367 和 0.477(在 0.01 水平显著相关)。它们均低于各发音人声门振动次数与重音级别的相关性。

表2: 重音级别和音高上限的相关性

音高上限和重音的相关性 (F001) ——所有声调			
		重音级别	音高上限
重音级别	Pearson 相关性	1	.367**
	显著性(双侧)		.000
	平方与叉积的和	5768.532	150320.550
	协方差	.658	17.158
	N	8762	8762

音高上限和重音的相关性 (M001) ——所有声调			
		重音级别	音高上限
重音级别	Pearson 相关性	1	.477**
	显著性(双侧)		.000
	平方与叉积的和	6429.746	179208.255
	协方差	.734	20.462
	N	8759	8759

考虑到汉语四个声调中,上声(3声)和其他声调的重音表现不一样(上声的典型音高特征是

“低”,所以其重读时音高一般不升反降),所以有必要排除上声音节后,再看看其余声调音节该项相关性数据是否会升高。具体结果如下。

表3: 重音级别和音高上限的相关性

音高上限和重音的相关性 (F001) ——去掉上声			
		重音级别	音高上限
重音级别	Pearson 相关性	1	.433**
	显著性(双侧)		.000
	平方与叉积的和	5070.975	150482.186
	协方差	.697	20.693
	N	7273	7273

音高上限和重音的相关性 (M001) ——去掉上声			
		重音级别	音高上限
重音级别	Pearson 相关性	1	.553**
	显著性(双侧)		.000
	平方与叉积的和	5610.037	175901.609
	协方差	.772	24.206
	N	7268	7268

上表显示排除上声后,F001 的音高上限与重音相关系数从 0.367 提高到 0.433,M001 则从 0.477 提高到 0.553。这说明在表达重音时,上声的音高表现的确不同于其他几个声调。去掉它后,音高上限与重音相关系数有了很大提高,基本接近于声门振动次数和重音级别的相关性水平。

3.4 音节调域(基频范围)和重音的相关性

下面我们再看一下音节调域与重音级别的相关性情况。本处所说的调域即音节的基频范围,也就是音节最高基频值和最低基频值的差值。具体相关性数据如下表。

表4: 重音级别和调域的相关性(所有声调)

调域和重音的相关性 (F001) ——所有声调			
		重音级别	调域
重音级别	Pearson 相关性	1	.087**
	显著性(双侧)		.000
	平方与叉积的和	5768.532	71510.842
	协方差	.658	8.162
	N	8762	8762

重音级别和调域的相关性 (M001) ——所有声调			
		重音级别	调域
重音级别	Pearson 相关性	1	.385**
	显著性(双侧)		.000
	平方与叉积的和	7720217.731	85721.164
	协方差	881.505	9.788
	N	8759	8759

数据显示,在包括所有声调音节时,F001 调域和重音级别的相关系数只有 0.087(在 0.01 水平显著相关)。M001 调域和重音级别的相关系数为 0.385(在 0.01 水平显著相关)。

以上相关性不算高,似乎并不符合传统句重音理论的观点。那么改用半音值(ST)单位是否能提高相关性呢?我们利用公式: $St=12*\log_2(F0 / Fref)$ 将赫兹值转化为半音值进行了相关性分析。发现采用半音作语调单位后,F001的相关系数上升为0.147,M001为0.320(在0.01水平显著相关),但总体相关性仍然不是很高。

此外可能造成该相关系数较低的原因是阴平调音节的“干扰”。考虑到阴平调(1声)音节本身基频范围很窄,即使重读也不容易扩展其调域范围。这可能导致整体数据相关系数下降,因此我们去除了阴平调音节,又对剩下的其他声调音节作了一次基频范围和重音级别的相关性分析。结果显示去掉阴平调后,F001基频范围和重音级别的相关性结果从0.087提高到0.123,M001的相关系数从0.385提高到0.497。

具体数据如下表所示。

表5:重音级别和调域的相关性(排除阴平调)

调域和重音的相关性(F001) 去掉阴平调			
		重音级别	pitchRng
重音级别	Pearson 相关性	1	.123**
	显著性(双侧)		.000
	平方与叉积的和	4946.651	92123.737
	协方差	.697	12.988
	N	7094	7094

调域和重音的相关性(M001) 去掉阴平调			
		重音级别	pitchRng
重音级别	Pearson 相关性	1	.497**
	显著性(双侧)		.000
	平方与叉积的和	5454.157	95909.071
	协方差	.770	13.533
	N	7088	7088

以上不管采用ST作单位,还是排除阴平调音节,似乎都不能证明调域与重音级别强相关,这一结果不支持以往主流句重音理论所认为的重音级别和调域变化有密切关系的结论。我们认为造成这一现象的原因是:

1) 自然语言存在音高下倾(或调域渐窄)等现象,以往句重音理论更依赖于数据规整以排除这种下倾干扰。而我们本研究所使用的所有数据(包括调域数据)均为原始数据,未对韵律位置等内容作数据规整;

2) 以往句重音理论中的调域延展侧重于局部范围内单元间的对比,而我们此处统计范围是整个发音人数据而非局部语句;

3) 以往句重音理论中,调域参数更适合在较大的韵律单元(如韵律词或)中反映重音变化,而我们此处是以音节为单位来统计相关性。

以上情况造成了在本项数据分析中,音节调域和重音级别间并未表现出较高的相关性,但这也正说明

调域参数在反映音节轻重时仍有很多局限。

3.5 时长和重音的相关性

接下来看一下音节时长和重音级别的关系。

表6:重音级别和时长的相关性

音节时长和重音的相关性(F001)			
		重音级别	音节时长
重音级别	Pearson 相关性	1	.462**
	显著性(双侧)		.000
	平方与叉积的和	5768.532	191.168
	协方差	.658	.022
	N	8762	8762

音节时长和重音的相关性(M001)			
		重音级别	音节时长
重音级别	Pearson 相关性	1	.540**
	显著性(双侧)		.000
	平方与叉积的和	6429.746	223.062
	协方差	.734	.025
	N	8759	8759

在上表中,F001和M001音节时长和重音级别的相关系数分别为0.462和0.540(在0.01水平显著相关),仍低于声门振动次数的相关性。

3.6 音强和重音的相关性

接下来,我们看一下音强和重音的相关性。在音乐中,音强是轻重感觉的决定性因素。但在以往语音学研究中,一般认为音强对轻重的影响力没有音高和时长那么明显。

表7:重音级别和音强的相关性

重音级别和音强相关性(F001)				
		indexofST 重音级别	intMax 音强最大值	intMean 音强均值
重音级别	Pearson 相关性	1	.061**	.051**
	显著性(双侧)		.000	.000
	平方与叉积的和	5768.532	1647.355	1612.972
	协方差	.658	.188	.184
	N	8762	8762	8762

重音级别和音强相关性(M001)				
		indexofST 重音级别	intMax 音强最大值	intMean 音强均值
重音级别	Pearson 相关性	1	.015	.008
	显著性(双侧)		.158	.450
	平方与叉积的和	6429.746	738.692	436.941
	协方差	.734	.084	.050
	N	8759	8759	8759

我们在上表中分别统计了音强的最大值和均值与重音的相关性。总的看来,音强最大值和音强均值与重音的相关性都很低,F001的重音与音强最大值、音强均值相关系数分别为0.061和0.051,M001则分别为0.015和0.008。我们认为音强和重音相关性之所

以这么低，除了声学相关特性外，还在于音强数据很难在绝对标准条件下收集，因为录音时发音人离话筒距离的远近变化、发音人发音状态的变化，都会对声音音强产生很大影响，这都造成了音强数据的不稳定性。

3.7 能量和重音的相关性

接下来，我们看一下音节总能量和重音的关系。具体统计数据如下。

表8: 重音级别和能量的相关性

F001 重音级别和能量的相关性			
重音级别		重音级别	音节能量
	Pearson 相关性	1	.046**
	显著性 (双侧)		.000
	平方与叉积的和	5768.532	1124.542
	协方差	.658	.128
N	8762	8762	

M001 重音级别和能量的相关性			
重音级别		重音级别	音节能量
	Pearson 相关性	1	.016
	显著性 (双侧)		.133
	平方与叉积的和	6429.746	359.953
	协方差	.734	.041
N	8759	8759	

F001 和 M001 音节能量和重音级别的相关系数分别为 0.046 和 0.016(在 0.01 水平显著相关)。这组数据的相关系数也非常低。我们认为其原因和音强类似。除了声学上也许的确和重音关系不大外，能量也是很容易受录音和发音情况影响的，这进一步造成了能量数据的不可靠性。

4. 结论与讨论

最后，我们把音节各个声学特征和其重音级别的相关系数排序如下。

表9: 音节各特征量和重音级别相关性排序

各特征量和重音级别相关性	F001	M001
声门振动次数	0.498	0.553
时长	0.462	0.540
音高上限 (去掉上声)	0.433	0.553
音高上限 (所有声调)	0.367	0.477
音高范围/调域 (去掉阴平后)	0.123	0.497
音高范围/调域 ST (所有声调)	0.147	0.320
音高范围/调域 (所有声调)	0.087	0.385
音强	0.051	0.008
能量	0.046	0.016

从表中可以看出，在各基本声学/生理特征量中，声门振动次数与重音相关性最强，其次是时长和音高上限、调域，然后是音强和能量等。

虽然声门振动次数与重音级别的相关性数据仍不算很高 (不超过 0.6)，但是考虑到所有参加分析的数据都是原始数据，没有作过数据规整 (或归一) 处理，所以该结果还是令人满意的。

我们之所以在此强调原始数据未作规整，是因为数据规整的可以使分析结果更加准确而优化。但是不同的声学特征量需要进行的数据规整方法会有所不同。对于声门振动次数而言，我们认为需要规整掉的较大干扰因素有声韵母类型和韵律位置。

声、韵母类型的干扰在于：即使在相同发音状态下，单韵母和复韵母的元音个数和类型差异，也会使其声门振动次数差异较大。如把它们放在一起统计会引入很大误差，而如分类别后对相似的同类韵母进行分类内对比，则可以减少这种系统影响；

韵律位置因素的影响在于：语言中存在音高下倾及末尾延长等现象，这些现象使处于不同韵律位置的音节在音高和时长特征上具有“先天”的差异性，并影响到声门振动次数的多少。如果可以按照韵律位置分类规整后再对同位置音节进行统计，将可减少韵律位置对音节声门振动次数的影响。

在以后更深入的工作中，我们将对数据进行规整后再予以分析，相信声门振动次数与重音级别的相关性将有进一步提高。

总之，在现有基本声学因素中，声门振动次数和重音级别之间表现出了最高的相关性。可以认为某音节重读程度越高，该音节的声门振动次数 (相比参考状态) 越多。这使声门振动次数这一特征量在一定程度上可以看作反映音节轻重的参考，从而大大简化了轻重问题的复杂性。

5. 参考文献

- [1]叶军, 2008, 《现代汉语节奏研究》, 上海世纪出版集团, 2008年8月
- [2]殷治纲, 2011, 《汉语普通话朗读语篇节奏研究》, 中国社会科学院研究生院 2010 博士学位论文, 《中国博士学位论文全文数据库》电子期刊, p46
- [3]赵元任, 1968/1979, 《汉语口语语法》, 商务印书馆
- [4]沈炯, 1985, 北京话声调的音域和调域, 《北京语音实验录》, 北京大学出版社
- [5]沈炯, 1994, 《汉语语调构造和语调类型》, 《方言》, 3, 221-228
- [6]曹剑芬, 2002, 汉语声调与语调的关系, 《中国语文》, 3, 195-202
- [7]王蓓, 2002, 《汉语韵律知觉的研究》, 中国科学院心理所博士论文
- [8]林焘, 王理嘉, 1992, 《语音学教程》, 北京大学出版社
- [9]林茂灿, 颜景助, 孙国话, 1984, 北京话两字组重音的初步实验, 《方言》第 1 期
- [10]曹剑芬, 1990, 《现代语音基础知识》, 人民教育出版社
- [11] 仲晓波, 王蓓, 杨玉芳, 吕士楠, 2001, 普通话韵律

词重音知觉，《心理学报》第 31 卷 6 期

[12] 仲晓波, 郑波, 杨玉芳, 2002, 关于普通话韵律短语重音的实验研究, 《声学学报》第 27 卷第 2 期

[13] 王韞佳, 初敏, 贺琳, 2006, 汉语焦点重音和语义重音分布的初步实验研究, 《世界汉语教学》, 2, 86-98

[14] 曹文, 2010, 《汉语焦点重音的韵律实现》, 北京语言大学出版社, 2010 年 4 月

[15] 蔡莲红, 吴宗济, 蔡锐, 陶建华, 2001, 汉语韵律特征的可计算性研究, 《新世纪的现代语音学》, 清华大学出版社

[16] 许洁萍, 初敏, 贺琳, 吕士楠, 2000, 汉语语句重音对音高和时长的影响, 《声学学报》第 25 卷第 4 期

[17] 仲晓波, 杨玉芳, 1999, 国外关于韵律特征和重音的一些研究, 《心理学报》第 31 卷第 4 期

(原载第 10 届中国语音学会 PCC2012)