汉语韵律特征的可计算性研究

蔡莲红[1] 吴宗济[2] 蔡锐[1] 陶建华[1]
[1]清华大学计算机科学与技术系 100084
[2]中国社会科学院语言研究所 100372

摘 要

本文主要介绍我们在语音分析的研究中，对语音的韵律特征的分析与计算，并给出一些探索性的实验结果。研究中，先给出该语音数据库的统计特性。然后分别选取一个韵律参数，研究它的变化对听感的影响及其影响程度。最后探讨韵律特征的计算方法。

引 言

语音学是令人瞩目的研究课题，众多专家从不同的角度研究纷繁复杂的语音和语言现象，形成了多个语音学分支。计算机为语音学研究提供了方便的计算工具。语音学家也纷纷利用计算机进行语音的研究。

人机语音交互是人机交互最自然的方式，语音处理已成为智能计算机必不可少的课题。要想让语音“进出”计算机就要“计算”。然而语音现象属感知范畴，如何能计算呢？多年来，计算机界与语音界进行了友好紧密的合作，试图找到妥善解决语音计算的问题。

韵律是语音研究的基本问题。韵律特征，也称超音段特征，通常指的是言语中除音色之外的音高、音长和音强三个特征。吴宗济曾指出，一个人所说语言，不论其经意与否，其表达的语气、情调都和韵律有关。韵律特征在自然语言中起着非常重要的作用。韵律特征的变化可以帮助听者更好地理解说话人的语义。说话人的语气、态度、感情色彩、个人特点在句子的韵律特征中都有所体现，从而使句中音节的韵律特征产生各种各样的变化。而说话人的思想正是通过韵律特征的变化得到确切的体现。对于韵律特征表现的研究已有很多成果，但较少见到对韵律参数的数学描述。我们对大容量语音数据库进行分析计算，期望通过计算自动得到韵律参数的描述。

1 试验材料和方法

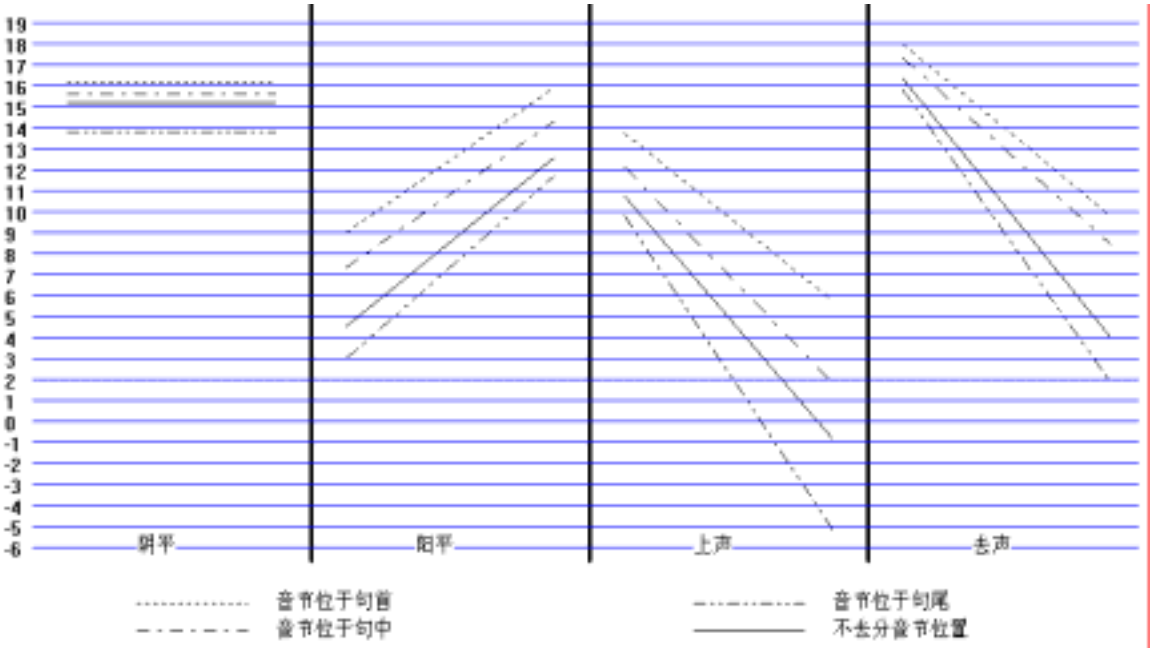
本实验采用的数据库是一个女声数据库，采样率为 16KHz、量化精度为 16 位。数据库中共有两千多个句子，包含近两万个音节。覆盖汉语的 417 种有调音节以及多种声学特征的搭配关系。该数据中的所有音节利用语音分析软件 Speech 进行了音节边界和基频的标注。此外，Speech 软件可以方便地显示语音数据的波形和语谱、测听选中的任何语句，因此可以对以上分析结果进行人工修正，基本杜绝了音节切分和标注的错误。

研究中，先给出该语音数据库的统计特性。然后分别选取一个韵律参数，研究它的变化对听感的影响及其影响程度。最后探讨韵律特征的计算方法。参与听音实验的是三名有一定经验的实验室成员。

2 特定人音域计算

基于上述数据库，计算语音的统计特性，下表给出计算结果：

	句首(Hz)	音阶	量化	句中(Hz)	音阶	量化	句尾(Hz)	音阶	量化	平均(Hz)	音阶	量化
阴平平均值	332	E4	16.1	290	D4	13.8	322	E4	15.6	**315**	**D#4**	**15.2**
阳平最大平均值	327	E4	15.9	257	C4	11.7	298	D4	14.3	271	C#4	12.6
阳平最小平均值	220	A3	9.0	156	D#3	3.0	199	G3	7.3	169	E3	4.5
阳平调域	**107**		**6.9**	**101**		**8.6**	**99**		**7.0**	**104**		**7.5**
上声最大平均值	288	D4	13.7	231	A#3	9.8	263	C4	12.1	242	B3	10.7
上声最小平均值	182	F#3	5.7	97	G2	-5.2	145	D3	1.8	124	B2	-0.9
上声调域	**106**		**7.9**	**134**		**15.0**	**118**		**10.3**	**119**		**10.5**
去声最大平均值	371	F#4	18.0	326	E4	15.8	356	F4	17.3	335	E4	16.3
去声最小平均值	229	A#3	9.7	145	D3	1.8	213	G#3	8.4	164	E3	3.9
去声调域	**142**		**8.4**	**181**		**14.0**	**143**		**8.9**	**155**		**10.1**



分析计算结果，从统计特性来看：

1．句首音节基频的平均值高于句尾，句中最低，且各种声调均有此特点。本语料库中多数语句为下倾语调。

2．从曲线上看，不同声调音节的基频变化斜率不同，去声音节的基频变化最快，其调域最宽。而同一种声调的句首音节、句中音节、句尾音节的基频曲线基本上是平行的，表明基频变化速度也基本相同。

3．观察该发音人的音高变化，句首音节的平均音高变化接近一个八度；句尾音节的平均平均音高变化大于一个八度；句中音节的平均平均音高变化比一个八度大得多，主要是基频下限比较低。这可能是句中的音节轻读或上声音节重读造成的。

为了分析该发音人的音域，首先学习了专家的论述。沈约说“若以文章之音韵，同弦管之声曲，则美恶妍媸，不得顿相乖反”。赵元任先生对音域的估计是：“说不出 1 到 5 的间隔有多大，或者音域有多宽，但在实践中，平均值是在增 5 度和 8 度高之间（每一阶段在一度和一度半之间），并且 1 的音高大约是讲话者嗓音的最低限（赵元任,音高的乐律表示）。吴宗济也指出：一个人的正常发音在一个八度范围内。而且以第一声或第二声的高点为上限，以第三声的转折点为下限。

本语料库都是陈述句，可以算做是中性语调。如赵元任所述 “即没有色彩的连贯话语中的语调。”由于是自然发音，故句中带有不是故意强调的重音、轻读音节。

对比专家的论述和本语料库的特点，我们进行计算：

1．将基频转换成乐律表示，而非频率表示，如上表。这样在比较同一发音人的多次发音的音高差异、不同发音人的音高差异时更符合感知特点。我们可以选阴平基频的平均值 315（D#4）为该发音人基调。

2．为了计算机处理方便，利用下文所列公式，将基频转换成数值表示，如上表。这样在比较音节的高低时，便成为简单的数值计算。如阴平音节，高于基调的音节为重读音节，低于基调的音节为轻读音节。

3 韵律特征的量化表示

3.1 音高频率的量化表示

对于音高的度量，传统上都是以“赫兹”为计量单位的。但是实验证明，赫兹坐标上的线性变化并不对应于听觉感知上的线性变化。研究表明，如果把频率转换成音乐的半音程来计量，频率的变化的半音程数与听感上的距离是比较一致的。

为便于计算机处理，对任意基频值 f，均按下式量化：

$$f_l = \left[12 \log_2 \frac{f}{f_0} \right]$$

其中 f_0 ＝131Hz，为标准低音 C 大调所对应的赫兹数。量化结果 f_l 即 f 相对 f_0 的半音程数。

3.2 基频均值的感知试验

基调（基频均值）是音高的最基本特点，每个人发音的基本特征是由基频均值决定的。我们常说有的人说话声特别尖锐，有的人则特别低沉，其主要原因就是这个人发音的基频较高或者较低的缘故。

实验方法如下：指定句中的某个音节，在保持其他声学参数不变的条件下，按一定比例提升该音节的基频值。保持句中的其他音节不变，让实验员听上面的句子，感受被修改的音节轻重变化的程度。

实验中，我们共选取了 100 个音节作为实验对象，实验结果记录如下：

	提高半个音阶			提高一个音阶			提高一个半音阶			提高两个音阶		
程度	不变	稍重	很重	不变	稍重	很重	不变	稍重	很重	不变	稍重	很重
A	91	9	0	45	55	0	15	69	16	0	42	58
B	88	12	0	32	68	0	3	72	25	0	21	79
C	95	5	0	61	39	0	23	73	4	0	47	53

3.3 基频调域的感知试验

实验方法和前面类似。不同的是修改调域时，采用的方法是保持该音节的基频均值不变，将各点的基频分情况作修改：大于平均基频的点基频值按比例提高；小于平均基频的点基频值按比例减小。

仍然选取 100 个音节作为实验对象，实验结果记录如下：

	展宽半个音阶			展宽一个音阶			展宽一个半音阶			展宽两个音阶		
程度	不变	稍重	很重	不变	稍重	很重	不变	稍重	很重	不变	稍重	很重
A	96	4	0	65	35	0	21	73	6	0	52	48
B	90	10	0	58	42	0	13	69	18	0	39	61
C	100	0	0	79	21	0	28	71	1	0	66	34

参考基频均值的量化方式，我们给出基频调域的量化公式：

$$R = 12 \times (\log_2 \frac{\max\limits_{n_1 \le i \le n_2} f_i}{\min\limits_{n_1 \le j \le n_2} f_j}) \qquad 81$$

同时，我们从实验结果中可以看到，改变调域对听感的影响没有改变均值的效果明显。

量化音高、音长、音强的根本目的是为了对改善语音合成系统的合成质量有所帮助。重音的判别和表示是目前文语转换系统中的一个难点。我们进行了一些尝试性的实验，考察量化后的句子和重音的关系。

4 韵律特征的可计算性研究

4.1 重音的音高表现

一般来说，基频均值最大的音节感知为重音。但研究表明，重音感知与基频并不是完全对应的关系。重音是在特定语境下轻重感知。在语调下倾的语句中，重音音节在句首时的基频普遍比句尾重音的基频高。句中或句尾的重音音节即使绝对音高不高，但已经足够让人感到强调了。通常在陈述句中有此表现。

为了分析计算本语料库的韵律特性，考虑到语调的影响，我们给出下面的粗略的修正公式：

$$f_{NEW} = f_{OLD} \times \delta_f \qquad \text{其中：} \qquad \delta_f = \begin{cases} \delta_{f1} & \text{该音节在句首或短语首} \\ 1.0 & \text{该音节在句中或短语中} \\ \delta_{f2} & \text{该音节在句尾或短语尾} \end{cases}$$

其中“首”是指句子或短语开始的两个或一个音节，“尾”指句尾或短语尾的两个或一个音节。通常$\delta_{f1} > \delta_{f2}$。

4.2 重音的音长表现

试验表明适当的增加时长能够提高听者的注意力，从而产生语言变重的感觉。超过一定量以后，时长再增加对听者的影响就变得不再明显，但该量的大小与原音节的时长有关。时长的改变对听感的影响没有改变音高明显。

一般来说重音音节是那些时长较长的音节，决不会是最短的。此外，时长和该音节所在句中的位置也是有关，普通语句中句尾的音会较长。量化音长时，给出下面的粗略的修正公式：

$$t_{NEW} = t_{OLD} \times \delta_t \qquad \text{其中：} \qquad \delta_t = \begin{cases} 1 & \text{该音节不在句尾} \\ \delta_{t1} & \text{该音节在句尾} \end{cases}$$

其中句尾的范围仅指句尾的一个音节。

4.3 重音的音强表现

音节稳音段幅度的最大值能较好的体现幅度的听感特性。改变幅度可以改变听感，只是改变幅度对听感的影响比较弱。

4.4 重音的线性公式模拟

我们采用基频、时长、调域作为评价一个音节的基本参数。下面我们将采用线性模拟的方式来进行讨论。

通过对几个声学参数的量化，可以近似的认为，人对这些声学参数的感知程度按公式线性增加。我们假设如下的公式成立：

$$L = \alpha \times F + \beta \times T + \gamma \times R$$

其中 F 和 T 均是经过δ_f和δ_t修正后的基频均值和音长值，R表示音域。我们选取 600 个句子，进行人为的重音标注。我们取出其中的 400 句作为训练集，其余的 200 句作为测试集。

对于训练集中的每个句子，可以得到一个方程组：(n 为句中的音节个数)

$$\begin{cases} \alpha \times A_1 + \beta \times B_1 + \gamma \times C_1 < \alpha \times A_k + \beta \times B_k + \gamma \times C_k \\ \alpha \times A_2 + \beta \times B_2 + \gamma \times C_3 < \alpha \times A_k + \beta \times B_k + \gamma \times C_k \\ \Lambda \\ \alpha \times A_n + \beta \times B_n + \gamma \times C_n < \alpha \times A_k + \beta \times B_k + \gamma \times C_k \end{cases}$$

解该方程组，可以得到α、β、γ的一个取值范围。

用 400 个语句，可以得到 400 组α、β、γ的取值范围。取这些α、β、γ的范围的交集的最大的部分，可以给出一个α、β、γ的大致取值。

经过计算，我们取$\alpha = 1.5$、$\beta = 0.95$、$\gamma = 0.65$。这组取值，能满足 342 组语句的取值范围。

即：

$$L = 1.5 \times F + 0.95 \times T + 0.65 \times R$$

将该公式用于剩余的 200 组测试集的重音位置标记上，我们发现能有 147 句的标注与人为标注吻合，约占 73.5%。结果基本让人满意。

5 结束语

在韵律特征的量化实验中，我们探求一种韵律特征的计算方法，取得了一些可喜的结果。但是的实验还是比较初步的。实验的数据还不够充足。涉及的参数还不够多。同时，参与听音实验的人群还应该扩大。量化公式的形式也有待进一步完善。

在重音的判定方面，线性模拟公式具有一定的可信度，我们也在尝试采用其他的方法，诸如神经元网络、数据挖掘等，去寻求声学参数和重音的关系。

参 考 文 献

[1]赵元任，语言问题，商务印书馆，2000
[2]吴宗济，从声调与乐律的关系提出普通话语调处理的新方法，中国语文，1997，P243-258
[3]杨玉芳，语句韵律结构知觉 ，声学学报(中文版) ，1998. 02