汉语音高模式及参数化描述的研究

张 章、贾 珈、蔡莲红、吴志勇

摘要: 汉语是声调语言，体现不同声调的基频复杂多变，而对于同声调的基频包络却有着许多相似。本文针对汉语单音节，研究汉语的音高模式及其参数化描述。通过分析汉语语音基频曲线的变化规律，从基频复杂的变化中归纳汉语四个声调的音高共性，提出了汉语的音高模式。为更全面的描述声调特性，区分不同发音人的音高特点，进一步提出了一种基频特性参数化描述的方法。基于该参数化描述方法，不仅能够体现基频曲线的变化规律，而且能够直观的反映基频表现的音高特点。

关键词: 音高模式、基频、参数化

1. 引言

汉语的音高变化承载了丰富的语音、语言学信息。研究表明，不同的声调，音高是不相同的。音高是一种主观心理量，当声音的频率由小到大变化时，听觉便产生一种与此相应的由低到高的不同音高的变化[2]。不同人发的相同的字音，其音高是复杂多变的，这些复杂的变化，也正反应了说话人的音高特点。音高的变化由声调的频率决定，声调的频率值是声带振动的基本频率，简称基频，基频是语音最为重要的声学特征之一，在语音学的很多领域都有着广泛的应用。本文希望通过对汉语语音信号基频的变化规律进行分析，给出一种能够反映发音人音高特点的音高模式，并给出描述发音人音高特点的参数化方法，将复杂的基频信息数字化，简化基频曲线的表示方法，直观的反应发音人的音高特点，为语音合成、声音转换等言语工程领域的应用提供便利。

对于汉语音高的研究，赵元任先生提出了五度标调法，他提出用 5 个数字来表示声调的不同音高调值，1 表示最低的调值，5 表示最高的调值[5]；周俏峰提出了汉语声调模型的参数描述，用[*H,B,I,R,D*]分别表示音域上限、音域下限、调值中心的基线、调域、调型时长[8]；陶建华进一步完善了声调模型，提出了音高基频规格化参数 SPIS，用[*B,H,N1,N2,F,E*]分别表示基频最小值、基频最大值、基频最小值位置、最大值位置、基频起始的值和终止的值[2]。这些方法用绝对的频率值作为参数描述基频，虽然能够刻画基频曲线的变化，但是不同发音人的音域存在差异，因此这些参数描述方法不便于评价和对比不同发音人的音高特点。

石锋进而提出了声调格局的概念，声调格局就是由一种语言中全部单字调所构成的格局[5]。利用语音实验得到一种语言或方言的一位发音人的测量数据，如每个声调取 9 个测量点，取好测量点后用如下 T 值公式对基频数据进行归一化和相对化分析：

$$T = \frac{\lg x - \lg min}{\lg max - \lg min} \times 5 \qquad (1)$$

其中 max 为该发音人各点平均值中的最大值，调域的上限；min 为最小值，调域的下限；x 表示待归一化的测量点值；T 为归一化后的值。

这种 T 值计算方法，得出的是某个测量点在该发音人的整个调域中的相对位置，从而实现了音高数据的相对化。相对化后的数据用平滑曲线连接各点，就得出了声调格局图形[5]。石锋的这一方法对基频数据进行了归一化处理，声调格局反映的是发音人在

各自调域内的相对关系，更有利于评价发音人的音高特点和对比不同人的差别。

本文主要借鉴了石锋提出的声调格局的概念，以带有完整声调信息的汉语单音节为语料库，提取汉语语音基频参数，研究汉语语音基频曲线的变化规律，给出了汉语的音高模式；为了更方便的对比不同发音人的音高特点，更全面的描述这些特点，提出了音高的参数化描述方法，直观的说明发音人的音高特点。

2. 汉语音高模式抽取

汉语的音高模式是包含汉语四个声调的全部音高信息的一种模式，其反应汉语的音高特点。对于一个人的语音信息，首先进行基频的提取和处理，得到基频变化曲线。然后通过归一化处理，抽取该发音人的音高模式。本章给出了基于标准语料库，抽取出的汉语四个声调的音高模式。

2.1 基频的提取和平滑

语音信号处理中，基频信号提取的研究一直是热点问题，相关研究人员提出了不少有效的基频提取算法，最为基本的就是基于二元激励模型的自相关算法。本文采用 Cheveigné 和 Kawahara 提出的准确率较高的 YIN 算法[1]来提取语音信号的基频曲线。虽然这一算法提取的基频曲线已经有了较高的准确度，但是仍然不能完全吻合实际的曲线，因此我们对得到的基频进一步做人工手动标注，修改基频标注点，从而修改基频曲线，得到较为准确的基频标注序列。

经过校准的基频曲线，基本符合了实际的曲线，但仍然会存在一些跳变点，使基频曲线不够平滑，所以还需要对其做平滑处理[3]，去掉曲线中的毛刺。我们采用改进的中值平滑方法：在被平滑的点的左右各取 1 个样点，取这两点连线的中点作为测试点，比较测试点与被平滑的点，当两者之差大于某一阈值时，则用测试点取代原来的点，否则被平滑的点不变。通过反复的试验，确定平滑效果较好的阈值。

2.2 基频的归一化

对于每一个音节，经过基频提取和标注后，对基频曲线进行了平滑处理。但是音节之间基频序列的长度并不相同，还需要进行长度上的归一化处理。为了避免长度的差异太大而对结果造成影响，对音节做如下归一化处理：统计同一声调的音节基频序列长度的均值 u，舍弃基频长度大于 $1.2u$ 或基频长度小于 $0.7u$ 的音，避免基频长度过长或过短带来的影响。对于剩下的长度适中的音，设基频的采样点为 10 个，也就是取基频序列中时间平均的 10 个点的基频值来代表整条基频曲线。对于同一声调的每一个音，都得到 10 个采样点，对这些点分别取平均，得到这一声调的 10 个基频点。

除了长度上的归一化，还需要对基频点的频率值做归一化处理。通常，女生的基频高于男生，小孩的基频高于成人，每个人的基频范围有很大差异，所以，直接用基频的值来进行音高模式抽取，不利于观察汉语声调的音高共性。因此，对得到的四个声调的基频点做如下归一化处理：我们沿用了赵元任先生五度的概念，将一个八度归一化到 0-5；结合石锋 T 值计算的方法[5]，并结合音乐中，半音程数的变化与听感上的距离比较一致的结论[1]，设计如下公式来进一步做归一化处理：

$$F = [\log_2 x - \log_2(\overline{f_1}/2)] \times 5 \qquad (2)$$

其中，$\overline{f_1}$ 为阴平的 10 个基频点的均值，x 为待求点基频值，F 为归一化后基频值。

以往的研究中，多把上声的最小值点作为 0 来计算相对值。而数据表明，上声的最小值点通常无法准确的标注出来，变化较大，不够稳定。而阴平的基频曲线近似为一条直线，基频序列的标准差在 1-4Hz 之间，变化相对很小，较为稳定，所以我们用阴平的均值作为 5 来计算相对值。借用石锋 T 值计算公式中取对数的归一化思想，归一化后的 F 表示基频在调域中的相对关系，取以 2 为底的对数，则借鉴了半音程的概念，使基频的变化更近似于听感的变化。

2.3 汉语音高模式

汉语音高模式应当尽可能完整的表示汉语四个声调的音高特点。

语音信号经过上述的基频提取、平滑和归一化处理，得到 40 个（每个声调 10 个）归一化后的基频点。接下来对于每一个声调的 10 个点，做三次曲线拟合，得到 16 个模式参数 $[a_i,b_i,c_i,d_i]$，其中 i = 1,2,3,4。拟合曲线公式为：

$$g_i(x)=a_ix^3+b_ix^2+c_ix+d_i \qquad (3)$$

其中，x 表示基频点的标号 0-9；a_i-d_i 表示模式参数，$g_i(x)$表示对应 x 点的 F 值。

这 16 个参数和阴平均值 $\overline{f}_1$ 共同组成了汉语音高模式。$[a_i,b_i,c_i,d_i]$ 反应了四个声调的相对变化关系，而 $\overline{f}_1$ 表示了发音人的音高水平。图 1 是汉语音高模式的曲线图，抽取该模式的语料库为标准发音的女生的语音数据，模式参数如表 1。

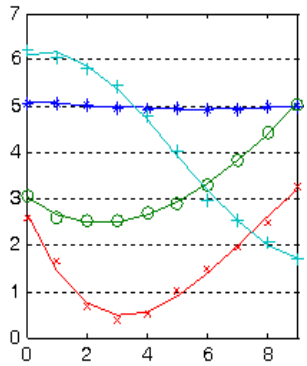

图 1：汉语音高模式图。其中横坐标为 x，即基频点的标号；纵坐标为 F，即归一化后的基频值；标记“*”“o”“x”“+”分别表示阴平、阳平、上声和去声的基频曲线。

表 1: 汉语音高模式参数。

$\overline{f}_1$ = 281.4	a	b	c	d
阴平	0.00	0.00	-0.04	5.10
阳平	0.00	0.10	-0.44	3.02
上声	-0.01	0.29	-1.46	2.65
去声	0.01	-0.20	0.23	6.12

3. 汉语音高的参数化描述

上一节我们从发音人的语音信号计算得到了汉语的音高模式，这一模式携带了发音人比较完整的音高信息，并将音高信息数字化，为语音合成和声音转换等领域的应用提供了便利。

模式中的 17 个参数虽然在工程上可以方便的应用，通过参数的计算可以完整的知道发音人的音高信息，但参数本身并不具有明显的物理意义，音高的特点只能从模式图中看出。为了更直观的描述发音人的音高特点，便于对比不同发音人的音高差别，我们进一步提出了一种参数化描述音高的方法，通过这一参数化描述方法，直观的反应发音人音高上的特点，包括音高水平、基频范围、四个声调的相对关系等。

对于单个发音人，在得到其汉语音高模式的 17 个参数之后，进一步计算其音高的描述参数，用如下的 10 个参数来描述发音人的音高特点：

1. $\overline{f}_1$：模式参数之一，阴平的基频均值。表示发音人的音高水平，例如男生的 $\overline{f}_1$ 约为 180Hz，而女生的音高要高于男生，$\overline{f}_1$ 约为 280Hz。
2. g_3：归一化后上声最小值，基频下限。
3. g_4：归一化后去声最大值，基频上限。这两个参数分别为基频的最小和最大值，通过这两个参数，可以看出发音人在自己的音高水平上的音域。例如，g_3=0、g_4=5 表示发音人的音域为一个八度；发音人的音域越宽，表示其发音变化越丰富。计算如公式(4)(5)。
4. p_2：阳平起始值。
5. p_3：上声起始值。
6. p_4：去声起始值。这三个参数表示了发音人除阴平外三个声调起始时的相对关系，这种相对关系主要用于对比不同发音人的基频差别。计算如公式(6)。
7. k_2：阳平转折点后半部分斜率。
8. k_3：上声转折点后半部分斜率。
9. k_4：去声转折点后半部分斜率。

这三个参数表示三个声调的上升或下降速度，主要用于对比不同发音人的基频差别。计算如公式(7)。

10. p：上声最小值点位置。
    阴平、阳平和去声都可以用直线来近似，而上声要用折线，因此转折点的位置是上声的一个重要特征，用参数 p 表示。值为公式(5)中 x_3。

$$g_i = g_i(x_i) \tag{4}$$

$$x_i = \frac{-b_i \pm \sqrt{b_i - 3a_i c_i}}{3a_i} \quad (x_i > 0) \tag{5}$$

$$p_i = d_i \quad (i = 2,3,4) \tag{6}$$

$$k_i = \frac{g_i(9) - g_i(x_i)}{9 - x_i} \quad (i = 2,3,4) \tag{7}$$

上述公式中，x_i为各声调的转折点值。

对于不同发音人，从上面介绍的参数，可以直接对比得出不同发音人音高水平的差异、音域的差异和四个声调细节上的差别。除此之外，还需要评价不同发音人音高差别的大小，以便后续研究哪些参数对音高差别的影响力更强。

对于每个发音人的音高模式，阴平的曲线都是接近 5 的一条近似直线，这是因为在做归一化时以阴平均值作为了基准。对于归一化后的音高模式曲线，阴平几乎没有差别，所以只需比较其余三个声调，用模式中同一声调两个发音人的两条曲线围成的面积，评价这一声调两个发音人的音高差别大小。对于两个发音人，用阳平、上声和去声三个面积来评价其音高差距。面积的计算公式如下：

$$s_{ab} = \sum_{k=0}^{9} \left| g_{ai}(x_k) - g_{bi}(x_k) \right| \quad (i = 2,3,4) \tag{8}$$

其中，s_{ab}表示发音人 AB 的音高差距，$g_{ai}(x)$和 $g_{bi}(x)$分别表示两个发音人音高模式中的三次曲线拟合函数。

综上所述，我们完整的给出了描述发音人音高特点的参数。通过参数描述，可以直观的得出每个发音人音高的细节特点，还给出了描述不同发音人音高差别大小的参数，用于评价两个发音人的音高差距。

4. 实验与讨论

本章通过对比性实验，验证本文用于提取汉语音高模式的语料库的准确性，证明上文中提出的汉语音高模式的可信度。进一步，选取多个语料库，对比不同发音人的音高差别，用参数化方法描述不同发音人的音高特点，对比观察不同发音人的音高差距。

4.1 实验语料库

实验所用语料库为安静环境下发音人录制的单音节语料。选取如下四个语料库用于实验：

1. 语料库 A：女生 A 录制的音，选取四个声调都有的音节，每个声调 150 个音，共 600 个音，发音标准。
2. 语料库 B：女生 B 录制的音，选取四个声调都有的音节，每个声调 150 个音，共 600 个音，上声发音不太完整。
3. 语料库 C：男生 C 录制的音，选取表现力强的音，共 395 个音，发音音域宽。
4. 语料库 D：男生 D 录制的音，选取较为标准的发音，共 192 个音，发音具有强调去声的特点。

其中，语料库 A 作为标准语料库，第二章中给出的汉语音高模式即为这一语料库抽取的结果。

4.2 实验方案设计

实验一：验证语料库 A 的标准性。对于语料库 A 中的音，按照石锋所提出的 T 值计算公式，即公式(1)，进行基频的归一化和相对化处理，画出相应的基频曲线图，与石锋的声调格局图进行对比，如果两者十分近似，说明语料库 A 的语料是标准的，根据语料库 A 抽取出的汉语音高模式是可信的。

实验二：基于语料库 A、B、C、D，分别按照第二章所述方法抽取其音高模式，画出其音高模式的曲线图，并按本章的方法进行提取音高的描述参数，通过对比描述参数，总结四个发音人音高的差别和差距，体现提出的参数化描述的实用性。

4.23 实验结果及分析

图 2 为实验一中验证性实验的对比结果。对比可以看出，两个图的四条曲线形态和值都非常接近，说明语料库 A 是标准语料库，第二章中给出的汉语音高模式是可信的。从图 1 的汉语音高模式曲线中可以看出，汉语的音高归一化后，音高的范围大约为 0-6，前面已经分析过 0-5 表示发音人的音高在一个八度之内，所以 0-6 表示现代人的发音普遍高于一个八度，上声最小值和阴平之间大概跨了一个八度，而去声会高于阴平。

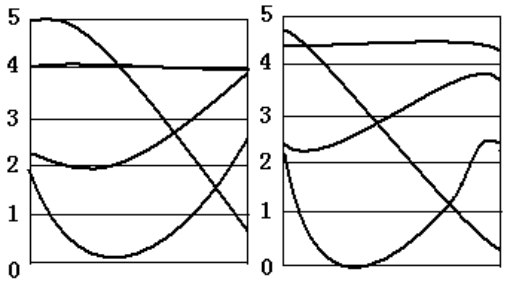

图 2：语料库 A 音高模式（左，计算方法为公式 1）与石锋声调格局图（右）对比。

图 3 为实验二中，四个语料库抽取的音高模式的曲线图。从图中可以看出，四个发音人的曲线形状基本相同，四条曲线的相对位置关系也很类似。

四个语料库的曲线存在的差异为：语料库 AB 比较近似，语料库 B 的上声后半段上升不够高，类似“半上”的曲线形状；语料库 C 音域较宽，近似为3-7，明显宽于另外三个语料库的音域；语料库 D 去声较高。

表 2 给出了四个语料库描述参数，从表中可以看出：

1. 语料库 AB 的 $\overline{f_1}$ 参数明显大于语料库 CD，说明 AB 的音高水平高于 CD，因为 AB 为女生的发音，CD 为男生的发音，女生的音高普遍高于男生。
2. 参数 g_3、g_4 表示音域，语料库 B 的音域略比 A 窄，语料库 C 的音域明显宽于另外三个语料库。可以看出语料库 B 音域略窄是因为 g_3 稍高，是由于上声发音不太完整，类似“半上”造成的；而语料库 C，录音时故意使声调富于变化，所以音域明显变宽，g_3=-2.81、g_4=6.53。
3. 从参数 p_2-p_4、k_2-k_4 和 p 可以看出发音人三个声调的细节，实验一中已经验证了语料库 A 较为标准，语料库 B 相比于 A，上声转折点参数 p 值略大，且上声斜率 k_3 略小，表现了上声发音的“半上”特点；语料库 C 由于音域差别过大，暂不考虑；语料库 D 相比于语料库 A，去声的起始的参数 p_4 略高，表现了其强调去声的特点。

可见，表中的描述参数基本反应了各个发音人的不同声调的音高细节信息，并与音高模式曲线图中所反应的信息一致，所以这种参数化描述方法能够很好的描述发音人的四个声调的音高特点和相互关系。

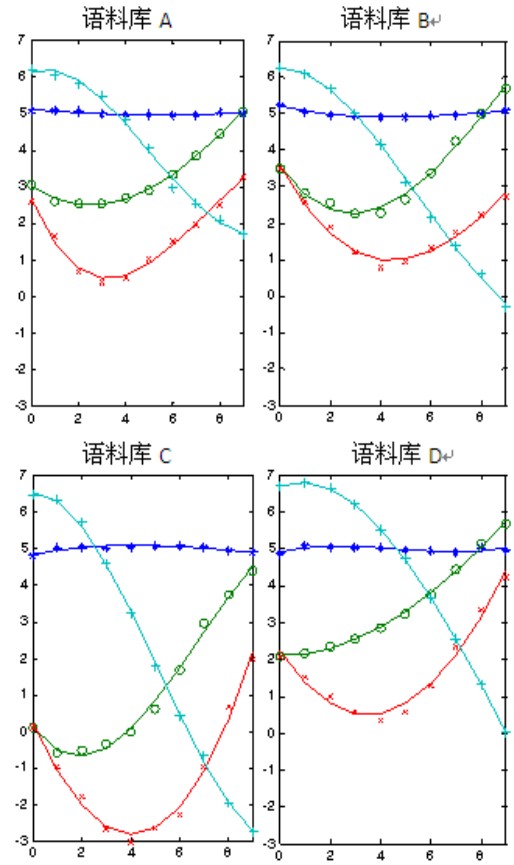

图 3：语料库 ABCD 音高模式的曲线图。

表 2: 语料库 ABCD 的参数描述。

语料库	f	g_3	g_4	p_2	p_3
A	281.4	0.49	6.17	3.01	2.64
B	280.6	1.01	6.26	3.59	3.59
C	179.5	-2.81	6.53	0.24	0.24
D	177.2	0.53	6.82	2.24	2.24
语料库	p_4	k_2	k_3	k_4	p
A	6.12	0.43	0.48	-0.65	3
B	6.27	0.60	0.37	-0.78	4
C	6.52	0.78	1.00	-1.13	4
D	6.71	0.41	0.65	-0.87	4

表 3: 语料库 ABCD 音高差距。

	B	C	D
A	2.9/4.6/5.7	20.8/27.5/19.5	4.6/2.6/7.3
B	/	22.2/30.4/11.9	4.3/7.9/10
C	/	/	22.2/28.9/20.9

表 3 给出了四个语料库的音高差距，表中每个三元组表示两个语料库的阳平、上声和去声的音高差距。从表中可以看出，语料库 C 由于音域较宽的原因，与其他语料库差距较大；语料库 A 与 B 最为接近，语料库 A 与 D 去声差别略大，语料库 B 与 D 上声和去声差别都略大。参数描述的差距符合模式图中所反应的差距关系，并进一步将这种关系量化了。

综上所说，我们给出了汉语音高模式的参数化描述方法，这些描述参数都具有其物理意义，能够直观的描述音高特点。

5. 结论与展望

人类语音的表现是纷繁复杂的，对于人类语音的研究也有着诸多的方向，例如语音合成、语音识别、声音转换等等。这些研究，都要借助于声学参数的转换来实现，因此声学参数的研究是研究语音的各个领域的基础。而音高是语音的一个最为重要的声学参数，所以研究语音的音高变化规律，能够为其他众多语音研究领域提供基础。

汉语是声调语言，其音高的变化由声调的频率决定，本文通过研究汉语语音四个声调基频的变化规律，提出了一种抽取汉语语音音高模式的方法，基于标准语料库抽取了汉语音高模式，总结了汉语音高的特点，并设计实验，验证了音高模式的可信性。为了直观的描述发音人音高的特点 ，本文还进一步提出了一种参数化的描述方法，通过参数，描述发音人音高的细节，并总结对比不同发音人音高细节上的差别。为方便对比和后续研究，本为还给出了评价两个发音人音高差距的方法，并做了对比实验。

汉语音高模式还有很多东西值得研究，本文只是针对汉语单音节语料做了研究，在连续语流中，音节的音高还会产生变化，将来的研究将进一步研究汉语的双音节词和连续语流的音高模式。另外，本文对不同发音人的音高差别做了比较，将来还需要进一步分析哪些因素的影响会较为明显的体现到听觉感知上。

6. 致谢

本文研究得到了国家自然科学基金（编号：60805008,，60928005，90920302）的经费支持。

7. 参考文献

[1] Boersma P. 1993 Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the Institute of Phonetic Sciences 17 97-110

[2] 蔡莲红 黄德智（2003）《现代语音技术基础与应用》。北京：清华大学出版社。

[3] 江太辉（2002）《一种改进的语音基频轮廓提取算法》五邑大学学报，自然科学版。

[4] 石锋 冉启斌 王萍（2009）《论语音格局》。南开语音年报 第三卷

[5] 吴宗济（2004）《吴宗济语言学论文集》。北京：商务印书馆。

[6] 周俏峰（1995）《音节数据库基音自动标注工具的研究》

张 章 清华大学计算机科学与技术系 100084
贾 珈 清华大学计算机科学与技术系 100084
蔡莲红 清华大学计算机科学与技术系 100084
吴志勇 清华大学深圳研究生院 518055