

汉语无标记疑问句的语调分析与建模*

赖玮^{1,2}, 李雅², 许小颖^{1,2}, 丁星光², 陶建华²,

(1. 北京师范大学文学院, 普通话水平培训测试研究中心, 北京 100875;

2. 中国科学院自动化研究所模式识别国家重点实验室, 北京, 100190)

摘要: 前人研究多关注疑问句韵律的局部特征, 而语音工程中疑问句基频的生成, 则有赖于对疑问句语调走势的全局把握与量化建模。文章以疑问句与陈述句的音高曲线的回归线代表语调走势, 结合实验语料与自然语料对二者语调进行分析。文章发现音高曲线的回归斜率是区别陈述句与疑问句语调的重要韵律特征, 该特征受句子长度、时域范围等因素影响, 在两种句型中有不同的表现。文章最后用反函数对疑问句语调斜率的变化规则进行建模, 取得了较高的拟合优度, 为疑问句语调走势的预测与生成提供了可靠依据。

关键词: 无标记疑问句; 语调; 韵律; 建模

***基金项目:** 本研究得到了北京师范大学 (11-30, 10-12-02)、中央高校自主研究基金 (2010105565004GK)、国家语委十二五科研计划项目(YB125-41)、中国国家高技术研究发展计划(863 计划)(No.2015AA016305)、国家自然科学基金(NSFC) (No.61273288, No.61233009, No.61203258, No.61305003, No. 61332017, No.61375027, No.61425017, No.61403386), 以及国家社会科学基金重大项目(13&ZD189)的支持。

作者简介: 赖玮 (1990-), 女, 汉族, 江西宜春人, 硕士研究生, E-mail: laiwei_0508@126.com

通讯联系人: 许小颖, 副教授, E-mail: xuxiaoying2000@bnu.edu.cn

疑问语调的分析与合成一直是人机交互与对话系统领域中的重要问题。以往的汉语疑问句韵律研究数量众多,且结论较为成熟,例如:疑问句整体音高大于陈述句^[1-3],有更高的音高起点^[4],有向上倾斜的边界调^[3,5],句调起伏度更大^[6],句末能量更强^[2,7]等。以上要素是疑问句在产出与感知方面的重要特征,也是疑问句识别与合成的重要依据。

然而,上述结论更多是对句子局部的韵律特征进行离散地分析,而对于整句基频如何发生即时的、连续的变化,则关注较少。江海燕^[8]观察不同类型疑问句的音高中线,发现其走势由高到低分别为:无标记疑问句、副词疑问句、语气词疑问句、代词疑问句、陈述句、选择疑问句、X不X疑问句。Jiahong Yuan^[7]将疑问句与陈述句相应音节的音高相减,发现疑问句音高在陈述句基础上有整体提升,但越接近句末升幅更大。Fang Liu^[9]分别用线性函数与指数函数对疑问句与陈述句的音高差进行拟合,认为疑问句的音高上升不是线性,而是成指数的。Wei Lai^[10]观察不同句型疑问句的韵律词最大值的连线,认为有标记的疑问句语调走势与陈述句相似,表现为整句下倾与终端下倾的叠加。

总体来看,以往针对疑问句句内音高趋势的研究相对薄弱,且多以陈述句音高为基准来分析疑问句与陈述句的音高之差,缺乏对疑问句的音高变化全貌的直接展现。此外,上述研究大多采用同声调音节组成的实验语料,所得结论与自然声调组合状态下的疑问句语调还有一定差别。

本文的研究对象是无标记的疑问句,即不具备任何文本的疑问标记的、与陈述句同形的疑问句。这类疑问句由于缺乏表达疑问的词汇或语法手段,只能凭借韵律手段表达疑问语气,是承载疑问语调的典型句型。在方法上,本文将自然语料与实验语料相结合、定性分析与定量分析相结合,试图弥补以往研究中自然语料的缺失,更精确地对疑问句语调走势进行量化,为提高疑问句合成语音的自然度与表现力做出贡献。

1 实验材料

1.1 相同调型组合的实验语料

为排除不同声调组合对音高曲线的干扰,更直观、清晰地观察语调走势,本实验采用了沈炯^[11]设计的四个声调的音节分别组成的实验语料:

- 1) 该孙英开飞机;
- 2) 由国华来完成。
- 3) 请小宝逮老鼠;
- 4) 让树庆去种菜。

以上语料由3名男发音人与2名女发音人分别朗读,发音人年龄在20~30岁左右。共有实验语料4(目标句)*10(发音人)*2(疑问/陈述)=80句。

1.2 不同调型组合的自然语料

为观察自然声调组合状态下疑问句与陈述句语调走势的区别,实验还采用了由不同声调的音节组合而成的同形疑问句与陈述句100句。语料的长度分布见表1。

表1 100句自然调型组合语料的长度分布

音节数	≤5	6	7	8	9	≥10
语料数	8	20	17	20	15	17

以上语料由3名男发音人与2名女发音人分别朗读,共有自然语料100(目标句)*5(发音人)*2(疑问/陈述)=1000句。由于发音人在说出某些句子时的嘎裂声过于严重,某些句子缺失中间音节的音高,无法计算回归斜率,所有声音文件的数据中有30句不可用,缺失率为3%。

2 实验一:疑问句、陈述句语调走势归一化

2.1 实验目的

本实验采用同调型组合的实验语料,以经验的方式观察同形的疑问句与陈述句在语调走势上的差异,测量、对比二者的音高回归线斜率;并在消除斜率差异后,考察二者音高的重合度,从而检验语调回归线斜率在疑问句音高预测方面的作用。

2.2 实验方法

所有声音文件在Praat中用Xu的脚本文件^[12]对声带的振动周期进行手工校对,然后该脚本文件根据振动周期自动转化为基频值并保存为文本文件。每个音节的边界也结合共振峰及声带振动变化手工标记出来。脚本自动为每个音节均匀提取10个音高赫兹值及这10个音高点所对应的时间点。

为消除不同发音人基频的音阶差异,所提取的赫兹值用公式(1),以发音人的音高最低点为参考值将Hz转为半音(St)。

$$F_{st} = 12 \times \log_2 \left(\frac{F_0}{F_{0_base}} \right) \quad (1)$$

四个目标句由分别用疑问、陈述语气朗读所得的音高半音曲线见图1中(a)~(d)四幅子图。由于疑问句与陈述句的时域分布模式不同,在图中加入音节的绝对时长的因素反而不利于语调曲线的对比。因此我们在每个音节内均匀地提取10个音高值,以便将时长归一化。图1中横坐标上的每个音节都由均匀的10个时间点组成,纵坐标对应着10个发音人朗读该句子的音高平均值。

以归一后的时间点为自变量,以音高值为因变量,在Excel中做线性回归,得公式(2),其中x代表时间点,k代表语调走势的斜率:

$$y = kx + b \quad (2)$$

用句子音高半音值减去其回归线上对应点的值，即可得到消除语调走势差异之后的音高值，见公式(3)。其中 F_x 代表时间点 x 处所对应的原始音高值，减去回归线上相应的值 $kx+b$ ，即得到新的水平状态下 x 处的音高值 F'_x 。

$$F'_x = F_x - (kx + b) \quad (3)$$

消除语调走势差异之后的陈述句与疑问句音高曲线的对比图见图 1(e)~(h)四幅子图。

2.3 实验结果

由图 1(a)~(d)中疑问句与陈述句的原始音高对比可以看出，疑问句的音高总体在陈述句之上；起点音高有的比陈述句高，有的与陈述句差别不大。在音高变化趋势方面，越接近句末，疑问句与陈述句的音高差异越大。但这样的变化趋势并不是由于疑问句句末音高加速上升导致的，而是由陈述句句末音高加速下降导致的，结合回归线斜率 k 的取值，我们发现疑问句的音高只呈缓慢的上升或下降，而陈述句音高下倾较为陡峭，而且还包括了降阶与句末音高下降等多重效应。

而由图 1(e)~(h)可见，当消除了疑问句与陈述句的语调差异后，二者音高曲线的一致性与重合度

是比较高的，基本只在高点、低点或句末处略有出入。表 2 展示了语调走势归一化前后疑问句与陈述句音高曲线的皮尔逊相关系数，可以看到，将语调走势归一化后，第一、二、四声的疑问句与陈述句音高曲线相关度都有提高，并达到了 .85 以上。

表 2 语调走势归一化前后四个声调的疑问句与陈述句音高曲线的皮尔逊相关系数

句子声调	Tone 1	Tone 2	Tone 3	Tone 4
归一化前	.621**	.724**	.752**	.641**
归一化后	.948**	.851**	.739***	.911***

注: *sig.<.05, **sig.<.01, ***sig.<.001。

此外，不同的声调组合对疑问句与陈述句的语调走势也有一定影响。例如，在图 1(c)中，第三声组成的疑问句音高曲线在低音点处与陈述句较为接近，这与 Yuan Jiahong^[7]所观察到的结论一致；而在语调走势归一化之后，疑问句与陈述句反倒在了低音点处形成了较大的音高差异，如图 1(g)中所示；另外，由于受到了边界处连续变调不一致的影响，“宝”字在两个句子中的声调形态区别较大：这些因素导致第三声的两条音高曲线在消除语调走势后相关度反而降低了($r=.739$)。

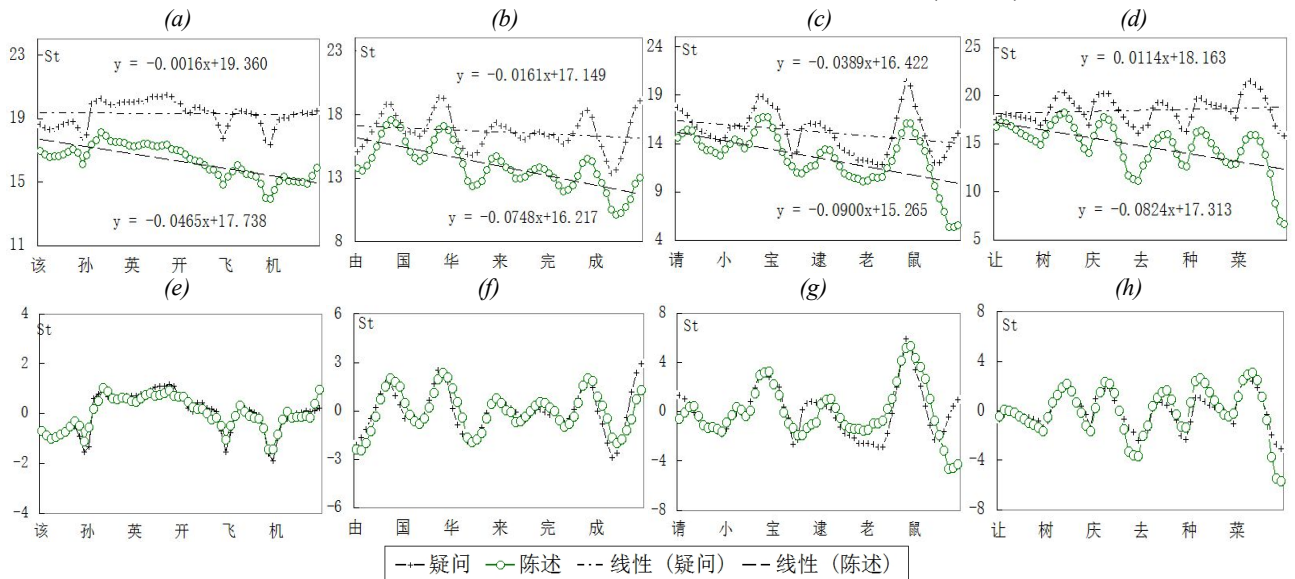


图 1 四个声调的句子的疑问句与陈述句的基频曲线及语调走势归一化情况 (图(a)~(d)为四种声调情况下五个发音人的平均音高曲线、趋势回归线与回归方程，图(e)~(h)为语调走势归一化之后疑问句与陈述句的音高曲线)

3 实验二：疑问句、陈述句语调分析与建模

3.1 实验目的

本部分采用不同声调音节组合而成的自然语料，通过音高回归线斜率这一参数将疑问句与陈述句在语调走势方面的区别进行量化，分析其在不同长度的句子中与不同时间域范围内的表现，并对两种句型的音高回归线斜率变化规则进行建模。

3.2 参数提取

对于每个句子，本实验提取如下参数：

- 1) $f_{st_initial}$ 、 f_{st_final} : 每个句子的句首音节、句末音节的平均音高值(St)；
 - 2) f_{st_SYLmax} : 句内每个音节的音高最大值(St)；
 - 3) d : 每个句子的时长(s)；
- 此外，将每个句子的每个音节提取 10 个音高点，并提取每个音高点所对应的时间点，以时间

点值为自变量、音高值为因变量做线性回归，回归系数 k 代表语调的斜率，具体参数如下：

4) k : 整句的音高回归线斜率(St/s);

5) $k_{initial}$, k_{final} , k_{middle} : 句首二音节、句末二音节以及中间剩余音节的音高回归线斜率(St/s);

6) k_i , $i=2\sim 10$: 句子前 x 个音节的音高回归线斜率(St/s)。

3.3 实验结果

3.3.1 疑问句、陈述句语调斜率与句子长度

图 2 左小图为 1000 个疑问句与陈述句的音高回归线斜率误差条形图，右小图为不同句长条件下的疑问句与陈述句音高回归线斜率误差条形图（按句长最接近的横坐标值归类）。

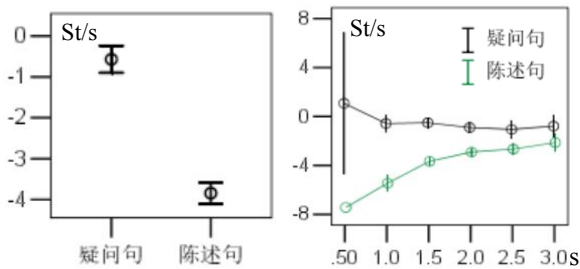


图 2 疑问句与陈述句语调回归线斜率(左)及二者受句长影响变化情况(右)

由图 2 左图可见，疑问句的语调回归线接近水平稍稍下倾，语调下倾斜率为每秒 0~1St；而陈述句的语调下倾斜率为每秒 3~4St 左右，比疑问句陡得多。由右图可见，对陈述句而言，句子越长，语调下倾坡度越缓，句子越短，语调下倾坡度越陡，这与前人的观察^[13, 14]一致。而这一规则并不适用于疑问句。句长为 0.5 秒左右的疑问句，其语调可上倾也可下倾，且斜率变化范围较大；当句长大于 1.0 秒时，疑问句的语调开始出现下倾现象，且随着句长增长下倾坡度稍有变陡，但变化不大。在 SPSS 中测量句子长度与语调回归线斜率之间的皮尔逊相关系数，发现：在陈述句中，句长与语调斜率呈现出显著的正相关($r=.324$, $\text{sig} < .001$)，而在疑问句中二者没有显著的相关性。

相关性检验还表明，语调斜率与句首、句末音高之间关系密切，不论疑问句和陈述句，其语调斜率都与句首音高成显著负相关($r=-.171$, $-.154$, $\text{sig} < .001$)，与句末音高成显著正相关($r=.467$, $.357$, $\text{sig} < .001$)。因此，本实验又分析了句长对疑问句与陈述句句首、句末平均音高的影响。由图 3 可见，随着句子长度的增加，疑问句与陈述句的句首音高都有一定上升，其中疑问句的句首音高变化更加明显，而陈述句的句首音高则变化不大；时长对句末音高没有太大影响，疑问句的句末音高比陈述句高 7~8St，但二者的句末音高都没有因受句长影响而出现非常规则的上升或下降。相关性检验表明，疑问

句的句首音高与句长之间有显著的正相关性($r=.324$, $\text{sig} < .001$)，而陈述句中二者没有显著的相关性。不论疑问句和陈述句，句长与句末音高之间也没有显著的相关性。

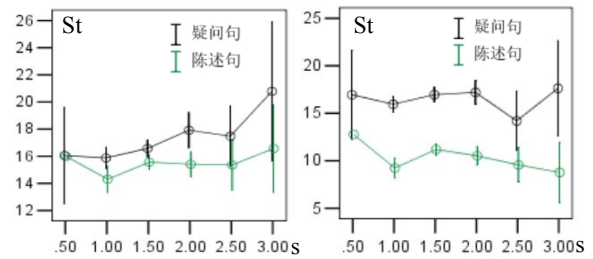


图 3 句长对疑问句与陈述句的句首(左)、句末(右)音节平均音高的影响

3.3.2 疑问句、陈述句语调斜率的时域分布

句子音高的变化是灵活多样的，语调走势在句中不同的时域范围内也会呈现出不同的特点。一般而言，在陈述句的音高最大值连上能观察到三个时序性的组成部分：句首持平/上倾(initial plateau/heightening)，句中下倾(middle inclination)，终端下降(final lowering)^[13]。

图 4 中，子图(a)~(c)分别呈现了由 6 音节、8 音节、9 音节组成的疑问句与陈述句的音高最大值连线(N=20, 20, 15)。可以看到，陈述句的高线虽然具体形态不同，但总体轮廓都由上述三部分构成；而同形的无标记疑问句也较为明显地分为三部分，具体表现为句首音高的快速上升，句中音高微微下倾，句末处音高在出现一个小幅回升后缓缓下降。

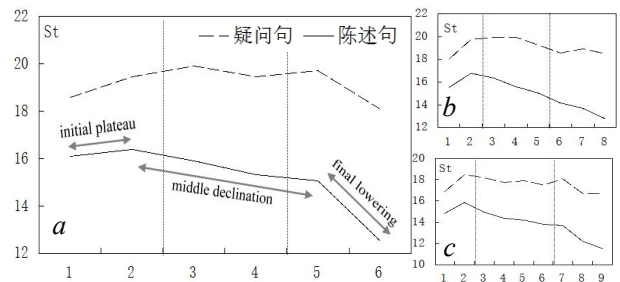


图 4 不同音节个数的疑问句与陈述句的音高高线

对 1000 句疑问句、陈述句前两个音节、后两个音节及中间剩余音节分别计算语调回归线斜率，两种句型在三段时域范围内的音高回归线斜率的误差条形图见图 5。

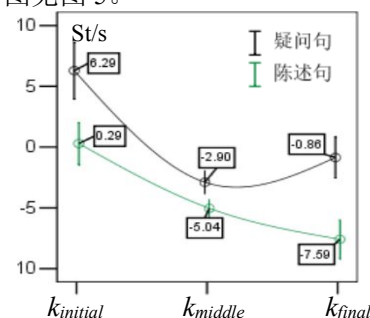


图 5 疑问句与陈述句句内不同时域范围内的语调斜率

由图可见, 陈述句句首部分语调走势接近水平, 斜率平均值为 0.29St/s, 中间部分呈下倾趋势, 每秒音高下降 5St 左右, 句末部分音高下降速率加剧到每秒 7St 左右, 形成终端下倾; 而疑问句中句首音高呈上倾趋势($k=6.29\text{St/s}$), 句中开始下倾, 斜度比陈述句稍缓, 每秒下降 2.9St 左右, 句末由于边界音高升高, 语调回归线也大致回调到水平位置($k=-0.86\text{St/s}$)。将两种句型的语调斜率变化情况对比来看, 最大区别处的在于句末, 其次在于句首, 句中下倾部分二者的语调斜率则趋于接近。

3.3.3 疑问句、陈述句语调走势斜率变化建模

图 6 为疑问句与陈述句语调回归线斜率的变化情况, k_i 表示句子前 i 个音节的音高回归线斜率。图中疑问句与陈述句语调回归线斜率随时间变化的情况与 3.3.2 的结论有呼应处, 开头上倾, 在句中渐落, 越接近句末越趋于平稳。图中虚线标志出了水平状态($k=0$), 可以看到, 疑问句的前 5 个音节处语调回归线由上倾缓缓变为水平, 此后则一直保持或接近水平状态; 陈述句的语调回归线则大概在第 3 个音节处由上倾降为水平, 其后则开始下倾。

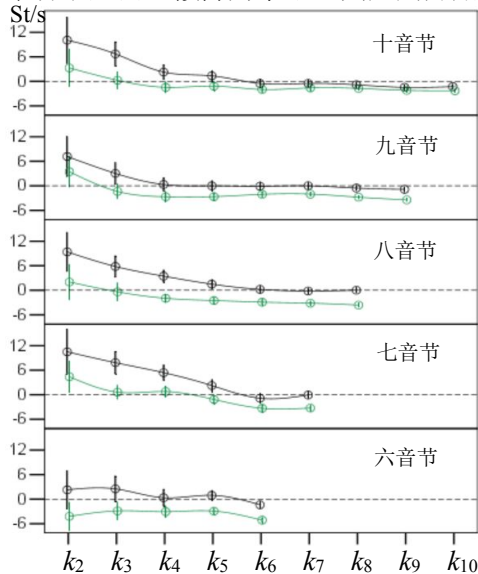


图 6 疑问句与陈述句句内不同位置的语调回归线斜率变化

为了将疑问句与陈述句的语调回归线斜率的变化规律进行量化, 提高这一结论的预测性与应用性, 我们选取每个 k_i 值 95% 置信区间内的数据, 以 i 为自变量, 分别采用线性函数、对数函数与反函数对因变量 k_i 进行拟合, 所得模型的 R^2 值、常数与系数见表 3。

表 3 三种模型的 R^2 值、常数与系数

句型	模型	R^2 值	系数	常数
疑问句	线性函数	.762***	-1.120	7.986
	对数函数	.901***	-6.073	11.573
	反函数	.942***	25.061	-3.841

句型	模型	R^2 值	系数	常数
陈述句	线性函数	.537***	-.485	.947
	对数函数	.723***	-2.843	2.867
	反函数	.855***	12.740	-4.566

注: *sig.<.05, **sig.<.01, ***sig.<.001。

比较 R^2 值大小可以看出, 不论疑问句还是陈述句, 都在用采用反函数进行拟合时达到了最优效率; 根据反函数模型的系数与常数, 可以得到根据音节数预测语调斜率的反函数模型, 疑问句见公式(4), 陈述句见公式(5):

$$y = 25.061/x - 3.841 \quad (4)$$

$$y = 12.740/x - 4.566 \quad (5)$$

公式(4)与(5)的 R^2 值分别达到了 .942 与 .855, ANOVA 检验结果均为显著 (sig.<.001), 说明拟合达到了良好效果, 根据音节数量预测疑问句与陈述句前若干个音节的语调回归线是可行的。

4 讨论

前人研究^[7,9]认为, 无标记疑问句语调上倾, 且在接近句末处音高上升加速, 这样的结论却较难得到生理发声上的支持。从发声的角度来说, 降低基频伴随声门张开与声带松弛, 是较为省力的发声活动; 反之, 提高基频伴随声门闭合与声带紧张, 则较为费力; 维持基频不变居于两者之间。若需加速音高上升来表达疑问似乎是一种低效的交流方式。此外, 如果句子够长, 音高上升不断加速, 也很容易达到生理承受极限。

本文认为, 以往疑问句语调上倾、句末音高上升加速等结论, 是以陈述句的基频为参照得出的, 因而忽视了陈述句自身的音高下倾与终端下降是导致二者音高差变大的主要原因。本文经测量发现, 无标记疑问句的语调走势在大部分情况下是水平或者微降的。该结论更容易获得生理发声上省力原则的支持。此外, 本文发现当句子较长时, 疑问句语调走势保持不变或者稍向下倾, 这也符合声带震动频率随时长衰减的生理规律。

然而在听感上, 无标记疑问句的语调的确是上升的, 这是人们获得疑问句语调上倾这一印象的经验基础。正如疑问句的音高下倾容易被感知为水平状态, 人们倾向于将疑问句趋于水平的语调感知为上倾, 这说明语调的产出与感知的关系是非线性的, 无标记问句语调的感知机制还需进一步探讨。

5 结论

本文结合实验语料与自然语料, 通过经验观

察、统计分析、数学建模的研究方法对无标记疑问句语调回归线斜率进行分析, 得出结论如下:

首先, 语调的回归线斜率是衡量陈述句与疑问句语调区别的重要韵律特征, 也是疑问句语调生成的重要依据。消除语调走势的差异后, 疑问句与陈述句的音高只在高点、低点与边界处略有出入;

在句长对语调的影响方面, 随着句子增长, 陈述句语调下倾变缓, 而疑问句语调则一直保持水平或稍稍下倾, 与句长的关系不大。此外, 句长对疑问句起点音高有一定影响, 句子越长, 疑问句起点音高越高; 句长对两种句型终点音高影响不大。

在语调走势的时域分布方面, 陈述句语调句首接近水平, 中间呈下倾趋势, 句末形成终端下倾; 而疑问句语调在句首呈上倾趋势, 从句中开始下倾, 到句末又向水平状态回调。

文章最后用反函数对疑问句、陈述句的语调斜率随音节数变化的规则进行建模, 取得了较高的拟合优度, 为语音工程中, 根据文本预测句内不同音节处的语调走势这一工作提供了依据。在未来的工作中, 我们会用更多人的语料来验证模型的效果。

参 考 文 献

- [1] 沈炯, 北京话声调的音域和语调[A]. 北京语音实验录[C], 林焯王李嘉等编. 北京: 北京大学出版社. 1985: 73 - 130.
SHEN Jiong. The pitch range of tone and intonation in Beijing Mandarin, in Lin, Wang, Beijing Yuyin Shiyuanlu, Beijing University Press, Beijing 1985: 73 - 130. (in Chinese).
- [2] Yuan J, Shih C, Kochanski G P. Comparison of declarative and interrogative intonation in Chinese [A]. The Proceeding of Speech Prosody 2002 International Conference [C]. 2002.
- [3] 伍艳红,陶建华,路继伦. 汉语疑问语调的韵律分析[A]. 第七届中国

- 语音学学术会议论文集[C]. 2006.
WU Yanhong, TAO Jianhua, LU Jilun. An prosodic analysis of intonation of Chinese interrogative sentences [A]. The Proceeding of 7th National Conference on Modern Phonetics [C]. 2006. (in Chinese).
- [4] Shen X S. The Prosody of Mandarin Chinese[M]. University of California Pr, 1990.
- [5] 林茂灿. 疑问和陈述语气与边界调[J]. 中国语文, 2006(4): 364-376.
LIN Maocan. Interrogative and declarative mood and boundary tone [J]. *Chinese Language and Writing*, 2006(4): 364-376. (in Chinese).
- [6] 王萍,石锋. 汉语北京话疑问句语调的起伏度[J]. 南开语言学刊, 2010(2):14-22+185.
WANG Ping, SHI Feng. Pitch fluctuation of interrogative mood in Beijing Mandarin [J]. *Nankai Linguistics*, 2010(2): 14-22+185.
- [7] Yuan J. Mechanisms of question intonation in Mandarin[M]. Chinese spoken language processing. Springer Berlin Heidelberg, 2006: 19-30.
- [8] 江海燕. 汉语陈述、疑问基本语调的调位表现[A]. 第七届中国语音学学术会议论文集[C]. 2006.
JIANG Haiyan. Representation of the intonation of Chinese interrogative and declarative sentences at the register level [A]. The Proceeding of 7th National Conference on Modern Phonetics [C]. 2006. (in Chinese).
- [9] Liu F, Xu Y. Parallel encoding of focus and interrogative meaning in Mandarin intonation[J]. *PHONETICA-BASEL-*, 2005, 62(2-4): 70.
- [10] Lai W, Li Y, Che H, et. al. The discovery of final lowering effect in questions and statements of Chinese Mandarin based on a large-scale natural dialogue corpus [A]. The Proceeding of Speech Prosody 2014 International Conference [C]. 2014.
- [11] 沈炯. 北京语音实验录[M]. 北京: 北京大学出版社, 1985.
SHEN Jiong. A collection of phonetic experiments on Beijing Mandarin. Beijing University Press, Beijing 1985. (in Chinese).
- [12] Xu, Y. TimeNormalizeF0.praat. <http://www.phon.ucl.ac.uk/home/yi/tools.html>. 2005-2008
- [13] Yuan J, Liberman M. F0 declination in English and Mandarin broadcast news speech [C]. The Proceeding of INTERSPEECH. 2010: 134-137.
- [14] Ladd D R. Declination.: a review and some hypotheses [J]. *Phonology*, 1984, 1: 53-74.

The Analysis and Modeling of the Intonation of Chinese Unmarked Questions

Wei Lai^{1,2}, Ya Li², Xiaoying Xu^{1,2}, Xingguang Ding², Jianhua Tao²

1. Chinese Language and Literature Department, Beijing Normal University/PSC Research Center, 100875, Beijing

2. Institute of Automation, Chinese Academy of Sciences, 100090, Beijing

Abstract: Previous studies largely concerned about the local prosodic features of questions. However, generation of questions in speech technology needs the overall understanding and accurate quantification of the intonation trend of questions. This article used the regression line of pitch contour as the trend of intonation and analyze it with experimental and natural questions. The article found that the slope of regression line of pitch contour was an important prosodic feature for distinguishing unmarked questions from statements. This feature was influenced by factors such as length of sentences, location in the sentence, etc., and had different representations in the two types of sentences. In the end the article uses inverse to model the changing of intonation slope within the sentence, and achieved good effect, which provided reliable reference for the prediction and generation of interrogative prosody.

Key words: unmarked questions; intonation; prosody; model