语调研究中心理和声学等价单位

李爱军

中国社会科学院语言研究所

摘要*

语调研究是韵律研究的热点问题，语调研究的物理实体一般是基频曲线 F0，不同发音人的基频变化范围有很大的差异。一般语调音高曲线 F0 以赫兹（Hz）为单位，研究中也采用与心理-声学（psycho-acoustic）相关的等价单位，如半音 semitones，镁 mel，巴克 Bark 以及 ERB 等。到底在语调研究中使用哪个单位更好？本文目的就是介绍国内外关于这个问题的研究，说明半音是最好的反应心理-声学的语调研究单位。最后还用半音分析了汉语情感语音库中几种情感语音音高曲线的特征。

关键词：语调、基频、音高、半音、心理-声学单位

1. 引言

话语的韵律特征是当今语音学、言语工程界十分关注的研究课题，其中语调又是研究的核心。现代语言学中“prosody 和 prosodic 是指 properties of speech that cannot be derived from the segmental sequence of phonetics underlying human utterances ”。[6]音高（Pitch）是感知到基频相关量。语调指音高变化，这种变化受到语言和语用信息影响，而这些信息从词和词组中是得不到的。感知上，音高的“高”和“低”的对立可以强调某个词，区分短语边界等。[8]

由于人耳听觉系统非常复杂，迄今为止人类对它的生理结构和听觉特性还不能从生理解剖角度完全解释清楚。所以，对人耳听觉特性的研究目前仅限于在心理声学和语言声学。人们对语调的特性研究目前也只能从心理感知和声学分析方面来揭示。一般音高曲线 F0 以赫兹为单位，也有使用与心理-声学相关量来研究音高的，如半音、镁、巴克以及 ERB 等。这些单位都被认为是能够反映心理感知的心理-声学感知量，到底在语调研究中使用哪个单位更好？下面将介绍国内外近两年关于这个问题的研究，说明半音是最好的语调研究单位。文章最后分析了汉语情感语音库的几种情感语音的音高变化特征。

2. 语调研究中基频 F0 对应的各种心理-声学单位

男性的 F0 范围一般在 80-200Hz; 女性的范围在 180-400 Hz。音高感知中第 3-6 谐波最重要，物理上基频是否存在对音高的感知没有太大影响[8]。下面列举语调研究中基频 F0 对应的各种心理-声学单位。

赫兹（Hz）：每秒钟声带振动的次数。Hz 是线性刻度，而我们对语音的产生和感知却不是线性的。

半音(semitone)是一种尺度衡量单位。它将用 Hz 表示的基频频率通过变换转换成对数域中的值。在乐律中，将一个倍频程等分为 12 个音程，每一音程称为半音，一个半音就相当十二分之一个倍频程。其转换关系为：

St=12*log2*（F0/Fref)，其中，St 是半音值，F0 是以 Hz 为单位的基频数值，Fref 是以 Hz 为单位的参考基频值。从这个公式可以看出，Hz 为单位的基频数值与半音值之间不是线性变换，而是图 1 中所示的非线性变换。

巴克（Bark）也是实验得到的感知量，通过建立影响声调感知最大带宽掩蔽宽带噪声得到。如果纯音刚好能被听到时的功率等于这一频带内噪声的功率，那么这一带宽称为临界频带宽度。临界频带的单位叫巴克(Bark)，1Bark＝一个临界频带宽度。这个特别的带宽在高频范围更大，Bark近似于对数关系，500Hz以下，是语音基频的主要范围，对音高的感知起重要作用，这时Bark与Hz的关系近似线性关系，见图 2。Hz与感知的声谱频率巴克（Bark）的对应关系为：$F_{bark} = 7 \ln (F0/650 + \mathrm{sqr}(1 + (F0/650)2))$。

镁/美（Mel）也是通过实验得到的 Hz 与感知的纯音音高的关系。人们利用受试者能相当准确地说出某一纯音音高是高一倍，还是低半倍，从而研究了主观感觉的音高与客观测量的频率之间的关系。图 3 表示以 Mel 为单位的音高与以 Hz 为单位纯音的频率之间的关系。响度级 40 方、1000 Hz 纯音的音高被定义为 1000 美。一个频率的声音若听起来是 40 方、1000 HZ 纯音音高的两倍，它的音高就是 2000 美。500 美的音高是 1000 美音高的一半。低于 500Hz 时，以美为单位的音高和以 HZ 为单位的频率之间近似成线性关系，但是，高于 500 Hz 时，美的增长速度逐渐小于 Hz 的增长速度，当频率达到一般听

* 《声学技术》，2005 年 3 期 （西部声学会议）

觉的最高值 16KHz 时，音高却只有 3400 美左右。

Hz与感知到的音高镁/美的计算公式：F_{Mel}= 550 ln (1 + F0 / 550)。

ERB(equivalent rectangular bandwidth)获得的原理同 Bark, 但蔽噪声用矩形波的“notch” 置于声调附近，建立 notch 的最小带宽使得声调的感知不受影响，在高频部分是对数关系，500Hz 以下处于直线与对数之间，如图 4 所示。Hz 到 ERB 的转换公式：Ferb = 11.17 ln ((F0 + 312) / (F0 + 14680)) + 43。

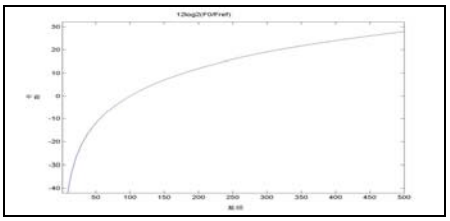


图 1， 50-500Hz 范围里，Hz 与半音的对应关系

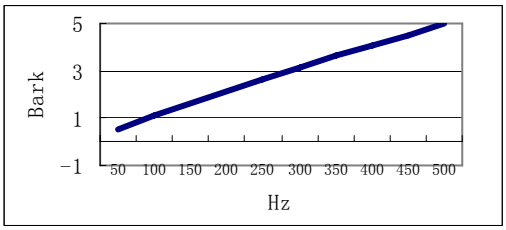


图 2， 50-500Hz 范围里 Hz 与 Bark 的对应关系

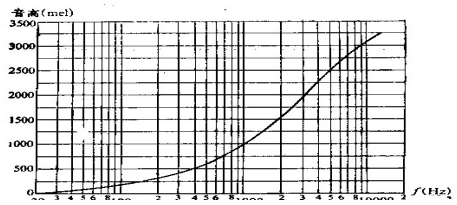


图 3， Mel 与 Hz 之间的音高对应关系

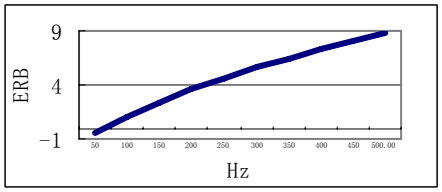


图 4，50-500Hz 范围里 Hz 与 ERB 的对应关系

3. 各种实验研究介绍

Ladd 指出[3]，音域(pitch range)是指音高变化的最高阈限与最低阈限之间的频率差，即两者之间的频率变化范围，也叫嗓音音域（voice range）。音高范围按照 Ladd 说法叫 Pitch Range，涉及音高变化的两个参数，一个是音高水平 level，另一个则是音高变化范围 span。音高水平是指一句话总体的音阶（key）值；音高范围（span）指说话人的声调空间，基频 F0 最大值和最小值之间的距离。研究语调的心理-声学量的方法有很大不同，有的知觉实验是针对音高范围、有的针对感知到凸显/重音，得到的结论也不同。

A.C.M. Rietveld 和 C. Gussenhoven[1] 通过让受试者判断频率不同、轻重不同的一对合成的刺激，当 F0 的偏移和变化用 Hz 表示，判断的正确性高于 semitone，这个结论似乎与 psycho-acoustic 相悖。

D.J. Hermes 与 J.C. van Geste[2]，则让受试者调整合成刺激的高度，让它与重音的参考高音阶或低音阶匹配，结果发现 ERB-rate scale 比 Hz 或 semitone 都要好。

Francis Nolan[4]就这个问题对英语进行了实验，他认为以前的知觉实验主要是对感知到的凸显进行判断，而不是对整个的旋律或韵律（melodic）进行感知。他提出了一种新的实验方法，能够应对不同性别的发音人确定音高等价量(pitch equivalence)。首先确定一些样板语音，然后让受试按照自己的发音来模仿这些样板发音，要求受试模仿听到的语调。受试模拟模板声音时音高范围 pitch span 尽量一致，而不用去考虑语音样板的声域(tessitura)。然后计算受试者语音和样板语音的音高范围的误差，分别按照不同的单位来计算误差的大小，包括半音 semitones, 镁 mel, 巴克 Bark 以及 ERB-rate 误差。统计结果发现，几个不同计量单位得到的误差值之间有显著的差异，而受试的性别没有对误差产生显著的影响；音高范围与受试的性别相关；半音计算得到的误差最小，其次是 ERB。实验说明对数（St）或接近对数（ERB）的单位能够最好地表达语调。

Rolf Carlson, Kjell Elenius 以及 Marc Swert[7]，利用语音库对瑞典语的音高范围进行了知觉实验。发音人的典型音高变化范围定义为 F0 累计频率曲线的 25%-75%两点间的距离，F0 中值为 50%对应的值。图 5a 和 5b 分别是 F0 采用 Hz 值和半音 St 得到的 498 名发音人的音高变化范围与中点值的情况，可以清楚看到，采用半音时，发音人的音高变化与中点值并行，就是说音高变化范围与中点值无关，而采用赫兹 Hz 为单位，中点值与音高变化范围相关性很大。所以在语调研究中使用半音 St 可以提高正交性。

Fant 等也在近两年的语调建模研究中提倡使用半音为研究单位，以消除平均声调级，使得不同发音人的语调模式不但反映了个人特性而且反映了整体平均值[5]。

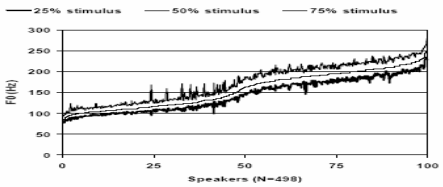


图 5-a, F0 采用 Hz 时发音人的音高范围和中值分布[7]

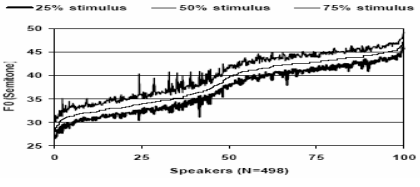


图 5-b, F0 采用 Hz 时发音人的音高范围和中值分布[7]

4. 汉语情感语音库的音高特征分析

我们认为上面的知觉实验尽管是对英语、瑞典语等进行的，但完全适用于汉语。因此我们在汉语的语调研究中也应该使用半音或 ERB 这样的对数比例单位。沈炯对传统的 5 度制规整方法也进行了论证[10]。

我们用上面提到的分析方法，对汉语情感语音库 CASS-EMC 的发音的音高特征进行了分析，CASS-EMC 是中国社会科学院语言所语音研究室制作的情感语音库[9]，主要包括字组、短句等各种语料的 6 种基本情绪，情绪分类为嘲讽(scorn)、高兴(happiness)、害怕(fear)、难过(sadness)、生气(anger)、厌恶(disgust),为了对比还录制了中性状态（neutral)。此外，还有短篇故事 3 篇。发音人为专业演员，3 男 1 女。我们对两位男发音人 MengY 和 ZWK 的相同语料的 F0 进行分析。

将所有基频数据按照不同情感状态统计，图 6-a 是发音人 MengY 的所有 F0 的累积变化频率分布曲线，图 6b-f 是 5 种情感状态下 F0 的累积变化频率分布曲线。

图7a和图7b分别是发音人MengY 的累积频率25%，50%，75%点对应的F0的Hz与St值。状态“All”是所有F0统计结果，代表了发音人真正的F0变化范围。下表是两位发音人在这3点上的值和“音域”(75%-25%)，可以看到两位发音人的典型音域分别是101-149Hz, 141-205Hz, 在线性坐标空间分别相差48和64Hz，但在对数空间若以ST(100Hz参考频率)为单位，分别是0.17-6.9St,5.9-12.4St,相差6.7和6.4个半音，非常接近。这说明用半音研究音高变化范围可以直接反映发音之间的异同。

Hz	*MengY*	*ZWK*		*ST*	*MengY*	*ZWK*
25%	101	141		25%	0.172264	5.948342
50%	121	161		50%	3.300085	8.244728
75%	149	205		75%	6.903748	12.42749
75%-25%	48	64		0.5	6.731484	6.479145

图 8a 和 8b 分别是按照传统方法计算的发音人的 F0 高点和低点以及 F0 均值。我们可以看到，两位发音人的音高分布模式有相似之处，音高的总体变化水平都是(高兴和愤怒)> 中性 >(害怕和悲伤)。但是两位发音人也有不同，MengY 的高兴与愤怒的 F0 变化范围有差距，Zwk 的高兴与愤怒 F0 变化范围没有显著差异；MengY 的害怕和难过与中性音高变化范围比 Zwk 的略大。说明，从音高上看，两位发音人使用的情感表达策略有一定的差别。从发音人基频变化的均值看，两位发音人有差别，MY 由高到低依次是：Happy〉Angry〉Neutra〉fear〉Sad；ZWK 由高到低依次是：Angry〉Happy〉Neutral〉fear〉Sad。从发音人音高上限看，两位发音人没有差别，由高到低依次是：Angry=Happy=4.9, Sad=Neutral=fear=4。从发音人音高下限看，两位发音人有差别，MY 由高到低依次是：Fear〉Happy〉Neutral〉Angry〉Sad；;ZWK 由高到低依次是：Happy〉Angry〉fear〉Neutral〉Sad。从发音人音高变化范围看，两位发音人有差别，MY 由高到低依次是：Sad〉Angry〉Happy〉Neutral〉fear；ZWK 由高到低依次是：Sad〉Neutral〉Angry〉fear〉Happy。

5. 小结

本文总结了当前关于语调研究中关于F0心理-声学相关量的各种知觉实验，提出在汉语语调研究中应该使用实验中验证的与基频F0有对数关系的半音、ERB等分析单位。

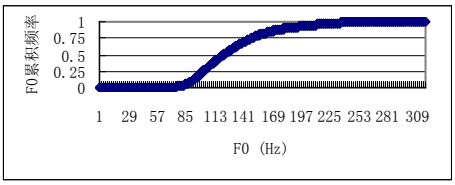


图 6-a, Mengy 的 F0 累计分布（HZ）

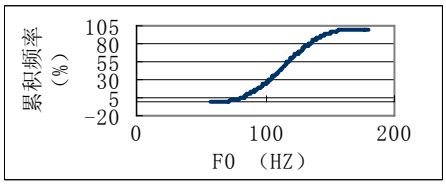
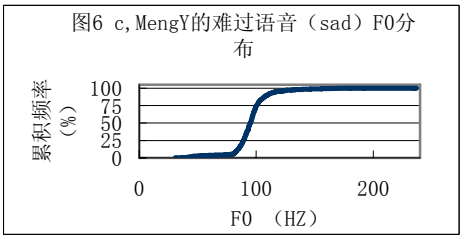
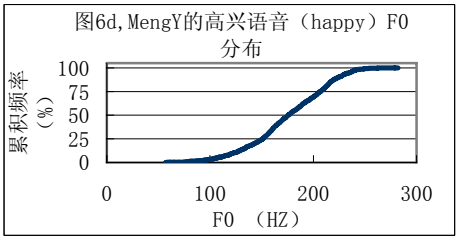
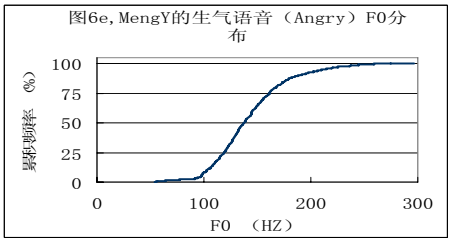
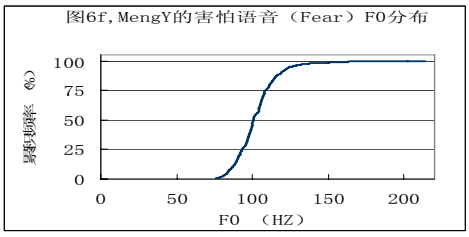


图 6-b, MengY 的中性语音（Nuetral）F0 分布









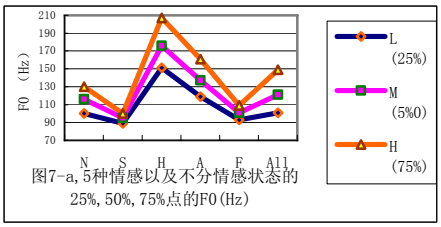


图7-a, 5种情感以及不分情感状态的25%, 50%, 75%点的F0(Hz)

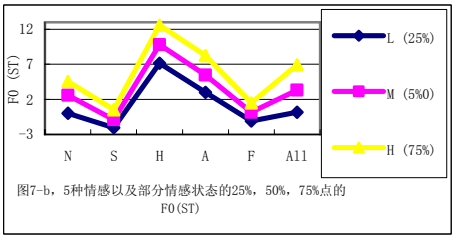


图7-b，5种情感以及部分情感状态的25%，50%，75%点的F0(ST)

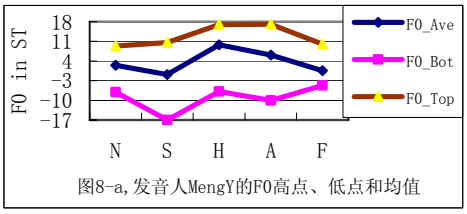


图8-a, 发音人MengY的F0高点、低点和均值

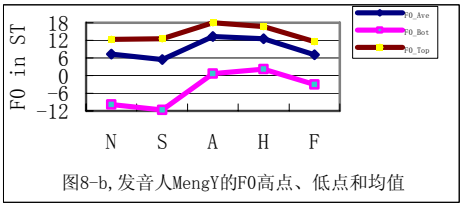


图8-b, 发音人MengY的F0高点、低点和均值

参考文献

[1] A.C.M. Rietveld and C. Gussenhoven，“On the relation between pitch excursion size and prominence”, Journal of Phonetics, 13, pp. 299–308, 1985..

[2] D.J. Hermes & J.C. van Gestel, “The frequency scale of speech intonation”, Journal of the Acoustical Society of America, 90, pp. 97–102, 1991.

[3] D. Robert Ladd, 1996, Intonational phonology, Cambridge University Press.

[4] Francis Nolan, Intonational equivalence: an experimental evaluation of pitch scales , ICPHS2003.

[5] Gunnar Fant, Anita Kruckenberg, Kjell Gustafson, Johan Liljencrants, A New Approach to Intonation Analysis and Synthesis of Swedish, Sp2002, Aix-en-Provence, France.

[6] John Clark and Colin Yallop,语音学与音系学入门, An introduction to phonetics and phonology, Blackwell publishers,外研社。

[7] Rolf Carlson, Kjell Elenius and Marc Swert ,Perceptual Judgments of Pitch Range , SP2004, Nara, Japan

[8] William J. Hardcastle and John Laver, Prosody of speech– Melody and Rhythm, from The handbook of phonetic sciences,, Blackwell publishers.

[9] 王海波，李爱军（2003），普通话情绪语音库的建立及听辨实验，第六届全国现代语音学学术会议论文集。

[10]沈炯，美标度是比例标度──纠正一种流行的误解，《第四届全国语音学研讨会论文集》，金城出版社1999。

The Psycho-acoustic Units for Intonation Study

Aijun LI

Institute of Linguistics, CASS

ABSTRACT ：This paper describes the current research on what unit that we should use in intonation research. Analysis is made for different units: Hz and other psycho-acoustic correlated units such as Semitone, Mel, Bark and ERB. And the result tells that semitone is the best for intonation analysis. The paper also analyzes the F0 for several basic emotional speeches as demonstration.

Keywords: intonation, pitch, F0, semitone, psycho-acoustic unit