从简单的音高目标到复杂的基频曲线

许毅
美国西北大学
2299 N。Campus Dr., Evanston, IL 60208, USA
xuyi@nwu.edu

摘 要

本文介绍我们新近提出的一种生成基频曲线的理论模型[38]。这个模型假定，普通话语句中的基频曲线是在主观和客观因素的共同作用下形成的。主观因素包括音高目标和调域。音高目标对应于声调，它决定每个音节里的底层音高形式。调域决定音高目标的实现范围，它取决于语调，情感等高层次的语言功能。客观因素包括多种发音局限，这些发音局限在相当程度上决定音高目标实现的方式和程度。

1 引 言

从事声调和语调研究的人们现在已经普遍意识到，表层的基频曲线跟底层的声调和语调的关系是十分复杂的。虽然如此，我们最近的一系列研究发现，这种关系是有一定规律可循的 [35, 36, 37]. 在这些研究的基础上，我们新近提出一种生成基频曲线的理论模型[38]。这个模型试图从理论上解释如何可以由简单的底层音高目标生成复杂的表层基频曲线。

2 基本框架

这个模型的基本假设是，影响基频曲线的因素虽然很多，但是可以分为根本不同的两大类。一类源于语言交际的功能，它们对基频曲线的影响起转达语言学或情感上的意义的作用，是说话人有意产生并期望别人听到的，所以可以称为主观因素。另一类因素源于发音局限，它们对基频曲线的影响是说话人不得已而为之的，是无意产生并且也无意让别人听到的，所以可以称为客观因素.

主观因素对基频曲线的控制是通过实现各种音高目标(pitch target) 和调域(pitch range)。音高目标是最小的可操作的音高单位，它是由声调或音高重音(pitch accent) 等低层次的语言学单元规定的. 调域是音高目标实现的范围，它是由语调、情感等高层次的语言功能决定的。音高目标和调域的实现受到多种发音局限的制约。这些制约直接影响表层基频曲线的形状。

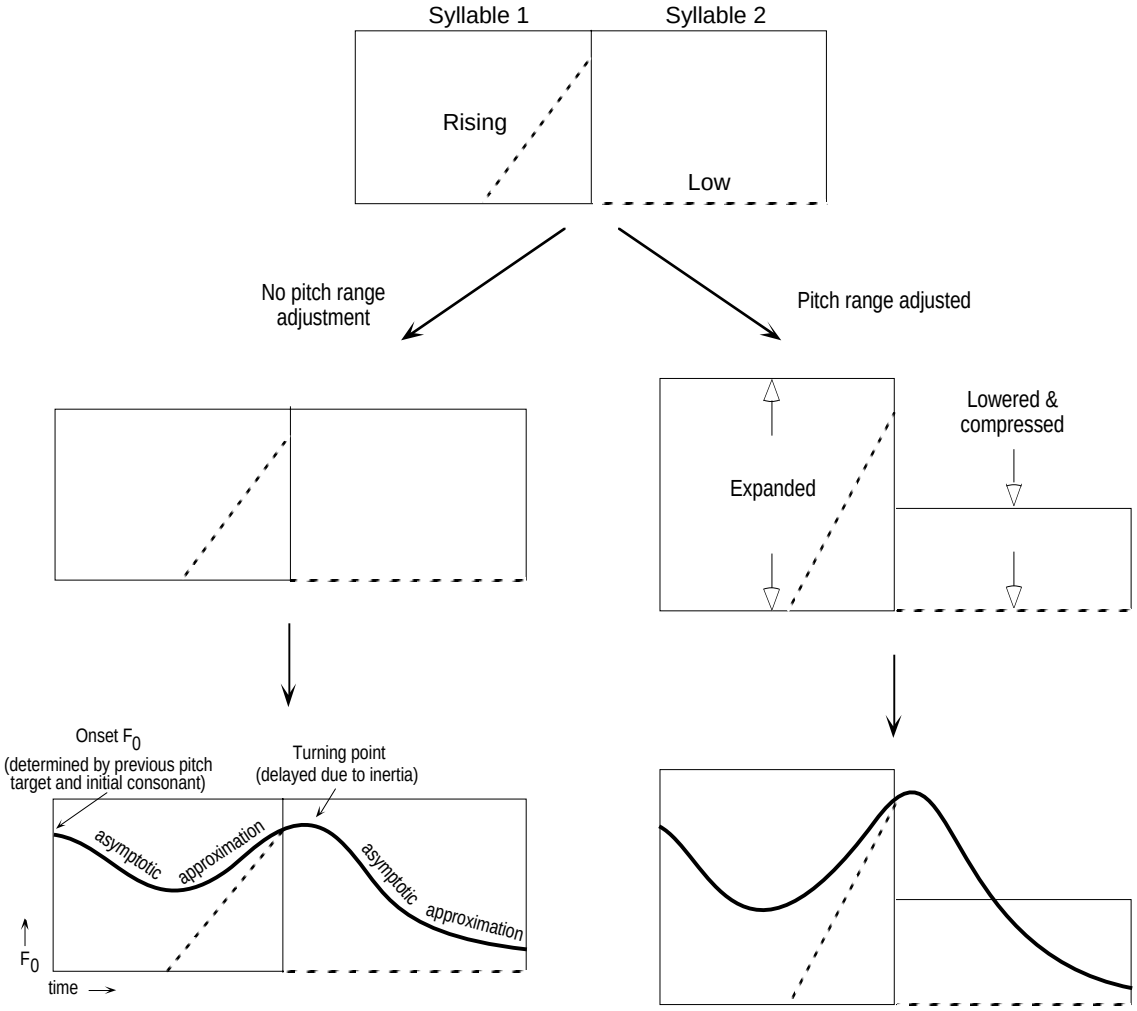
图一是该模型的图形示例。在图一里，垂直线表示音节边界，虚线表示音高目标。在该模型里，音高目标分为静态和动态两种。静态目标只规定音高的高度，比如高、中、低。动态目标只规定音高的变化，比如升、降。在图一的双音节里，前一个音节里的音高目标是动态的，后一个音节里的音高目标是静态的。普通话里的阴平和非单念的上声音高目标都是静态的，分别为高和低。阳平和去声的音高目标则是动态的，分别为升和降。所以，图一里前音节的音高目标对应于普通话里的阳平，后音节的音高目标对应于普通话里非单念的上声。

图一里的水平实线表示调域的上端和下端。在图一[上] 里，调域是中性的。在图一[中] 里，左边的双音节调域保持中性，右边的双音节的调域受到两种不同的调节：在前音节里，调域被加宽，即上端增高，下端降低。在后音节里，调域被同时降低和压缩。

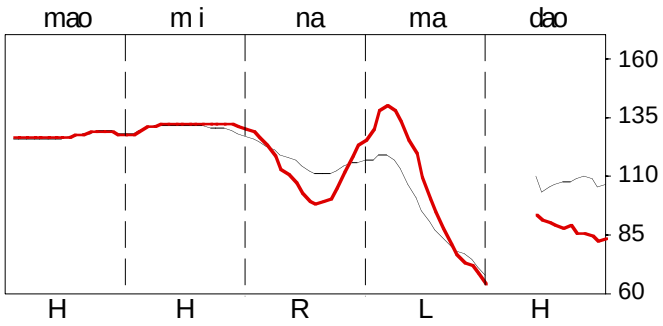
图一[下] 表示的是音高目标如何最后实现为基频曲线(粗实线)。可以清楚地看到，基频曲线跟底层的音高目标很不一样，这是因为音高目标的实现受到发音局限的制约。我们把这些制约的规律归纳如下:

一、每一个音高目标都有自己固定的载体;
二、音高目标与其载体同步实现，也就是两者的实现同时开始同时结束;
三、音高目标的实现贯穿于载体的始终，并且是以渐进线方式逼近;
四、逼近音高目标的速度和程度受限于说话人升、降音高的最快速度;
五、音高运动转变方向的速度受限于说话人音高转向的最快速度。

我们将在后边讨论这些规律的来由。现在先解释它们如何将音高目标实现为基频曲线。



图一 音高逼近模型示例。



图二 “猫咪拿马刀”的平均基频曲线（四位男发音人）。细曲线: 无焦点;
粗曲线: 焦点在“拿”字上。

首先，在普通话里，因为音高目标跟声调直接对应，而声调的载体是音节，所以音高目标的载体是音节. 第二，由于音高目标与其载体同步实现，并且以渐进线方式逼近，一个音节里的基频曲线始终处于过渡状态。其起点取决于前一个声调的终点值以及本音节的声母。过渡的方向是朝本音节的音高目标逼近。第三，由于音高目标的这种实现方式，音节后部的基频曲线最接近于声调的本值。因此，在图一[下] 里，基频曲线上升最快的部分(即最典型的"升") 是在第一个音节的后部；基频的最低点(即最典型的"低") 则在第二个音节的后部。第四，在图一[下] 里，因为前一个音节里的音高目标是"升"，后一个音节的音高目标是"低"，基频曲线在音节交界处由升变降。但是，基频曲线的转折点(即最高点) 并不是落在音节交界点上，而是落在交界点之后，也就是在后一个音节里。这种"滞后"是由于音高转向的速限造成的。在实现一个"升" 目标时，由于持续逼近，最快的基频上升发生在音节的尾部。在音节交界处，虽然基频上升已经减慢，但是还没有停止，因为彻底的转向受音高转向最快速度的限制，需要花一些时间。结果真正的转向发生在音节交界之后。

3 主观因素

3.1 音高目标

以往的许多研究认为所有的曲拱(contour) 声调都应该分析为复合调[6, 10, 17, 33]，比如普通话里的阴、阳、上、去在非单念时应为HH、LH、LL、HL。但是其他一些研究认为曲拱声调不应分析为复合声调[1, 3, 23, 31]。我们自己最近的一些研究发现，说话人是把每个声调作为一个整体而不是作为两个单独的目标来实现的[35, 36, 37]。所以，在我们的模型里，音高目标无论是静态或动态的都是单一的而不是复合的。而且，和其他反对复合声调的理论不一样，我们认为普通话声调的底层形式并不是复杂的曲线，而是简单的音高目标。如图一所示，由于多种因素的共同作用，简单的音高目标经过发音实现便变成了复杂的基频曲线。

3.2 调域

调域可以受到多种因素的调节。这些调节可以是全局性的，即影响到整个甚至多个句子，也可以是局部性的，即只影响句子的一部分。调域的调节有两种方式。一种是调节调域的范围，即上端和下端之间的距离。一种是调节调域的高低。这两种调节可以同时进行，也可以分别进行。以焦点(focus) 为例。根据已有的研究[4, 7, 37]，当一个句子里某个成分成为焦点时，该成分的调域加宽，即上端升高，下端降低；该成分之后的调域降低并且变窄；但是该成分之前的调域基本保持中性。图二显示的是"猫咪拿马刀" 的平均基频曲线。其中细曲线的句子没有任何焦点。粗曲线的句子焦点在"拿" 字上。图中的垂直虚线是音节边界；H、R 和L 分别代表阴平、阳平和上声。从图中可以清楚地看到，当"拿" 字为焦点时，它的基频最低点降低，最高点提高(并由于前面提到的滞后落在后边的上声音节前部)。在"马" 字里，基频迅速下降，到音节尾已低于中性句里的最低点。在"刀" 字里，基频更是大大低于中性句子里的句尾基频。图一里模拟的就是"猫咪拿马刀" 里"拿马" 二字的基频曲线生成过程。图中左右分支的差别在于有没有调域调节. 其中左分支的调域保持中性. 在右分支里，第一音节的调域加宽，第二音节的调域同时降低和压缩。

还有一种已知的影响调域的因素是引入新的话题。在这种情况下，句子里第一个重读词的调域提高，但并不一定加宽。其结果是，句首的基频大大提高，随后基频以渐进线方式下降。这种由于引入新话题而造成的句首调域提高已经在一些语言里发现[18, 20, 30]，它跟篇章结构有关。由它造成的调域变化还有待于今后进一步的研究。

4 客观因素

如前所述，客观因素指的是影响基频的各种发音局限。我们现在已经知道的发音局限至少有六种: (1) 说话人的总调域、(2) 元音的内在音高、(3) 辅音对基频的影响、(4) 喉部与声门上发音器官的协调、(5) 音高升降的最快速度、(6) 音高转向的最快速度。下面对它们逐一作简要的讨论。

4.1 总调域

据Fairbanks [8]，一个人在说话时所用的调域可跨越两个八度音。比方说，如果某一男声的最低基频为80 Hz，他的最高基频应该可以达到320 Hz 。一般的声调和语调研究很少涉及到这么大范围的调域。所以对此我们所知甚少。至于总调域会如何影响音高目标的实现我们现在还不很清楚。因为不同人的总调域会有差异，所以应该会有一定的影响。

4.2 元音内在音高

在其他条件一切相等的情况下，不同元音的基频有所不同，这已在多种语言里发现[19, 32]。其差异主要取决于元音高度. 高元音的基频略高于低元音。普通话也不例外，同一个声调在不同元音里的基频不一样[26]。也有一些研究发现，在连续的话语里，元音内在音高的差别大大减小[15]。

4.3 辅音对基频的影响

这种影响有两种，一种是对后接元音的影响，一种是对前接元音的影响。对前者已经有多项研究。主要的发现是，清辅音会抬高后接元音的基频，某些浊辅音会降低后接元音的基频 [11, 19, 24]. 不过这种影响是很局部的。范围一般不超过 30 ms。对辅音对前接元音基频的影响研究较少。我们自己非系统的观察注意到，塞音和擦音会降低前接元音的基频 (Shih [27] 也注意到类似现象)。不过这种影响也是很局部的。

以上三种发音局限对本模型的基本设想都没有太大的影响，而且它们的作用将来可以很容易地补入。所以它们现在还都不是本模型的基本成分。相反，另外三种发音局限直接影响音高目标的实现方式，所以它们的作用都直接体现在本模型里。

4.4 喉部与声门上发音器官的协调

这方面的直接研究现在还很少。不过我们可以从两类有关的研究里略见一斑。一个是关于肢体运动协调的研究。已有的研究发现，人的肢体的任何两个部位若要协调运动，它们之间最稳定的相位差为零[12, 25]，也就是两者的动作同时开始同时结束。当运动的频率快到一定程度时，两者协调的唯一可能是保持相位差为零。另一类研究是我们对普通话声调的系统观察 [35, 36, 37]。观察的结果表明声调的实现跟音节是共时的。当一个声调跟在不同的声调后面时，越接近音节尾，它的各条基频曲线越趋向于集中，并且越接近该声调的深层模式。而当不同的声调跟在某一个声调后面时，在音节一开头它们的基频都很接近，然后才分道扬镳，各自逼近自己的本调。所以，在我们的模型里，音高目标与其载体同步实现。(现有的许多音系学理论都认为在很多语言里声调的载体是不固定的，一个声调可以从它的载体音节蔓延到后边的音节。但我们的研究发现，至少某些非洲语言 (如 Yoruba [16]) 里的声调蔓延现象非常类似于普通话里的声调实现方式[37]。)

4.5 音高升降的最快速度

Ohala & Ewan [21] 和 Sundberg [29] 对此作了专门研究。他们的主要发现是，一、音高下降的速度快于音高上升的速度；二、女声改变音高的速度快于男声；三、音高变化跨越的幅度越小，速度越慢。例如，音高上升 12 st (semitone)，即一个八度音，需要 120 ms。而音高上升 6 st 则需要 70-80 ms. 在普通话里，由四声造成的基频变化常常在 6 st 左右，而一个正常重音的普通话音节的平均时长为 180 ms [37]。这样要在一个音节里实现去声，如果音节开头处基频很低的话，基频从低到高的过渡至少要占将近半个音节。而去声本身所需要的 " 降" 则只好用剩下的半个音节。我们对普通话声调基频模式的分析证明了这种推测 [36, 37]。由于音高变化的速限，如果音节时长过短，音高目标的实现程度也会受到影响。这个推测也已有以往数据的一定支持 [27, 34]。

4.6 音高转向的最快速度

这里指的是音高变化从升到降或从降到升的转变速度。Ohala & Ewan [21] 和 Sundberg [29] 的研究并没有测量这个速度。但是我们知道，由于声门的运动必然有惯性，音高转向肯定要花时间。这在我们观察过的基频曲线里可以清楚地看到。尤其是类似于图二里阳平最高点的滞后现象已经在多种语言里观察到 [2, 5, 14, 35, 36, 37]。现在仍然不够清楚的是各种音高转向具体要花多少时间。这有待于今后进一步的研究。

4.7 音高目标的逼近方式

我们假设逼近音高目标是以渐进线方式首先是根据对实际基频曲线的观察。至于背后的机制，我们的猜测是，这是精确控制目标逼近的必然结果：随着基频曲线越来越接近音高目标，逼近音高目标的速度必须减慢以保证逼近目标的精确性。在研究伸手触摸物体的实验里发现，手指的运动速度是先快后慢的 [13]。而且，被触摸的物体越小，也就是说对精确度的要求越高，手指达到物体要用的时间就越长。所延长的时间主要是花在手指逼近目标的最后阶段。所以，以渐进线方式逼近目标很

可能是受控运动的一条基本规律。当然，这也有待于今后更近一步的研究。

5 与现有模型的比较

现有的大部分基频曲线模型作的都是表层模拟。表层模拟的最大问题是缺乏预测能力，因此很难用于语音合成和识别。在此就不作详谈。深层模型有 Pierrehumbert 模型 [22] 和 Fujisaki 模型 [9]. Pierrehumbert 模型是用来模拟英语语调的。它以重读音节里的基频高峰和低谷为模拟的主要目标，而非重读音节里的基频则只用简单的下垂 (sagging) 曲线来模拟。这种模型用于声调语言显然是不够的。Fujisaki 的模型最初是为模拟日语语调而设计的。它假设每个句子的语调都是由两种成分构成的：重音 (accent) 和短语成分。重音的底层形式是台阶形指令；短语成分的底层形式是脉冲形指令。表层的基频曲线产生于对重音和短语指令的响应。对两种指令的响应都是临界阻尼振荡：一旦指令结束，基频曲线便回归底线。在诸多现有的模型里，Fujisaki 的模型属于最深层的。但是，它跟我们的模型相比，有两点重要的不同。第一，在我们的模型里，基频曲线在逼近一个音高目标之后，并不需要回归底线，而是立刻开始逼近下一个目标。第二，在我们的模型里，音高目标与其载体是完全同步实现的。在 Fujisaki 的模型里，重音指令跟深层的重音并不完全同步，而且重音指令的数量跟深层重音的数量也往往不一样。这就会使模型的预测能力受到影响。

6 数学模拟

以上描述的模型还只是一个理论框架。如要用于语音合成或识别就必须把它量化。这方面的研究我们也已经开始进行，并已有初步进展 [39]。

参 考 文 献

[1] Abramson, The phonetic plausibility of the segmentation of tones in Thai phonology, The 12th International Congress of Lingusitics, Vienna, 1978.

[2] Arvaniti, A., Ladd, D. R. & Mennen, I., Stability of tonal alignment: the case of Greek prenuclear accents, Journal of Phonetics, 36, 1998.

[3] Clark, M., A Dynamic Treatment of Tone, with Special Attention to the Tonal System of Igbo, Ph.D. dissertation, University of Massachusetts, Amherst, 1978.

[4] Cooper, W. E., Eady, S. J. & Mueller, P. R., Acoustical aspects of contrastive stress in question-answer contexts, Journal of the Acoustical Society of America, 77, 1985.

[5] de Jong, K., Initial tones and prominence in Seoul Korean, OSU Working Papers in Linguistics, 43, 1994.

[6] Duanmu, S., Against contour tone units, Linguistic Inquiry, 25, 1994.

[7] Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R. & Lotts, D. W., Acoustic characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments, Language and Speech, 29, 1986.

[8] Fairbanks, G., Voice and Articulation Drillbook, Harper & Row, New York, 1959.

[9] Fujisaki, H., Modeling the process of fundamental frequency contour generation. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), Speech Perception, Production and Linguistic Structure, IOS Press, Amsterdam, 1992.

[10] Gandour, J., On the representation of tone in Siamese, UCLA Working Papers in Phonetics, 27, 1974.

[11] Hombert, J.-M., Consonant types, vowel quality, and tone, In V. A. Fromkin (Ed.), Tone: A linguistic survey, Academic Press, New York, 1978.

[12] Kelso, J. A. S., Phase transitions and critical behavior in human bimanual coordination, American Journal of Physiology: Regulatory, Intergrative and Comparative, 246, 1984.

[13] Kelso, J. A. S., Southard, D. L. & Goodman, D., On the nature of human interlimb coordination, Science, 203, 1979.

[14] Kim, S.-A., Positional effect on tonal alternation in Chichewa: Phonological rule vs. phonetic timing, Proceedings of Chicago Linguistic Society, 34, 1999.

[15] Ladd, D. R. & Silverman, K. E. A., Vowel intrinsic pitch in connected speech, Phonetica, 41, 1984.

[16] Laniran, Y., Intonation in Tone Languages: The phonetic Implementation of Tones in Yoruba, Ph.D. Dissertation, Cornell University, 1992.

[17] Leben, W. R., Suprasegmental Phonology, Massachusetts Institute of Technology, 1973.
[18] Lehiste, I., The phonetic structure of paragraphs, In A. Cohen & S. E. G. Nooteboom (eds.), Structure and process in speech perception, Springer-Verlag, New York, 1975.
[19] Lehiste, I. & Peterson, G. E., Some basic considerations in the analysis of intonation, Journal of the Acoustical Society of America, 33, 1961.
[20] Nakajima, S. & Allen, J. F., A study on prosody and discourse structure in cooperative dialogues, Phonetica, 50, 1993.
[21] Ohala, J. J. & Ewan, W. G., Speed of pitch change, Journal of the Acoustical Society of America, 53, 345(A), 1973.
[22] Pierrehumbert, J., Synthesizing intonation, Journal of the Acoustical Society of America, 70, 1981.
[23] Pike, K. L., Tone Languages, University of Michigan Press, Ann Arbor, 1948.
[24] Rose, P. J., On the non-equivalence of fundamental frequency and pitch in tonal description. In D. Bradley, E. J. A. Henderson & M. Mazaudon (Eds.), Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg, Pacific Linguistics, Canberra, 1988.
[25] Schmidt, R. C., Carello, C. & Turvey, M. T., Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people, Journal of Experimental Psychology: Human Perception and Performance, 16, 1990.
[26] Shi, B., & Zhang, J., Vowel intrinsic pitch in Standard Chinese, The 11th International Congress of Phonetic Sciences, Tallinn, Estonia, 1987.
[27] Shih, C., Tonal variations in connected speech, Presented at International Workshop on Tone, Stress and Rhythm in Spoken Chinese, Prague, 1999.
[28] Shih, C. & Sproat, R., Variations of the Mandarin rising tone, The IRCS Workshop on Prosody in Natural Speech No. 92-37, Philadelphia, The Institute for Research in Cognitive Science, University of Pennsylvania, 1992.
[29] Sundberg, J., Maximum speed of pitch changes in singers and untrained subjects, Journal of Phonetics, 7, 1979.
[30] Umeda, N., "F0 declination" is situation dependent, Journal of Phonetics, 10, 1982.
[31] Wang, W. S. Y., Phonological features of tone, International Journal of American Linguistics, 33, 1967.
[32] Whalen, D. H. & Levitt, A. G., The universality of intrinsic F0 of vowels, Journal of Phonetics, 23, 1995.
[33] Woo, N., Prosody and phonology, Massachusetts Institute of Technology, 1969.
[34] Xu, Y., Production and perception of coarticulated tones, Journal of the Acoustical Society of America, 95, 1994.
[35] Xu, Y., Contextual tonal variations in Mandarin, Journal of Phonetics, 25, 1997.
[36] Xu, Y., Consistency of tone-syllable alignment across different syllable structures and speaking rates, Phonetica, 55, 1998.
[37] Xu, Y., Effects of tone and focus on the formation and alignment of F0 contours, Journal of Phonetics, 27, 1999.
[38] Xu, Y. & Wang, Q. E., Pitch targets and their realization: Evidence from Mandarin Chinese, Speech Communication, in press.
[39] Xu, C. X., Xu, Y. & Luo, L.-S., A pitch target approximation model for F0 contours in Mandarin, To be presented at the 14th International Congress of Phonetic Sciences, San Francisco, 1999.