

音节时间结构与拉萨藏语的声调起源*

胡 方

[摘要] 本文认为声调不是与载调单位抽象连接的自主音段，而是音节产生的有机组成部分。通过对拉萨藏语音节产生中的辅音、元音、声调等发音动作的时间结构关系的考察，让声调重新回归音节，并以此为基础讨论藏语的声调起源问题。

[关键词] 音节时间结构 发音动作 声调起源 自主音段

一 引言：声调与音节产生

自从 Goldsmith (1976) 以来，音系学界一般便采用自主音段的理论来讨论声调等超音段的问题，音段归音段，声调归声调。音段指元音和辅音，它们之间的区别是声音的本质区别，即音质或称音色的区别。超音段主要是语言的韵律特征，表现在音高（声调）、音强（重音）、时长（音长）等的不同。除此之外，元音和谐、鼻化等施加于语言的基本发音之上的种种可以提取出来的发音修饰特征，在音系上均可以被处理为自主音段。一方面，自主音段理论强调声调行为的独立性。在生理与物理层面，基频由喉部的声带振动产生，在控制上相对独立于喉上的发音 (*supralaryngeal articulations*)，即音段。在音系层面，声调语言中普遍存在的一些声调现象，比如浮游调 (*floating tones*)、声调蔓延 (*tonal spreading*) 等都不以声调所基于的音段为条件。也就是说，声调与音段处于一种平行关系，因此，音段层 (*segmental tier*) 与声调层 (*tonal tier*) 便构成了相互自主的关系。另一方面，除了规定自主音段层之间存在着抽象的符合形式合法条件 (*Well-Formedness Condition*) 规约的线条连接，自主音段音系学并未解释声调层与音段层之间存在着什么样的关系；而所谓形式合法条件，其实只有一条，即连接线不交叉（参见 Yip 2002:65-104）。

在言语韵律研究的框架中，语言中的基频行为也是与音段分开处理的，无论是在建模的还是在经验的研究中均是如此 (Fujisaki et al. 2005; Tao et al. 2006; Tseng et al. 2005)。不过，音段与声调或更广义的基频行为之间的关系并非是抽象的。相反，语音学的实验研究揭示，无论是在声调语言中 (Xu 1998, 1999, 2005)，还是在非声调语言中 (D'Imperio et al. 2007; Mücke, Grice, Becker et al. 2009)，声调或基频行为与音段之间的连接关系在时间结构上呈现出稳定、具体的特点。简言之，实验研究的结果是支持声调的传统视角的，即声调是音节的组成部分 (Chao 1968)。

近些年，声调在音节中的组织架构问题，在发音音系学 (Articulatory Phonology) 框架内

* 本文在“《民族语文》第十四届学术研讨会”(贵阳 2021.7.17-18) 上宣读过，感谢与会专家、匿名审稿人的意见和建议。

得到了重新审视。发音音系学(Brownman & Goldstein 1986, 1988, 1992)将发音动作(articulatory gesture)视为音系的基本单位，每一个单独的发音动作便是一个行动单位，其涉及声腔内某一特定部位收缩的形成与解除。不同于传统的音系概念声称属于语言学内部，是自治的(autonomous)，发音音系学的基本单位为发音动作，它遵循运动学(kinematics)与动力学(dynamics)的一般原则，因而可以将其定义为一个拥有若干时间与空间属性的动态任务系统(a task dynamic system with temporal and spatial properties)(Saltzman 1986; Saltzman & Kelso 1987; Saltzman & Munhall 1989)。通过观察发现，当发音动作处于同相位关系(in-phase coupling)——即同时开始(synchronized)时，或者逆相位关系(anti-phase coupling)——即次序开始(sequential)时，发音动作之间在时间结构关系上是稳定的(Brownman & Goldstein 1988, 2000; Goldstein et al. 2006)。Brownman & Goldstein (1988)指出，一个音节中的发音动作遵循两个一般原则：(1) 元音(vowel, V)前的辅音(consonant, C)与元音是同相位的，即与元音同时发音；(2) 元音前的辅音个体之间是逆相位的，即它们之间次序发音。因而，在一个C₁C₂V的复辅音声母的音节中，由于C₁与C₂均要和V同时发音，而C₁与C₂之间又要次序发音，一个可能的结果就是C₁的发音相对于V提前，同时，C₂的发音则相对于V推后(Brownman & Goldstein 2000)。如图1所示意，在理想的状态下，元音的发音动作恰好大致会在两个辅音的中间开始。这个“辅音中心效应(the C-center effect)”便成为描述音节时间结构的一个重要手段，在诸多语言的复辅音声母的音节中可以观察到(Brownman & Goldstein 1988, 2000)。

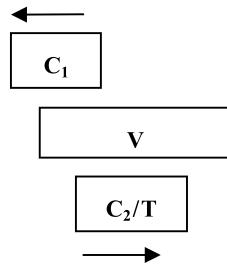


图1 音节时间结构的辅音中心效应

声调(tone, T)也不例外。声调是一个喉部发音动作，也可以纳入音节产生(syllable production)的时间结构关系进行观察。Gao (2008)检视了汉语普通话音节产生中的发音协同关系，发现普通话的CV音节中的辅音C、元音V、声调T三个发音动作展现了辅音中心效应。也就是说，在普通话CV音节中，声调就像是另一个辅音声母C₂，就像把图1中的发音动作C₂换成T。因此，普通话的带调音节CV也可以被表征为CTV音节，这个结构与带辅音从语言中的复辅音声母音节C₁C₂V是类似的。这个发现更新了声调的音系表征问题：声调不是抽象的自主音段，而是有机结合在音节时间结构中的一个喉部发音动作。在Gao的声调表征中，拱度调可以分解为平调的结合。因此，普通话阴平调由一个高调发音动作H组成，阳平调由一个低调发音动作L与一个高调发音动作H组成，上声调是一个低调发音动作L，去声调由一个高调发音动作H与一个低调发音动作L组成。有意思的是，Gao发现，阳平调升调的声调发音动作LH之间是同相位结合的，即L与H同时发音；而去声调降调的声调发音动作HL之间则是逆相位结合的，即H与L次序发音。沿着Gao的思路，Hsieh (2011)

进一步指出，普通话的上声调应该被表征为 LH 的逆相位结合，这样才能解释其在单念时实现为曲折调，而在[上声+上声]变调时实现为升调。其后，Yi & Tilsen (2015) 又进一步检视了普通话的上声变调升调与单字调升调之间的区别，并细化了发音音系学框架下的声调表征问题。此外，从发音动作的视角检视声调现象的研究也包括了泰语等其他东亚、东南亚声调语言 (Karlin 2014)。

Hu (2012) 发现拉萨藏语音节发音动作间的时间结构关系也支持将声调表征为音节产生的内在组成部分。音节产生的发音动作分析对于藏语研究来说具有特别的意义，因为这将打破声调语言与非声调语言之间的类型壁垒，为二者之间建立内在的沟通桥梁，并为真正理解声调起源问题提供理论上的可能性。众所周知，藏语方言呈现了极其丰富的从无声调语言至声调语言之间的演变梯度：有些方言是典型的声调语言，有些方言是完全的无声调语言，而更多的方言则处于二者之间（黄布凡 1994；Sun 2003）。拉萨藏语属于典型的声调语言，其声调拥有完善的高低调的对立，不过，学者们对于拉萨藏语究竟有多少个高低调的对立却存在着争议（参见瞿靄堂 1981），这主要是因为拉萨藏语声调的调形与音节类型高度相关（Hu & Xiong 2010）。本文重新分析 Hu (2012) 中的材料，并结合新近的研究讨论拉萨藏语音节的时间结构与声调起源问题。

二 研究方法

本文以拉萨藏语单念单音节为研究对象，检视其相关的声学与发音属性。根据藏文单念的拉萨藏语有四种音节结构，每一种都能与高调或者低调相连接。这样便有八类，每类若干例字，代表八种目标音节与声调的组合。为了便于观察发音动作，我们尽量选取双唇辅音声母 [p m]、低或半低元音 [a ε] 的有意义的单音节词作为例字，如表 1 所示。

表 1 测试音节与例字

音节类型	调类	例字		
长音节 CVS	高调	[peɪ] 丰富、吉祥	[par] 照片	[paŋ] 草坪
	高调	[meɪ] 脊宿	[mar] 红色	[maŋ] 根基
	低调	[peɪ] 羊毛	[par] 烧	[pam] 粗壮结实 _{将来时}
		[meɪ] 住处	[mar] 酥油	[maŋ] 多
复杂喉塞尾音节 CVN?	高调	[pan?] 高度		[man?] 平民
	低调	[pam?] 粗壮结实 _{过去时}		[maŋ?] 多
喉塞尾音节 CV?	高调	[pa?] 推测 _{过去时}		[ma?] 战役
	低调	[pa?] 沾染		[ma?] 女婿的首音节
尾送气短音节 CVh	高调	[pah] 勇敢		[mah] 低
	低调	[pah] 藤		[mah] 妈

关于字表，有三点需要说明：第一，发音人朗读的时候所使用的藏文读书音保留了较多的文字信息，与日常口语有所不同。比如，在藏文读书音中的长音节保留了正字法中所带的流音韵尾 [l] 或 [r]，而在日常口语中则实现为长元音，没有韵尾。不过，这个并不影响本文

所讨论的内容，因此，下文不赘。第二，在文献中，CVN? 常常归类于 CV? 音节。本文将其单列，是因为 CVN? 音节拥有复杂韵尾，其音节时长也长于 CV? 音节，这可能影响其声调的实现。第三，尾送气音节 CVh 在文献中一般处理为开音节。在语流中，确实不易听见尾送气成分，不过，在单念的读书音中，这个音节结尾的送气成分——即喉擦音——还是很清晰的。

测试音节用藏文书写，放入一个用藏文书写的载体句内，用 LCD 显示器随机呈现在发音人面前。所用的载体句为：“X, ji ke? (藏文) ti (这个) X sa (是)”，意为“X，这个藏文是 X 字”。这样，发音人一遍发音便有两个目标音节 X：其中一个在句首单念，适合分析单念音节声调的基频曲线；另一个位于句中，适合观察发音器官的运动轨迹。

我们采用三维电磁发音仪（德国 Carstens 公司的 AG500 EMA 系统）同步录取发音人的语音与发音器官运动信息。共有三位女性发音人参加了实验，均为西藏拉萨当地人，无言语或听力障碍，录音时为中央民族大学大一或大二学生，20 岁或 21 岁。录音重复 10 至 15 遍。如图 2 所示，沿着发音人的中矢平面，将实时监测发音器官的传感器粘贴在发音人的上唇（UL）、下唇（LL）、下齿龈脊（代表下颌运动，Jaw）、舌尖（TT）、舌体（TB）。除此之外，还有三个传感器分别粘贴在鼻梁中间与左右耳后作为参考，以去除说话时头部移动对采样点数据的影响。发音数据的采样率为 200 赫兹，采集后的数据经过 12 赫兹低通滤波处理、头部移动校准等预处理步骤。为了使不同说话人之间的数据在显示上呈现一致性，我们录制了每个发音人的咬合面，在预处理步骤中，我们将发音数据旋转至与各自发音人的咬合面平行。

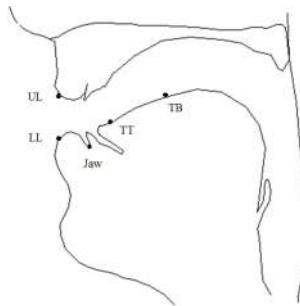


图 2 传感器粘贴示意图

由于选取了双唇音声母例字，本文用唇开合度，即上下唇采样点之间的距离来定义辅音声母的发音动作。由于选取了低、半低元音例字，因此，元音的发音动作可以用舌体采样点的运动特性来定义，即表述为舌体的下降。声调的发音动作则用基频 (F0) 来定义。根据 Gao (2008)，我们将高调的基频曲线前一个基频低点定义为高调的起始，而将低调的基频曲线的前一个基频高点定义为低调的起始。

图 3 显示的是对预处理之后的发音数据分析与标注的一个实例 [mar]。图中显示了四层标注、六层信号。信号层位于图的下部，从上至下分别为：声波 (audio)、基频 (F0)、唇开合速度 (L|vel)、唇开合位置 (L|pos)、舌体位置 (tb|pos)、舌体运动速度 (tb|v)。标注层位于图的最上方。其中，两层是声学标注，依据相应的声学信号所做的标注；两层是发音标注，依据相应的发音运动信息所做的标注。第一层是目标音节的声学段，在这个例子中就是 [mar]，标注的依据是 [mar] 的声学边界。第二层是声调，在这个例子中是一个高调 H，依据的是 [mar] 音节的韵母段的基频曲线。第三层是唇开合度，在这个例子中由声母 [m] 的双唇关闭 (close)、

双唇打开(open)动作构成,依据的是唇开合的位置与相应的速度信息。第四层是舌体动作,在这个例子中就是音节核心元音[a]的舌体下降(lower)动作,依据的是舌体位置与相应的舌体运动速度信息。发音标注只应用于句中的目标音节,因为无法厘定单念的目标音节的发音边界。在发音标注中,我们运用相应的发音器官采样点的最小速度准则(the criterion of tangential velocity minimum)来确定发音边界。如图3所示,当采样点的切线速度处于波谷的时候,即是其发音动作的边界处。如有必要,还可以进一步参考相应采样点的加速度信息(未在图中显示)来帮助确定发音边界。声学标注则应用于两个目标音节。下文分析中的声调部分基于单念的目标音节,而发音部分的音节时长则基于句中位置的目标音节。为了画面整洁,图3中没有显示目标声调的起始点。按照定义,高调H的起始点是其前的基频折点最低处,从图3中可以看到,这个位置刚好与句中目标音节[mar]的左声学边界重合,也就是位于载体句中前一音节[ti]的结束处。

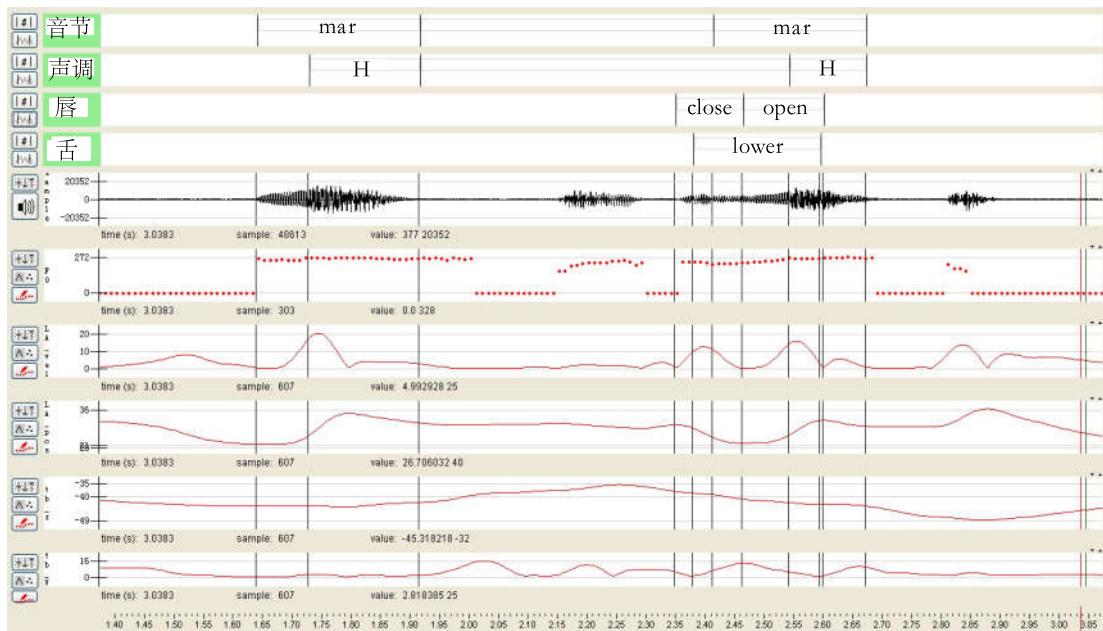


图3 拉萨藏语音节[mar](高调)的声学与发音标注

图中显示了四层标注(图的最上方从上至下):目标音节的声学段(syllable)、目标音节韵母段的基频即声调(tone)、唇开合度(LA)、舌体动作(TB)。

三 研究结果

(一) 声学研究结果

先看声学研究的结果。在检视了采样个体的变异情况之后,图4显示了拉萨藏语三位发音人的基频曲线均值。图中的每一条曲线均基于每个发音人单念位置目标音节的所有采样(2个例字×10至15次重复=20-30个样本);因为未对时长进行归一,因此,图中某些曲线的结尾处有些不稳定的扰动,忽略即可。从图中可以看到,拉萨藏语的八种音节结构对应于各自的基频曲线,即拥有不同的调形,而且,这种调形之间的区别模式在三位发音人间是一致的。

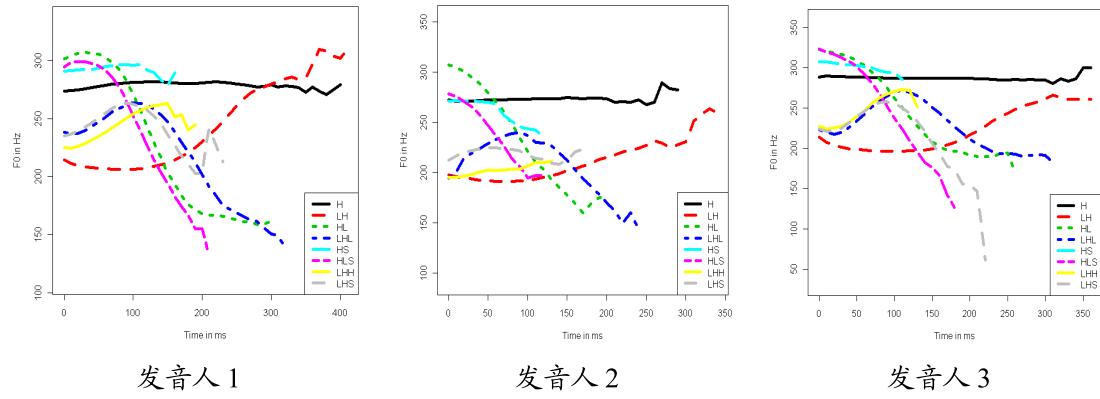


图 4 拉萨藏语声调

我们将拉萨藏语的音节类型、调类、调形总结如表 2。首先，比较每一种音节结构的高低调，可以清晰地观察到其中的区别。在声学上，高低调的区别在声调的开头部分特别明显，高调的起始基频在 270-320 赫兹，而低调的起始基频在 190-240 赫兹。此外，低调的最高基频也会比相应的高调略低一些。也就是说，相对于高调来说，低调拥有一个显著的低基频的开头，然后升至比相应的高调的基频略低的位置。其次，声调的调形与音节结构是紧密相关的。从音系分析的角度讲，拉萨藏语的缺省调值是一个高的调形 H：如果是高调类，那么，就实现为 H；如果是低调类，那么，就在前面加一个低的起始调形 L，实现为 LH。此外，如果有喉塞尾，那么，增加一个低的调尾 L。

表 2 拉萨藏语的音节类型与声调调形

音节类型	调类	调形	标签
CVS	高	长平	H
	低	长升	LH
CVN?	高	短（？）降	HL
	低	短（？）升降	LHL
CV?	高	短降	HLS
	低	短升降	LHS
CVh	高	短平	HS
	低	短升	LHH

本文揭示的上述两个事实能够帮助我们更好地理解文献中关于拉萨藏语声调的争论。一般认为拉萨藏语拥有高低调的区别，但是对拉萨藏语究竟有几个声调则有不同观点。拉萨藏语的声调调形与音节结构之间存在着紧密关联这一事实为不同的音系解释提供了多种可能性。“两调说”认为拉萨藏语只有高低两个声调，因此是一种只有调域区别的平调语言 (Sprigg 1954, 1990)。“四调说”则进一步将音节时长或喉塞尾作为音位对立，要么在区分高低调的基础上再区分长短调，要么根据有无喉塞尾将高低调再进行两分 (参见 Sun 1997:491-492)。而“六调说”则在区分高低调的基础上同时区分音节长短与喉塞尾 (胡坦等 1982)。

本文认为关键不在于声调的数量，而是在于分析调形与分化，以帮助厘清拉萨藏语声调起源与发展的脉络。首先，拉萨藏语声调起源符合 Hombert et al. (1979) 所提出的内在基频干扰理论 (the intrinsic F0 perturbation theory)。我们知道，拉萨藏语的低调音节来自于古藏语的浊声母音节，拉萨藏语的浊阻塞音声母虽然已经清化，但是，浊声母对音节首的基频干扰，即低基频起始，被保留了下来，与相对应的清声母音节，即现在的高调音节^①，构成了声调上的音位对立。其次，调形的对立正在发展之中。首当其冲的是低调本身引起的调形变化。如前所述，低调只影响音节的起始部分，因此，所有的低调都有一个升调的调形。那么，这个升调的调形在语言学上有没有意义？以长音节为例，拉萨藏语低调音节的 LH 与其相应高调音节 H 之间的声调对立究竟是仅区别调域的平调系统 (register tones)，还是拥有调形区别的拱度调系统 (contour tones)？对这个问题的完整回答需要进一步的语音与音系分析。就本文的材料来看，拉萨藏语往往拥有调形区别的拱度调系统发展。除了低调引起的升调，韵尾也对声调的调形产生影响。如前所述，拉萨藏语的喉塞尾音节的尾部带有下降的基频曲线。也就是说，喉塞尾导致基频下降。虽然在汉语声调的发展历程中常常观察到喉塞尾与高调关联在一起，Hombert et al. (1979) 也表示喉塞尾引发基频升高，但是，喉塞尾降低基频在语言中也并非是少见的现象。这是因为喉塞往往实现为嘎裂声 (creaky voice)，让所在音节带上一个降调尾是非常自然的。而且，正如谭克让、孔江平 (1991) 所观察到的，拉萨藏语的喉塞音带有紧喉色彩；这便印证了喉塞尾音节的降调尾。不过，需要注意的是，喉塞尾对基频的降低作用有时候是急剧的，因此，反过来，一个急剧下降的基频曲线往往标示着喉塞音的存在 (Zee & Maddieson 1979)。也就是说，基频下降的幅度非常关键。急降的基频曲线是喉塞音附带的一个冗余特性，听话人不易感知到降调的存在；而缓降的基频曲线则容易被感知为降调。就拉萨藏语来说，CV? 音节时长较短，而 CVN? 音节的时长较长，虽然二者拥有相同的基频曲线——高调 HL、低调 LHL，但是，高调 CV? 音节容易被感知为短高调，而高调 CVN? 音节则容易被感知为降调 (参见 Hu & Xiong 2010)。同理，低调 CV? 音节容易被感知为短升调，而低调 CVN? 音节则容易被感知为升降调。而且，拉萨藏语 CVN? 音节中的喉塞尾是非常容易脱落的 (金鹏 1983:13)，所以，韵尾脱落之后的降调尾的发展问题便成了一个关键。此外，从图 4 中可以观察到，尾送气也可以造成基频下降，不过，这个效果没有喉塞尾那么明显，尤其是在发音人 1 和发音人 2 的材料中。前文提到，尾送气音节前人处理为开音节，与长音节以时长区别，在声调上也就不存在调尾的问题。如前所述，胡坦等 (1982) 总结拉萨藏语拥有六个声调。本文与之唯一的区别就是，我们进一步区分了 CVN? 与 CV? 音节。因此，本文总结了八种调形。但是，我们并不认为拉萨藏语的声调发展犹如汉语声调发展历程的再现。基于目前的分析，我们认为拉萨藏语的声调还在发展过程之中，将来是否如汉语一般走向成熟的拱度调系统，还有待进一步观察 (Hu & Xiong 2010)。而且，就目前的情况看，拉萨藏语的声调发展已经呈现了与汉语的声调发展历史不同的特点：汉语是先四声（调形）再八调（高低），而拉萨藏语则是先高低后调形。

（二）发音研究结果

毫无疑问，拉萨藏语已经产生了具有音位对立意义的声调，而且，声学研究证实了拉萨

^① 拉萨藏语的高调音节还包括原来带上加字、前加字的响音声母音节，本文不讨论响音声母之间的声调对立问题。

藏语声调的高低对立来自于古浊音声母遗留下来的对低调音节起始基频的干扰。浊辅音声母，尤其是浊阻塞音声母降低音节起始部分的基频是语言中的一个普遍的现象，无论在声调语言，还是在非声调语言中都可以观察到 (Hombert et al. 1979)。也就是说，所有的浊阻塞音声母都会引起音节起始部分基频的下降，但是，并不是所有的浊阻塞音声母的清化都会导致声调起源。也就是说，辅音声母引起的基频干扰属于浊辅音的伴随特性，音节起始部分的低基频特性为什么没有随着浊辅音的清化而消失，反而是被保留了下来，浮现为声调的区别特性，这是需要进一步考察的。

发音研究检视了拉萨藏语音节产生中发音动作之间的时间结构关系 (the intergestural timing of syllable production)，尤其是音节开始部分的辅音声母、元音、声调的发音动作之间是否存在稳定的结构关系。图 5 中的条状图总结了来自三位发音人的发音数据。图中的横条标示了音段或发音动作的时长均值，单位：毫秒。其中，音节层 (syllable) 标示拉萨藏语音节产生的声学界标：辅音声母音段 [p] 或 [m]、韵母 (rime)；此外，用虚线条标示了声调发音动作的起始 (tone onset: t_on)。唇开合 (lip) 标示了塞音 [p] 或鼻音 [m] 所需的双唇发音动作：双唇关闭 (close) 然后打开 (open)。舌发音 (lingual) 标示了发低元音 [a] 或半低元音 [ɛ] 所需的舌体下降 (lower) 动作。

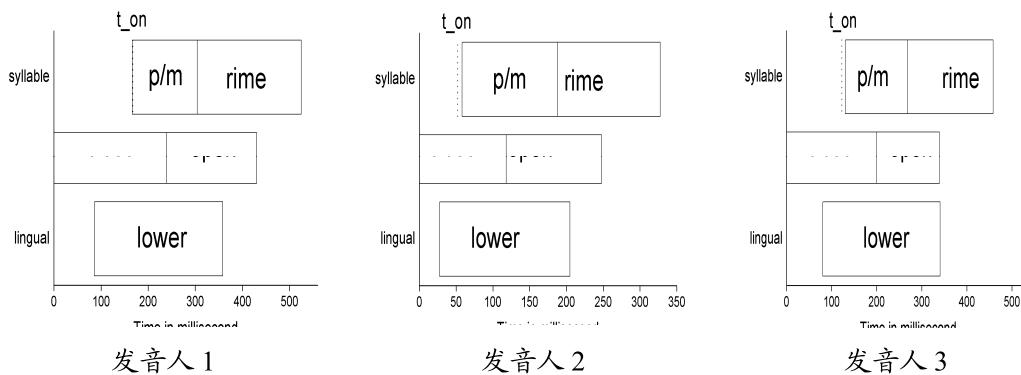


图 5 拉萨藏语音节产生中发音动作间的时间结构关系

从图 5 中可以看到，三位发音人所展示的发音动作间的时间结构关系是一致的，即使发音人 2 的语速比其他两位发音人快了很多。首先，无论是双唇发音动作还是舌发音动作，它们在辅音声母音段 [p] 或 [m] 之前就开始了。也就是说，辅音声母与元音的发音动作开始于音节的声学起始之前，事实上是在双唇闭塞持阻阶段之前。其次，声调发音动作略早于音节的声学起始，或者二者基本同时。第三，双唇打开动作大约开始于 [p] 或 [m] 闭塞持阻阶段的中间，结束于韵母的中间。第四，大概在韵母的前半段，舌体已经下降到最低的位置，也就是到达了低或者半低元音的发音目标位置 (articulatory target)。

从图 5 中也可以看到：在拉萨藏语的音节产生中，辅音声母、元音、声调的发音动作依次开始。因此，元音在辅音声母与声调之间开始的时间架构总体上符合“辅音中心效应”。本文的材料大致符合三因素设计：音节结构 (CVS、CVN?、CV?、CVh)、辅音声母 ([p m])、声调 (高、低)。为了进一步检视上述三因素是否影响辅音声母、元音、声调的发音动作之间

的时间结构关系，我们计算了每一位发音人的辅元间隔（即辅音声母至元音的时间间隔，the Consonant-to-Vowel lag）、元调间隔（即元音至声调的时间间隔，the Vowel-to-Tone lag），并用三因素方差分析评估辅元间隔与元调间隔之间的差异。

来自发音人1的发音数据研究结果显示：音节结构 ($F(3,320) = 0.6463, p = 0.5858$)、辅音声母 ($F(1,320) = 0.0272, p = 0.8692$)、声调 ($F(1,320) = 0.0519, p = 0.8200$) 均没有显著效应；音节结构与辅音声母的交互 ($F(3,320) = 0.5314, p = 0.6611$)、音节结构与声调的交互 ($F(3,320) = 1.0751, p = 0.3598$)、辅音声母与声调的交互 ($F(1,320) = 0.1604, p = 0.6890$)，以及音节结构、辅音声母与声调的三者交互 ($F(2,320) = 1.2462, p = 0.2890$) 也均没有显著效应。

来自发音人2的发音数据研究结果显示：音节结构 ($F(3,208) = 0.7506, p = 0.5231$)、辅音声母 ($F(1,208) = 1.0917, p = 0.2973$)、声调 ($F(1,208) = 0.0855, p = 0.7703$) 均没有显著效应；音节结构与辅音声母的交互 ($F(3,208) = 1.6034, p = 0.1897$)、音节结构与声调的交互 ($F(3,208) = 0.2495, p = 0.8616$)、辅音声母与声调的交互 ($F(1,208) = 0.1123, p = 0.7379$) 也均没有显著效应；但是，音节结构、辅音声母与声调的三者交互在 <0.05 的水平上拥有显著效应 ($F(2,208) = 3.3392, p = 0.0374$)。

来自发音人3的发音数据研究结果显示：音节结构 ($F(3,214) = 3.7871, p = 0.0112$) 与辅音声母 ($F(1,214) = 8.9015, p = 0.0032$) 均拥有显著效应，但是声调 ($F(1,214) = 1.4429, p = 0.2310$) 没有显著效应；音节结构与辅音声母的交互 ($F(3,214) = 2.0283, p = 0.1109$)、音节结构与声调的交互 ($F(3,214) = 0.9170, p = 0.4335$)、辅音声母与声调的交互 ($F(1,214) = 0.0282, p = 0.8668$)，以及音节结构、辅音声母与声调的三者交互 ($F(2,214) = 0.0339, p = 0.9667$) 均没有显著效应。

综上分析，在拉萨藏语的音节产生中，声调类别并不影响发音动作间的时间结构关系；而且，在不同的音节结构与辅音声母类别之间，辅音声母、元音、声调的发音动作间的时间结构关系也基本上呈现稳定的“辅音中心效应”。

拉萨藏语音节产生中的发音动作之间存在着稳定的、呈现“辅音中心效应”的时间结构关系，这为进一步解释拉萨藏语的声调起源问题提供了强有力的生理—物理证据。发音动作的时间结构关系直接反映发音器官之间的协同，这说明声调作为一个发音动作本身就是音节的有机组成部分。如图1所示，从音节产生的时间结构关系看，拉萨藏语的声调作为发音动作的行为完全就像是音节首的第二个辅音。从历时演变的角度讲，浊音作为辅音声母的一个特征在拉萨藏语中消失了，但它所引起的音节开始部分的基频干扰作为属于辅音的一个发音动作（声调）被保留了下来。因此，声调在音节时间结构关系上就像是音节首的一个额外的辅音这一事实为声调语言与非声调语言之间建立了直接的联系，突破了二者之间的类型区隔。在历时层面，以发音动作为语言音系的生理—物理基本单位为我们理解音系演化，尤其是跨类型的演化提供了一个直接的观察手段（Goldstein et al. 2006）；在共时层面，以发音动作为基础的研究范式为我们在完全不同的语言之间进行音节复杂性的比较提供了可能。

四 结论与讨论

本文的声学研究确认拉萨藏语拥有清晰的高低调的声调区别，不过，拉萨藏语声调的调形与音节结构紧密关联，显示其调形系统还在发展过程之中。从历时的角度看，拉萨藏语声

调来自于遗留下来的浊辅音——尤其是已经清化了的浊阻塞音声母——对音节起始部分基频的扰动。发音动作间的时间结构关系揭示了拉萨藏语的辅音声母、元音、声调发音动作呈现“辅音中心效应”，也就是元音的发音动作大体上介乎于辅音声母与声调的发音动作之间的中点位置。声调并非只是概念化的自主音段，与载调单位（Tone Bearing Unit, TBU）进行抽象连接。在拉萨藏语的音节产生中，发音动作间拥有稳定的时间结构关系，这便说明声调作为喉部发音动作属于音节的有机组成部分，受到音节产生时间结构关系的制约。声调发音动作犹如音节首的另一个辅音这一事实则很好地解释了作为浊辅音附带韵律特性的低基频扰动如何作为一个发音动作浮现为语言中的声调区别。

在声调语言中，声调是词汇层的表征，尤其是在像汉语普通话这类音节节奏型的语言中，声调作为喉部发音动作受到音节产生机制的控制，成为音节的内在组成部分，在时间结构上与其他发音动作之间呈现出稳定的结合关系，是一个比较容易理解的现象。但在非声调语言中，基频行为属于句子层面的韵律特性。因此，来自加泰罗尼亚语（Catalan）与德语的研究表明，当基频行为与某一音节连接，比如音高重音落在某一音节时，基频与音节的连接仍然属于句子层面，并不影响音节内辅音与元音的时间结构关系（Mücke, Grice, Becker et al. 2009; Mücke, Nam, Hermes et al. 2012）。因此，从宏观的角度去理解声调起源是错的，声调与语调虽然都属于基频行为，但声调并非起源于宏观的、句子层面的基频行为的词汇化。相反，根据本文的研究结果，声调来源于音节内部辅音特征的转化，浊辅音声母对于音节开始部分的基频扰动虽然在物理表现上属于韵律的内容，但在发音生理上仍然类似于一个辅音——一个喉部发音动作。

根据 Geissler et al. (2021) 近期对二男四女总共六位生活在美国并都操类似于拉萨话的藏语卫藏方言的成年藏族人的语音采样，只有四位发音人有声调对立，另外两位发音人没有声调；不过，无论声调对立是否存在，六位发音人的辅元间隔（CV lags）是一致的，均遵循声调语言的音节内的发音动作时间结构模式。这便说明，相比较于通过基频这个物理属性显现的声调对立而言，在发音生理层面，声调作为喉部发音动作是更为基础的内在属性。

参考文献

- [1] 胡 坦、瞿靄堂、林联合. 1982.《藏语（拉萨话）声调实验》，《语言研究》第 1 期.
- [2] 黄布凡. 1994.《藏语方言声调的发生和分化条件》，《民族语文》第 3 期.
- [3] 金 鹏主编. 1983.《藏语简志》，北京：民族出版社.
- [4] 瞿靄堂. 1981.《藏语的声调及其发展》，《语言研究》第 1 期.
- [5] 谭克让、孔江平. 1991.《藏语拉萨话元音、韵母的长短及其与声调的关系》，《民族语文》第 2 期.
- [6] Browman, C. P. & L. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook*, 3: 219-252.
- [7] Browman, C. P. & L. Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica*, 45: 140-155.
- [8] Browman, C. P. & L. Goldstein. 1992. Articulatory Phonology: An overview. *Phonetica*, 49: 155-180.
- [9] Browman, C. P. & L. Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5: 25-34.
- [10] Chao, Y.-R. 1968. *A Grammar of Spoken Chinese*. Berkeley & Los Angeles: University of California Press.
- [11] D'Imperio, M., H. Loevenbruck, C. Menezes, N. Nguyen & P. Welby. 2007. Are tones aligned to articulatory

- events? Evidence from Italian and French. In Jennifer Cole & José Ignacio Hualde (eds.), *Laboratory Phonology 9*, pp. 577-608. Berlin & New York: Mouton de Gruyter.
- [12] Fujisaki, H., C. Wang, S. Ohno & W. Gu. 2005. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command-response model. *Speech Communication*, 47: 59-70.
- [13] Gao, M. 2008. Mandarin Tones: An Articulatory Phonology Account. Yale University Ph.D. dissertation.
- [14] Geissler, C., J. Shaw, M. Tiede & F. Hu. 2021. Eccentric C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts. In Mark Tiede, Doug Whalen & Vincent Gracco (eds.), *Proceedings of the 12th International Seminar on Speech Production*, pp. 162-165. New Haven: Haskins Press.
- [15] Goldsmith, J. A. 1976. Autosegmental Phonology. MIT Ph.D. dissertation.
- [16] Goldstein, L., D. Byrd & E. Saltzman. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael A. Arbib (ed.), *Action to Language via the Mirror Neuron System*, pp. 215-249. Cambridge: Cambridge University Press.
- [17] Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language*, 55: 37-58.
- [18] Hsieh, F.-Y. 2011. A gestural account of Mandarin Tone 3 variation. In *Proceedings of International Congress of Phonetic Sciences XVII*, pp. 890-893, Hong Kong.
- [19] Hu, F. 2012. Tonogenesis in Lhasa Tibetan—Towards a gestural account. In P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer & B. Kühnert (eds.), *Consonant Clusters and Structural Complexity*, pp. 231-254. Berlin & Boston: De Gruyter.
- [20] Hu, F. & Z. Xiong. 2010. Lhasa tones. In *Proceedings of Speech Prosody 2010*, 100163: 1-4, Chicago.
- [21] Karlin, R. 2014. The articulatory TBU: Gestural coordination of lexical tone in Thai. Zenodo. <https://doi.org/10.5281/zenodo.3726325>.
- [22] Mücke, D., M. Grice, J. Becker & A. Hermes. 2009. Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *Journal of Phonetics*, 37: 321-338.
- [23] Mücke, D., H. Nam, A. Hermes & L. Goldstein. 2012. Coupling of tone and constriction gestures in pitch accents. In P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer & B. Kühnert (eds.), *Consonant Clusters and Structural Complexity*, pp. 205-230. Berlin & Boston: De Gruyter.
- [24] R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [25] Saltzman, E. 1986. Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (eds.), *Generation and Modulation of Action Patterns*, pp. 129-144. Berlin: Springer Verlag.
- [26] Saltzman, E. & J. A. S. Kelso. 1987. Skilled actions: A task dynamic approach. *Psychological Review*, 94: 84-106.
- [27] Saltzman, E. L. & K. G. Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4): 333-382.
- [28] Sprigg, R. K. 1954. Verbal phrases in Lhasa Tibetan. *Bulletin of the School of Oriental and African Studies*, 16(1): 134-156, (2): 320-350, (3): 566-591.
- [29] Sprigg, R. K. 1990. Tone in Tamang and Tibetan, and the advantage of keeping register-based tone systems separate from contour-based systems. *Linguistics of the Tibeto-Burman Area*, 13(1): 33-56.

- [30] Sun, Jackson T.-S. 1997. The typology of tone in Tibetan. *Chinese Languages and Linguistics IV: Typological Studies of Languages in China* (Symposium Series of the Institute of History and Philology, Academia Sinica, Number 2), pp. 485-521. Taipei: Academia Sinica.
- [31] Sun, Jackson T.-S. 2003. Variegated tonal developments in Tibetan. In D. Bradley, R. LaPolla, B. Michailovsky & G. Thurgood (eds.), *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff*, pp. 35-51. Canberra, Australian National University: Pacific Linguistics.
- [32] Tao, J., Y. Kang & A. Li. 2006. Prosody conversion from neutral speech to emotional speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4): 1145-1154.
- [33] Tseng, C., S. Pin, Y. Lee, H. Wang & Y. Chen. 2005. Fluent speech prosody: Framework and modeling. *Speech Communication*, 46: 284-309.
- [34] Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, 55: 179-203.
- [35] Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27: 55-105.
- [36] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46: 220-251.
- [37] Yi, H. and S. Tilsen. 2015. Gestural timing in Mandarin tone sandhi. *Proceedings of Meetings on Acoustics*, 22, 060003.
- [38] Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.
- [39] Zee, E. & I. Maddieson. 1979. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *UCLA Working Papers in Phonetics*, 45: 93-129.

On Temporal Organization of Syllable Production and the Emergence of Tones in Lhasa Tibetan

HU Fang

[Abstract] This paper disagrees with autosegmental phonology, and argues that tones are not autosegmentals that are associated with tone bearing units (TBU) in an abstract fashion. Rather, tone gesture is coupled with consonant and vowel gestures in syllable production. This paper describes gestural coordination of initial consonants, vowels, and tones in Lhasa Tibetan, and gives a syllable-based account for the emergent tones or tonogenesis.

[Keywords] temporal organization of syllable production gesture the emergence of tone (tonogenesis) autosegmental

(通信地址: 100732 北京 中国社会科学院语言研究所)

【本文责编 胡鸿雁】