

基于语料库的藏语韵律短语声学特征研究

曲 珍^{1,2} 德庆卓玛^{1,3} 拉巴顿珠^{*1,3}

(1. 西藏大学信息科学技术学院 西藏拉萨 850000

2. 西藏大学藏文信息技术创新人才培养示范基地 西藏拉萨 850000 3. 西藏信息化省部共建协同创新中心 西藏拉萨 850000)

摘要 语句和篇章韵律结构分析与建模,是提升语音合成自然度、降低语音识别错误率的关键环节之一。文章基于韵律标注语料,系统分析了藏语韵律短语的时长与音高等声学特征,并深入探讨了其分类问题。研究表明,在藏语朗读过程中,主韵律短语与次韵律短语在边界停顿时长和音高走向上呈现出显著的系统性差异。其中,边界处是否出现显著停顿是区分两类短语的核心声学依据;与此同时,短语末端普遍出现音高重置现象,但主韵律短语的末端音高下降幅度更为显著。上述发现不仅深化了对藏语韵律层级的认识,并为语音合成和语音识别等技术提供声学特征依据。

关键词 藏语;主韵律短语;次韵律短语;停顿时长;音高下倾;音高重置

DOI:10.16249/j.cnki.2096-4617.2026.03.012

中图分类号 TP391.1

文献标识码 A

文章编号 2096-4617(2026)03-123-07

引言

在语音合成系统的性能评估中,可懂度与自然度是衡量听觉感知质量的两大核心指标。当前,藏语语音合成在可懂度方面已取得显著进展^[1],能够基本实现语义信息的清晰传递,但在自然度上仍有待进一步优化。提高合成语音的自然度,不仅需要准确划分韵律结构,更为关键的是实现对韵律短语边界位置的准确预测^[2-4]。

在语音韵律层级体系中,韵律短语是连接上下文的重要韵律单元,通常呈现相对稳定的韵律组织模式。其边界的准确识别直接影响基频曲线的建模和音节时长的拟合,进而影响合成语音的自然度与流畅性^[5]。因此,提高韵律短语边界预测的准确度,是改善藏语语音合成性能的关键技术路径。

在其他语言的韵律短语研究中,通常将韵律短语划分为主韵律短语(major phrase)和次韵律短语(minor phrase)两类。其中,主韵律短语由一个或多个次韵律短语构成,具有短语重音,但其后常伴随着明显的停顿,并呈现音高重置与音高下倾的现象^[6];次韵律短语则由一个或多个韵律词构成,同样具有短语重音,但其后常伴有较短的停顿,表现为无声停顿或轻微的有声停顿。

收稿日期:2026-01-07

基金项目:教育部人文社会科学研究项目(20YJA740035);国家自然科学基金项目(62406256);西藏自治区科技计划重点研发及转化项目(XZ202502ZY0034);西藏自治区自然科学基金项目(XZ202501ZR0113)。

第一作者简介:曲珍,女,藏族,西藏日喀则人,西藏大学信息科学技术学院副教授,主要研究方向为藏文信息处理。

通讯作者简介:拉巴顿珠,男,藏族,西藏日喀则人,西藏大学信息科学技术学院副教授,主要研究方向为藏文信息处理。

在藏语韵律层级的研究中,通常将藏语的韵律单位划分为韵律词、韵律短语和语调短语3个层次^[7]。其中,韵律短语被定义为处于同一节奏群内、相互联系紧密的若干韵律词组合,其边界通常伴随音高重置以及边界前音段延长等声学特征^[8]。然而,目前专门针对藏语韵律短语的研究较为匮乏,使得其内部层级结构难以得到精细刻画。

为此,本文参照其他语言韵律短语的分类方法^[9],将藏语韵律短语划分为主韵律短语和次韵律短语两个子类;进而基于藏语韵律标注语料库,从时长、音高走向等声学特征入手,对两类韵律短语进行对比分析,以获得更为细致的分类依据,并为藏语韵律短语的结构描写和声学特征的系统刻画提供科学的解释框架。

1 实验数据

本研究所使用的藏语合成语料库共收录了5 000条藏文句子,均由播音专业的本科女生在自然状态和正常语速下进行朗读并录制。从中挑选出内容完整、长度适中的500条陈述句作为实验材料。基于Praat语音标注工具完成音素和音节的自动切分与标注后进行人工校对,并对韵律词及韵律短语的边界进行标注。其中,韵律词边界标记为“1”^[7]。

在其他语言的韵律结构边界识别中,各类标点符号可作为韵律边界的显性标识,为韵律短语、语调短语等边界划分提供一定辅助^[10]。然而,藏语并未形成汉语般完备、丰富的标点符号体系。藏语句子内部以“隔音符”(Tsheg)分隔音节,句末则借助单垂符“|”、双垂符“||”等分隔符号标示,这对语调短语边界识别具有重要参考价值,却难以用于韵律短语的划分^[11]。因此,藏语韵律短语边界只能依托听觉感知,并结合句法与语义信息进行标注,倘若听觉感知结果同句法、语义信息存在显著矛盾,一般优先采信听觉感知的判断。具体标注需遵循以下准则:

- (1)句末标记:在每个句子结尾处标注韵律短语的边界。
- (2)长时停顿标记:在出现较长无声停顿的地方标注韵律短语的边界。
- (3)感知间断标记:即使没有明显的无声停顿,只要听觉上能感受到语音的间断,也需在该处标记韵律短语的边界。

在上述基础上,进一步结合听觉感知与藏文句法、语义信息,对韵律短语进行层次划分。划分方法如下:次韵律短语边界标记为“3”;主韵律短语边界标记为“4”或“5”(二者在语言学上无本质区别,仅用于区分不同层次的主要短语)。若听觉感知与句法、语义信息出现冲突,仍以听觉感知作为判断韵律短语边界类型的主要依据。因此,本研究的标注以听觉感知为核心,句法和语义信息仅作参考。标记“3”“4”“5”分别对应次韵律短语及两个层次的主韵律短语,从而实现对藏文语音韵律结构较为细致的划分。

图1为实验语料中第0266句:“བུ་རྒྱུ་ཆོག་གི་པ་ལུ་བཟ་ཤིས་སྒང་ནི་ལུ་ལྗོངས་མཛེས་སྤྱུག་ལྷན་པ་དང་དམངས་སྤྱོད་ཆོས་ལྷན་པ་འི་ལུ་ལྗོངས་ཆེན་མོ།”(平措的家乡扎西岗是一个风景秀丽、民风淳朴的好地方)的韵律标注实例。

完成韵律标注后,再借助专门编写的Praat脚本,对语料库中的主韵律短语和次韵律短语进行系统特征提取。具体而言,针对每一个韵律短语,本研究统计了其内部包含的音节数量和韵律词数量,测量了整体时长,提取了短语边界处的停顿时长以及对应的基频(音高)走向等声学参数。在此基础上对上述声学参数进行了系统分析,以深入探讨两类韵律短语之间的声学关系。

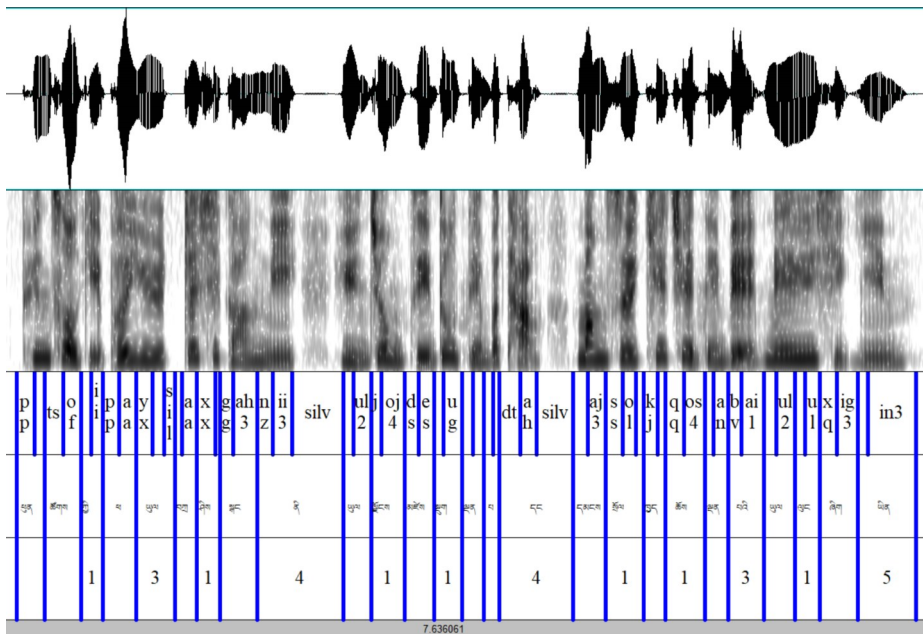


图1 “0266”句的韵律标注
Fig.1 Prosodic annotation of “0266” sentence

2 韵律短语的时长特性

时长是语音中重要的韵律特征之一,对语音的自然度和表现力具有决定性影响。在自然语音中,韵律单元的时长并非固定不变,而是受到其特性与更高层级韵律特征之间复杂交互作用的影响,呈现系统性变化。这种变化既反映了语音产生过程中的生理机制,也承载了丰富的语言学和副语言学信息^[12]。

在语音合成任务中,仅依靠固定或简单的规则设定时长难以生成自然、流畅且富有表现力的语音,因此需要构建能够精准刻画上述复杂变化的时长模型。理想的时长模型应依据文本上下文、句法结构、语义重点以及情感色彩等因素,准确预测各发音单元的时长。因此,系统性地分析影响韵律时长的各类因素,并基于此构建稳健高效的时长模型,对提升合成语音的自然流畅度与表现力具有重要意义^[13,14]。

2.1 韵律短语的平均时长和音节数

根据对500条朗读语句的韵律标注,共切分出3 725个韵律短语。其中,主韵律短语边界1 245条,次韵律短语边界2 480条,表明藏语句子内部存在较为明显的层级结构。

时长分析结果表明,主韵律短语的平均时长为2.29 s,次韵律短语平均时长为1.09 s。与汉语和蒙古语的对应单元相比,藏语韵律短语的时长介于两者之间,即显著长于汉语,而略短于蒙古语^[9,10,15]。

在音节数量方面,主韵律短语平均包含9.6个音节和4.1个韵律词;次韵律短语平均包含5.6个音节和2.4个韵律词。跨语言比较表明,藏语韵律短语的平均音节长度略高于汉语,但低于蒙古语^[9,10,15]。

进一步的结构分析发现,每个主韵律短语大约包含2~4个次韵律短语,揭示了藏语韵律组织在层次结构上的规律性。

2.2 韵律短语边界处的停顿时长

在连续语流中,停顿既满足说话人换气的生理需求,也与语言的结构组织密切相关。说话人通过停顿维持肺部气压,同时利用停顿表现语义单元之间的松紧关系,并起到消除歧义、突出重点、调节节奏以及实现含蓄委婉等语用功能,从而帮助聆听者更准确理解语义,并形成自然流畅的韵律感^[10]。

在韵律层面,停顿不仅是识别短语边界的重要声学线索,更是语义组织与信息结构的关键标记。通过停顿的有无、时长以及音高走向,听者能够感知意群的起止、层次的递进以及信息重点的突出,从而在连续语流中自然地划分出句法和语义单元^[15]。因此,准确捕捉并标注韵律短语边界的停顿信息,对实现自然、流畅的语音合成至关重要。

本研究统计结果显示,在所有语料的次韵律短语边界处,约有 60% 的边界末出现显著停顿;其余 35% 的停顿时长均低于 0.15 s,仅有约 5% 的停顿时长超过 0.15 s。针对主韵律短语,句中出现的 745 处边界(不含句末位置的 500 条),停顿时长低于 0.15 s 者占比仅约 8.7%;而停顿时长在 0.15 s 至 0.45 s 之间的边界则占约 91%。

基于上述统计结果,本研究将 0.15 s 设定为显著停顿的判定阈值:句中无声停顿若超过 0.15 s 的视为显著停顿,等于或低于 0.15 s 的则归为非显著停顿。本文进一步统计两类韵律短语边界处显著停顿出现情况,结果见表 1。

表 1 韵律短语边界处有无显著停顿的数据
Table 1 Number of the presence or absence of salient pauses at prosodic phrase boundaries

韵律短语类型	非显著停顿	显著停顿	合计
主韵律短语(条)	65	680	745
次韵律短语(条)	2 356	124	2 480

统计结果显示,停顿时长是区分主韵律短语和次韵律短语的重要声学线索。研究发现,两者在边界停顿方面存在显著差异:次韵律短语的边界通常无停顿,或停顿时长不超过 0.15 s;主韵律短语的边界普遍伴随显著停顿,时长往往超过 0.15 s。

以图 1 中的句子“**ཕུན་ཚོགས་ཀྱི་ཡུལ་བར་གྱིས་སྒང་ནི་ཡུལ་ལྗོངས་མངོན་སུག་ལྷན་པ་དང་དམངས་སྲུང་ཆོས་ལུན་པའི་ཡུལ་ལྔ་ཞིག་ཡིན།**”(平措的家乡扎西岗是一个风景秀丽、民风淳朴的好地方)为例,其中共标注出 5 个韵律短语:“**ཕུན་ཚོགས་ཀྱི་ཡུལ་**”“**བར་གྱིས་སྒང་ནི་**”“**ཡུལ་ལྗོངས་མངོན་སུག་ལྷན་པ་དང་**”“**དམངས་སྲུང་ཆོས་ལུན་པའི་**”和“**ཡུལ་ལྔ་ཞིག་ཡིན།**”。各韵律短语边界处的语音停顿特征如下:在第 1 个边界停顿约 0.1 s,在第 2、第 3 个边界分别出现显著停顿,时长分别约 0.35 s 和 0.22 s;在第 4 个边界无语音停顿;第 5 个边界则位于句末。因此,将第 1 和第 4 个韵律短语的末端边界标记为“3”,将第 2 和第 3 个边界标记为“4”,第 5 个边界标记为“5”。通过上述标记,将原有的 5 个韵律短语合并为 3 个主韵律短语。

以图 2 中的句子“**ཁོང་ཚོར་བྱ་སྤྱད་རྩོམ་འདི་ཐོག་ནས་དུས་རབས་གསར་པའི་ལུང་ལྷོད་གི་སྤྱད་བཟང་དར་སྤེལ་གྱི་དགོས་པའི་སྒྲུལ་ཕྱག་གནང་བ།**”(鼓励他们以实际行动弘扬新时代的雷锋精神)为例,其中共标注出 6 个韵律短语:“**ཁོང་ཚོར་**”“**བྱ་སྤྱད་**”“**རྩོམ་འདི་ཐོག་ནས་**”“**དུས་རབས་གསར་པའི་**”“**ལུང་ལྷོད་གི་སྤྱད་བཟང་**”“**དར་སྤེལ་གྱི་དགོས་པའི་**”和“**སྒྲུལ་ཕྱག་གནང་བ།**”。各韵律短语边界处的语音停顿特征如下:第 1 个边界停顿约 0.11 s,第 2 个和 5 个边界分别出现显著停顿,时长分别约 0.21 s 和 0.24 s;第 3 个和第 4 个边界无语音停顿;第 6 个边界位于句末。因此,将第 1、第 3 和第 4 韵律短语的末端边界标记为“3”,将第 2 和第 5 边界标记为“4”,将第 6 个边界标记为“5”。通过上述标记,将原有的 6 个韵律短语合并为 3 个主韵律短语。

3 韵律短语的音高特征

在语言学中,音高本质上是一个相对概念,其关注点并非个体噪音的绝对高低,而是语音在连续语流中呈现出的相对起伏、变化及其整体模式。这种音高模式是构成语言韵律(亦称超音段特征)的核心要素之一^[15,16]。

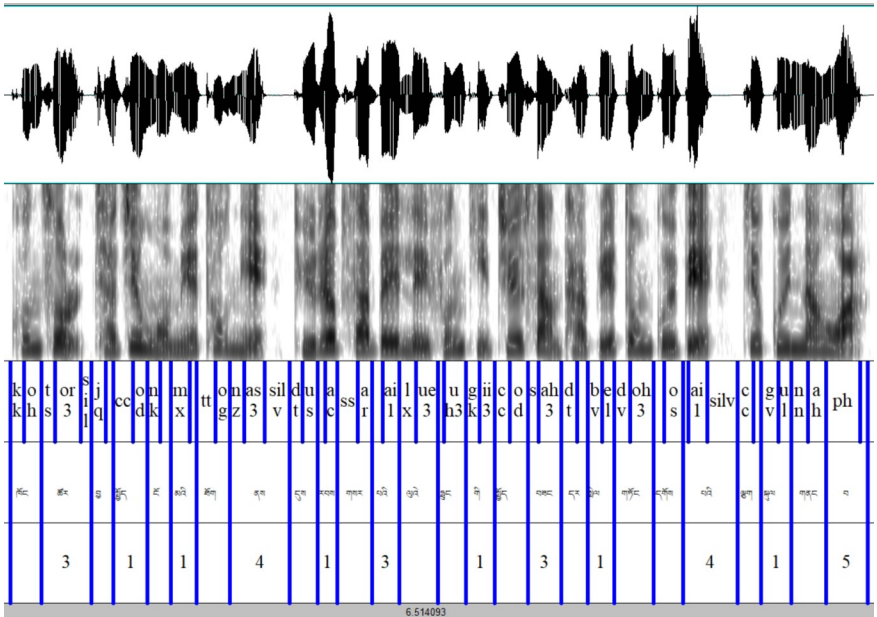


图2 “0312”句的韵律标注
Fig.2 Prosodic annotation of “0312” sentence

音高重置是指在较长的语音单元(如句子或语调短语)中,当音高下降到一定程度时,说话人有意识地将音高“重置”至较高水平,以开启新一轮的音高变化区间,从而确保后续表达的清晰度^[10]。这一现象通常出现在语法或语义单元的边界处,向聆听者传递“一个意群结束,另一个意群开始”的信息。若缺乏音高重置,整句话的音高可能会持续下降,使语音听感显得单调、不自然,甚至妨碍聆听者对句子的内部结构有效辨识^[17]。

在非声调语言(如英语)中,音高重置使语调自然流畅、富有表现力^[18];在声调语言(如汉语)中,音高重置则确保连续语流中的声调得以正确实现^[17]。而藏语缺乏词汇层面的声调,其语流中的音高起伏可以较为直接地反映句子的韵律组织结构。基于此,下文通过具体实例,进一步分析藏语音高模式与韵律短语之间的对应关系。

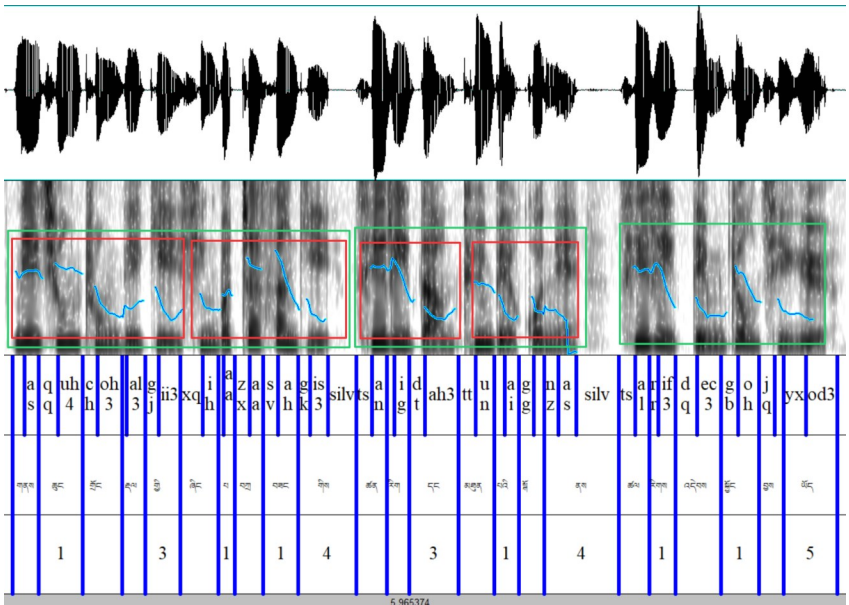


图3 “0435”句的韵律标注
Fig.3 Prosodic annotation of “0435” sentence

以图3中的句子“གནས་ཚུང་གི་ཐུང་ལྷོ་ཞིང་པ་བཟ་བཟང་གིས་ཚན་རིག་དང་མཐུན་པའི་སློན་ཆལ་རིགས་འདེབས་སྒྲིང་བྱས་ཡོད།”(乃琼镇的农民扎桑科学种植蔬菜)为例,该句音高曲线整体呈先上升后下降的变化趋势,峰值出现在重音附近,随后逐渐回落直至整个音高曲拱结束。研究发现,藏语句子在朗读过程中,每个音高曲拱大多与韵律短语相对应。基于此特征,该句可划分为5个韵律短语:“གནས་ཚུང་གི་ཐུང་ལྷོ་ཞིང་པ་བཟ་བཟང་གིས་”“ཚན་རིག་དང་”“མཐུན་པའི་སློན་ཆལ་”“རིགས་འདེབས་སྒྲིང་བྱས་ཡོད།”。同时,该句揭示了主韵律短语内部的音高下倾特征。如图3所示,该句第2和第4韵律短语边界处存在明显的语音停顿,实测时长分别约0.18 s和0.23 s;第1和第3韵律短语边界则未见显著停顿;第5韵律短语位于句末。依据边界停顿时长的差异,可判定第2、第4韵律短语末端对应更高层级的韵律边界。为此,本研究将第1、第3韵律短语的边界标记为层级“3”,第2、第4韵律短语边界标记为层级“4”,从而可将全句划分为3个主韵律短语。其中,前两个主韵律短语均包含2个次韵律短语。进一步观察以上4个次韵律短语的音高走势(图3中红框标示部分)可以发现,同一主韵律短语内部的次单元之间,音高从左至右逐步递减,呈现出明显的音高下倾模式。

韵律短语的边界类型与其所在位置的音高变化密切相关。研究表明,随着右侧边界韵律层级的升高,边界末端基频呈逐步下降趋势,形成明显的音高递减特征,这一声学模式被视为韵律层级结构的直接反映。该音高递减现象具有重要的功能意义:在区分主韵律短语与次韵律短语时,边界处的音高变化起到了关键识别作用。由于不同层级边界在声学上表现出系统性的音高差异,音高变化因而成为识别两类韵律短语的有效依据,具体机制表现为:在同一主韵律短语内部,各次韵律短语的音高曲线在趋近边界时呈递减趋势,末端基频逐级降低,形成自然的音高坡度;而在相邻主韵律短语之间,边界处则通常出现音高“重置”现象,即基频在交界点骤然回升至较高水平,之后才进入下一阶段的下降过程。这种“递减”与“重置”的交替出现,既清晰反映了韵律结构的层次性,也为语音合成和韵律分析提供了重要的声学线索。

结语

本研究基于500句藏语语料,对韵律短语的时长、停顿和音高特征进行了系统分析。通过实例观察与统计检验发现:藏语韵律短语可被划分为主韵律短语和次韵律短语两个层级;主韵律短语的末端边界普遍伴随显著的语音停顿,次韵律短语的末端边界则多为无停顿或仅出现极短的停顿;在同一主韵律短语内部,各次韵律短语的音高呈现左高右低的递降趋势,表现出显著的音高下倾的特征;不同主韵律短语之间的交界处则常出现明显的音高重置现象。综上,停顿和音高变化是区分藏语主韵律短语和次韵律短语的重要声学特征。

由于目前本研究尚处于探索阶段,存在一定的局限性:其一,语料规模有限,用于核心声学特征筛选的语料仅包含500句藏语语料,数量偏少;其二,发音人数量不足,仅选取一名女性播音员,其发音带有一定的个体特征,从而对研究结论的普适性产生影响。未来研究将从以下两方面拓展:一是扩大实验发音人规模,涵盖不同性别、年龄及方言背景的发音人,以考察不同个体和方言背景下韵律特征的共性与差异;二是进一步扩充有效分析语料,在更大规模、多样化的语料基础上验证和完善藏语韵律短语的划分规律,提供更具普适性的韵律切分依据。

参考文献

- [1] 拉巴顿珠,官政先,德庆卓玛,等. 完全端到端的藏语语音合成方法[J]. 中文信息学报,2024,38(9):82-92+116

- [2] Zhang W Z, Yang H W. Improving sequence-to-sequence Tibetan speech synthesis With prosodic information[J]. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2023, 22(9): 1–13.
- [3] 罗敏. 藏语安多方言端到端语音合成方法研究[D]. 兰州: 西北师范大学, 2023.
- [4] 陆晨, 祖漪清, 刘晨宁. 藏语拉萨话韵律词库——基于语音合成的实验研究[P/OL]. 中国科学院科技论文预发布平台. (2025-08-07) [2026-03-09]. <https://chinaxiv.org/abs/202508.00194>.
- [5] 倪崇嘉, 刘文举, 徐波. 汉语韵律短语的时长与音高研究[J]. *中文信息学报*, 2009, 23(4): 82–87.
- [6] 敖敏, 熊子瑜, 白音门德. 蒙古语韵律短语的分类研究[J]. *民族语文*, 2014(1): 76–81.
- [7] 曲珍, 春燕. 藏语语音合成中语料数据标注规则的设计[J]. *西藏大学学报(自然科学版)*, 2012, 27(1): 62–66.
- [8] Barrientos F, Bajuk L, Gökkaya M. The role of the Prosodic Hierarchy on learning phonological rules[J]. *Linguistic Approaches to Bilingualism*, 2026, 16(2): 233–250.
- [9] 敖敏. 基于韵律的蒙古语语音合成研究[D]. 呼和浩特: 内蒙古大学, 2012.
- [10] 殷治纲. 再论韵律边界的声学特征及其形成机制[J]. *中国语音学报*, 2020(1): 38–50.
- [11] 邓戈. 藏语语音研究[M]. 拉萨: 西藏藏文古籍出版社, 2013.
- [12] 马宁, 于洪志, 李永宏, 等. 藏语韵律词和韵律短语的时长特征研究[J]. *西北民族大学学报(自然科学版)*, 2014, 35(3): 14–17.
- [13] 艾金勇. 藏语韵律单元的时长与音高研究[J]. *西藏民族学院学报(哲学社会科学版)*, 2011, 32(2): 53–56.
- [14] Zhang W Z, Yang H W. Meta-learning for Mandarin-Tibetan Cross-Lingual speech synthesis[J]. *Applied Sciences*, 2022, 12(23): 12185.
- [15] 包通拉嘎. 蒙古国新闻播音连续话语韵律特征的声学研究[D]. 呼和浩特: 内蒙古大学, 2025.
- [16] 黄贤军, 郑海洋, 吕士楠, 等. 韵律短语边界对降阶和焦点后音高骤降的影响[J]. *声学学报*, 2016, 41(4): 529–536.
- [17] 冯胜利. 汉语的韵律、句法与词法[M]. 北京: 北京大学出版社, 2009.
- [18] 马丽娜·内斯波, 艾琳·沃格尔. 韵律音系学[M]. 宫齐, 译. 上海: 商务印书馆, 2021.

A Corpus-Based Study on the Acoustic Characteristics of Prosodic Phrases in Tibetan

Quzhen^{1,2} Dechen-Droma^{1,3} Lhakpa-Dondrub^{*1,3}

(1. School of Information Science and Technology, Xizang University, Lhasa 850000, China;

2. Tibetan Information Technology Innovative Talent Cultivation Demonstration Base, Xizang University, Lhasa 850000, China;

3. Provincial and Ministerial Collaborative Innovation Center for Informatization in Xizang, Lhasa 850000, China)

Abstract: Analyzing and modeling the prosodic structure of sentences and passages are essential for making speech synthesis sound more natural and for reducing speech recognition errors. Based on prosodic annotation corpora, this paper systematically studies the duration and pitch characteristics of prosodic phrases in the Tibetan language and thoroughly explores their classification. The study shows significant systematic differences in the duration of boundary pauses and pitch progression between primary prosodic phrases and secondary prosodic phrases during Tibetan reading. The presence or absence of a significant pause at the boundary serves as the core criterion for distinguishing between the two types of phrases. Additionally, pitch reset often occurs at the end of a phrase, while the drop in pitch at the end of a main prosodic phrase is especially noticeable. These findings not only deepen our understanding of the Tibetan prosodic hierarchy, but also provide valuable acoustic feature references for technologies such as speech synthesis and speech recognition.

Keywords: Tibetan, primary prosodic phrase, secondary prosodic phrase, duration of pause, pitch descent, pitch reset

[责任编辑: 索郎桑姆]