Abstract

Temporal Articulatory Stability, Phonological Variation, and Lexical Contrast Preservation in Diaspora Tibetan

Christopher Alden Geissler

2021

This dissertation examines how lexical tone can be represented with articulatory gestures, and the ways a gestural perspective can inform synchronic and diachronic analysis of the phonology and phonetics of a language. Tibetan is chosen as an example of a language with interacting laryngeal and tonal phonology, a history of tonogenesis and dialect diversification, and recent contact-induced realignment of the tonal and consonantal systems. Despite variation in voice onset time (VOT) and presence/absence of the lexical tone contrast, speakers retain a consistent relative timing of consonant and vowel gestures.

Recent research has attempted to integrate tone into the framework of Articulatory Phonology through the addition of tone gestures. Unlike other theories of phonetics-phonology, Articulatory Phonology uniquely incorporates relative timing as a key parameter. This allows the system to represent contrasts instantiated not just in the presence or absence of gestures, but also in how gestures are timed with each other. Building on the different predictions of various timing relations, along with the historical developments in the language, hypotheses are generated and tested with acoustic and articulatory experiments.

Following an overview of relevant theory, the second chapter surveys past literature on the history of sound change and present phonological diversity of Tibetic dialects. Whereas Old Tibetan lacked lexical tone, contrasted voiced and voiceless obstruents, and exhibited complex clusters, a series of overlapping sound changes have led to some modern varieties that have tone, lack clusters, and vary in the expression of voicing and aspiration. Furthermore, speakers in the Tibetan diaspora use a variety that has grown out of the contact between diverse Tibetic dialects. The state of the language and the dynamics of diaspora have created a situation ripe for sound change, including the recombination of elements from different dialects and, potentially, the loss of tone contrasts.

The nature of the diaspora Tibetan is investigated through an acoustic corpus study. Recordings made in Kathmandu, Nepal, are being transcribed and forced-aligned into a useful audio corpus. Speakers in the corpus come from diverse backgrounds across and outside traditional Tibetan-speaking regions, but the analysis presented here focuses on speakers who grew up in diaspora, with a mixed input of Standard Tibetan (*spyi skad*) and other Tibetan varieties. Especially notable among these speakers is the high variability of voice onset time (VOT) and its interaction with tone. An analysis of this data in terms of the relative timing of oral, laryngeal, and tone gestures leads to the generation of hypotheses for testing using articulatory data.

The articulatory study is conducted using electromagnetic articulography (EMA), and six Tibetan-speaking participants. The key finding is that the relative timing of consonant and vowel gestures is consistent across phonological categories and across speakers who do and do not contrast tone. This result leads to the conclusion that the relative timing of speech gestures is conserved and acquired independently. Speakers acquire and generalize a limited inventory of timing patterns, and can use timing patterns even when the conditioning environment for the development of those patterns, namely tone, has been lost.

Temporal Articulatory Stability, Phonological Variation, and Lexical Contrast Preservation in Diaspora Tibetan

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Christopher Alden Geissler

Dissertation Director: Jason Anthony Shaw

June 2021

Table of Contents

Figures and tables

Acknowledgement

It takes a village to write a dissertation.

Jason Shaw was not yet at Yale when I arrived, and he only took me on after the departure of my previous advisor. We met in his office, we talked for three hours, and by the end of that meeting we had a plan. The five chapters of this dissertation follow directly from that plan. Jason treated me as an apprentice, mentoring me in laboratory procedure, teaching, thinking through problems, and several genres of writing. Most of all, I am grateful to him for patiently and unwaveringly sticking by me through all the times when I was not the easiest person to advise. I could not have hoped for a better advisor.

A number of other faculty mentors have contributed enormously to my graduate career. Ryan Bennett modeled much of what I aspire to in scholarship, and Claire Bowern gave me a lot of practical help but also made me feel that someone always had my back. More recently, teaching with Raffaella Zanuttini and Maria Mercedes Piñango was a most joyful and reinvigorating experience. Among my committee members, Natalie Weber helped me put my dissertation and writing process into perspective; Mark Tiede brought methodological wisdom and gently righted my course more than once; Fang Hu graciously helped a stranger on the opposite side of the world with unique insight and thoughtful advice; and Lisa Zsiga inspired confidence in the value of this work.

That community also includes my fellow graduate students. Luke Lindemann welcomed me with open arms on my first visit to Yale and joined me for a most enjoyable and productive (see Chapter 3) summer in Kathmandu. E-Ching Ng, Sara Sanchez-Alonso, Dolly Goldenberg, and Rikker Dockum were like the cool older siblings I looked up to and tried to emulate. Sammy Andersson, Sarah Babinski, Marisha Evans, Martín Fuchs, Sophie Hao, Vivian Guo Li, Josh

Phillips, Sirrý Sigurðardóttir, Matt Tyler, Andy Zhang, and all the rest—it was a wild ride, and I can't begin to list all the ways I have leaned on you over the years.

Perhaps the most joyful part of graduate school has been participating in the broader scholarly community—faculty and graduate students, my "conference buddies." There are so many, and I don't think they have any idea how important they are to me. The same goes for my students.

Fieldwork in Nepal would not have been possible without the dedicated effort of interviewers Tenzin Norbu and Sonam Bhuti, or the advice and connections from Nawang Tsering and Dorje Tsering. I am also grateful to all my Tibetan language teachers: Tashi Tsering, Tseten Chonjore, Tenzin Tinley, Dekyi Lhamo, Lama Tsondru Sangpo, and Sonam Tsering. The work itself was possible thanks to the generosity of all the speakers who lent their voice to recordings, and their lips and tongues for EMA experiments.

Of course, there is no way I would have gotten here without so many others—most notably my parents, Ann and Bill. They distinctly felt the limits on how they could support me, and so did everything they could. I also distinctly felt the presence of those important people who were no longer around to celebrate this accomplishment—my grandmother Madeleine Sierakowski, great-aunt Roberta Geissler, and teachers William McCrystal and Donald Thieberger.

The final completion of this dissertation is largely thanks to my writing buddies, including Catarina Soares, Suzanne McFate, Rashad Ullah, Emily Kluver, and Luke Lindemann. I may not have finished without these companions, and I thank Willa Miller and Christine Lidz for the conversations in which this idea originated. Other communities also supported me in ways I cannot begin to name—notably New Haven Sacred Harp, the Natural Dharma Fellowship, and the Branford College Pottery Studio.

Dedication

To my parents

To my advisor

To my colleagues

To my participants and collaborators

སེམས་ཅན་ཐམས་ཅད་བདེ་ཞིང་སྐྱིད་པར་ཤོག།།

Note on nomenclature and conventions

In this dissertation, I use “Standard Tibetan” to refer to the variety of Tibetan that serves as a *lingua franca* across Tibetan-speaking regions. It is similar, but not identical, to the forms of Tibetan spoken in the main cities of the U-Tsang region such as Lhasa, Shigatse, and Gyantse. This similarity and the prominence of Lhasa have led some to call this “Lhasa Tibetan,” although that term more narrowly refers to the variety unique to Lhasa city itself. In Tibetan, this “Standard Tibetan” is known as *spyi skad*, which might be more accurately translated as “Common Tibetan,” the term used by Caplow (2017). However, I use “Standard Tibetan” because this term is more widely recognized (e.g. Tournadre and Dorje 2003) and to avoid confusion with the use of “Common Tibetan” to refer to the reconstructed ancestor of modern Tibetan dialects (e.g. Hill 2010). Likewise, I use “Diaspora Tibetan” to refer to Standard Tibetan as spoken in the Tibetan diaspora communities in India, Nepal, and around the world.

Orthographic forms of Written Tibetan are presented in *italics*. Written Tibetan is generally treated as similar to Old Tibetan, but see sections 2.3, 2.4, and 2.9 for discussion on this relationship. In transcribing Tibetan orthography, I use the Wylie system[1] with the following variations. I omit capitalization of the “head letter” for proper names, and I demarcate syllables within a word using a period <.>, reserving the space for demarcating word boundaries. I also use <ḫ> for the letter <འ་> rather than the Wylie <‘>, following its recent use by Hill (2019). This choice is meant to clarify for readers unfamiliar with Tibetan that this likely refers to a post-velar fricative, while remaining agnostic about its specific phonetic realization. In other cases, I use IPA transcription throughout,

[1] Wylie, Turrell V. 1959. A standard system of Tibetan transcription. *Harvard Journal of Asiatic Studies* 22(261–267).

in accordance with general practice across phonetics and phonology. Tone is marked as [V́] for high-level tone, and low [V̀] or rising [V̌] for low tone in disyllabic and monosyllabic words, respectively. When forms are cited from a source that uses Chao numbers, they are reported as in the source text (a scale of 1-5, low to high, e.g. 55 for a high level tone and 13 for a low-rising tone).

1 Introduction

1.1 General introduction

This dissertation explores the ways in which tone conditions the timing of consonants and vowels in Tibetan as spoken in diaspora. The overall theme that emerges is that inter-gestural timing is consistent despite variation in other phonetic parameters within and across speakers. Onset voicing and aspiration is variable (with low tone), while speakers differ in their laryngeal stop contrasts and even in whether or not they produce a tone contrast. Nevertheless, all speakers exhibit similar relative timing of consonant and vowel gestures, indicating temporal uniformity for speakers and for the language as a whole.

An improved understanding of inter-gestural timing is an empirical contribution with a number of implications. As a description of Tibetan, this work begins to document the diversity of Tibetan speakers in diaspora, including the novel finding that some speakers do not produce a tone contrast. The present state of the language adds another step to the ongoing history of tonal and laryngeal contrast realignment in Tibetan, and also contributes to our understanding of how language change can occur in diasporic, internally-diverse communities. For phonological theory, gestural timing represents an important part of language-specific knowledge. The findings presented in this dissertation challenge the competitive-coupling model of tone (Gao 2008 et seq.). It also extends the idea of uniformity in production (Chodroff 2017) to the temporal domain, and suggests that that speakers may seek to match the timing of those around them.

This dissertation makes use of Articulatory Phonology (Browman & Goldstein 1986 et seq) in order to make empirical phonetic predictions based on phonological facts. Rather than discrete, linearly-ordered segments, Articulatory Phonology decomposes speech into gestures, controlled movements of the vocal tract that unfold over time. Since gestures can overlap in time, any gesture may be realized in different ways depending on the overlapping gestural context. In particular, Articulatory Phonology allows for contrasts in supralaryngeal gestures (as in [b] vs. [d], [b] vs. [m]), laryngeal gestures ([b] vs. [p]) or a combination of the two. Unlike other approaches to phonology, Articulatory Phonology includes relative timing in the lexical representation, which allows for predictions to be made about how intergestural timing might differ in different environments. Since gestures are not locked into a linear order, a mechanism is needed to explain their temporal coordination. Coupling relations among gestures describe a range of timing patterns (Nam & Saltzman 2003, Nam et al. 2009), and make empirical predictions—including how timing between two gestures is affected by the presence or absence of a third gesture.

In particular, the gestural timing under investigation is that of oral and laryngeal gestures as influenced by tone. Following Gao (2008) and subsequent work (Karlin 2014, 2018; Zhang et al 2019; Zsiga 2020), the nature of the tonal gestures and their relationship to oral gestures can be observed through the effect of tone on the relative timing among oral gestures. Evidence comes from acoustic and articulatory studies, but also from an analysis of changing phonological contrasts in historical time. Tibetan provides a useful source of timing data because of its interacting tone and laryngeal contrasts, the phonological diversity of its dialects, and its documented history showcasing tonogenesis, dramatic consonant cluster simplification, and laryngeal reanalysis.

Studying Tibetan also affords the chance to expand the typology of tonal languages investigated using instrumental observation of articulation.

1.2 Target Uniformity

Features and gestures serve as units of contrast, and tend to recur across different segements in a language. For example, Maddieson (1996) found that doubly-articulated consonants in Ewe use very similar articulatory movements as singly-articulated consonants. This makes sense if a finite number of gestures are deployed in a variety of configurations, an idea that was termed "gestural economy." This is similar to "feature economy," the related idea that a phonological inventory tends to use a minimal number of features for a maximal number of phonemes. However, Clements (2003) distinguishes between the two: gestural economy predicts the recurrence of specific gestures (e.g. bilabial closures), while feature economy predicts broader classes (e.g. any articulation involving the lips for [LABIAL]). Finding that such clustering of phonemes at the level of features does occur, Clements concludes that gesture economy is not adequate to explain these cases, and predicts gesture economy to be active for non-contrastive properties rather than phonemic ones.

However, contrastive properties may still be subject to varying degrees of "uniformity," the consistency in phonetic detail across productions for a given speaker. Keating (2003) found that English speakers varied in the uniformity of their VOT production across contexts. Some speakers exhibit variation in acoustic and/or articulatory output that reflects articulatory ease, while other speakers maintain uniform acoustic and/or articulatory output even when doing so requires more effortful articulation. The relative importance of uniformity can differ across speakers, and the uniformity is more fine-grained than the presence or absence of a feature or gesture.

Chodroff (2017) expanded on the notion of uniformity by investigating the structured covariation of phonetic parameters in American English and explaining this data as the result of three types of uniformity constraints. The first, “pattern uniformity”, enforces consistency across speakers: it requires that the distances between phonetic targets be equivalent for different speakers. The second, contrast uniformity, requires that speech sounds that instantiate a phonological contrast be produced with phonetic differences comparable to other sounds instantiating that contrast. Finally, “target uniformity” requires similar phonetic realization of a distinctive feature value.

The explanatory value of uniformity is shown by the differences predicted by the different kinds of uniformity. Chodroff (2017) found greater support for target uniformity than contrast uniformity in the covariation of VOT in American English stops, and in the mid-frequency peak of Czech and American English fricatives. Target uniformity predicts correlations among stops sharing a laryngeal feature (i.e. /p t k/ vs. /b d g/) and among fricatives sharing a value of [anterior] (i.e. /s z/ and /ʃ ʒ/). Contrast uniformity predicts further correlation between sounds sharing a different feature: place of articulation for the stops, and voicing for the fricatives, as well as similarities in the magnitude of differences across speakers. In all cases, the predictions of target uniformity were borne out but predictions of contrast uniformity were not.

Another example of target uniformity is provided by the structured variation in Suzhou Chinese fricative vowels (Faytak 2018, 2020). Suzhou Chinese has fricative ([$\text{i}_{\text{ʑ}}$],[$\text{y}_{\text{ʑ}}$]) and apical vowels ([ɿ],[ʮ]) that contrast in rounding but whose place of articulation is determined by the preceding consonant. The apical vowels [ɿ ʮ] occur after apical fricatives and affricates /s t͡s t͡sʰ/, while the fricative vowels [$i_ʑ$ $y_ʑ$] occur after palatal fricatives and affricates /ɕ t͡ɕ t͡ɕʰ/. Faytak (2018) found that tongue shape remained constant through these consonant-vowel sequences. Unsurprisingly, the center of gravity of the unrounded fricative-vowel pairs covaried. However, the center of gravity of the rounded pairs differed: rounding lowered the center of gravity of the vowel by an unpredictable amount. These results show that speakers can retain consistent articulation even with variation in the acoustic output.

Why should a speaker maintain a consistent articulatory posture at the cost of permitting greater variability in the acoustics to arise? As an alternative, Suzhou speakers could adjust their articulation to maximize acoustic consistency (i.e. center of gravity) across matched rounded-unrounded vowel pairs [ɿ $i_ʑ$] and [ʮ $y_ʑ$]. Prioritizing acoustic consistency in each pair would reflect the featural contrast in the acoustics, which would be consistent with pattern uniformity and appear to aid recoverability. Instead, the articulation remains constant at the cost of acoustic variability, which is consistent with target uniformity. In this way, target uniformity can be seen as a constraint favoring articulatory similarity as opposed to listener-oriented perceptual factors. This dissertation builds on target uniformity by extending it into the temporal domain in order to account for observed consistencies in relative timing of articulators.

1.3 Tibetan

The Tibetan language offers a promising set of characteristics for investigating the temporal relationships between laryngeal, supralaryngeal, and tonal gestures. Some Tibetan varieties lack consonant clusters and contrast lexical tones; others permit large clusters but do not use tone. Onset consonants can contrast for varying combinations of voicing and aspiration, and speakers of all varieties coexist and interact in diverse diasporic speech communities. This diversity permits comparison across differences in the co-occurence of these various traits. Such comparison may take place across languages, across speakers of a language, and across contexts for a given speaker. For across-language comparison, Tibetan adds to the limited inventory of tonal languages for which measurements of gestural timing have been reported. It also invites typological study across the diversity of Tibetan varieties and their diachronic phonology. Importantly, this includes both varieties that have consonant clusters but no tone, and varieties with tone but not clusters. This confluence of diachronic, synchronic, and sociolinguistic factors have further created conditions for the rarely-studied tone loss.

Across-speaker comparison is used in this dissertation to investigate which characteristics are specific to individual speakers, and which are shared across the speech community. Not only is the there substantial variation across dialect regions, but speakers of all those dialects have come into close contact in the post-1959 diaspora. As a result, two generations of speakers raised in diaspora have acquired Tibetan while exposed to speakers of many varieties of the language. There is a high degree of variation among speakers in the same speech community, a fact highlighted in Chapter 4. While Mandarin has allowed study of tonal and toneless environments for a given speaker (Zhang et al 2019), in Tibetan there are also tone-contrasting and non-contrasting speakers of the same speech community. Finally, different contexts for a given speaker are also investigated, particularly for the effect of tone on VOT (Chapter 3) and C-V timing (Chapter 4).

1.4 Theoretical motivation: Articulatory Phonology

Rather than segments or features, the primary representational unit of Articulatory Phonology is the gesture, a controlled, dynamic movement of the vocal tract (Browman & Goldstein 1986, 1988). Gestures are specified for a constriction location and degree (analogous to place and manner of articulation), and unfold over the course of a period of time. Since gestures can overlap in time to various degrees (Öhman 1966, Fowler 1983), the arrangement of gestures in time is depicted in a gestural score. A separate module, implemented as Task Dynamics (Saltzman & Kelso 1987, Saltzman & Byrd 2000), models how gestures become physical trajectories of articulators. This is accomplished by modeling articulatory movements as critically damped mass-spring systems, following work on kinematics of limb movement (Haken et al. 1985).

A key question in Articulatory Phonology is how to determine the temporal arrangement of multi-gesture productions—in other words, where does a gestural score come from? Building on the expression of relative timing in terms of phase (Browman & Goldstein 1988), the coupled oscillator model (Saltzman & Byrd 2000, Nam & Saltzman 2003, Nam et al. 2009), treats each gesture as a separate oscillator with a particular relationship to another gesture expressed in terms of phase. Planning oscillators can settle at any value from 0° to 360°; the number of degrees reflects the point in the first gesture that the second gesture begins. In other words, phasing relations quantify the relative timing of the starts of the two gestures. However, on analogy with other movements such as bimanual tapping (Turvey 1990, Haken et al. 1996), in-phase (0°, synchronous) and anti-phase (180°, sequential) coupling are most likely; any other phasing is called “eccentric coupling” (Goldstein 2011). When more than two gestures are coupled in a configuration that would place competing demands on the kinematics (“competitive coupling”), the phasing relationships conflict and an intermediate phasing results. Articulatory Phonology thus allows for the relative timing of gestures to be explained with reference either to the number and type of coupling relations or to the strength of those coupling relations. In the case of competitive coupling, modulating the coupling strength can produce a continuous range of phasing values, including those that would approximate in-phase and anti-phase coupling.

An important use of competitive coupling has been to account for the “C-center effect”. As described by Browman & Goldstein (1988), the C-center refers to the global timing of complex onset gestures to a vowel gesture. Onset consonant gestures partially overlap with each other and, taken as a whole, are timed to the vowel gesture in a similar way to how a singleton onset is timed to a vowel gesture (see Fig. 1.1(a,c), below). Mücke et al (2020) observe that previous implementations of competitive coupling have assumed that the gestures are coupled with equal strength, but this need not be the case. While the explanatory value of coupling relations merits further attention, coupling strength will not play a role in the analyses in this dissertation. As will be shown, the coupling diagrams under consideration either consist of only two gestures, or of comparisons between a two-gesture representation and a three-gesture representation. In either case, the crucial comparisons involve different sets of coupling relations rather than the strengths of those relations. Fig. 1.1 schematizes the relationship between different coupling relations and gestural timing.

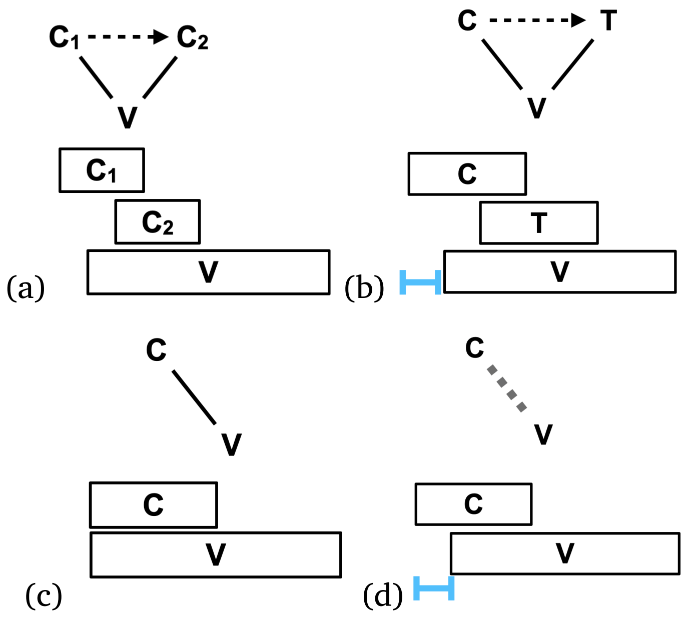


Figure 1.1. Coupling graphs and gestural scores. (a) C-center timing achieved through competitive coupling; (b) C-center-like timing in a CV syllable with tone; (c) In-phase C-V timing; (d) C-center-like timing achieved through eccentric coupling.

The gestural scores in Fig. 1.1 illustrate that the same observed timing can result from different sets of coupling relations. C-center timing results from competitive coupling is shown in Fig. 1.1(a), and Fig. 1.1(b) illustrates this for a CV syllable with tone, as in Gao (2008). While a C-V syllable would be expected to have in-phase coupling as in Fig. 1.1(c), eccentric coupling could yield different timing. Note that the C-V timing is identical in Figs. 1.1(b) and 1.1(d): this shows that eccentric coupling can produce the same C-V timing as would be predicted by the competitive-coupling model of tone. More detail on C-V timing is presented in section 4.1.1, as part of the motivation for the EMA study.

Given the range of factors that could affect gestural timing, particularly coupling relations and coupling strength, it is necessary to determine how to uniquely identify coupling relations. Mücke et al. (2020) do not seek to explain interspeaker variation, but they propose coupling strength as a locus of variation among individuals with a shared phonology. Since coupling relations are phonological in Articulatory Phonology, speakers with similar phonological systems but some differences in timing may share coupling relations but differ in coupling strength. Empirical evidence for the type of coupling relation as opposed to coupling strength can come from covariation between temporal measurements. For example, Shaw et al. (2019) used covariation between gesture duration and relative timing to distinguish between clusters and complex segments. A similar approach is adopted in this dissertation (Chapter 4) to investigate the coupling relations between gestures.

1.5 Dissertation roadmap

In the following chapters, three studies present converging evidence on the relationship between tone and gestural timing in Tibetan. Chapter 2: Diachrony traces the history of Tibetan, including the emergence and evolution of the laryngeal and tonal contrasts of the modern language dialects. It is shown how some dialects underwent a dramatic restructuring in syllable shape and phonological inventory, setting the scene for the range of inventories seen among speakers today. An acoustic corpus study is presented in Chapter 3, which establishes that speakers vary in whether or not they produce a tone contrast, and that word-initial VOT is sensitive to a word's phonological tone category. An EMA study presented in Chapter 4 focuses on articulatory timing itself, presenting evidence for consistent phonetic timing across speakers with different laryngeal and tonal contrasts.The implications of these findings for the place of articulatory timing in phonology are discussed in Chapter 5.

1.6 Chapter bibliography

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2019. The interaction of timing and scaling in a lexical tone system: an example from Silluk. In *Proceedings of the 19th International Congress of the Phonetic Sciences*. Melbourne, Australia.

Browman, Catherine P. & Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252. https://doi.org/10.1017/S0952675700000658.

Brunelle, Marc, Thành Tấn Tạ, James Kirby & Lư Giang Đinh. 2019. Obstruent Devoicing and Registrogenesis in Chru. In *Proceedings of the 19th International Congress of Phonetic Sciences*, 5. Melbourne.

Brunnelle, Marc & James Kirby. 2020. Relative cue weighting in the production and perception of register. Presentation at the 17th Annual Meeting of the Association for Laboratory Phonology. Vancouver.

Chodroff, Eleanor R. 2017. *Structured variation in obstruent production and perception*. Johns Hopkins University.

Chomsky, Noam & Morris Halle. 1968. The sound pattern of English. Harper & Row New York.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2(1). 225–252. https://doi.org/10.1017/S0952675700000440.

Clements, G. N. 2003. Feature economy in sound systems. *Phonology* 20(3). 287–333. https://doi.org/10.1017/S095267570400003X.

Clements, George N. & Rachid Ridouane. 2006. Quantal phonetics and distinctive features. In Antonis Botinis (ed.), *Proceedings of 1st Tutorial and Research Workshop on Experimental Linguistics*, 17–24. Athens, Greece: ExLing Society.

Ernestus, Mirjam. 2011. Gradience and categoricality in phonological theory. In Marc van Oostendorp (ed.), *The Blackwell companion to phonology*. Malden, MA: Wiley-Blackwell.

Faytak, Matthew Donald. 2018. Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese. University of California, Berkeley.

Faytak, Matthew. 2020. Articulatory, but not acoustic, target uniformity in Suzhou Chinese. Poster presentation presented at the 2020 Annual Meeting of the Linguistics Society of America, New Orleans.

Fowler, Carol A. 1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* 112(3). 386.

Gao, Man. 2008. Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics), Yale University*.

Gick, Bryan, Ian Wilson, Karsten Koch & Clare Cook. 2004. Language-Specific Articulatory Settings: Evidence from Inter-Utterance Rest Position. *Phonetica* 61(4). 220–233. https://doi.org/10.1159/000084159.

Goldstein, Louis. 2011. Back to the Past Tense in English. In Rodrigo Gutiérrez-Bravo, Line Mikkelsen & Eric Potsdam (eds.), 69–88. Santa Cruz, CA: Linguistics Research Center, UC-Santa Cruz Department of Linguistics.

Haken, H., C.E. Peper, P.J. Beek & A. Daffertshofer. 1996. A model for phase transitions in human hand movements during multifrequency tapping. *Physica D: Nonlinear Phenomena* 90(1–2). 179–196. https://doi.org/10.1016/0167-2789(95)00235-9.

Haken, Hermann, JA Scott Kelso & Heinz Bunz. 1985. A theoretical model of phase transitions in human hand movements. *Biological cybernetics* 51(5). 347–356.

Iskarous, Khalil. 2017. The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics* 64. 8–20. https://doi.org/10.1016/j.wocn.2017.05.003.

Karlin, Robin. 2018. Towards an articulatory model of tone: a cross-linguistic investigation. Cornell University Doctoral Dissertation.

Keating, Patricia. 2003. Phonetic and other influences on voicing contrasts. In *Proceedings of the 15th international congress of phonetic sciences*, 375–378.

Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language*. Linguistic Society of America 70(3). 419–454.

Maddieson, Ian. 1996. Gestural economy. *UCLA Working Papers in Phonetics* 94. 1–6.

Mücke, Doris, Anne Hermes & Sam Tilsen. 2020. Incongruencies between phonological theory and phonetic measurement. *Phonology* 37(1). 133–170. https://doi.org/10.1017/S0952675720000068.

Myers, S., S. Namyalo & A. Kiriggwajjo. 2019. F0 Timing and Tone Contrasts in Luganda. *Phonetica* 76(1). 55–81. https://doi.org/10.1159/000491073.

Nam, Hosung, Louis Goldstein & Elliot Saltzman. 2009. Self-organization of Syllable Structure: A Coupled Oscillator Model. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.), *Approaches to Phonological Complexity*. Berlin, New York: Walter de Gruyter. https://doi.org/10.1515/9783110223958. http://www.degruyter.com/view/books/9783110223958/9783110223958/9783110223958.xml (23 September, 2020).

Nam, Hosung & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of the Phonetic Sciences*.

Öhman, Sven EG. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39(1). 151–168.

Prince, Alan S. & Paul Smolensky. 2002. Optimality Theory: Constraint Interaction in Generative Grammar. No Publisher Supplied. https://doi.org/10.7282/T34M92MV. https://rucore.libraries.rutgers.edu/rutgers-lib/42030/ (13 December, 2020).

Remijsen, Bert & Otto Gwado Ayoker. 2014. Contrastive tonal alignment in falling contours in Shilluk. *Phonology* 31(3). 435–462. https://doi.org/10.1017/S0952675714000219.

Saltzman, Elliot & Dani Byrd. 2000. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19(4). 499–526.

Saltzman, Elliot & J. A. Kelso. 1987. Skilled actions: A task-dynamic approach. *Psychological review*. American Psychological Association 94(1). 84.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics*. LONDON: Elsevier Ltd 17(1–2). 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7.

Tạ, Thành Tấn, Marc Brunelle & Quý Nguyễn Trần. 2019. Chrau register and the transphonologization of voicing. In *Proceedings of the 19th International Congress of Phonetic Sciences*, 2094–2098. Melbourne.

Turvey, Michael T. 1990. Coordination. *American psychologist* 45(8). 938.

Zhang, Menghan, Shi Yan, Wuyun Pan & Li Jin. 2019. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569(7754). 112–115.

Zsiga, Elizabeth. 2020. Tonal Timing in Articulatory Phonology: Evidence from Igbo Vowel Reduction. Presented at the Association for Laboratory Phonology 17, Vancouver.

2 Diachrony

2.1 Chapter overview

This chapter situates the phonetics and phonology of Tibetan in the context of historical sound change and contemporary variation. The common ancestor of all Tibetan varieties, Old Tibetan, resembled other older Sino-Tibetan languages: it had consonant clusters, monosyllabic stems with some prefixes, and no lexical tone. Over time, some dialects preserved many of these features, while others reduced clusters, developed largely disyllabic words, and innovated lexical tone (section 2.1-2.2). From the perspective of Articulatory Phonology (see section 1.3), these changes in clusters, word shape, and tonality would have involved the gain and loss of articulatory gestures, leading to substantial changes in the relative timing of gestures. More recently, speakers of diverse dialects have come into close contact in the post-1959 Tibetan diaspora (2.5), where speakers are now raised with an input consisting of multiple dialects. Sections 2.6-2.8 focus on the phonological inventory of Common Tibetan (*spyi skad*), the variety spoken with variation in Central Tibet and dominant in diaspora. Laryngeal contrasts and their timing are discussed in section 2.9, setting up for the phonetic study of laryngeal and oral timing in chapters 3 and 4.

2.2 Sino-Tibetan

Tibetan is a Bodic language of the Bodish branch of the Sino-Tibetan family. The internal subgrouping of this family has presented a vexing problem for historical linguists, though the most common classification assigns the Sinitic varieties to one primary-level branch and the rest of the family, including Tibetan, to the other branch, called Tibeto-Burman (e.g. Shafer 1955, Benedict 1972, Bradley 1997, Matisoff 2003. Thurgood 2003). However, this classification is not universal, and alternative tree proposals are often signaled by alternative names for Sino-Tibetan. In the “Tibeto-Burman” proposal, van Driem (2002) places Sinitic as one of several branches rather than a highest-level division. Those who favor “Trans-Himalayan” may remain agnostic about the internal subgrouping of the family (Owen-Smith and Hill 2014, van Driem 2014), and rightly draw attention to the uncertain genetic relationships between subgroups. Indeed, Blench and Post (2014) argue that Sinitic, Bodish, and Burmish should be grouped as a single sub-clade. The latter have rightly drawn attention to the uncertain position of many sub-groups, particularly a number of less-studied languages from the eastern Himalayan region. As it remains the most widely-recognized term, this dissertation uses “Sino-Tibetan” to refer to the language family as a whole.

Hypotheses about the geographical origin of Sino-Tibetan fall into two main groups. According to the Southwestern Hypothesis, Proto-Sino-Tibetan would have been spoken in Sichuan province in southwest China (van Driem 2005), and/or the Himalayas (Matisoff 1991). This hypothesis is primarily supported by the fact that most Sino-Tibetan subgroups are found in this region; additionally, the expansion of Sino-Tibetan from this region can be linked with the spread of rice cultivation. By contrast, the Northern Hypothesis places the origin of the family in Yellow River valley of northern China (Thurgood 2003), and the expansion of Sino-Tibetan from this region has been linked with millet cultivation. Other hypotheses have placed Proto-Sino-Tibetan speakers in the Eastern Himalaya (Blench and Post 2014), radiating outward following varying modes of subsistence, or central China between the Yellow and Yangtze Rivers (Bellwood 2005).

Progress on understanding the internal classification of Sino-Tibetan has been made using both the traditional Comparative Method and phylogenetic tools adapted from evolutionary biology, in dialogue with archaeology and population genetics. These have included reconstruction of a number of subgroups within Sino-Tibetan, including Sinitic (e.g. Baxter & Sagart 2014), Tibetic (Hill 2011), and TGTM (Mauzaudon 1994). Recently, this work has been compiled and updated by Hill (2019) for the Bodish, Burmish, and Sinitic branches, along with preliminary reconstructions of the phonological system of Proto-Sino-Tibetan.

Recent research using Bayesian phylogenetic methods by Sagart et al (2019) and Zhang et al (2019) has generated tree structures for Sino-Tibetan languages based on cognate sets. While these studies differ in some aspects of the internal subgroupings they propose, both support variations of the traditional view of Sino-Tibetan in which the family consists of two primary branches: the Sinitic languages on one branch, and most of the other languages in the family on the other. The time-depth offered by these analyses are both consistent with an origin among millet farmers in Northern China, followed by a migration and radiation to the southwest. Sagart et al. date the initial split of Proto-Sino-Tibetan to around 7500 B.P. (years before the present) and identify the language with the Cishan and Yangshao cultures, while Zhang et al. (2019) offer the overlapping range of 4200-7800 B.P., corresponding to the Yangshao and Majiayao cultures.

2.3 Tibetan language history

“Tibetan" refers to a broad range of different language varieties spoken by around six million people, traditionally spoken in what is today five countries. The diversity of the Tibetan varieties has been compared to that of Romance by Tournadre (2008), who identified twenty-five languages on analogy with nineteen Romance languages. In this, “Common Tibetan” (*spyi skad*) is used to describe a range of mutually-intelligible varieties spoken in Central Tibet and the Tibetan Diaspora (Denwood 1999, Tournadre and Dorje 2003).

Of approximately six million total speakers of Tibetan, the substantial majority continue to live in China, where the language has official status in the Tibet Autonomous Region and a number of autonomous areas in four provinces covering traditionally Tibetan-speaking areas. Language policy has varied over time and in different regions, but Tibetan is still widely used—though media and official publications are often in regional standards with influence from Classical Tibetan (Dwyer 1998, Tournadre and Jiatso 2001). However, as Mandarin has taken an increasingly important role in government, education, and daily life, the use of Tibetan has declined. The use of mixed Tibetan and Chinese has risen over time, gaining the derisive term *ra ma lug skad* ‘half-goat-half-sheep language’ (Tournadre 2002/2003), a term I have also encountered in Diaspora to describe code-mixing with, or extensive lexical borrowing from, South Asian and European languages. Extensive language contact, however, is not new: notably, the Amdo region of the northeastern Tibetan Plateau has seen extensive lexical and structural borrowings among Tibetic, Sinitic, Mongolic, Turkic, and other languages (Janhunen 2007, Dwyer 2013, Sandman & Simon 2016).

While the diversity of Tibetan varieties is undeniable, the most common way dialects are grouped is influenced by geographic and culturally-salient social categories that do not always correspond well to the distribution of linguistic features. Geographically, the three traditional cultural regions of U-Tsang, Kham, and Amdo (Central, Eastern, and Northeastern, respectively) are used as broad macro-dialect groupings by linguists such as Denwood (1999), Tournadre and Dorje (2003), and Tsering (2011), among many others. Additional dialects not generally grouped with these three can be found along the Himalayas, including Balti, Ladakhi, Dzongkha, and many others. Alongside geographical groupings, local dialects are often identified as belonging to either sedentary farmers (*rong skad* 'valley speech') or nomadic pastoralists (*'brog skad* 'nomad speech'); though there are linguistic differences among these groups in many regions, the cultural perception that all nomadic pastoralists speak identically has been criticized (Denwood 1999). While linguists do continue to use these geographic and social-economic labels, they are generally understood more as a convenient shorthand than as an accurate classification.

In this perspective, the entire Tibetan-speaking region can be seen as a dialect continuum, where changes have arisen and diffused without clear divisions that could be modeled as a tree. Indeed, the features identified as characteristic of the geographic groupings tend not to be shared innovations, which are necessary criteria for establishing a subgroup. Sun (2014) lists these characteristics as follows: "most current classifications of Tibetic rely on typological similarity (e.g. tonality, phonation, syllable structure), common phonological changes (e.g. cluster simplification, loss of codas), as well as regular shared inheritance (e.g. obstruent voicing, complex onsets, consonantal codas)." Such groupings have also neglected to consider other domains of languages besides phonology; for instance, the auxiliary verbs catalogued by Tournadre and Jiatso (2001) vary tremendously across dialects.

Nevertheless, individual local varieties do tend to follow typological generalizations. The phonology of "archaic" or "cluster" dialects more closely resembles that of Old Tibetan in having complex consonant clusters and lacking tone. In contrast, "innovative" or "non-cluster" dialects tend to exhibit simplified clusters, fewer codas, and contrastive lexical tone. Geographically, the "innovative" varieties tend to be located in the central regions such as U-Tsang, while the "archaic" varieties tend to be located around the periphery—including the far Western dialects such as Balti, Ladakhi, and Purik and the Northeastern Amdo dialects (see Fig. 2.1 for map). This would appear to be consistent with sound changes arising in various locations and spreading across the Tibetan Plateau, such that the geographically (and, often, politically and culturally) central regions would undergo the most changes.

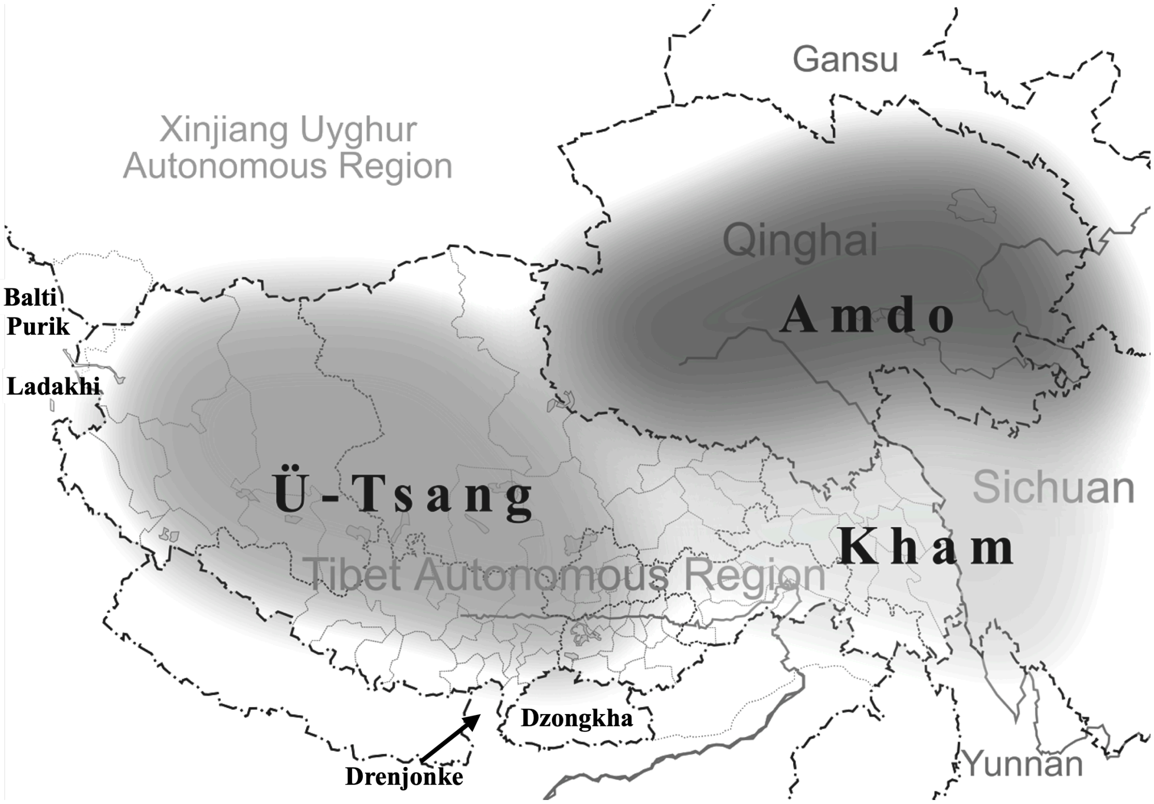


Figure 2.1. Select Tibetan language regions (adapted from Musser 2011)

The Tibetan orthography, traditionally attributed to the 7th-century scholar Thönmi Sambhota (ཐོན་མི་སམྦྷོ་ཊ་), is adapted from Indic scripts. It includes characters representing voiced (ཟ་ཞ་ལ་ར་) and voiceless (ས་ཤ་ལྷ་ཧྲ་) fricatives and liquids, and voiced (ག་གྱ་གྲ་ད་བ་ཇ་ཛ་), voiceless unaspirated (ཀ་ཀྱ་ཀྲ་ཏ་པ་ཅ་ཙ་), and voiceless aspirated (ཁ་ཁྱ་ཁྲ་ཐ་ཕ་ཆ་ཚ་) stops and affricates. The orthography also reflects the phonotactics of the language at the time, such as in its lack of voicing/aspiration contrast for stops in syllable codas. Since only the characters for voiced stops (not aspirated or voiceless stops) occur as codas, Benedict (1972) infers that these were the closest match for the pronunciation of stop codas as well. Presumably the same could be said for the continuants, in that only voiceless fricatives and voiced sonorants appeared as codas. It should be noted that the forms of these graphemes are adapted from their Indic counterparts; digraphs are used for Tibetan sounds not present in Sanskrit (such as the voiceless lateral ལྷ་ and rhotic ཧྲ་). Importantly, the graphemes for Sanskrit sounds not present in Tibetan of the time are written either with digraphs similar to *bh, dh, jh,* and *gh* (བྷ་དྷ་ཛྷ་གྷ་), or by mirroring graphs for coronal sounds to depict retroflex consonants (ཥ་ཊ་ཋ་). Conversely, native Tibetan retroflex stops were a later development, so are written according to their etymological origin (e.g. *kr, tr*)

2.4 Tone and laryngeal contrasts in Tibetan varieties

The classification of Tibetan varieties into “archaic” and “innovative” dialects is long-established in the study of Tibetan by European and international linguists (see Denwood 1999 for the relevant history, dating back to work by Róna-Tas and Jäschke in the nineteenth century). Though intuitive, the archaic/innovative grouping is based on typology rather than genetic relationship, and largely reflects shared retentions. Caplow (2009, 2013) highlights similarities between "archaic" dialects of Baltistan and Amdo, located in the far West and Northeast. The mainstream view holds that these peripheral regions failed to adopt "innovative" sound changes due to geographic, political, and cultural isolation; however, Denwood (2006) has proposed that contact between the northeastern and western regions remained possible across the north-central Changthang (བྱང་ཐང་) region, which has since become extremely dry and effectively uninhabited.

Particularly salient among these shared retentions are the presence of clusters and the lack of contrastive tone, as illustrated in Table 2.1. These two characteristics are closely related, as some contrasts formerly maintained by clusters were progressively reanalyzed as laryngeal and tonal contrasts.

Written (Classical) Tibetan	Balti (Western)	Rebkong (Northeastern)	Tokpe Gola (Central)	Gloss
khrag	[kʂʌk]	[t͡ɕɤɣ]	[ʈʰʌ́k]	'blood'
drug	[druk]	[ɖɯɣ]	[ʈʰùk]	'six'
spyang ki	[spjaŋ.ˈku]	[xt͡ɕaŋ.ˈkʰɤ]	[t͡ʃáŋ.gú]	'wolf'
zam pa	[zam.ˈpa]	[sam.ˈpa]	[sàm.pá]	'bridge'

Table 2.1. Onset clusters across dialects, adapted from Caplow (2013)

While not comprehensive, the examples in 2.1 illustrate the retention of clusters in Western and Northeastern varieties, and the presence of tone in Central Tibetan. The tone contrast is the result of reanalysis of historical onset consonant voicing, as shown by low-tone 'six' and 'bridge' and high-tone 'blood' and 'wolf'. Low tone in modern varieties corresponds to Old and Classical Tibetan voiced obstruent and simplex voiced sonorant onsets. High tone in modern varieties corresponds to Old and Classical Tibetan voiceless obstruents (aspirated or unaspirated) and voiceless liquids onsets, as well as sonorant onsets preceded by a minor-syllable prefix (Sprigg 1972). This description holds for Common Tibetan (Tournadre and Dorje 2003), U-Tsang dialects such as Lhasa (Chang and Shefts 1964, Dawson 1980, Denwood 1999, Tsering 2011), Shigatse (Haller 2000, Tsering 2011), Tokpe Gola (Caplow 2009) and Sherpa (Kelly 2004), and at least some Kham dialects, of which Tsering (2011) provides examples from Dege and Bathang. The tonogenesis pathway just described—reanalysis of onset voicing—is cross-linguistically common and well-supported by phonetic studies (e.g. Hombert et al 1979, Maddieson 1984). Many other Tibetan varieties lack tone, as in Amdo (e.g. Tsering 2011) and Balti (Lobsang 1995). or may have undergone independent tonogenesis, as in Drenjongke (Sikkimese; Yliniemi 2005, 2019; Lee et al. 2019)

2.5 Tibetan in diaspora

In addition to traditional Tibetan regions in China and across the Himalayas, around 150,000 Tibetans currently live in a diaspora which began in 1959, following the Chinese annexation of the Tibetan Plateau. Most of these live in India, Nepal, and Bhutan, with several thousand now residing in North America, Australia, and Europe. Especially in South Asia, their communities form a web of interconnected enclaves: formally-designated settlements, monastic institutions, boarding schools, and urban neighborhoods. While they often use the Tibetan language among themselves, Tibetans are often multilingual, regularly using the dominant languages of their host countries such as Nepali, Hindi, or English, and many also speak Chinese (Denwood 1999, Roemer 2008). Although spread over a geographically wide area, Tibetans in diaspora frequently move between these areas and continue vibrant use of the Tibetan language, thus defining a speech community separate from those inside China. Over the past six decades, several waves of migrants have come from all Tibetan regions of China into diaspora. However, Roemer (2008) reports that 70% of the first wave of 85,000 Tibetans to arrive in India and Nepal in 1959-1962 came primarily from the central U-Tsang region. Of the remainder, 25% came from the eastern region of Kham, and only 5% from the northeastern region of Amdo. Subsequent rates of immigration varied over time, but with increasing proportions of Tibetans from Kham and Amdo regions. U-Tsang, however, remains prominent in diaspora, particularly due to the presence of its historic capital city, Lhasa, and since it was the region where the Dalai Lamas and many other elites traditionally lived.

2.6 Common Tibetan Consonants

While Old Tibetan and many modern varieties are famous for their large consonant clusters, Common Tibetan is an “innovative” variety with no clusters and a maximal (C)V(C)(ʔ) syllable, as well as lexical tone. While the consonant inventory in Table 2.2 is generally consistent with Common Tibetan as spoken in Diaspora and Tibet, there is substantial variation across these and other dialects.

	Bilabial	Dental/Alv	Retroflex	Palatal	Velar	Glottal
Stops	p p^h (b)	t̪ t̪^{h} (b)	ʈ ʈ^h (ɖ)	c c^h (ɟ)	k k^h (g)	ʔ
Affricates		t͡s t͡s^h		t͡ç t͡ç^h		
Fricatives		s		ç		h
Approx.	w	l l̥	ɻ ɻ̥	j		
Nasals	m	n̪		ɲ	ŋ	

Table 2.2. Consonant phonemes of Common Tibetan. Parentheses indicate that voiced stops are variably voiceless and only present for some speakers. Aspiration/voicing of stops, affricates, and approximants is only contrastive in word-initial position (adapted from Chang and Chang 1978 and Tournadre and Dorje 2003).

Position in the syllable significantly limits the realization of consonants. All contrasts are present in word-initial position. In word-medial syllable-initial position, the aspiration contrast for stops and affricates and the voicing contrast for approximants is neutralized. Only a limited set of segments is possible in syllable-final position: /p k m n ŋ l ɻ/, though /n ŋ/ are often heavily reduced and appear as nasalization on the preceding vowel, and /l ɻ/ are often reduced, conditioning greater length of the preceding vowel. Coda /k/ may also be realized as [ʔ]. Palatal stops /c c^h/ are listed here, which are derived from what are historically clusters [kj k^{h}j]; since there is not contrast between simplex palatal stops and clusters, it is possible that speakers may vary in their realization of these items. If /c c^h/ is really /kj k^{h}j/, they are the only surface tautosyllabic clusters in the language.

Voiceless nasals are not attested in Common Tibetan, but have been attested in some dialects. They are described for Lhasa speakers by Chang and Shefts (1964) and Dawson (1980), who list voiceless counterparts of the velar, palatal, and bilabial nasal (but not the alveolar/dental) occurring in the onsets of words with high tone. These include verb root alternations corresponding to perfective aspect, and on the negative verbal prefix [mə-] when the verb root begins with an aspirated high-tone consonant. Denwood (1999) does mention that some speakers, particularly those of higher social status, have voiceless [m̥] in addition the the voiced [m], but does not mention other voiceless nasals. Interestingly, Denwood claims these speakers with voiceless nasals comprise a subset of those who lack prevoicing in stops and affricates. This is consistent with the description of Chang and Shefts (1964) and Dawson (1980), who were working in the United States with two Lhasa speakers of educated, upper-class background.

The Lhasa voiceless nasals correspond orthographically to Old Tibetan *sm-, sɲ-*, and *sŋ-* onset clusters. These historical clusters also yielded high tones in the U-Tsang dialects, including Lhasa and Shigatse. Synchronically, these voiceless nasals are thus similar to voiceless liquids /l̥ r̥/ in only co-occuring with high tone. However, the voiceless liquids are attested in Old Tibetan, prior to any tonogenesis, while the voiceless nasals are a comparatively recent development. Both tonogenesis and the loss of *s-* have taken place across U-Tsang dialects, but the relative chronology is not clear. It is possible that the *s-* was retained long enough after tonogenesis, at least in Lhasa, to condition the voicelessness in those nasals. Alternatively, the *s* + nasal onset clusters may have been voiceless prior to tonogenesis. Under this scenario, they would have merged with the voiced nasals in most U-Tsang dialects, but have been retained in Lhasa. Interestingly, voiceless [m̥] is also described for the closely-related Shigatse dialect by Haller (2000), but only for the negative prefix when affixed to a verb root beginning with a voiceless fricative or aspirated stop or affricate. Haller (2000) specifically notes not finding the other voiceless nasals described for Lhasa in Chang and Shefts (1964). Given the geographic proximity and historic ties between the the two cities, it is plausible that this form was borrowed into Shigatse Tibetan from Lhasa.

In the Kham dialects of the eastern Tibetan plateau, a full set of voiced and voiceless nasal pairs is characteristic of the region. These have been observed in several Kham dialects, such as Batang (Gesang 1989), Dege (Häsler 1999), Dongwang (Bartee 2014) and Kami. Of the latter, Chirkova (2014) says the voiceless nasals "normally have homorganic voiced nasal onsets but voiceless, slightly aspirated release, i.e. respectively, [m͡m̥ʰ],[n͡n̥ʰ], [ɲ͡ɲ̥ʰ], [ŋ͡ŋ̊ʰ]." While Kham Tibetan voiceless nasals have yet to be studied with instrumental phonetics, the description above differs from observations on voiceless nasals in Burmese (Dantsuji 1984) and Mizo (Bhaskararao and Ladefoged 1991), which show a short period of breathy or voiced phonation at the end the nasal, rather than the beginning. Dantsuji (1986) further finds that the voiced (or, perhaps, breathy) portion of the Burmese nasals contains more acoustic cues to place of articulation than the voiceless portion. In Angami, however, voiceless nasals conclude not with voiced nasal airflow, but with a weak oral release (Bhaskararao and Ladefoged 1991).

2.7 Common Tibetan Vowels

Common Tibetan has eight phonemic vowels /i y e ø ɛ a o u/. A mid-central vowel [ə] may also appear as a reduced alternant of /a/ in some open syllables and before bilabial codas (though, notably, the [a] ~ [ə] distinction is not mentioned by Tsering (2011)). Furthermore, a lower alternant of /o/ appears in closed syllables (Tournadre and Dorje 2003). The latter is listed as a phonemic vowel for the Lhasa speakers recorded by Chang and Shefts (1964) and Dawson (1980), as well as by Tsering (2011). However, the open syllables with [ɔ] in those sources correspond to closed syllables in Common Tibetan, meaning that either the codas are synchronically deleted, or the phonemicization of [ɔ] is unique to the Lhasa dialect. The vowels of Common Tibetan are shown in Table 2.3, below.

	Front		Central	Back
	Rounded	Unrounded		
High	i	y		u
Mid	e	ø	[ə]	o [ɔ]
Low	ɛ			a

Table 2.3. Vowels of Common Tibetan. Brackets indicate non-phonemic status; [ə] *is an allophone of /*a*/, and* [ɔ] *is an allophone of /*o*/.*

The vowel inventories of other dialects can vary substantially from the eight phonemic vowels listed in Table 2. Other dialects of the U-Tsang region have largely similar inventories (e.g. Shigatse, Haller 2000 and Tsering 2011). However, the common ancestor of the modern Tibetic varieties had fewer vowels: Old Tibetan had six phonemic vowels, /i e a o u ɨ/, which had collapsed to five /i e a o u/ in Classical Tibetan. These are preserved in some five-vowel varieties such as Balti (Lobsang 1995), while the U-Tsang dialects developed the front vowels /y ø ɛ/ from fronting conditioned by underlyingly tautosyllabic coronal codas, some of which have since been lost. Amdo dialects are characterized by vowel shifts, such as the merger of historical /i/ and /u/ to /ə/ in many environments (Tsering 2011, Denwood 1999), while Kham dialects merge /a/ and /o/ before velar nasal codas (Tsering 2011).

2.8 Common Tibetan tone

Different authors have produced different descriptions of the Tibetan tone system. Part of this can be ascribed to the variation across dialects, many of which lack tone contrasts entirely. Another factor is a reliance on impressionistic description rather than instrumental study or spoken corpora, which has made comparison between studies difficult. Finally, different traditions of linguistics have led different authors to draw different conclusions about the tone system.

Despite various analyses, sources agree that Common Tibetan and similar U-Tsang varieties contrast high and low tonal registers. These are simply called "tones" in this dissertation, but “register” is used here to highlight the difference between the high/low contrast and the contours discussed below. Both tone registers co-occur with all vowels and most consonants, with the following exceptions. Initial /l̥ ɻ̊ h/ only occur with high tone, and the voiced allophones of unaspirated stops, affricates, and fricatives only occur with low tone. The presence of voicing is variable where it does occur. Words have from one to three syllables, but only one tone, which is determined by the tone of the initial syllable.

In addition to tonal register, Tibetan words exhibit tonal contours that have been described in various ways. One common description lists four tones that result from the interaction of register and contour: high-level, high-falling, low-level, and low-falling; this has been described for the dialects of Lhasa (Chang and Shefts 1964) and Shigatse (Haller 2000), though Chang and Chang (1978) add that the falling contours are accompanied by a degree of glottalization. Another description for Lhasa delineates six tones predictable based on coda, as shown in Table 2.4.

Coda	High	Low
Sonorant or long vowel	55	13
Stop coda	52	121
Short vowel, open syllable	54	12

Table 2.4. Six-tone description of tone on Lhasa monosyllables, from Hu et al (1982; reported in Duanmu 1992).

For Duanmu (1992), the falling contours of high tones or closed low-tone syllables are a phonetic consequence of glottalization associated with glottal stops or glottalized coda stops. This leaves a simpler phonological description of two tones: a high tone (H) and a rise (LH). For polysyllabic words, H or L tones associate with syllables one-to-one, with the last spreading rightward; the H component of a LH tone is realized later for long (sonorant-final) syllables, resulting in tonal patterns matching those reported in Qu and Tan (1983):

Coda	High (H)	Low (LH)
Disyllable, final sonorant	55 55	11 14
Disyllable, other	55 53	11 53
Trisyllable, final sonorant	55 55 55	11 55 55
Trisyllable, other	55 55 53	11 55 53

Table 2.5. Tone contours of Lhasa polysyllables, from Qu and Tan (1983; reported in Duanmu 1992).

Hu and Xiong (2010) describes eight surface variants of two phonological tones in Lhasa monosyllables. Echoing Duanmu's H and LH tones, they list high and low phonological tones, which usually surface as high-level and rising tones. However, final glottal stops cause H tones to surface as falling contours, and low/rising tones to appear as rising-falling contours. Finally, they describe tones as "long" in sonorant-final syllables, intermediate in nasal + glottal stop coda syllables, and "short" elsewhere. However, if length is treated as a property of the segmental string and not the tone, then this combination of high and low/ rising tones, plus a syllable-final fall in the presence of a glottal stop or glottalized coda, matches closely with the descriptions of Duanmu (1992).

In all these descriptions, the pitch value at the beginning of any tonal contour is presented as being of equivalent level for the same register, and for any length of syllable or word: high tones of any contour begin at 5, and low/ rising tones of any contour begin at 1. This is supported by the pitch tracks reported in Hu and Xiong (2010), Hu (2016) and Chang and Chang (1978), for which contours of the same phonological tone (high or low/rising) begin at very similar values. For present purposes, then, Tibetan tones are taken to present only two pitch values at the onset of voicing, with later contours determined by length of syllable/word and the presence/absence of glottal codas.

The papers cited above offer various analyses of tonal representation in Tibetan. Qu and Tan (1983) and Chang and Chang (1978) treat tone as a property of a syllable, with sandhi rules applying in different environments. Duanmu (1992) uses tonal autosegments, with high tone represented as H, low tone represented as LH, and rightward spreading rules to explain the tone of polysyllabic words. Hu (2012, 2016), using tonal gestures (discussed in the next section), also uses separate H and L gestures. By approaching the problem of representing these tones in a gestural framework, a number of hypotheses will be formulated and tested in the following chapters. Chapter 3 will investigate the acoustics of the tones across a number of speakers, and the relationship between tone and VOT. Chapter 4 will use articulatory evidence to explore the predicted effects of these tonal representations on C-V timing.

2.9 Laryngeal contrasts in Tibetan language history

Identifying sound correspondences across Sino-Tibetan languages has been made difficult due to extensive language contact and a lack of shared inflectional morphology across languages, but this is especially true when investigating laryngeal contrasts. As discussed in Handel (2008), researchers have struggled to identify phonological correspondences in terms of voicing and aspiration in onsets; codas often do not contrast along these dimensions. Reconstruction efforts such as Benedict (1972) and Matisoff (2003) have had greater success identifying cognates with stop onsets by place of articulation, but often must remain agnostic about reconstructing laryngeal contrasts. The variation across Sino-Tibetan languages has been attributed to sound changes conditioned by prefixes or “pre-initials” that were lost in most daughter languages.

Laryngeal contrasts have changed substantially in the history of Sino-Tibetan. Among the many languages in the family, it is possible to find voiced, voiceless, aspirated, breathy, prenasalized, and glottalized obstruents, as well as voiced and voiceless sonorants and many examples of consonant contrasts reanalyzed through tonogenesis. Nevertheless, sources generally agree that Proto-Sino-Tibetan was non-tonal (Benedict 1972, Matisoff 2003, Hill 2019), with tone arising independently in several branches (though tonality may spread through language contact in the Mainland Southeast Asia linguistic area, e.g. Post 2015).

However, the reconstruction of laryngeal contrasts in Proto-Sino-Tibetan has been elusive, made difficult in part by the widespread prefixes in many daughter languages that have conditioned changes in phonation and tone. While Hill (2019) declines to reconstruct laryngeal contrasts for Proto-Sino-Tibetan, both Benedict (1972) and Matisoff (2003) have posited that Proto-Tibeto-Burman had a simple two-way contrast between voiced and voiceless stops, fricatives, and affricates. The prefixes that conditioned much subsequent phonology have been reconstructed as monosyllabic, often an open syllable, and have served many functions and changed in diverse ways throughout Sino-Tibetan. Phonologically, they have affected voicing, glottalization, aspiration, and tone, and triggered other processes such as palatalization, metathesis, fusion, cluster formation, and segmental deletion in various languages (Benedict 1972, Matisoff 2003).

By the time of Old Tibetan, some of the remaining prefixes may have remained productive in verbal paradigms, but many were no longer productive. The Tibetan orthography, developed during this period, does not mark a vowel for these prefixes, but distinguishes tautosyllabic clusters from those derived from prefixes. These historically-derived clusters have been described as having a status intermediate between a tautosyllabic cluster and a full disyllable: they have been reconstructed by Hill (2019) as having the form *Cə- (the *ə vowel having otherwise merged with *a); cognates in Old Chinese are sometimes written with two characters (Baxter and Sagart 2015); and Matisoff (2003) refers to them as “sesquisyllables.” In the Tibetan orthography, tautosyllabic onset clusters are written with letters stacked atop each other, while the historically-derived clusters are written with letters in sequence. This can be illustrated with the minimal pair གྱང་ *gyang* *gjaŋ 'earth wall' and གཡང་ *g.yang* *gə.jaŋ 'auspicious'. In modern Tibetan varieties, some historically-derived clusters are retained as clusters, while in others, including the Central Tibetan dialects, they have been lost (Hill 2012). Thus the Central Tibetan reflexes of these etyma are [kjàŋ] or [càŋ] for གྱང་ *gyang* 'earth wall', but [jáŋ] for གཡང་ *g.yang* 'auspicious'.

Gestural timing plays an important role in understanding these differences. Butler (2015) convincingly shows that the term 'sesquisyllable' has been used to describe words with at least two different types of articulatory timing. In Khmer, purported sesquisyllables in fact resemble monosyllables with complex onsets whose gestural timing results in excrescent vocoids within the cluster. In contrast, the purported sesquisyllables of Bunong resemble iambic disyllables, which have a proper vowel target. Whether Old Tibetan forms like གཡང་ *g.yang* *gə.jaŋ 'auspicious' had an excrescent vocoid or a vowel gesture cannot be determined with certainty. If these forms were disyllabic as in Bunong, *g.yang* would include two vowel gestures and take the form CV.CVC. However, if they were monosyllables, then a tantalizing possibility presents itself: perhaps the difference between *g.yang* and *gyang* lies only in the timing of their articulatory gestures. In this case, the sesquisyllable-like *g.yang* would have a looser cluster allowing for an excrescent vocoid, while the monosyllable-like *gyang* would have a cluster with tighter overlap.

Characterizing this difference in terms of overlap—that is, in relative timing of gestures—can be further analyzed using Articulatory Phonology. As discussed in section 1.3, this framework can represent timing and overlap in a gestural score, and provides means such as coupling graphs from which to derive gestural scores (e.g. Fig. 1.1). The different kinds of clusters can be represented with different gestural scores, which would in turn correspond to a different coupling graph. Fig. 2.2 proposes gestural scores and coupling graphs for each type of cluster.

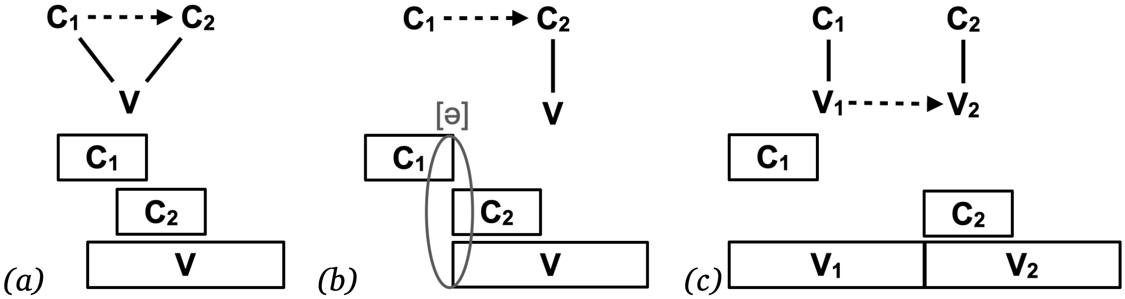


Figure 2.2. Gestural coordination options for Old Tibetan onsets. (a) competitive coupling; (b) simplex timing with excrescent vocoid; (c) two syllables

Of the three options shown in Fig. 2.2, the competitive coupling in (a) corresponds to a standard Articulatory Phonology view of a complex onset, as presented in section 1.4. The simplex timing of (b) has been proposed to model certain onsets, including /s/-stop clusters in Italian (Hermes et al. 2012) and onset clusters in Moroccan Arabic (Shaw et al. 2009, 2011) and Tashlhyit Berber (Goldstein et al. 2007, Hermes et al. 2017). The three coupling modes make different predictions concerning excrescent vocoids. The competitive coupling of (a) results in substantial overlap between the two consonantal gestures, preventing any transitionary vocalic material from being expressed. The simplex timing in (b) would allow for some excrescent vocalic material to emerge under the right circumstances, such as continuous voicing through the onset and the appropriate stiffness values. Finally, the third structure in (c) simply represents two syllables, so the vowel articulation in the first syllable would be an actual vowel (possibly contextually reduced) rather than an excrescent vocoid.

The orthographic conventions of Tibetan, having been standardized based on the phonology of Old Tibetan, can offer insights into the syllabic structure.

Excrescent vocoids are not depicted in the writing system, but the arrangements of prevocalic consonants differ by position. While all consonants can immediately precede a vowel, only a limited number can come before the prevocalic consonant. Of those, the consonants written <*s r l*> are written above the prevocalic consonant, while those written <*b d g m*> are written before the prevocalic consonant. Importantly, orthographic <*bC-*> and <*dC-*> are in complementary distribution: <*b-*> appears before coronal prevocalic consonants and <*d-*> before other prevocalic consonants. This presents two configurations which might inhibit the realization of an intervening vocalic element: a sequence of [d-] followed by a coronal stop, which could coalesce into a single closure and release, and a sequence of [d-] followed by a coronal fricative, which could be confused with an affricate. In addition, all of the consonants that are written before a prevocalic vowel are voiced, which facilitates voicing between them. By contrast, the remaining superscribed letters correspond to continuants which would be less likely to lead to the production of an excrescent vocoid.

Unfortunately for the present study, sound changes in the subsequent history of Tibetan mean that the gestural timing of Old Tibetan is of little relevance for understanding the modern varieties. Firstly, Old Tibetan lacked tone and, as Hill (2010) argues, also predates the split between voiceless aspirated and unaspirated stops. Moreover, the consonants that have survived the simplification of consonant clusters are the immediately prevocalic ones. Prefixed consonants survive to varying degrees in Northeastern and Western dialects, but even in Balti (Lobsang 1995) these are described as lacking any intervening vowel. For the Central Tibetan dialects most relevant to this study, only the prevocalic consonants remain. For any of the gestural coordination schemas depicted in Fig. 2.2, the prevocalic consonants are simply timed in-phase to the vowel, which would predict that the preceding stops of Old Tibetan would have no effect on the timing of present-day onsets. Even for those dialects with clusters, however, the lack of intervening vowels predict that these dialects have shifted away from the Old Tibetan system that allowed for these vocalic elements.

However, some traces of the clusters remain due to their influence on aspiration and tone, even in Central Tibetan dialects. While sonorant-initial syllables do not occur with onset clusters, the presence or absence of other preceding consonants conditioned tonogenesis. Words with simplex sonorant onsets developed low tone, and those in clusters developed high tone: *mi > [mì] 'person'; *rmi > [mí] 'dream'. Despite widespread attestation (surveyed in Denwood 1999, Caplow 2013), I am not familiar with any proposals for how this took place. Perhaps, as voiceless onsets were becoming high-tone and voiced onsets low-tone, the pre-initial cluster consonants were sufficiently devoiced to condition high tone.

In addition to tone, Old Tibetan clusters participated in the phonologization of aspiration. While the aspirated and unaspirated stops were contrastive in Classical Tibetan, this contrast was not present in Old Tibetan (Shafer 1955, Benedict 1972, Hill 2007): Old Tibetan voiceless stops are generally written as aspirated word-initially and unaspirated word-medially, which suggests an allophonic alternation (e.g. the item *khol~kol* 'servant' is written as aspirated in *khol-yul* 'fief-land' and unaspirated in *gnam-kol* 'servant of heaven'). However, by the time of Classical Tibetan, aspirated and unaspirated stops were contrastive in simplex onsets and predictable in clusters: aspirated stops appeared in clusters following <*m*> and <*h̬*>, while unaspirated stops appeared in clusters following <*b d g s r l*>. Of these, <*s r l*> would have been in tighter clusters with no intervening vowel or vocoid, while <*b d g*> and <*m* *h̬*> would have had a vocalic interval. It thus appears that voiceless stops were unaspirated in the tighter clusters and when preceded by another stop, but aspirated in other environments, namely <*m h̬*> and simplex onsets. Some voiceless-unaspirated simplex onsets also exist, though many are loanwords or etymologically related to forms occurring in historical clusters.

From a gestural perspective, the glottal spreading gesture associated with aspiration was present word-initially, but eventually lost through coarticulation in <*s r l*> clusters and following another stop. Laryngeal gestures have been shown to be shared across a cluster (Löfqvist & Yoshioka 1980, Hoole et al 1999), so it is not surprising that this glottal spreading gesture overlapped with a preceding <*s r l*> in a tight cluster. This overlap would have allowed subsequent generations to reinterpret those stops as unaspirated. However, the fact that aspiration was lost following another stop is telling: it indicates these <*b d g*> units were behaving more as single segments in a cluster than as CV syllables. The remaining environments involved <*m h̬*>, voiced continuants with looser timing that, as voiced segments, would be considered in Articulatory Phonology to lack a laryngeal gesture. These patterns were preserved in the orthography: aspirated stops appear only as orthographic simplex onsets, such as *phar* *p^{h}ar 'over there,' *chung* *t͡ʃ^{h}uŋ 'small,' and *khang* *k^{h}ang 'house,' while unaspirated stops appear only in orthographic clusters: *spar* *spar 'picture,' *gcung* *t͡ʃuŋ 'younger sibling', and *bkang* *kaŋ 'filled'. Thus, as contrastive aspiration was emerging in the transition between Old Tibetan and Classical Tibetan, the former prefixes would have been behaving as clusters, supporting the analysis of them as simplex-timing clusters with excrescent vocoids (Fig. 2.2(b)) rather than Cə syllables (Fig. 2.2(c)). This example also shows that changes in gestural timing were crucially important in the development of Tibetan laryngeal contrasts.

The clusters were subsequently lost to varying degrees in modern dialects, but the aspiration and voicing contrasts remained. Some clusters have remained even in dialects with greatly simplified syllable inventories. This can be seen in compounds such as [mèndá] 'gun', which is composed of [mè] 'fire' and [tà~dà] 'arrow'; the [n] is part of the underlying representation of [dà] 'arrow' and surfaces when permitted by the phonotactics. In tonal dialects, underlying clusters preserved voicing in words such as [dà] 'arrow' after the contrastive role of voicing was transferred to the tone; in Eastern dialects these are generally prevoiced or even prenasalized while simplex onsets became voiceless. Some Central Tibetan dialects took this further, with etymologically simplex onsets becoming aspirated and etymological clusters becoming voiceless-unaspirated (Tsering 2011). These stages are summarized in Table 2.6, below:

	Etymological onsets				Innovative features
Orthography	སྤ་	ཕ་	བ་	སྦ་	
Old Tibetan	s^{ə}pa	p^{h}a	ba	s^{ə}ba	
Classical Tibetan; Northeastern and Western dialects	spa	p^{h}a	ba	ʁba	consolidation of clusters aspirated/unaspirated contrast
Eastern dialects	pá	p^{h}á	pà	bà	tonogenesis cluster simplification
Central dialects	pá	p^{h}á	p^{h}à	pà	voiced simplex > voiceless voiced clusters > aspirated

Table 2.6. Summary of changes in laryngeal contrasts, listed according to past and present varieties with these contrasts.

2.10 Chapter summary

This chapter has traced the historical development of Tibetan beginning with its Sino-Tibetan origins. Over the centuries, the language radiated across the Tibetan Plateau and adjacent regions. Rather than neatly splitting in a tree-like fashion, regional varieties diverged as sound changes and other historical processes took root and spread. In the phonological domain, dialects in more central regions underwent extensive changes: clusters radically simplified, tone was innovated, and average word length increased as many monosyllabic words became disyllabic. Fewer changes took place in peripheral Tibetan-speaking regions, where dialects retained more clusters and remained non-tonal. These typologically-diverse varieties were thrown into extensive contact in the Tibetan diaspora beginning in 1959, setting the stage for further contact-induced change. Two aspects of the phonological system were especially ripe for reanalysis: the diverse sets of voicing and aspiration contrasts, and the newly-innovated tones whose contrastive function was increasingly shared with greater word length. The following chapters investigate the laryngeal and tonal contrasts in diaspora speakers through acoustic and articulatory analysis.

2.11 Chapter bibliography

Bartee, Ellen. 2014. Dongwang. In Jackson T.-S. Sun (ed.), *Phonological Profiles of Little-Studied Tibetan Varieties* (Language and Linguistics Monograph Series 55). Taipei: Institute of Linguistics, Academia Sinica.

Baxter, William H. & Laurent Sagart. 2014. *Old Chinese: A new reconstruction*. Oxford University Press.

Bellwood, Peter. 2005. Examing the farming/language dispersal hypothesis in the East Asian context. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, 17–30. 1st edn. Routledge. https://doi.org/10.4324/9780203343685. https://www.taylorfrancis.com/books/9780203343685 (24 June, 2020).

Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus*. New York: Cambridge University Press.

Bhaskararao, P. & Peter Ladefoged. 1991. Two types of voiceless nasals. *Journal of the International Phonetic Association* 21(2). 80–88. https://doi.org/10.1017/S0025100300004424.

Blench, Roger & Mark W. Post. 2014. Rethinking Sino-Tibetan phylogeny from the perspective of Northeast Indian languages. *Trans-Himalayan Linguistics*. by Nathan Hill and Tom Owen-Smith 266. 71–104.

Butler, Becky. 2015. Approaching a phonological understanding of the sesquisyllable with phonetic evidence from Khmer and Bunong. In N.J. Enfield & Bernard Comrie (eds.), *Languages of Mainland Southeast Asia*. Berlin, München, Boston: DE GRUYTER. https://doi.org/10.1515/9781501501685-010. https://www.degruyter.com/view/books/

9781501501685/9781501501685-010/9781501501685-010.xml (14 June, 2020).

Caplow, Nancy J. 2013. Amdo and Balti Tibetan: Sister Languages, 12,00 Miles Apart. In *The Third International Conference on Tibetan Language*, vol. 1, 303–324. New York: Trace Foundation.

Caplow, Nancy J. 2017. Inference and deferred evidence in Tibetan. In Lauren Gawne & Nathan W. Hill (eds.), *Evidential Systems of Tibetan Languages*, 225–257. Walter de Gruyter GmBH & Co KG.

Chang, Kun & Betty Shefts. 1964. A manual of Spoken Tibetan (Lhasa Dialect). University of Washington Press.

Chang, Kun & Betty Shefts Chang. 1978. *Spoken Tibetan texts*. Vol. 1. Nankang, Taipei: Academia Sinica.

Chirkova, Katia. 2014. Kami. In Jackson T.-S. Sun (ed.), *Phonological Profiles of Little-Studied Tibetan Varieties* (Language and Linguistics Monograph Series 55). Taipei: Institute of Linguistics, Academia Sinica.

Dantsuji, Masatake. 1984. A Study on Voiceless Nasals in Burmese. *Studia phonologica*. Institution for Phonetic Sciences, University of Kyoto. 18. 1–14.

Dantsuji, Masatake. 1986. Some acoustic observations on the distinction of place of articulation for voiceless nasals in Burmese. *Studia phonologica*. Institution for Phonetic Sciences, University of Kyoto. 20. 1–11.

Dawson, Willa. 1980. *Tibetan Phonology* (Ph.D. Dissertation). University of Washington.

Denwood, Peter. 1999. *Tibetan* (London Oriental and African Language Library). Vol. 3. Philadelphia: John Benjamins Publishing Company.

Denwood, Peter. 2006. Early connections between Ladakh/Baltistan and Amdo/ Kham. In John Bray (ed.), *Ladakhi Histories: Local and Regional Perspectives*, 31–40. Leiden and Boston: Brill.

Driem, George van. 2002. Tibeto-Burman phylogeny and prehistory: Languages, material culture and genes. *Antiquity* 72. 885–97.

Driem, George van. 2005. Tibeto-Burman vs. Indo-Chinese: Implications for population geneticists, archaeologists and prehistorians. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. 1st edn. Routledge. https://doi.org/10.4324/9780203343685. https://www.taylorfrancis.com/books/9780203343685 (24 June, 2020).

Duanmu, San. 1992. An autosegmental analysis of tone in four Tibetan languages. *Linguistics of the Tibeto-Burman Area* 15(1). 65–91.

Dwyer, Arienne M. 1998. The texture of tongues: Languages and power in China. *Nationalism and Ethnic Politics*. Routledge 4(1–2). 68–85. https://doi.org/10.1080/13537119808428529.

Glover, Jessie R. & Deu Bahadur Gurung. 1979. *Conversational Gurung*. Pacific Linguistics, Research School of Pacific and Asian Studies, The ....

Glover, Warren W. & Jessie Glover. 1972. *A guide to Gurung tone*. Tribhuvan University [and] Summer Institute of Linguistics.

Haller, Felix. 2000. *Dialekt und erzählungen von shigatse*. Bonn, Germany: VGH Wissenschaftsverlag GmbH.

Handel, Zev. 2008. What is Sino-Tibetan? Snapshot of a Field and a Language Family in Flux. *Language and Linguistics Compass*. Wiley Online Library 2(3). 422–441.

Hermes, Anne, Doris Mücke & Bastian Auris. 2017. The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics* 64. 127–144.

Hildebrandt, Kristine A. 2004. A grammar and glossary of the Manange language. In Carol Genetti (ed.), *Tibeto-Burman languages of Nepal: Manange and Sherpa* (Pacific Linguistics 557), 1–192. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Hill, Nathan W. 2007. Aspirated and unaspirated voiceless consonants in Old Tibetan. 語言暨語言學/*Languages and Linguistics* 8(2). 471–493.

Hill, Nathan W. 2010. An overview of Old Tibetan synchronic phonology. *Transactions of the Philological Society* 108(2). 110–125. https://doi.org/10.1111/j.1467-968X.2010.01234.x.

Hill, Nathan W. 2011. An Inventory of Tibetan Sound Laws. *Journal of the Royal Asiatic Society* 21(4). 441–457. https://doi.org/10.1017/S1356186311000332.

Hill, Nathan W. 2019. The Historical Phonology of Tibetan, Burmese, and Chinese. Cambridge University Press.

Hombert, Jean-Marie, John J. Ohala & William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language*. JSTOR 37–58.

Hoole, Philip, Christer Gobl & Ailbhe Ní Chasaide. 1999. Laryngeal coarticulation. In William J. Hardcastle & Nigel Hewlett (eds.), *Coarticulation*, 105–143. 1st edn. Cambridge University Press. https://doi.org/10.1017/CBO9780511486395.006. https://www.cambridge.org/core/product/identifier/CBO9780511486395A017/type/book_part (26 October, 2020).

Hu, Fang. 2016. Tones are not abstract autosegmentals. In *Speech Prosody*, 302–306.

Hu, Fang & Ziyu Xiong. 2010. Report of Phonetic Research 2010, Institute of Linguistics, Chinese Academy of Social Sciences 5.

Jacques, Guillaume. 2014. Cone. In Jackson T.-S. Sun (ed.), *Phonological Profiles of Little-Studied Tibetan Varieties* (Language and Linguistics Monograph Series 55). Taipei: Institute of Linguistics, Academia Sinica.

Janhunen, Juha. 2007. Typological interaction in the Qinghai linguistic complex. *Studia Orientalia Electronica* 101. 85–102.

Jumian, Gesang. 1989. Phonological analysis of Batang Tibetan. *Acta Orientalia Academiae Scientiarum Hungaricae* 43(2–3). 331–358.

Kelly, Barbara. 2004. A grammar and glossary of the Sherpa language. In Carol Genetti (ed.), *Tibeto-Burman languages of Nepal: Manange and Sherpa* (Pacific Linguistics 557), 193–324. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Lee, Seunghun J., Shigeto Kawahara, Céleste Guillemot & Tomoko Monou. 2019. Acoustics of the four-way laryngeal contrast in drenjongke (Bhutia): Observations and implications. *Journal of the Phonetic Society of Japan* 23. 65–75.

Lobsang, Ghulam Hassan. 1995. *Short sketch of Balti grammar: a Tibetan dialect spoken in Northern Pakistan* (Arbeitspapier). Vol. 34. Bern: Institute für Sprachwissenschaft, Universität Bern.

Löfqvist, Anders & Hirohide Yoshioka. 1980. Laryngeal activity in Swedish obstruent clusters. *The Journal of the Acoustical Society of America* 68(3). 792–801. https://doi.org/10.1121/1.384774.

Maddieson, Ian. 1984. The effects on F0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics* 12(1). 9–15. https://doi.org/10.1016/S0095-4470(19)30845-9.

Matisoff, James A. 1979. Problems and progress in Lolo-Burmese: Quo vadimus. *Linguistics of the Tibeto-Burman area* 4(2). 11–43.

Matisoff, James A. 1991. Sino-Tibetan Linguistics: Present State and Future Prospects. *Annual Review of Anthropology* 20. 469–504.

Matisoff, James A. 2003. Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction. Univ of California Press.

Mazaudon, Martine. 1978. Consonantal mutation and tonal split in the Tamang sub-family of Tibeto-Burman. *Kailash* 6(3). 157–179.

Mazaudon, Martine. 1994. Problèmes de comparatisme et de reconstruction dans quelques langues de la famille tibéto-birmane. Paris: Université de la Sorbonne Nouvelle (Paris III) Thèse d'Etat.

Musser, Karl. 2011. *Tibet_provinces.png*. Retouched digital image. https://commons.wikimedia.org/wiki/File:Tibet_provinces.png.

Nishida, Tatsuo. 1977. Some Problems in the Comparison of Tibetan, Burmese and Kachin Languages. *Studia phonologica* 11. 1–24.

Owen-Smith, Thomas & Nathan W. Hill. 2014. Introduction. In Thomas Owen-Smith & Nathan W. Hill (eds.), *Trans-Himalayan Linguistics*, 1–10.

Post, Mark W. 2015. Morphosyntactic reconstruction in an areal-historical context: A pre-historical relationship between North East India and Mainland Southeast Asia? In N.J. Enfield & Bernard Comrie (eds.), *Languages of Mainland Southeast Asia: The state of the art*, 209–265. Boston: De Gruyter Mouton.

Qu, Aitang & Kerang Tan. 1983. Ali Zangyu (Ali Tibetan). zhongguo shehui kexue yan chubanshe (Chinese Academy of Social Sciences Press).

Roemer, Stephanie. 2008. The Tibetan government-in-exile: politics at large. New York: Routledge.

Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill & Johann-Mattis List. 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *PNAS* 116(21).

Sandman, Erika & Camille Simon. 2016. Tibetan as a “model language” in the Amdo Sprachbund: Evidence from Salar and Wutun. *Journal of South Asian Languages and Linguistics*. De Gruyter 3(1). 85–122.

Shafer, Robert. 1955. Classification of the Sino-Tibetan Languages. WORD 11(1). 94–111. https://doi.org/10.1080/00437956.1955.11659552.

Sun, Jackson T.-S. 2014. Preface. In Jackson T.-S. Sun (ed.), *Phonological Profiles of Little-Studied Tibetan Varieties* (Language and Linguistics Monograph Series 55). Taipei: Institute of Linguistics, Academia Sinica.

Thurgood, Graham. 2003. A subgrouping of the Sino-Tibetan languages: the interaction between language contact, change, and inheritance. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages* (Routledge Language Family Series), 3–21. New York: Routledge.

Tournadre, Nicolas. 2003. The Dynamics of Tibetan-Chinese Bilingualism: The Current Situation and Future Prospects. (Trans.) Peter Brown. *China Perspectives* (1). https://doi.org/10.4000/chinaperspectives.231. http://journals.openedition.org/chinaperspectives/231 (29 April, 2020).

Tournadre, Nicolas. 2008. Arguments against the concept of “Conjunct’/’Disjunct” in Tibetan. In Brigitte Huber, Marianne Volkart & Paul Widmer (eds.), *Chomolangma, Demawend und Kasbek. Festscchrift für Roland Bielmeier zu seinem 65.*, vol. 1, 281–308. International Institute for Tibetan and Buddhist Studies.

Tournadre, Nicolas & Sangda Dorje. 2003. *Manual of Standard Tibetan*. Ithaca, NY: Snow Lion Publications.

Tournadre, Nicolas & Konchok Jiatso. 2001. Final auxiliary verbs in Literary Tibetan and in the dialects. *Linguistics of the Tibeto-Burman Area* 24.1. 177–239.

Tourndre, Nicolas. 2002. The Dynamics of Tibetan-Chinese Bilingualism (Peter Brown, translator). *Perspectives chinoises* (74). 31–37.

Tsering, Tondup. 2011. *bod kyi yul skad rnam bshad (An examination of Tibetan regional dialects)*. Beijing: krung go'i bod rig pa dpe skrun khang (Chinese Tibetan Studies Publishing House).

Van Driem, George. 2014. Trans-Himalayan. *Trans-Himalayan Linguistics*. Berlin: Mouton de Gruyter 266. 11–40.

Yliniemi, Juha. 2019. *A descriptive grammar of Denjongke* (Ph.D. Thesis). Department of Modern Languages, University of Helsinki (with Sikkim University).

Yliniemi, Juha-Sakari. 2005. Preliminary phonological analysis of Denjongka of Sikkim (Master's Thesis). Helsinki: University of Helsinki.

Zhang, Menghan, Shi Yan, Wuyun Pan & Li Jin. 2019. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569(7754). 112–115.

3 Corpus study

3.1 Introduction

3.1.1 Toward phonetic predictions

The descriptions of Tibetan surveyed in Chapter 2 present a language with unique opportunities for the study of the interaction between tone, aspiration, and voicing. The approach adopted here uses phonetic data to make inferences about the representation of consonants and tones in Tibetan. In this chapter, a corpus of acoustic recordings are analyzed in terms of the primary phonetic correlates of aspiration and tone: voice onset time (VOT) and fundamental frequency (F0). The results motivate a constrained set of hypotheses regarding the articulatory timing of Tibetan consonant and vowel gestures, which are tested in Chapter 4.

3.1.1 Aspiration and VOT

Languages tend to exhibit either two, three, or four oral stop contrasts that make use of voicing and aspiration (Lisker & Abramson 1964). Four-contrast languages such as Hindi and Marathi make use of phonetically voiced, aspirated, unaspirated, and voiced-aspirated stops, while three-contrast languages like Thai and Eastern Armenian use phonetically voiced, unaspirated, and aspirated stops. However, two-category languages can either exhibit a contrast that is primarily voiced/voiceless as in Dutch and Spanish, or unaspirated/aspirated as in English and Cantonese (Lisker & Abramson 1964). These typologies are commonly explained through the co-occurence of two features, as schematized in Fig. 3.1, and this characterization will be adopted throughout this chapter. The use of separate [SG] and [voice] features follows the "laryngeal realist" approach (e.g. Halle & Stevens 1971, Lombardi 1991/1994, Iverson & Salmons 1995, Honeybone 2005). This approach uses distinctive features corresponding to stop categories, particularly [voice] for languages with a two-way voiced/voiceless contrast and [SG] for languages with a two-way aspirated/unaspirated contrast, rather than using different phonetic realizations of a single phonological feature. The laryngeal realist approach was adopted here because it can account for three- and four-category systems using only binary features (Schwartz et al 2019).

Two-way contrast (English)

		[+voice]	[-voice]
	[+SG]		aspirated
(a)	[-SG]		plain

Two-way contrast (Dutch)

		[+voice]	[-voice]
	[+SG]		
(b)	[-SG]	prevoiced	plain

Three-way contrast (Thai)

		[+voice]	[-voice]
	[+SG]		aspirated
(c)	[-SG]	prevoiced	plain

Four-way contrast (Hindi)

		[+voice]	[-voice]
	[+SG]	breathy	aspirated
(d)	[-SG]	prevoiced	plain

Figure 3.1. [Spread Glottis] and [Voice] features for four types of contrast systems, following a "laryngeal realist" perspective.

Of course, the four typologies presented in Fig. 3.1 do not include all the stops in the worlds' languages. Other series, such as implosives and ejectives, require additional features to instantiate the contrasts. Proposals include the articulatory laryngeal features of Halle and Stevens (1971) and Gallagher's (2011) [long VOT] feature that groups aspirated and ejective stops. The latter is grounded in acoustics rather than articulation, and motivated by the phonological patterns of Quechua. The tension between articulatory, acoustic, and abstract aspects of phonological representation has remained a consistent theme throughout the history of research on laryngeal phonology (e.g. Chomsky & Halle 1968, Keating 1984, Lombardi 1991, Iverson & Salmons 1995).

Phonetically, VOT has been used as a primary acoustic correlate of these contrasts, but languages instantiate the same phonological contrasts using very different VOT values (Lisker & Abramson 1964, Cho & Ladefoged 1999, Abramson & Whalen 2017, Hussain 2018). However, (Central) Tibetan is unique in that it has been described with three degrees of positive VOT length—aspirated, unaspirated, and an intermediate value—without ejective or breathy-voiced series (e.g. Denwood 1999, Tournadre & Dorje 2003, Tsering 2011). The research presented in this chapter provides the first acoustic measurements to quantify this claim.

3.1.2 Tone, F0, and VOT

Tone, and its primary phonetic exponent of F0, is closely related to consonantal laryngeal contrasts and VOT. Both F0 and VOT rely on the larynx for articulation, both are frequently involved as secondary cues for the other, and both can be reanalyzed as the other diachronically.

From an articulatory perspective, VOT and F0 are naturally related through sharing articulation at the glottis, and extensive research has investigated the effects of phonological voicing/aspiration and phonetic VOT on

F0. Broadly speaking, two major hypotheses have been proposed: vocal fold tension and aerodynamics. According to the vocal fold tension hypothesis (e.g. Halle & Stevens 1971, Ohde 1984), the vocal folds are slackened to facilitate voicing and stiffened to inhibit voicing; stiffer vocal folds vibrate faster, causing a higher F0. According to the aerodynamic hypothesis, F0 is correlated with the rate of airflow across the glottis, so F0 would rise near high-airflow productions such as aspirated stops, and lower near low-airflow productions such as voiced stops.

From an acoustic perspective, automatically-arising phonetic correlates can serve as cues to phonological contrast, as in the case of lowered F0 with voiced stops. However, Kingston and Diehl (1994) argue that speakers, aware of these associations, can marshal secondary cues to support a contrast. Taken further, this could lead to a later generation reanalyzing the cues, leading to tonogenesis, as surveyed in Sections 1.3.1 and 2.5.

The kind of precise modulation of phonetic parameters as discussed by Kingston and Diehl (1994), however, relies on speakers adjusting the details of highly-practiced articulations. In contrast, the form of enhancement discussed in Quantal Theory (Keyser & Stevens 2006, Stevens & Keyser 2010) explicitly involves the addition of an articulatory gesture. The role of enhancement in the present study is discussed in Section 3.4.1, and Section 3.4.2 analyzes the patterns in terms of gestures.

3.2 Corpus Methods

3.2.1 Participants

Data was collected from nineteen native speakers of Tibetan living in Kathmandu, Nepal, as part of a larger study of Tibetan in diaspora. Recruitment took place through social networks of Tibetans known to the author. Sixteen were born in Nepal, and three were born in the Tibet Autonomous Region (U-Tsang dialect regions: Lhasa and Kyirong) but came to Nepal as children. Eight were women and eleven were men; age ranged from 21-33 years (median 22; mean 23.8). All spoke at least some Nepali and many were fully bilingual; all also reported knowing at least some English, and several also had some knowledge of Chinese, Hindi, or another language.

3.2.2 Procedures

Interviews took place in a range of locations according to the comfort and availability of the speaker. Locations included speaker's homes, monasteries, a school, and a spare room in a Tibetan-owned apartment building. All interviews were conducted with both the author and one of two native-speaker interviewers present. As the author is not a native speaker, the interviewer was primarily responsible for interacting with the speaker, in order to facilitate communication, maximize speaker comfort, and minimize foreigner-talk. Recordings were made on a Zoom H4N recorder at 48 kHz sampling rate with an Audio-Technica ATM73a headset microphone; an Audio-Technica AT2020 microphone on a small tripod was also present, but only recordings from the ATM73a headset microphone were analyzed.

Interviews were conducted according to a standard sociolinguistic interview format, with tasks proceeding from more- to less-structured. Following basic demographic questions, the items used in this study appeared in a wordlist presented in the Tibetan orthography, for which speakers were asked to repeat each item twice. Following the wordlist, the interview continued with a choice task involving light verbs, a short reading passage, storyboard elicitations, and free speech/narration. Twenty-two words from the 64-item wordlist were used in this study. The order of items was randomized once and the same order used for all speakers.

3.2.3 Stimuli

The twenty-two items used in this study included examples of all combinations of register tone values (high and low) and word-initial aspiration values (aspirated and unaspirated). All attested places of articulation for stops were represented (bilabial, dental, retroflex, palatal, and velar), though unbalanced and in combination with varying vowels, tones, and aspiration categories, and vowels.

Taken together, 15 items were aspirated and 7 unaspirated, while 9 were high-tone and 13 were low-tone. Table 3.1 shows the distribution of these items across tone and aspiration categories: 7 items were low-tone and aspirated, 6 items were low-tone and unaspirated, 8 items were high-tone and aspirated, but only one item was high-tone and unaspirated. This distribution is depicted in Table 3.1, below.

	Aspirated	Unaspirated	Total
High-tone	8	1	9
Low-tone	7	6	13
Total	15	7	

Table 3.1. Items of interest by aspiration and voicing.

Each place of articulation was represented by four or five items, though the full four-way aspiration/tone contrast was only represented in the dental series, which comprised two minimal pairs for aspiration: the high-tone pair / tá.mák/ *rta dmag* 'cavalry' and /tʰá.mák/ *tha mag* 'cigarette,' and the low-tone pair /tǒm/ *dom* 'bear' and /dǒm/ *sdom* 'spider.' Further detail about the distribution of items is depicted in Table 3.2, below.

	Bilabial	Dental	Retroflex	Palatal	Velar	Total
High-tone, Aspirated	1	1	0	2	4	8
High-tone, Unspirated	0	1	0	0	0	1
Low-tone, Aspirated	3	2	2	0	0	7
Low-tone, Unaspirated	0	1	2	3	0	6
Total	4	5	4	5	4	

Table 3.2. Items of interest by tone/aspiration and place of articulation of initial consonant.

3.2.4 Data analysis

Measurements were taken using Praat (Boersma and Weeninck 2011). VOT values were measured from hand-selected TextGrid intervals, calculated by subtracting the time index of the first appearance of a release burst from of the beginning of periodicity in the waveform. Most VOT values were positive, though some negative values indicating pre-voicing were observed (see Fig. 3.2, below).

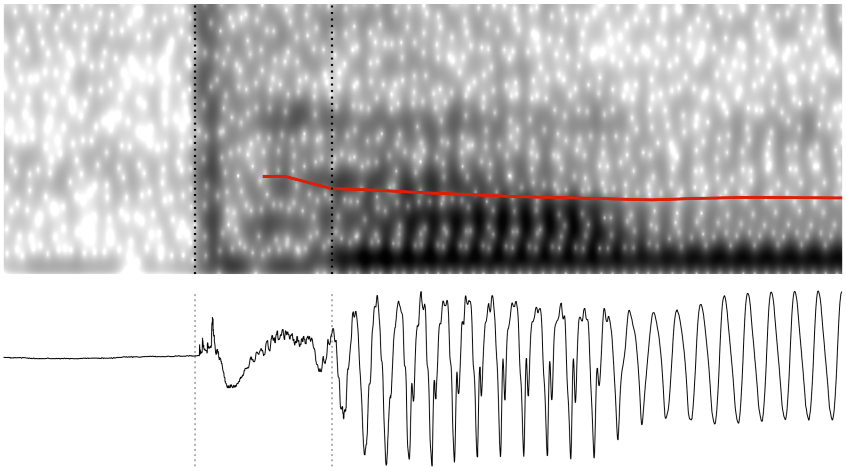


Figure 3.2. Sample spectrogram, waveform, and pitch track for the first syllable of [t^{h}á.mák] 'cavalry.' Vertical dotted lines indicate identified aspiration, and the red lines indicate the pitch track. F0 was measured at the end of aspiration, i.e. where the dotted line crosses the pitch track.

F0 was measured using Praat's built-in pitch tracker, with a time step of 0.01 seconds and a pitch range of 75-500 Hz for all speakers. Whenever possible, the pitch value recorded was that interpolated at the first period following the release burst; in most cases, this coincided with the endpoint of the VOT. When the pitch tracker had not interpolated a pitch value at this point, as well as for pre-voiced tokens, F0 was measured in the first period at which a pitch value was available.

Average values for VOT and F0 were calculated within and across the various phonological categories. Z-scores were also calculated by speaker in order to effectively compare VOT and F0 values relative to other tokens produced by the same speaker.

3.3 Results

3.3.1 F0 and Tonality

A first task was to establish the status of the tone contrast across speakers. Given the limited data analyzed so far, F0 at the onset of voicing was compared across the two tones for the various speakers. Since the high and low tones are phonetically high-level and low-rising, the contrast should be most robust at the beginning of a word; if there is a significant difference at this point, the speaker can be considered as using a tone contrast. F0 at the onset of voicing, by speaker and tone, is presented in Fig 3.3, below.

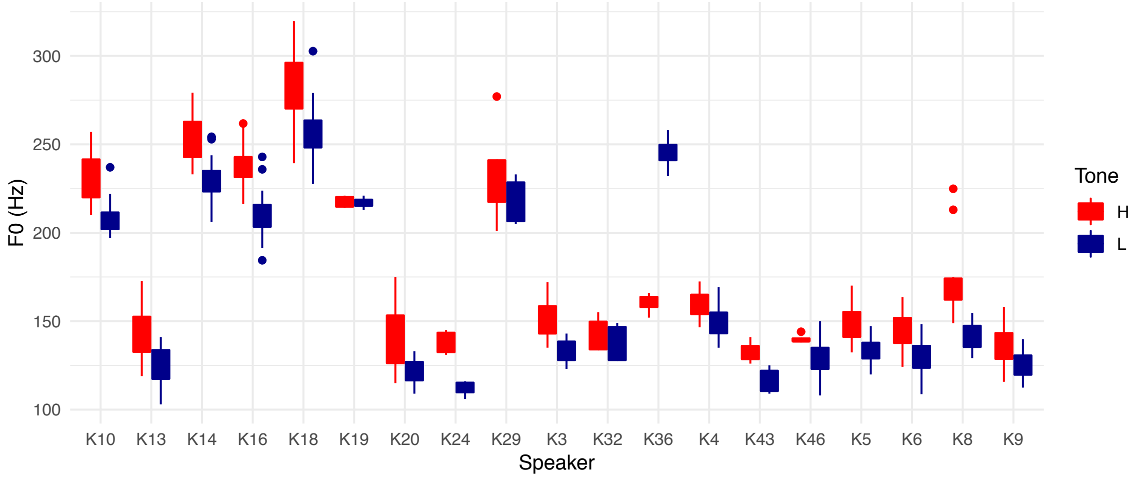


Figure 3.3. F0 at the onset of voicing by speaker and tone. Tokens presented are the same as analyzed for VOT later in this chapter. F0 values in Hertz.

Visual inspection of Fig. 3.3 shows that F0 at the onset of voicing is generally higher for high-tone words than for low-tone words, but with significant overlap and variability by speaker. In order to quantify this, a linear mixed-effects model was constructed using the *lme4* package (Bates et al. 2014) in R (R core team 2013) to predict these F0 values (in Hz). A random effect was included for word (lexical item), fixed effects for speaker and putative tone category, and an interaction of speaker and putative tone category. Post-hoc analysis (Chi-square tests with Holm-adjusted *p*-values) was conducted using the *phia* package (De Rosario-Martinez 2015) to determine which speaker*tone interactions were significant. The results are presented in Table 3.3.

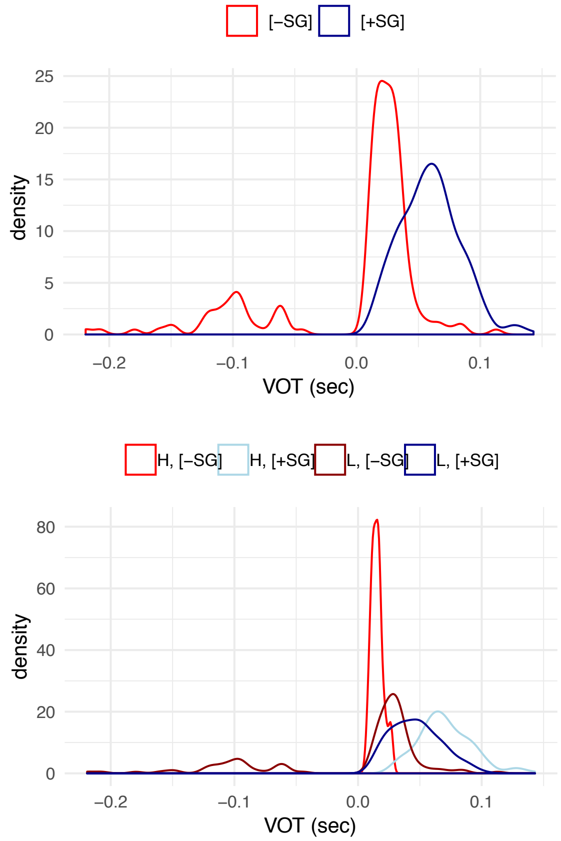
Speaker	Coefficient	Chisq	*p*-value	Speaker	Coefficient	Chisq	*p*-value
***K10**	23.144	28.5104	1.305E-06	K32	13.900	3.2099	0.1463923
***K13**	19.086	19.6039	0.0001143	K36	-76.100	96.2155	< 2.2e-16
***K14**	23.302	29.4728	8.506E-07	K4	10.931	6.3785	0.0412732
***K16**	28.916	45.0427	3.47E-10	K43	24.872	9.0594	0.0179229
***K18**	27.517	40.4372	3.249E-09	K46	19.900	6.5791	0.0412732
K19	9.400	1.4679	0.2256740	***K5**	17.356	16.0929	0.0006031
***K20**	28.900	13.8759	0.0017576	***K6**	14.513	10.6203	0.0089482
***K24**	34.900	20.2356	8.901E-05	***K8**	34.085	42.7435	1.061E-09
K29	23.400	9.0969	0.0179229	K9	12.730	8.5759	0.0179229
***K3**	18.768	19.3094	0.0001223				

Table 3.3. Post-hoc analysis of speaker*tone interactions. Chi-square values vary substantially, with larger values corresponding to larger differences between high and low tones. 11 of 19 speakers, indicated with asterisks (*) and in bold, have p > .01 and positive coefficients, indicating higher F0 for H tone than for L tone. Only one speaker, K36, has a negative coefficient, indicating unexpectedly lower F0 in H than in L.

As shown in Table 3.3, 17 of 19 speakers (all except K19 and K32) have a $p<.05$, though $p>.01$ for K29, K43, K46, and K9. Additionally, K36 shows a coefficient in the opposite direction, with H tones lower than L tones. The majority of speakers thus show evidence of the expected tone contrast, though for some speakers this may be attenuated or absent.

3.3.2 VOT

The data was hypothesized to exhibit clustering in VOT predictable by the phonological feature [Spread Glottis] (henceforth, [SG]) and High/Low tone. However, analysis of the data indicated that some stops were pre-voiced as well, as shown in Figure 3.4, below:



(a)

(b)

Figure 3.4. Density plot of voice onset time (sec) by onset category. (a) VOT by [SG]. (b) VOT by [SG] and tone.

As expected, most items exhibit positive VOT values, and [+SG] stops show longer VOT than [-SG] stops. Prevoicing (negative VOT) is observed in a subset of the [-SG] stops with low tone. Unsurprisingly, the [-SG] stops exhibit shorter VOT than the [+SG] stops. However, the [+SG] stops of Fig. 3.4(a) appear as two distinct clusters in Fig. 3.4(b): [+SG] stops with high tone show a longer VOT than their counterparts with low tone.

The prevoiced tokens were produced by different speakers, and no items were consistently voiced by all speakers. The frequency of voicing by item was as follows: 3/17 tokens of /cà.gé/ 'Chinese language'; 3/18 tokens of /cà.mí/ 'Chinese person'; 4/37 tokens of /cà.t͡ʃá/ 'pheasant'; 4/23 tokens of /tǒm/ 'spider'; 12/28 tokens of /ʈǔ/ 'barley'; and 10/24 tokens of /ʈǔk/ 'dragon', all of which were non-[SG], as well as two [SG] tokens: 1 token of /p^{h}à.t͡ʃúk/ 'cow' and 1 token of /t^{h}ǒm/ 'bear.' In both of these latter cases, the voicing may be interpreted as a reading error on the part of the speaker, since the same orthographic characters are used for both aspirated and unaspirated low-tone series. These two tokens have been excluded from all subsequent analyses. Otherwise, the voiced tokens were spread across all low-tone, [-SG] items measured. The frequency of voicing preceding retroflex consonants was somewhat higher than other places of articulation, though it is not clear that there is any principled reason for this.

Since tokens with prevoicing exhibit relatively long intervals of prevoicing, this is interpreted as a distinct alternate realization of [-SG] stops with low tone, rather than as incidental, limited vibration of the vocal folds during the closure. This alternative is available to many speakers, as 12 speakers produced at least one pre-voiced token. Treating pre-voicing as rooted in a different phonological form, three categories present themselves with regards to voicing and aspiration: "voiced" for low-tone, [-SG], pre-voiced tokens; "unaspirated" for [-SG] tokens with either tone but without pre-voicing; and “aspirated” for [+SG] tokens with either tone. The effect of this categorization on VOT is shown in Fig. 3.5:

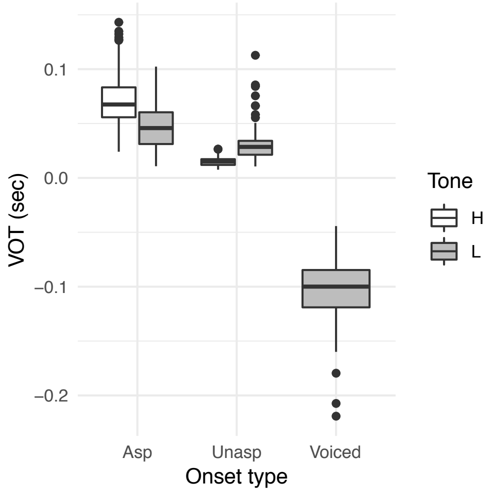


Figure 3.5. Effect of Phonological aspiration, voicing, and tone on VOT (sec). The x-axis depicts the three phonological categories of (left-to-right) aspirated, unaspirated, and voiced tokens, broken down into low-tone and high-tone tokens. Note that voicing only co-occurs with low tone.

Figure 3.5, above, shows the distribution of VOT across aspirated, unaspirated, and voiced categories. When tone category (high or low) is considered, as in Fig. 3.4(b), it becomes apparent that VOT is longer among high-tone aspirated tokens than low-tone aspirated tokens, which is in line with predictions of the gestural theory. However, there is an unexpected result of a shorter VOT for high-tone unaspirated tokens than low-tone unaspirated tokens, where no difference was predicted. It is worth noting that only one item, / tá.mák/ ‘cavalry,’ was high-tone and unaspirated.

In order to assess the robustness of these results, a series of linear mixed-effects models were fit to the data drawn from all places of articulation. The first model is a baseline model, with random effects for Speaker, Word, and Place (bilabial, dental, retroflex, palatal, and velar). The second model adds a factor for [SG], and a third model adds a factor for tone; these interact in the fourth model. The summary of the model comparison appears in Table 3.4.

(1) VOT ~ (1|Place) + (1|Speaker) + (1|Word)

(2) VOT ~ SG + (1|Place) + (1|Speaker) + (1|Word)

(3) VOT ~ Tone + SG + (1|Place) + (1|Speaker) + (1|Word)

(4) VOT ~ SG*Tone + (1|Place) + (1|Speaker) + (1|Word)

Model	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	*p*-value
Baseline	5	-2,844.69	-2,823.37	1,427.35	-2,854.69	NA	NA	NA
SG	6	-2,859.83	-2,834.24	1,435.92	-2,871.83	17.14	1	3.472E-05
SG + Tone	7	-2,863.84	-2,833.99	1,438.92	-2,877.84	6.01	1	0.0142
SG*Tone	8	-2,865.87	-2,831.75	1,440.94	-2,881.87	4.03	1	0.0447

Table 3.4. Summary of Model Comparison. All models include random effects of Place, Speaker, and Word.

Crucially, comparing these models reveals a significant effect of the interaction between SG and Tone. This interaction indicates that, for example, for a single value of SG, changing the value of Tone (from ‘0’ to ‘1’, i.e. from Low to High) leads to an improvement in the model—precisely what is predicted if High-tone conditioned a longer VOT in aspirated segments but not unaspirated segments..

3.3.3 VOT and F0

To what degree is the patterning of VOT and F0 in Tibetan under phonological control? Section 3.3.1 found that lexical tone predicts F0 at the onset of voicing for most speakers, with variation, while Section 3.3.2 found that VOT is affected by both the [SPREAD GLOTTIS] feature and by tone. This section investigates the covariation of the two phonetic parameters, VOT and F0. A scatterplot of the two, grouped by phonological categories of [SG] and tone, is presented in Fig. 3.6.

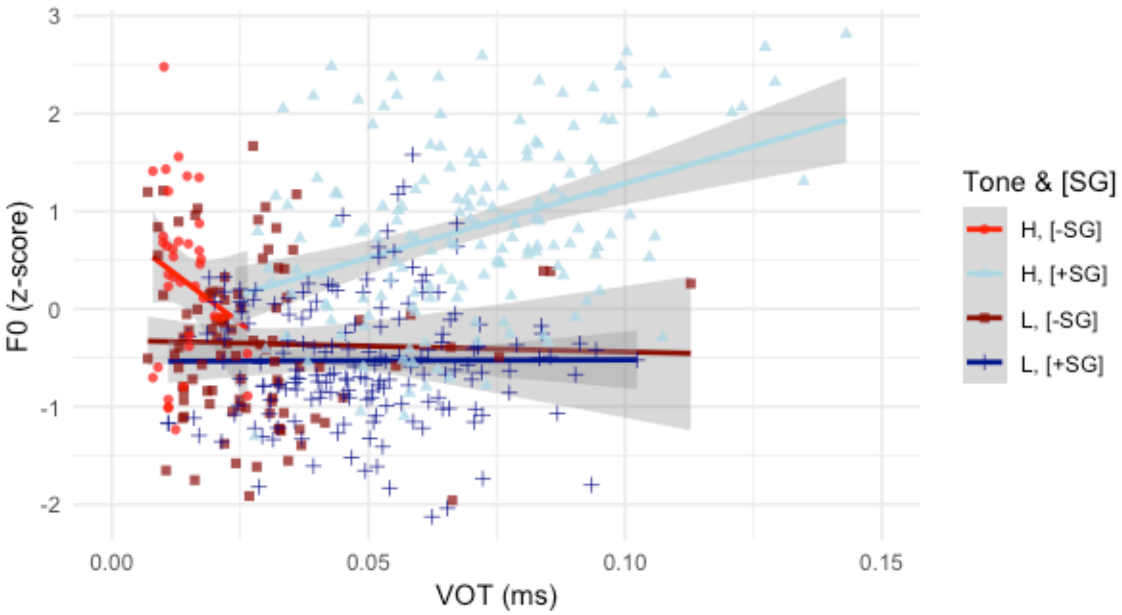


Figure 3.6. VOT and F0 (z-score by speaker) at onset of voicing. Negative-VOT items are excluded, and data from all nineteen speakers is aggregated. Linear regression lines and 95% confidence intervals are included..

The previously-established relationships are visible in Fig. 3.6. High-tone items have generally higher F0 than low-tone items, [+SG] stops have generally longer VOT than [-SG] stops, and low-tone, [+SG] stops have an intermediate VOT value. However, the covariation of VOT and F0 differs by the interaction of tone and [SG]. Only among high-tone aspirated stops is longer VOT associated with higher F0. No relation is found among the low-tone aspirated or low-tone unaspirated stops, and the trend is reversed among high-tone unaspirated stops.

3.4 Discussion

The present study investigates the relationship between VOT, tone, and F0 in Tibetan. Based on recordings of nineteen speakers, it was found that word-initial VOT varied significantly across phonological categories of both tone and aspiration. Prevoicing was only observed for unaspirated stops in low-tone words, and only variably. Positive VOT values were conditioned not only by whether the stop in question was aspirated or unaspirated, but also by the tone of the word. Among aspirated stops, VOT was shown to be longer in high-tone words than in low-tone words. However, among unaspirated stops, a small effect was observed in the opposite direction: VOT was slightly shorter in high-tone words than in low-tone words. Finally, the covariation of VOT with F0 differed according to the interaction of tone and [SG], providing evidence of different physiological mechanisms instantiating these phonological categories.

It is important to consider how the presentation of stimuli using the Tibetan orthography might affect results. Tibetan speakers are well aware that a word's spelling differs from its pronunciation, and that dialects vary in their pronunciation. The orthography is still related to pronunciation, however, and so bias is possible. Research on other languages indicate that orthographic factors can induce effects resembling phonetic/phonological processes such as incomplete neutralization (Warner et al. 2006). The context where orthographic effects seem most plausible is low-tone stops, which are all written with the same letter (the historically voiced series *b d g*) irrespective of aspiration.

Indeed, the aggregated VOT data presented in Fig. 3.4(b) and Fig. 3.5 show that low-tone aspirated stops have a shorter VOT than high-tone aspirated stops—that is, that low-tone aspirated stops fall between the other aspirated stops and the other low-tone stops. An orthographic explanation is unlikely for two reasons. First, the prevoicing observed on some stops only ever occurs with the low-tone unaspirated stops; speakers do not produce prevoicing on the low-tone aspirated stops even though they are written with the same graphemes. Second, interspeaker variation is likely, and this topic is explored in Chapter 4 using the larger number of tokens per speaker in the EMA experiment (see FIg. 4.5). Therefore, the apparent intermediate value is the result of data aggregation necessitated by the small number of tokens per speaker. Perhaps orthography has contributed to the fact that some speakers maintain short VOT for the low-tone "aspirated" stops, but this should be understood as a historical rather than synchronic factor. The long VOT with low tone is a relatively recent sound change (see section 2.9), so it is the speakers with long-VOT low-tone stops who have undergone a change. It is conceivable that the orthography contributed to other speakers not adopting this change, though the fact remains that the orthography does not appear to influence the synchronic phonology of the speakers in this study.

3.4.1 Enhancement account

Could the observed effects of VOT on tone be explained with reference to phonetic enhancement? According to the Quantal Theory account of enhancement (Keyser & Stevens 2006, Stevens & Keyser 2010), enhancement consists either of adding a gesture either to increase the perceptual saliency of a contrast, or to introduce a new parameter to support the existing contrast.

The latter provides an account of the prevoicing for unaspirated stops with low tone: the addition of prevoicing furnishes an additional phonetic parameter to aid the perception of the unaspirated + low tone items. Since no other word-initial stops are prevoiced, prevoicing helps distinguishes these items both from high-tone unaspirated stops and low-tone aspirated stops. The optionality of prevoicing is also consistent with its role in enhancement. The use of voicing as this parameter follows from the history of the contrast: the subphonemic lowering of F0 after voiced stops was reanalyzed as a tone contrast, but the voicing was able to remain and be recruited for the purpose of enhancement.

Why, then, would tone cause a difference in aspirated stops? Before tonogenesis, the words now produced with low tone and aspirated stops were voiced, but this voicing has been entirely lost among these speakers and those of other Central Tibetan dialects. Low-tone aspirated stops risk confusion with their low-tone unaspirated and high-tone aspirated counterparts; the high-tone unaspirated stops are already distinguished along two parameters. Languages may develop so this contrast is staggered along the VOT axis. Low-tone unaspirates are already at near-zero (or negative) VOT, so introducing a longer VOT would support the tone contrast; this appears to be the diachronic origin of aspiration with low tone. However, to avoid confusion with the high-tone aspirated stops, the low-tone aspirated stops could be produced with an intermediate value, longer than the unaspirated stops but not as long as the high-tone aspirated stops.

While perceptual distinctiveness motivates the intermediate VOT of low-tone aspirated stops, a complete account requires an additional mechanism to explain how the contrast is articulated. The examples used to illustrate enhancement by Keyser and Stevens (2006) involve the addition of a gesture, but it is not clear in this case what such a gesture would be. Instead, the difference could be one of intergestural timing.

VOT is not an articulatory measure; it is an acoustic consequence of articulatory timing. If tonal gestures are timed with onset consonant and vowel gestures in a similar way to the second member of a consonant cluster (as in Gao 2008, Karlin 2014, Hu 2016; see sections 1.4 and 4.1.1), this should cause the consonant gesture to begin earlier in time relative to the following vowel. No change in VOT would result if the glottal spreading gesture moved along with the oral consonantal gesture. However, in other languages it has been observed that a single glottal spreading gesture might be shared across a consonant cluster, overlapping with the oral gestures of multiple consonants. It is thus possible to hypothesize that the effect of a tone gesture on consonant-vowel timing might cause a VOT difference in aspirated stops.

The type of acoustic evidence presented in this chapter is not sufficient to test this hypothesis. If the effect of tone on VOT is mediated by consonant-vowel timing, articulatory evidence of consonant-vowel timing would be required. Chapter 4 presents the EMA study conducted to gather this evidence. Further implications for theories of temporal coordination will be discussed in Chapter 5.

3.4.2 Gestural scores

The results presented in Section 3.3 describe how VOT is conditioned by tone in Tibetan. In this section, those results are interpreted in terms of gestural scores, a mechanism for relating phonetic production with phonological representation. (More detail on intergestural coordination can be found in sections 1.4 and 4.1.1.)

Broadly speaking, three main aspects of a gestural score can be invoked to explain the differences in VOT. First, the gestures themselves might be different: different glottal gestures correspond to different laryngeal postures, such as a spread glottis (for aspiration), a critical opening (for voicing), and a partly open glottis (for voiceless unaspirated stops) (Esling & Harris 2003, Edmondson & Esling 2006). Second, the gestures may differ in in their duration. Third, the gestures may differ in temporal coordination with each other. In the remainder of this section, two possible accounts of the Tibetan data are discussed: one based on differing gestural activation durations, the other two based on different gestural coordination. The four Tibetan VOT categories according to the gesture duration account are presented in Fig 3.7(a-b) and (c-d), while those according to the gestural coordination account are presented in Fig. 3.7(a-b) and (e-f).

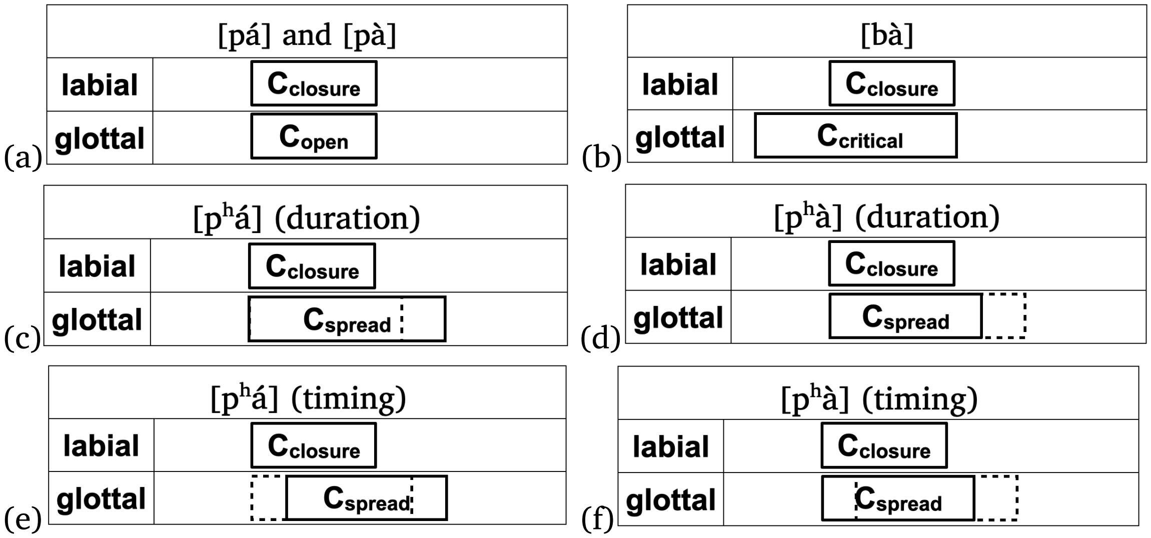


Figure 3.7. Partial gestural scores for Tibetan onset stops. For each pair of (c)-(d) and (e)-(f), dotted lines represent the temporal arrangement of the glottal gesture in the other member of the pair. (a) Short-VOT stop features glottal opening synchronous with oral closure. (b) Negative VOT stop features critical glottal gesture for prevoicing (this is the prevoiced alternant of the short-VOT stop with low tone). (c-f) The long-VOT and intermediate-VOT stops feature a glottal spreading gesture that could be (c) longer in the high-tone long-VOT stop and (d) shorter in the low-tone intermediate-VOT stop. Alternatively, the glottal spreading gesture could be the same duration in both cases, but (e) begin after the start of the oral closure for the long-VOT stop and (f) begin synchronous with the oral closure for the intermediate-VOT stop.

The gestural scores presented in Fig. 3.7 correspond to the four onset-tone categories of Tibetan. As in a Thai-style 3-way VOT contrast, three glottal gestures are used: a phonation-inhibiting glottal opening gesture for the short-VOT stops in Fig. 3.7(a), a critical glottis gesture for the voiced stop variant in Fig. 3.7(b), and a glottal spreading gesture for the long- and intermediate-VOT stops in Fig. 3.7(c-f). Fig.3.7(c-d) account for the difference between long and intermediate VOT as the result of different activation durations of the glottal spreading gesture. As the gesture is active for longer in 3.7(c) than in 3.7(d), the resulting VOT is longer. Alternatively, Fig. 3.7(e-f) accounts for the difference as a result of gestural timing. The glottal spreading gesture has the same activation duration for both stops; however, it begins and ends later in (e) than in (f), resulting in a longer duration of this gesture after the conclusion of the oral closure, which produces a longer VOT.

How might these accounts be evaluated? The most straightforward test would directly image the glottis to observe the nature and timing of its movements. However, this was found not to be feasible due to to the physical and technical difficulty of imaging the larynx. Instead, indirect tests are required. The remainder of this section presents and evaluates several predictions of the glottal gesture and glottal duration accounts.

In terms of phonology, the descriptions of Central Tibetan reviewed in Chapter 2 (including Denwood 1999, Tournadre & Dorje 2003, and Tsering 2011) group the stops and affricates into “aspirated” and “unaspirated” categories, both of which occur with high and low tone. In these descriptions, the different VOT by tone is a matter of surface-level phonetics, not the underlying contrast. This is most similar to the gestural timing account, which uses the same gesture—a unit of phonological contrast—for the high-tone long-VOT and low-tone intermediate-VOT stops. The different temporal coordination, then, instantiates the surface-level difference. The phonological descriptions are not consistent with the gesture duration account, as its four different gestures effectively posit four consonants rather than three. This abandons the attempt to maximize parsimony and even relegates tone to a redundant status. Previous phonological literature is thus more consistent with the temporal coordination account.

Another line of evidence comes from typological comparison. The world’s languages vary tremendously in the duration of their VOT contrasts (e.g. Lisker

& Abramson 1964, Cho & Ladefoged 1999), but a language with three positive VOT contrasts remains unattested. It is thus more consistent with the typological literature to derive the intermediate-VOT stops of Tibetan from another category of stops, as in the gestural timing account. Again, the gesture duration account would effectively create another class of consonants, which would be unique among known languages.

Diachronically, the intermediate-VOT stops are derived from the historically-voiced simplex onsets (see Section 2.5). As such, the voicing was reanalyzed as low tone, but the speakers at the time of this tonogenesis would not have heard this series produced with long VOT. This means that the low-tone intermediate-VOT stops would have passed through a stage where they had lost prevoicing, but not yet developed longer VOT—a stage retained in many Eastern (Kham) dialects such as Dege (Tsering 2011), Bathang (Tsering 2011), and Thebo (Lin 2014). (In these dialects, contrast with the low-tone unaspirated series is maintained by consistent voicing in the latter, rather than the variable voicing in Central Tibetan). How would this series have gained a longer VOT? For reasons of enhancement (see Section 3.4.1), a variant featuring a longer VOT and glottal spreading gesture could have increased distinctiveness and spread, particularly if prevoicing was lost or became a less reliable cue. The difference between this series and the long-VOT stops would be maintained by tone, and further enhanced by a different glottal gesture or temporal coordinaiton. In light of these historical developments, both duration and timing accounts are diachronically plausible.

Finally, the two accounts generate articulatory predictions that can be tested as indirect evidence for the glottal gestures. In particular, the gestural timing account involves different timing of the glottal gesture in the intermediate- and long-VOT stops, which could be associated with timing differences in other gestures as well. Given the implausibility of directly observing glottal gestures, a mechanism is needed to generate predictions for oral gestures instead. The competitive coupling model of tone surveyed in Section 1.4.3 (Gao 2008) furnishes these predictions.

The coupling graphs in Fig 3.8. depict possible sets of coupling relations among the following types of gestures in a CV syllable: oral consonantal (C), glottal consonantal (G), vowel (V), and tone (T). Fig. 3.8(a) depicts the coupling relations in a toneless CV syllable as per Goldstein et al. (2009): in-phase C-V and C-G coupling[2] leading to simultaneous start times of the three gestures. The kind of tonal syllable presented in Gao (2008) is shown in Fig. 3.8(b): it retains the coupling relations of Fig.3.8 (a), adding in-phase V-T and anti-phase C-T coupling to model partial overlap. Finally, Fig. 3.8(c) represents an alternative structure for the tonal syllable: the glottal gesture here has a second in-phase coupling relation with the tonal gesture, reflecting the cluster-like relationship of consonant and tone and sharing a glottal gesture across such a cluster.

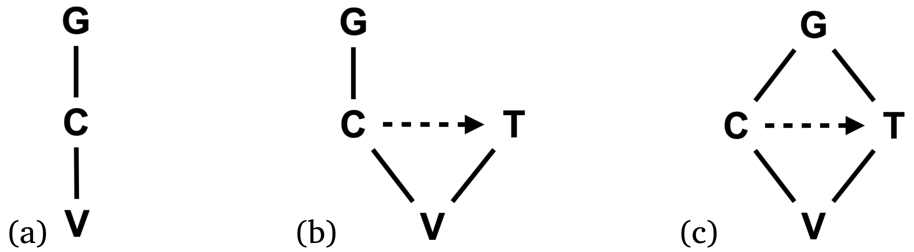


Figure 3.8. Predicted coupling graphs with competitive coupling of tone. C refers to oral consonant gesture; V refers to oral vowel gesture; T refers to tonal gesture; G refers to glottal gesture. Solid lines indicate in-phase coupling and dotted arrows indicate anti-phase coupling. (a) Mandarin-like tonal syllable. (b) Tonal syllable with

[2] In-phase C-G timing is used for voiceless and aspirated stops (e.g. Goldstein et al. 2009); voiced stops and other consonants require different treatment, but the present discussion concerns only the voiceless stops of Tibetan.

glottal gesture coordinated in-phase to both consonant and tone gestures. (c) Syllable without tone gestures.

The three coupling graphs of Fig. 3.8 offer three scenarios for the relationship between tonality, VOT, and C-V lag. Fig. 3.8(a) is a CV syllable with no tone gesture, and predicts in-phase C-V timing. Fig. 3.8(b) and Fig. 3.8(c) are CV syllables with tone, and predict the C gesture will begin before the V gesture, a difference known as C-V lag. This C-V lag should also covary with the duration of the C gesture as a result of anti-phase timing (Shaw et al. 2019). Where these two differ is the timing of the glottal gesture: in Fig. 3.8(b) the C and G gestures begin simultaneously, while in Fig 3.8(c) the C gesture begins before the G gesture. As a result, the difference in time between the end of the C and G gestures is longer for Fig. 3.8(c) than for Fig. 3.8(b)—a difference corresponding to a longer VOT for Fig. 3.8(c).

Finally, these coupling graphs and their predictions are applied to the partial gestural scores in Fig. 3.7, resulting in the more complete gestural scores in Fig. 3.9, below. The two accounts presented above can now be evaluated on the basis of phonetic predictions.

The gestural duration account, whose gestural scores are presented in Fig. 3.9(a-d), relies on the original gestural model of tone coupling graph, Fig. 3.8(b). As a result, all syllables are predicted to exhibit C-V lag that varies dynamically with C duration.

The gestural timing account, presented in Fig. 3.9(e-h), is instead based on the coupling graphs of Fig. 3.8(a) and (c). Here, C-V lag that covaries with C duration is predicted for three of the four conditions, but not the low-tone intermediate-VOT condition. The gesture timing difference between long-VOT and intermediate-VOT conditions in Fig. 3.9(g-h) is caused by the presence of a high tone gesture, but with no specified low tone gesture. Therefore, the high-tone syllables are predicted to exhibit a C-V lag (covarying with the C gesture duration), but the low-tone syllables are predicted to exhibit C-V simultaneity.

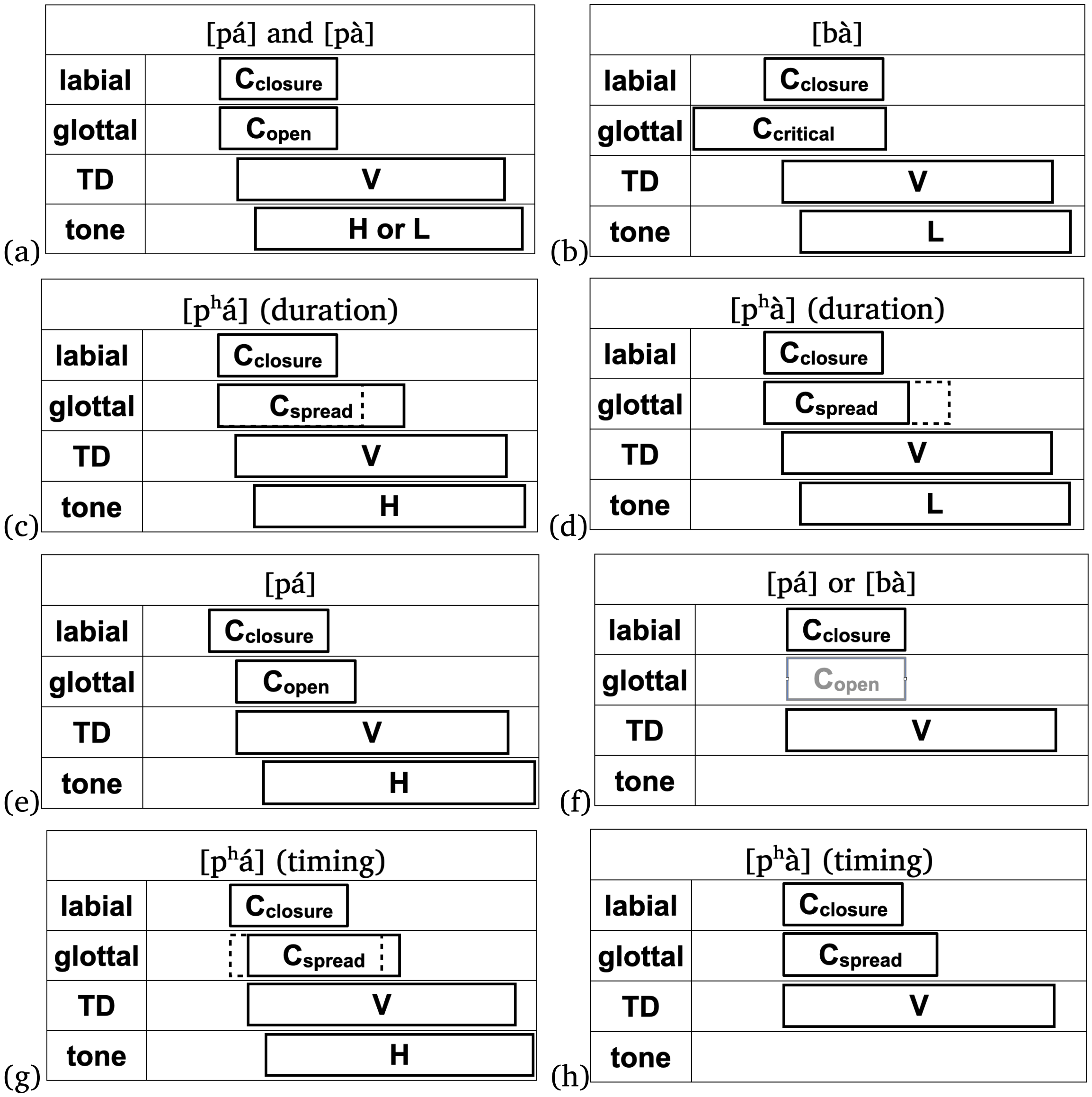


Figure 3.9. Gestural scores for Tibetan onset stops. (a-d) Gestural scores according to gesture duration account: (a) Short-VOT stop with C-G synchrony and C-V lag; (b) Negative-VOT stop with C-V lag; (c) High-tone long-VOT stop with long glottal spreading gesture; (d) Low-tone intermediate-VOT stop with shorter glottal spreading gesture. (e-h) Gestural scores according to gestural timing account. (e) High-tone short-VOT stop with C-V lag; (f) Low-tone short-VOT stop (as shown) or negative-VOT stop with C-V synchrony; (g) High-tone long-VOT stop with C-V lag (dotted lines

indicate timing of glottal gesture without competitive coupling); (h) Low-tone intermediate-VOT stop with C-V simultaneity.

The key difference in the gestural scores in Fig. 3.9 lies in C-V timing. Under the gesture duration account of Fig. 3.9(a-d), all words are predicted to exhibit similar, synchronous start times of C and V gestures. Under the gestural timing account of Fig 3.9(e-h), the timing would differ across tones: high-tone words would show a longer C-V lag than low-tone words due to the presence of a specified high-tone gesture. These predictions form the basis of the EMA experiment developed in Chapter 4.

The basic predicted difference is that C-V lag will be similar across tones for the gesture duration account, but will differ by tone under the temporal coordination account. This difference also applies to some variations of these accounts. For example, the version of the gesture duration account presented relies on the coupling graph in Fig. 3.8(b); if instead the coupling graph in Fig. 3.8(c) is used, the C-V timing would not be synchronous, but would still be consistent across tones. Likewise, the gesture duration account also does not predict differences in C-V lag by tone. However, such possibilities are not consistent with the established observation that VOT differs by tone in the aspirated stops.

The above discussion relies on an "H-L" or "H-LH" analysis of tone (see Section 2.10). If instead an "H-∅" analysis is used, C-V lag would be predicted to differ across tones because the ∅-tone condition, lacking a gesture, would remove the gesture coordinated anti-phase, and the remaining coupling relations would all be in-phase. Finally, the "L-∅" analysis is less plausible on phonological grounds, but it would result in longer VOT with *low-tone* words, the opposite of the results presented in Section 3.3. The relationship between the accounts and predictions are summarized in Table 3.5, below.

Onset series	**Tone**	**Glottal gesture (duration)**	**Glottal gesture (timing)**
short~negative VOT	L	open~critical	open~critical
short VOT	H	open	open
intermediate VOT	L	spread (shorter)	spread (earlier)
long VOT	H	spread (longer)	spread (later)
Predictions consistent with:			
Phonological description		No	Yes
Typology		No	Yes
Diachronic plausibility		Yes	Yes
Articulatory prediction		**C-V timing not different (simultaneous) by tone**	**C-V lag longer with H tone than with L tone**

Table 3.5. Summary of glottal gesture and predictions of proposed accounts.

The predictions presented here have dealt with the basics of gestural timing, but have abstracted away from a substantial amount of detail. For example, the timing-based accounts include gestural scores where the C closure begins before the glottal opening/spreading gesture. This would predict a short period of voicing leakage at the beginning of these consonants, which has not yet been observed. As for the tone gestures, this discussion only investigates them inasmuch as they interact with the other gestures, and does not make particular claims about the targets (e.g. F0 trajectories) of these gestures. Given the rising F0 trajectory of the low tone, the H-∅ analysis may still require a high tonal target in the "∅" or low-tone condition, perhaps coupled anti-phase to the vowel. All gestures were treated as unitary entities, though the coupling graphs and gestural scores could be constructed in a number of different ways, such as with the split-gesture hypothesis (Nam 2007). Nevertheless, the current framing is sufficient to establish viable hypotheses for the EMA experiment conducted in Chapter 4.

3.5 Ongoing corpus development

The analysis of this chapter has touched on only a small portion of the corpus: just the wordlist data from the nineteen diaspora-raised speakers. The remaining portions of these speakers' interviews, which include spontaneous-speech data, has not been analyzed, nor has the data from the other speakers. Thanks to a Doctoral Dissertation Research Improvement Grant from the National Science Foundation, I have been able to hire a native Tibetan speaker to transcribe the interviews, and another research assistant to help with forced-alignment of the corpus. This will allow the analysis of this chapter to be extended to a larger and more naturalistic set of data, as well as allow comparison across speakers of other dialects also living in Kathmandu.

3.6 Chapter summary

This chapter investigates the relationship between Tibetan speakers' phonetic parameters of F0 and word-initial VOT, and phonological contrasts of tone and aspiration. With a two-way contrast in tone and a two-way contrast between aspirated and unaspirated stops, it was found that tone affected VOT. Aspirated stops in high-tone words had a longer VOT than aspirated stops in low-tone words. Prevoicing was present, in a variable manner, only for unaspirated stops with low tone. The pattern of prevoicing was explained as a phonetically-natural and diachronically-plausible enhancement gesture. However, the tonal interaction that causes three positive VOT lengths is more difficult to explain. Two accounts are presented: one based on different gesture activation durations, the other based on different gestural timing caused by competitive coupling between consonant and tone gestures. The two accounts differ in their articulatory predictions: the first predicts no effect of tone on the timing of consonant and vowel gestures, while the second predicts different tones could affect C-V timing. An EMA study testing these predictions is described in chapter 4.

3.7 Chapter bibliography

Abramson, Arthur S. & Douglas H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of phonetics*. Elsevier 63. 75–86.

Boersma, Paul & David Weenink. 2018. Praat: Doing phonetics by computer [Computer software]. Version 6.0. 43.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27(2). 207–229. https://doi.org/10.1006/jpho.1999.0094.

De Rosario-Martinez, Helios, John Fox, R. Core Team & Maintainer Helios De Rosario-Martinez. 2015. Package 'phia.' *CRAN repository*. *Retrieved* 1. 2015.

Gallagher, Gillian. 2011. Acoustic and articulatory features in phonology – the case for [long VOT]. *The Linguistic Review* 28(3). https://doi.org/10.1515/tlir.2011.008. https://www.degruyter.com/doi/10.1515/tlir.2011.008 (1 September, 2020).

Gao, Man. 2008. Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics), Yale University, CT*.

Hu, Fang. 2016. Tones are not abstract autosegmentals. In *Speech Prosody*, 302–306.

Hussain, Qandeel. 2018. A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. *Journal of Phonetics*. Elsevier 71. 284–305.

Iverson, Gregory K. & Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12(3). 369–396. https://doi.org/10.1017/S0952675700002566.

Karlin, Robin. 2014. The articulatory TBU: gestural coordination of tone in Thai. In *Cornell Working Papers in Linguistics*.

Keyser, Samuel Jay & Kenneth N. Stevens. 2006. Enhancement and Overlap in the Speech Chain. *Language* 82(1). 33–63. https://doi.org/10.1353/lan.2006.0051.

Kingston, John & Randy L. Diehl. 1994. Phonetic Knowledge. *Language* 70(3). 419–454. https://doi.org/10.1353/lan.1994.0023.

Kingston, John, Randy L. Diehl, Cecilia J. Kirk & Wendy A. Castleman. 2008. On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics* 36(1). 28–54. https://doi.org/10.1016/j.wocn.2007.02.001.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*. Taylor & Francis 20(3). 384–422.

Ohde, Ralph N. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America* 75(1). 224–230. https://doi.org/10.1121/1.390399.

Stevens, Kenneth Noble & Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38(1). 10–19. https://doi.org/10.1016/j.wocn.2008.10.004.

Warner, Natasha, Erin Good, Allard Jongman & Joan Sereno. 2006. Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34(2). 285–293. https://doi.org/10.1016/j.wocn.2004.11.003.

4 EMA study

4.1 EMA study introduction

4.1.1 C-V lag

It has been observed that the articulation of word-initial consonants and vowels begin approximately simultaneously, overlapping with vowels which show minimal overlap with each other (Öhman 1966, Fowler 1983). Even though articulation is near-simultaneous, the listener can still hear sequential output in the acoustics because the consonant gesture is shorter than the vowel, so the acoustic consequence of the consonant is heard first, followed by that of the vowel. However, consonant and vowel gestures are not always simultaneous, notably in two contexts: consonant clusters and lexical tone.

In clusters, the two (or more) consonant gestures partially overlap with each other and with the vowel. Onset clusters, in particular, are timed to their associated vowel such that the mean of the midpoints of the onset cluster is timed to the start of the vowel, a finding termed the "C-Center effect" (e.g. Browman and Goldstein 1988). In addition to basic type of temporal coordination, the details of degree of gestural overlap have been explained with reference to the coupling strength of particular gestures (Browman and Goldstein 2000), which has been argued to vary by the type of consonant (Pastätter and Pouplier 2017) or other factors such as place of articulation (Mücke et al 2020).

In Articulatory Phonology, words are composed of a set of gestures, dynamic actions in the vocal tract that unfold over time and are coordinated with other gestures (Browman and Goldstein 1986 et seq.) Each gesture is modeled as an oscillator (Saltzman and Byrd 2000), and these oscillators are coupled with each other, most commonly in one of two coupling modes: in-phase and anti-phase (Saltzman and Byrd 2000, Nam and Saltzman 2003). In-phase coupling, which dictates that gestures that start simultaneously, is assigned in the theory to a syllable onset and nuclear vowel (C-V coupling) in order to explain the simultaneous onset of C and V gestures. Anti-phase coupling, which dictates that gestures start sequentially, is likewise assigned in the theory to sequences of consonants and for nuclear vowels with following consonants (C-C and V-C coupling). Coupling specifications in competition with each other dictate intermediate timing relations; for example, the C-center effect can be derived from the competition between in-phase coupling of consonant gestures to a vowel gesture and anti-phase coupling among the consonant gestures (Nam and Saltzman 2003).

The type of cluster timing predicted by the competitively coupled oscillator model—starting a vowel gesture at the center of an onset cluster—has been observed in a variety of languages. Aside from English, these include French (Kühnhert et al. 2006), Polish (Pastätter and Pouplier 2017), and Georgian (Byrd & Chitoran 2002, Goldstein et al. 2007, Kwon & Chitoran 2018). However, simultaneous C-V timing may be preserved in some clusters. Hermes et al. (2008, 2012) showed that while Italian onset clusters generally follow English-like C-center organization, /s/ as the first member of a cluster does not cause the rightward shift of other consonants or CC clusters. This is interpreted to mean that that the prevocalic consonant in these clusters is timed to the vowel as an onset, but the /s/ is not. Romanian also shows English-like C-center timing in sibilant-stop clusters, with variation in timing of other clusters affected by segmental identity and frequency, but Italian-like /s/+C clusters show different timing (Hermes et al. 2008, Hermes et al 2012). In Moroccan Arabic, Shaw et al. (2009) found the timing of vowels to immediately prevocalic consonants was unchanged when preceded by additional consonants, perhaps indicating a heterosyllabic parse of these clusters. Similar results have been obtained for Tashlhyit Berber (Goldstein et al. 2007, Hermes et al. 2017). While these cases do not show C-center timing, they can be accounted for in the coupled oscillator model by treating only the immediately prevocalic consonant gesture as in-phase coordinated to the vowel gesture. The remaining consonants are timed anti-phase to the prevocalic consonant, and no competition among coupling modes takes place.

Lexical tone has been posited as a second environment that conditions non-simultaneity in C-V timing. While near-simultaneous C-V timing has been observed in European languages without lexical tone such as German and Italian (Niemann et al. 2011) and Catalan (Mücke et al. 2012), a later start of the vowel relative to the consonant has been observed in languages with lexical tone, namely Mandarin (Gao 2008), Thai (Karlin 2014), and Lhasa Tibetan (Hu 2016). This difference in timing, reflected in the measure of “C-V lag” has been measured in the lab to be around 50 ms for the lexical tone languages and near zero for the non-tonal languages. In addition to the cross-linguistic evidence, contextually-toneless syllables in Mandarin show reduced C-V lag relative to their fully-tonal counterparts, indicating that tone conditions C-V lag within a language (Zhang, Geissler, and Shaw 2019), not just across languages. Differences in C-V lag across tones in Mandarin were documented by Gao (2008), but not found to be significant by Shaw and Chen (2019).

This evidence indicates that lexical and intonational tones behave differently with regards to gestural timing, in that lexical tones cause non-simultaneous C-V timing and intonational tones do not. Gao (2008) explains this by positing a tone gesture coupled in-phase to the vowel and anti-phase to the onset consonant (though note that Silverman (1995) treats Comaltepec

Chinantec tones as gestures for the specific purpose of adjusting their timing for auditory recoverability). The competitive coupling between anti-phase consonantal and tonal gestures, both of which are coupled in-phase to the vowel, causes the consonant gesture to begin earlier than the vowel gesture (see Fig. 4.1). In this chapter, we refer to this coordinative structure as the competitive-coupling model of gestural tone. Intonational tone gestures, which do not affect C-V timing, are treated by Katsika et al (2014) as coupled in-phase to the vowel gesture, with no coupling to the consonant gesture. Katsika et al. note that the post-lexical status of intonational tones may be related to this difference from lexical tones.

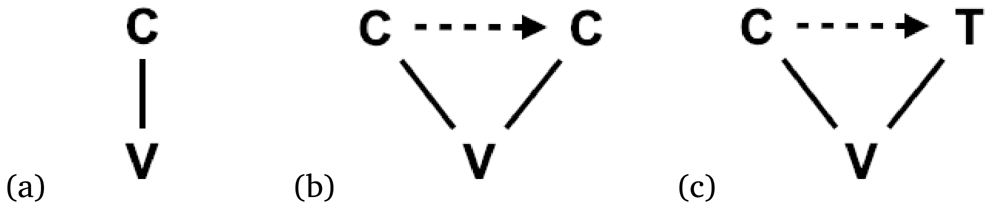


Fig. 4.1. Coupling graphs of CV and CCV syllables with tonal CV. Solid lines indicate in-phase coupling; dotted arrows indicate anti-phase coupling. (a) CV syllable without lexical tone; (b) CCV syllable without lexical tone; (c) CV syllable with tone.

This chapter introduces a third line of evidence for the effect of tone on C-V timing: comparison between speakers of the same language who do and do not use tone. This requires a speech community where some members exhibit a lexical tone contrast and others do not. One such example is the Tibetan-speaking population of the Tibetan Diaspora. Speakers from many dialect regions moved to new communities in Nepal and India beginning in 1959. With diverse origins, the Tibetan spoken by the descendants of these migrants includes a mix of features that does not directly reflect any one specific dialect

(Geissler 2018). While the numerically- and socially-dominant Central Tibetan varieties (including the Lhasa dialect) use lexical tone, some other dialects are non-tonal, and as shown below, the participants in this experiment include both tonal and non-tonal speakers.

We use electromagnetic articulography (EMA) to investigate C-V timing in Tibetan. We begin by identifying which participants do and do not contrast lexical tone. This allows us to ask a first question: is C-V lag different for speakers with and without lexical tone contrasts? Next, building on work showing an association between lexical tone and longer C-V lag across languages and across words within a language, we hypothesize that the C-V lag will be longer for speakers contrasting tone than for those who do not contrast tone. In other words, this investigates whether C-V lag differs systematically across individual speakers or of the speech community as a whole. The second aim of this study furthers this inquiry by asking whether C-V lag varies at the level of phonemic contrasts. This is done by comparing C-V lag across words beginning with aspirated and unaspirated stops. Since the voicelessness during aspiration prevents the realization of F0, lengthening C-V lag for aspirated stops could enhance the perception of the tone contrast, as discussed in section 4.1.2. Finally, we look to the covariation between consonant duration and C-V lag to test the kind of coupling modes present among these gestures.

4.1.2 Perceptual factors in C-V timing

The competitive-coupling model of tone gestures predicts a certain range of C-V timing values resulting from in-phase and anti-phase coupling modes and the competition among conflicting couplings of consonant, vowel, and tone gestures. However, listener-oriented perceptual factors may also play a role in C-

V timing. The role of perceptual salience in gestural timing was prominently explored by Silverman (1995). In his comparative study of the relative timing of laryngeal and supra-laryngeal gestures, he describes evidence for perceptual recoverability of gestures across languages. Perceptually sub-optimal patterns of gestural timing only emerged when perceptually optimal patterns are also present; for example, sub-optimal preaspirated stops only occur in languages with more-optimal post-aspirated stops. For Silverman, the drive for perceptual salience exists in tension with a general trend toward overlapping, parallel production of gestures; it is the need for perceptual recoverability that results in other timing patterns.

Related work on perception and gestural overlap by manner of articulation in Tsou was conducted by Wright (1996). While continuants provide acoustic information during the peak constriction, the cues necessary to distinguish a stop are found in formant transitions and the release burst. Therefore, Wright predicted that (1) speakers would produce initial stop-stop clusters with less overlap, in order to produce an audible burst for the first stop; (2) continuant-continuant clusters could show more overlap, since they contain internal cues; and (3) word-medial stop clusters could show more overlap, because transitions into the first stop provide cues as to the identity of the stop. Data from Tsou matched these predictions: stop release bursts were audible in nearly all initial stop-stop clusters but only 1/3 of word-internal clusters, and overlap was greater for continuants than for stops.

In the so-called “place order” effect, clusters with front-to-back ordering (e.g. /gd/, /gb/, or /db/) exhibit a greater degree of overlap than clusters with back-to-front ordering (e.g. /dg/, /bg/, or /bd/). This was found for Georgian stop + stop clusters by Chitoran et al. (2002), who explain it as the consequence of perceptual recoverability: since the first consonant in a stop-stop cluster requires a release burst in order to be acoustically identifiable, a back-to-front cluster must provide sufficient time for the posterior consonant to release before the anterior consonant closure begins. In a related observation, word-internal stop-stop onset clusters showed a greater degree of overlap than word-initial clusters; greater overlap was permitted because transitions into the first consonant would be audible word-interally, thus reducing the importance of an audible release. Chitoran et al. (2002) account for these perceptually-motivated patterns by positing that the coupling strength of C-C sequences must be able to vary based on order and position of the gestures.

While stops require transitions and/or a release burst for accurate perception, the cues for fricatives and sonorants are present during the consonant construction itself (Wright 1996). This would predict that the place-order effect would not be observable on stop + /l/ and stop + /n/ clusters, but Kühnhert et al. (2006) show just this in French. They suggest a biomechanical rather than perceptual explanation: since the velar stops share use of the tongue with these coronal sonorants while the labial stops do not, the labial consonants were more free to begin movement sooner. More free, early movement of bilabial stops reduced overlap. However, perceptual factors may still play a role in the timing of these clusters: observing less overlap for stop + /n/ clusters than stop + /l/ clusters, they posit that beginning nasal airflow too early would compromise perception of the stop.

These differences can be accounted for in Articulatory Phonology with reference to coupling strength. Rather than different coupling relations, different coupling strengths can cause different degrees of overlap (Saltzman and Byrd 2000). If coupling strength can differ across consonantal clusters to enhance perceptual recoverability, the same may be possible for the timing of tone gestures. F0, the primary acoustic correlate of tone, necessarily requires voicing in order to be realized. A syllable consisting of a voiced sonorant onset and vowel allows F0 to be heard throughout the syllable; however, a syllable with a voiceless consonant, and especially an aspirated stop onset, involves more voiceless material overlapping with the vowel. This reduces the amount of time during which F0 can be heard for the voiceless, and especially for the aspirated stops. However, if coupling strength could be modulated to reduce the amount of overlap between onset and vowel in these environments, more vocalic material would be present during which to realize F0. This would predict that C-V lag could lengthen for voiceless or aspirated onsets, and reduce for voiced or unaspirated onsets.

4.1.3 Research Questions

The present study investigates the timing of consonant, vowel, and tone gestures in Tibetan in three phases. First, the effect of tone on C-V lag is investigated by comparing the C-V timing of speakers with a tonal contrast to that of speakers without a tonal contrast. Second, segment-specific timing is tested using the C-V lag of words with aspirated and unaspirated stops. Finally, the third question looks to the covariation of C-V lag with consonant duration to investigate the coupling mode used in C-V timing.

These questions are chosen in order to investigate the competitive-coupling model of tone gestures. This model predicts that C-V lag should be long for tone-contrasting speakers and short for speakers who do not contrast tone, and that aspiration should not affect C-V lag. Secondly, if perceptual recoverability alone drove coordinative patterns like C-V lag, then segment-specific characteristics like aspiration should affect C-V lag. In order to adjudicate between these, we turn to the covariation of consonant duration with C-V lag.

4.2 EMA methods

In order to determine the relationship between tone and C-V timing, an experiment was conducted using electromagnetic articulography (EMA). Six speakers participated in this experiment (four female, two male), all of whom lived in or around New Haven, CT or New York, NY at the time of the experiment. All speakers were raised in the Tibetan Diaspora in India and Nepal, can read and write Tibetan, and use the language socially and in other contexts. All speakers were multilingual and use multiple languages in their daily life.

Stimuli were presented in the Tibetan orthography and displayed on a screen. Speakers were asked to read each target word first in isolation, then in a carrier sentence:

ཚིག་འདི་ ___ འདུག
t͡s^{h}ík t̪ǐ ___ t̪ǔk
‘This word is ___.’

Target words each contained a back vowel /u o a/ in order to identify the start of the V gesture following a front vowel in the preceding context, and a bilabial onset in order to identify the onset C gesture independently of the V gesture. Sonorant, unaspirated stops, and aspirated stops /m p p^{h}/ were included as onsets, and both high and low tones were used. Both CV and CVC syllables were used, and both monosyllabic and disyllabic words. With three onsets, two tones, two (initial) syllable shapes, and monosyllabic and disyllabic words, there were

72 items. We aimed to elicit ten repetitions of each item per speaker, and full list of 720 tokens was completed for four participants (F01, F03, M01, and M02), while only only 430 items were recorded for speaker F04 because of a malfunctioning sensor, and only 552 items were recorded for speaker F02 because the experiment was terminated early due to sensors repeatedly falling off.

EMA sensors were placed on the upper and lower lips, three on the tongue at approximately 1, 3, and 5 centimeters from the tongue tip, and the right lower incisor (to measure jaw movement), and reference sensors were placed on the right and left mastoids and nasion. After head movement correction, articulatory gestures were identified using the *lp_findgest* procedure in *Mview*, a Matlab-based program (Tiede 2005). The start time of gestures was identified as the point at which 20% of the maximum velocity toward the target was attained, and the attainment of gestural target was identified as the point at which velocity had reduced to 20% of the maximum velocity toward target. Consonant gestures were identified using lip aperture, the distance between upper and lower lip sensors, while vowel gestures were identified using the backmost tongue sensor, which we refer to as the tongue dorsum. C-V lag was calculated as the difference in start time between consonant and vowel gestures:

$$\text{C-V lag} = (\text{start of tongue dorsum retraction }) - (\text{start of lip aperture reduction})$$

The duration of gestures was calculated as the difference in time between the start of the gesture and the attainment of target.

$$\text{Gesture Duration} = (\text{attainment of target}) - (\text{gesture start})$$

Acoustic measurements were taken of the sound file recorded concurrently with the EMA data. Analysis was conducted in Praat (Boersma and Weenink 2018); VOT and time-normalized pitch were calculated using Praat scripts (DiCanio 2011, 2018).

4.3 EMA results

4.3.1 Question 1: Is C-V timing different for speakers with and without tone?

This hypothesis was motivated by the comparison of C-V lag values across tonal and non-tonal languages, as well as results of C-V lag in tonal and toneless syllables in Mandarin. If the presence of tone causes longer C-V lag, then this effect should also apply at the level of speakers. If some speakers in a speech community use lexical tone and others do not, then the speakers who use tone would be predicted to exhibit longer C-V lag values, like speakers of a tonal language; conversely, speakers who do not use tone should exhibit near-zero C-V lag values, like speakers of a non-tonal language.

The Diaspora Tibetan participants in the present research study represent such a group: some seem to use contrastive tone, and some do not. F0 was measured at ten time-normalized points across each speaker's /m/-initial CV syllables. Speaker-normalized F0 values were calculated by taking the z-score of each of the F0 measurements across all tokens of that speaker's production of all target words in the experiment (Rose 1987).

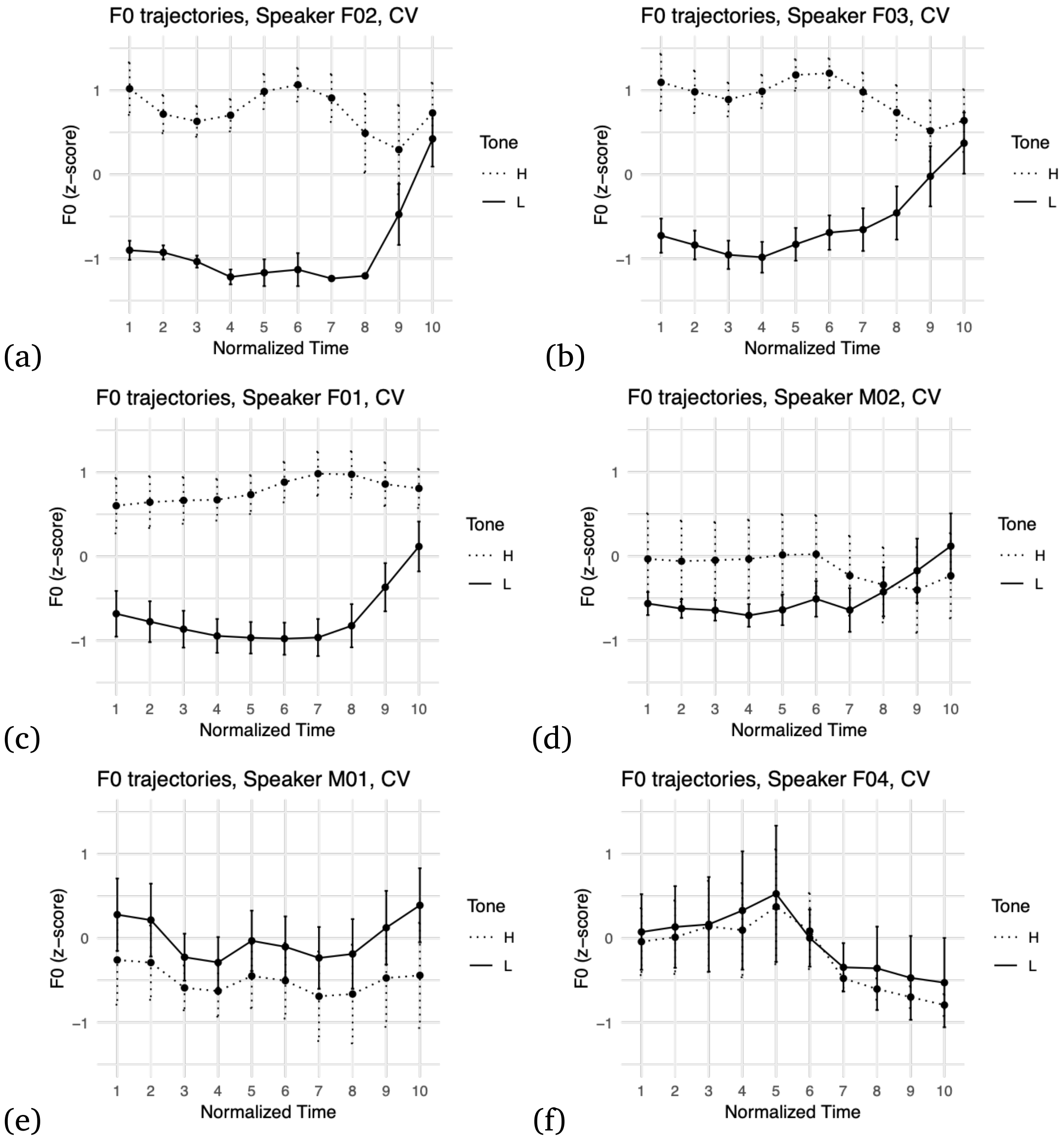
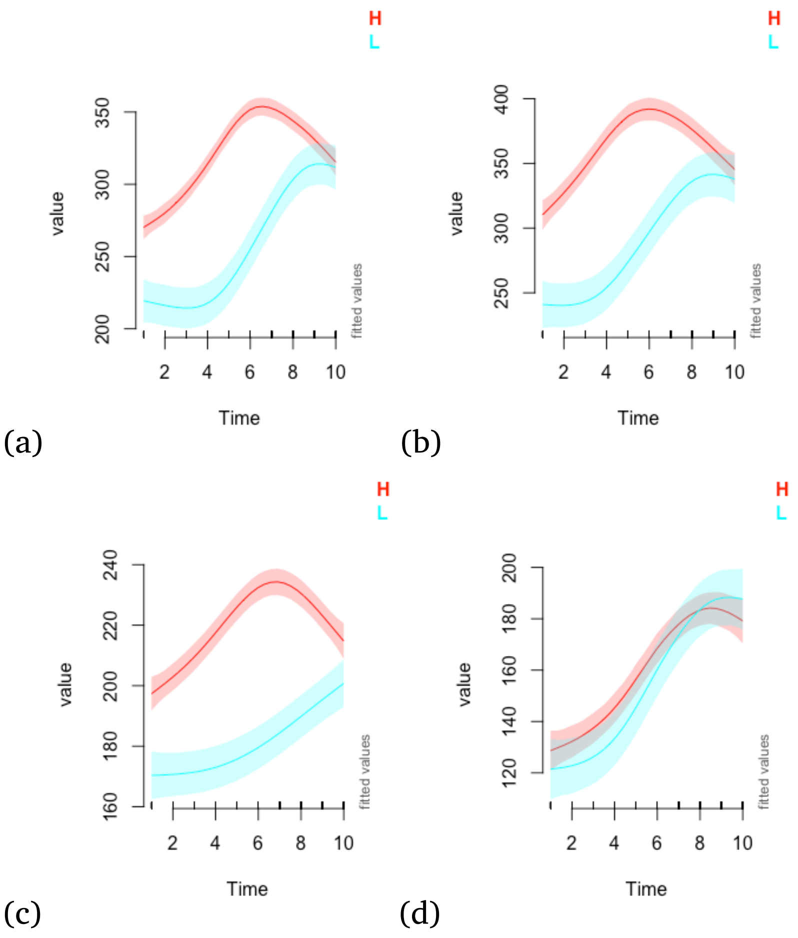


Fig 4.2: Time-normalized F0 in /mV/ syllables, z-scored by speaker across all target items in the experiment. (a)-(d) (Speakers F01, F02, F03, and M02)showed a significant interaction of tone category and time, while (e)-(f) (Speakers F04 and M01) did not.

The F0 trajectories by tone for each speaker are presented in Fig. 4.2. While While data for F01, F02, and F03 appear to show the expected high-level and low-rising trajectories, differences by tone are not as clear for the other three speakers.

The status of the F0 contrast was determined by Generalized Additive Mixed Modeling (GAMM) using the *mgcv* package in R (Wood 2017). Each model included a parametric term for tone, a smooth term for time, and a difference smooth term for tone. To account for variation by lexical item, by-word random smooths were included. Model diagnostics and residual analysis were conducted with *gam(check)* from the *itsadug* package (van Rij et al. 2016). Fig. 4.3 shows smooths for tone.



(a) (b)

(c) (d)

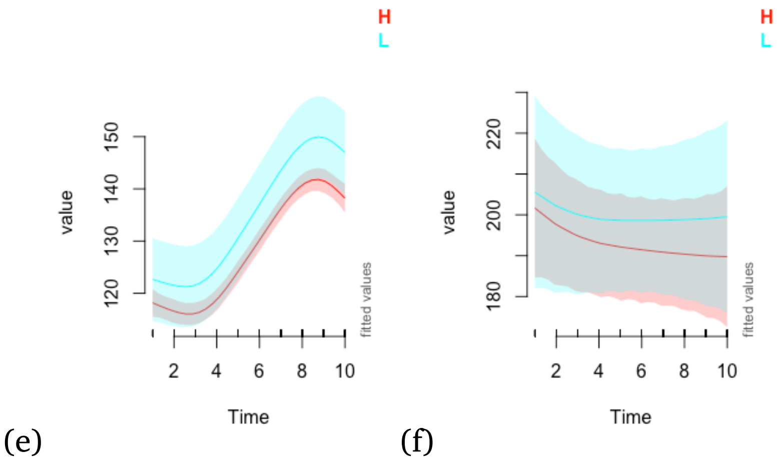


Fig 4.3: Smooths for tone from GAMMs, fitted to F0 values over time. Speakers are as follows: (a) F02; (b) F03; (c) F01; (d) M02; (e) M01; (f) F04 (a)-(d). Four speakers (a)-(d) (F01, F02, F03, and M02) showed a significant difference smooth by tone, while two speakers (e)-(f) (M01 and F04) did not.

Significance differed substantially across terms and speakers. The time smooth was significant for all speakers except F04, the difference smooth by tone was significant for all speakers except M01 and F04, and the parametric term for tone was significant for F01, F02, and F03, but not for F04, M01, and M02. Additionally, the deviance explained was over 65% for all speakers except F04, which was only 8.57%. These findings are summarized in Table 4.1, below.

term	F01	F02	F03	M02	M01	F04
tone (parametric)	*	*	*			
time smooth	*	*	*	*	*	
difference smooth by tone	*	*	*	*		

random smooths by word	*	*	*	*	*	*
Deviance explained	67.8%	94.8%	80%	71.6%	77%	8.57%

Table 4.1 Summary of GAMM results for each speaker.

These results are interpreted as follows. The non-significant time smooth for F04 indicates that F04 did not change systematically over the duration, while the significant values for the other five speakers show overall change over time. Significant difference smooths by tone (see Fig. 4.3) indicate that high- and low-tone words differ in F0 during at least some portion of the duration—that is, these speakers have a tone contrast. For three speakers (F01, F02, and F03), the significant parametric term for tone indicates that the F0 values for high-tone words are higher overall than the F0 values for low-tone words. For M02, the significant difference smooth but non-significant parametric smooth indicates that the F0 trajectories differ in parts of the duration, but that the overall F0 values are not significantly different between the two tone categories.

Taken together, this analysis reveals that four speakers (F01, F02, F03, M02) were found to contrast tone in production, though for M01 this contrast manifests in contour rather than overall height. Two speakers (M01 and F04) show no contrast in tone; among them the pitch tracks of M01 followed a consistent shape, while those of F04 did not.

With four speakers exhibiting a tone contrast and two lacking one, we can now test the prediction that tone would condition C-V timing. Based on previous research, we predicted that speakers with tone would exhibit longer C-V lag than speakers without tone, with median C-V lag of around 50ms for the former and 0-20ms for the latter. Observed C-V lag results, for all /m/-initial target items, are presented in Fig 4.4.

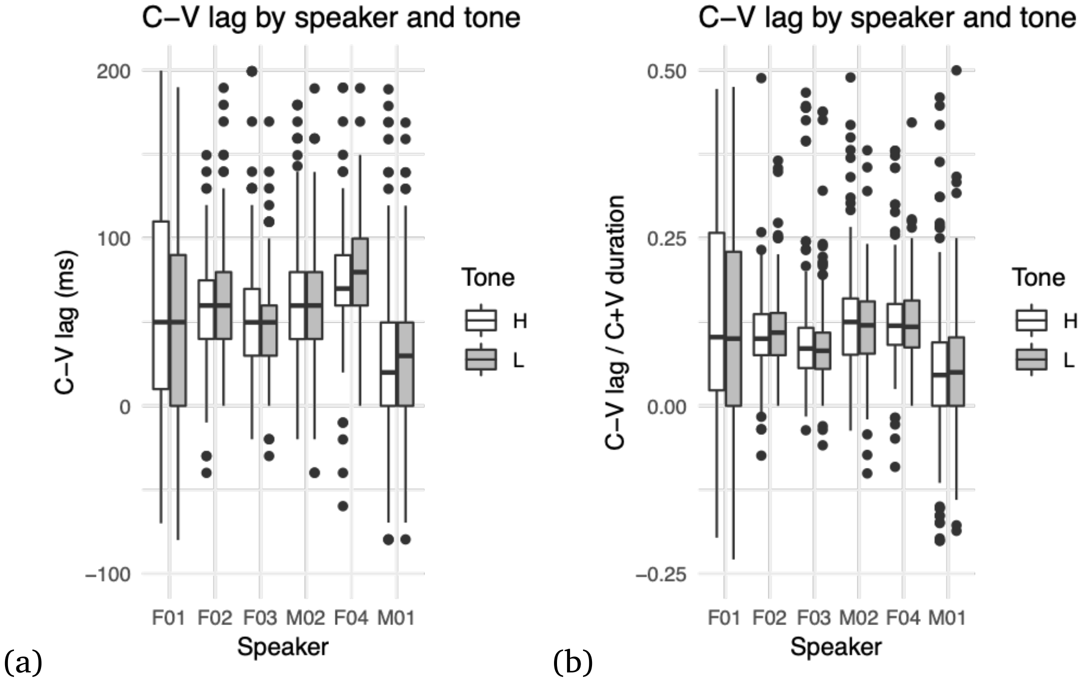


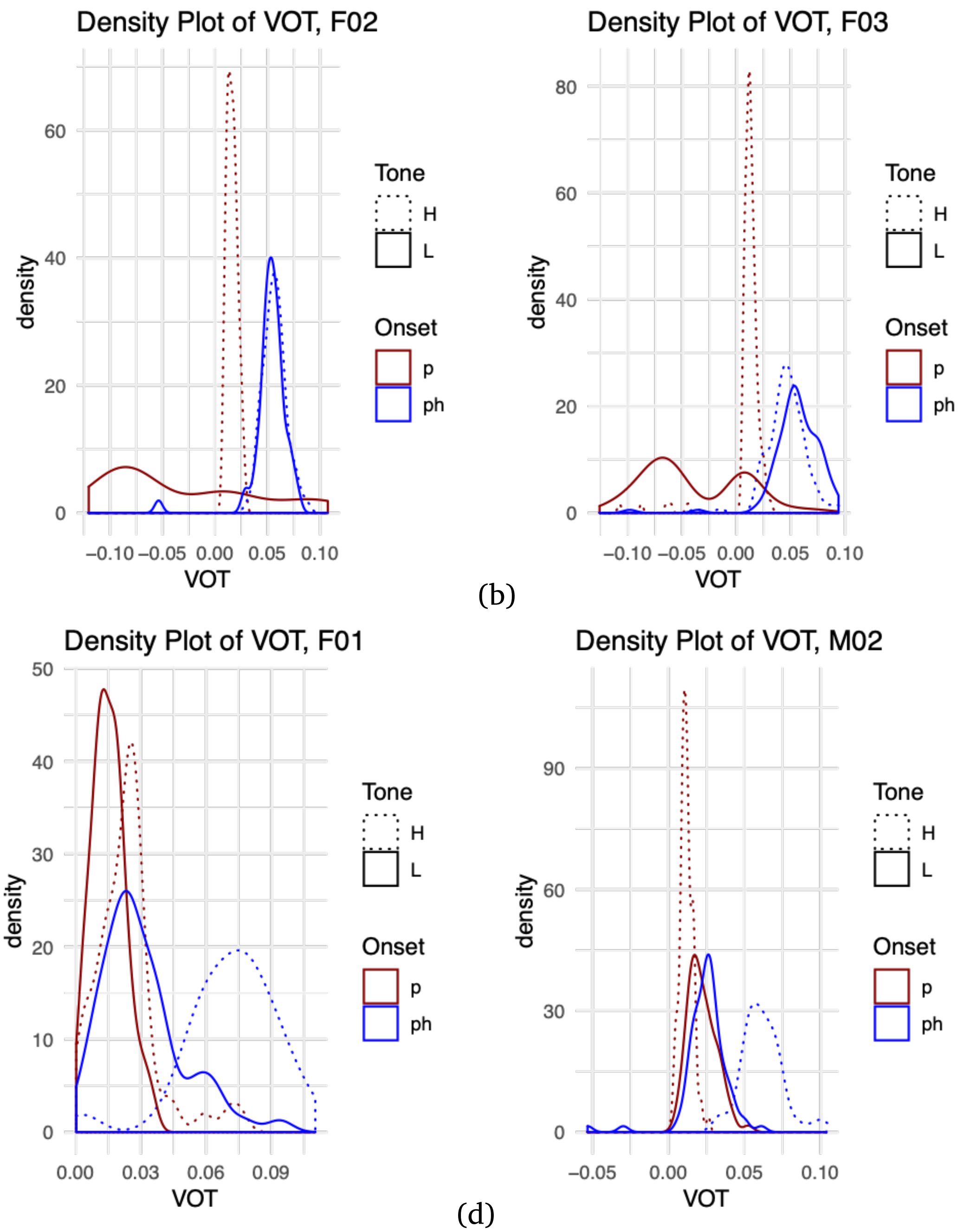
Fig 4.4: C-V lag by speaker and tone, presented as (a) raw values and (b) relative to overall duration of C and V gestures. The four tonal speakers (the four leftmost, F01-M02) exhibit C-V lag clustered around 50ms, while the two non-tonal speakers, F04 and M01, show the highest and lowest C-V lag, though these differences are attenuated in the relativized data.

C-V lag values by speaker and tone are shown in Fig. 3, with raw millisecond values in Fig. 3(a) and lag durations relative to the total C-V duration, measured from the start of the consonant gesture to the end of the vowel gesture. In each plot, lag values are shown by speaker, with the four tone-contrasting speakers indicated. It appears that the prediction was partially borne out: the four tonal speakers do exhibit a long C-V lag of around 50 ms., and one non-tonal speaker M01 has a shorter lag than the tonal speakers. However, the other non-tonal speaker, F04 has the longest lag of all six speakers in raw values, and similar lag in relative values. Furthermore, the similar values of C-V lag for the two tones shows that the tone category does not have a meaningful effect on C-V lag.

4.3.2 Question 2: Does aspiration affect C-V timing?

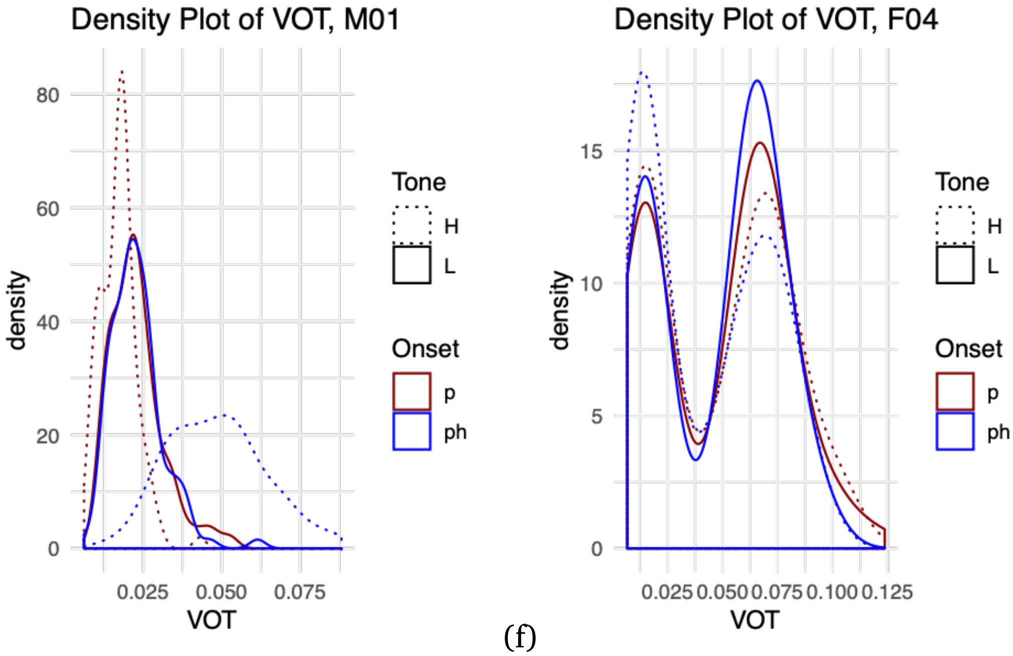
This question was raised to address a possible explanation for the cross-linguistic and alternation-based effects of tone on C-V lag: increasing C-V lag could increase the duration of a vowel gesture that does not overlap with an onset consonant, providing a greater duration in which to realize tone. If C-V lag is modified to help realize tone in this way, C-V lag could also be modified in other environments when a vowel overlaps with voicelessness, namely, in syllables with aspirated onsets as compared to unaspirated onsets. Thus, aspirated stops are predicted to have longer C-V lag than unaspirated stops for speakers who contrast two tones following aspirated stops. However, speakers for whom only one tone appears with aspirated stops may not contrast tone with aspiration, since that tone is predictable from the presence of aspiration.

VOT is plotted by onset consonant and the lexical tone category in Fig. 4.5. These are presented as density plots, with one plot per speaker. The six speakers appear to fall into one of three types. Two speakers, F02 and F03, exhibit four categories of stop and tone: both high and low tone with long-VOT aspirated stops, short-VOT unaspirated stops with high tone, and a mix of short- and negative-VOT unaspirated stops with low tone. Three speakers, F01, M01, and M02, exhibit three categories, with long-VOT aspirated stops only occurring with high tone, and all other stops showing short-VOT with both tones. Finally, one speaker, F04, exhibits both short and long VOT, but inconsistently mapped across lexical items, not systematically matching aspiration or tone categories.


Density Plot of VOT, F02
density
VOT
0
20
40
60
−0.10
−0.05
0.00
0.05
0.10
Tone
H
L
Onset
p
ph
Density Plot of VOT, F03
density
VOT
0
20
40
60
80
−0.10
−0.05
0.00
0.05
0.10
Tone
H
L
Onset
p
ph
Density Plot of VOT, F01
density
VOT
0
10
20
30
40
50
0.00
0.03
0.06
0.09
Tone
H
L
Onset
p
ph
Density Plot of VOT, M02
density
VOT
0
30
60
90
−0.05
0.00
0.05
0.10
Tone
H
L
Onset
p
ph

(a) (b)

(c) (d)



(e) (f)

Fig. 4.5: Density plots of VOT for all participants. (a)-b) Four-category speakers (F02, F03), show both high and low tone with long-VOT aspirated onsets; (c)-(e) three-category speakers (F01, M01, M02) show only high tone with long-VOT aspirated onsets; (f) one speaker shows two VOT lengths that do not correspond to aspiration or tone categories. Order of subjects is identical to that in Fig. 1.

C-V lag did not differ according to aspiration category, as shown in Fig. 4.5. It was shown in Fig. 4.4 using /m/-onset items that C-V lag values were similar for tone-contrasting speakers and for speakers without the tone contrast. The C-V lag of oral stops was also found to be similar for tone-contrasting speakers and speakers without the tone contrast, as is shown in Fig.4.6. As with Fig. 4.4, Fig. 5.6(a) depicts raw C-V lag values, and Fig.4.6(b) shows C-V lag relative to C-V duration. Data is presented by speaker, broken down by the lexical category of the onset, and tone-contrasting speakers are again indicated.

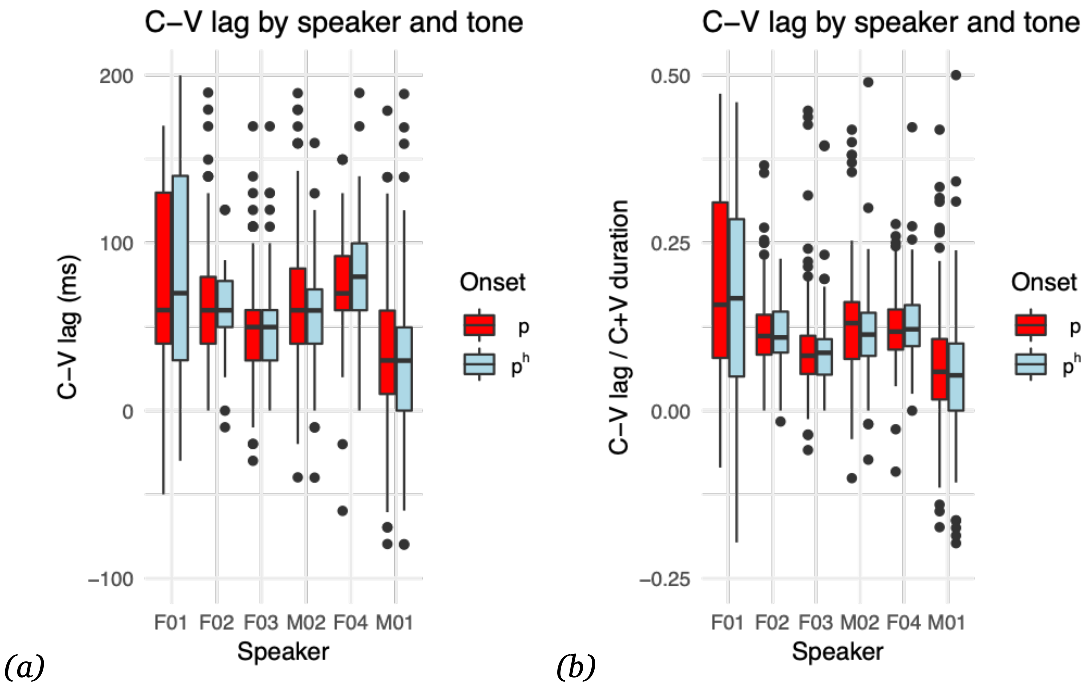


Fig. 4.6. C-V lag by onset category for all participants. (a) Raw C-V lag of all syllables. (b) C-V lag of all syllables, relativized by total duration of consonant and vowel gestures. "Onset" category based on the type of contrast exhibited by the speaker in Fig. 3, except for F04, who lacked this contrast and is presented according to the lexical categories of a four-category speaker.

These observations were corroborated through comparison of linear mixed-effects models fit to the relativized C-V lag data presented in Fig. 4.6. The relativized data was chosen over the raw data because it clarified that any difference in C-V lag would not simply be due to differences in overall duration of consonant and vowel gestures. A baseline model included random effects of speaker and lexical item and a fixed effect of whether or not a speaker contrasted tone. This was compared with a model that also included a fixed effect for onset type (i.e. /p/ vs. /p^h/; speaker F04, not producing a reliable aspiration contrast, was not coded for onset). As shown in Table 4.2, adding the fixed effect of onset did not result in a significant improvement in the model.

Model (fit to relative data)	AIC	BIC	logLikelihood	*p*-value
baseline	2209.4	2236.2	-1099.7	
baseline + aspiraiton	2211.3	2243.5	-1099.6	0.7552

Table 4.2: Comparison of linear mixed-effects models fit to C-V lag data relativized by the sum of the duration of C and V gestures.

Hypothesis 2 predicted that C-V lag would be longer for aspirated stops than for unaspirated stops. Table 4.2 shows that adding a factor for onset did not improve the model fit to either raw or relativized data, this hypothesis is not borne out, and aspiration does not appear to affect C-V lag.

4.3.3 Question 3: How are consonant, vowel, and tone gestures coordinated?

Having established that C-V lag in Tibetan is learned independently of tone contrast and aspiration, we turn to the third question: how are consonant, vowel, and tone gestures coordinated? C-V timing data presented in section 4.3.1 demonstrated that consonant gestures begin before vowel gestures, but further evidence is needed to determine the coupling relations among the gestures. In particular, is the consonant gesture coupled in-phase or anti-phase to the vowel gesture.

The answer to this question begins with C-V lag. If the consonant and vowel are timed in-phase to each other, with no other factors affecting their timing, C-V lag should be approximately zero. If consonant and vowel are timed anti-phase to each other, C-V lag should be positive. Raw C-V lag is presented in Kernel density plot form in Fig. 4.7(a). Furthermore, the ratio of C-V lag to C gesture duration may be illustrative: anti-phase C-V timing (with no other factors) should yield a C-V lag value approximately equal to the duration of the C gesture. A C-V lag value greater than zero but less than the C duration would thus require additional explanation. The ratio of C-V lag to C duration, here called C-V phasing, is presented in Fig. 4.7(b). In this figure, a C-V phasing value of zero means C-V lag is zero, and a value of C-V phasing value of one means that C-V lag and C duration are equal.

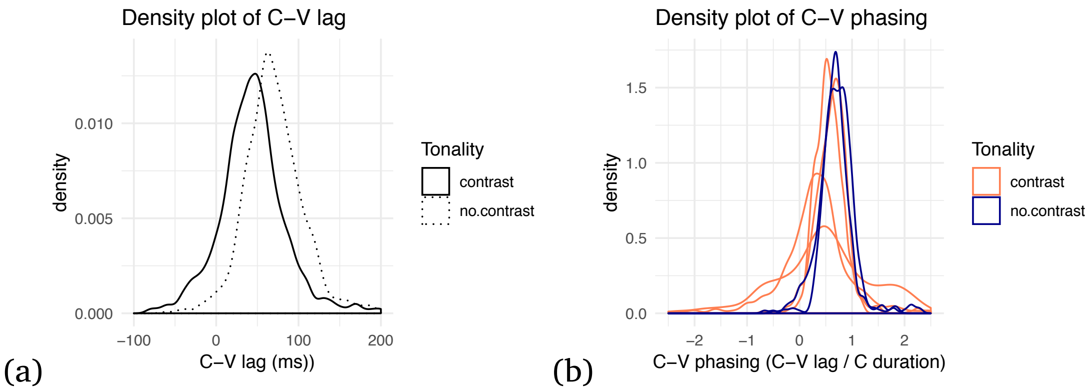


Fig. 4.7. Density plot of C-V lag and C-V phasing by tonality. (a) Raw C-V lag value plotted for tokens produced by speakers who contrast tone and do not contrast tone (see 4.3.1). (b) C-V phasing, the ratio of C-V lag to C duration, plotted by speaker with tone contrast status indicated by color.

As shown in Fig. 4.7(a), the values of C-V lag for speakers with and without tone contrasts are similar in value and positive, about 50ms (as in Figs.

4.4 and 4.6), indicating that a simple in-phase C-V coordination is not sufficient explanation. Fig. 4.7(b) shows that both groups of speakers show a C-V phasing between 0 and 1—that is, C-V lag is positive but less than the C duration. This indicates that a simple anti-phase C-V coordination is not sufficient explanation. What, then, accounts for the observed timing?

We test this question by investigating the covariation of consonant gesture duration and C-V lag (Shaw et al. 2019). If the consonant and vowel gestures are timed in-phase to each other, then the duration of the consonant gesture should have no effect on the duration of C-V lag. Conversely, if the consonant and vowel gesture are timed anti-phase to each other, then as consonant duration increases, C-V lag will also increase. This is because the beginning of the vowel gesture is timed to the end of the consonant gesture, so longer consonant gestures will cause later vowel start times. Covariation of consonant duration and C-V lag is presented in Fig. 4.8.

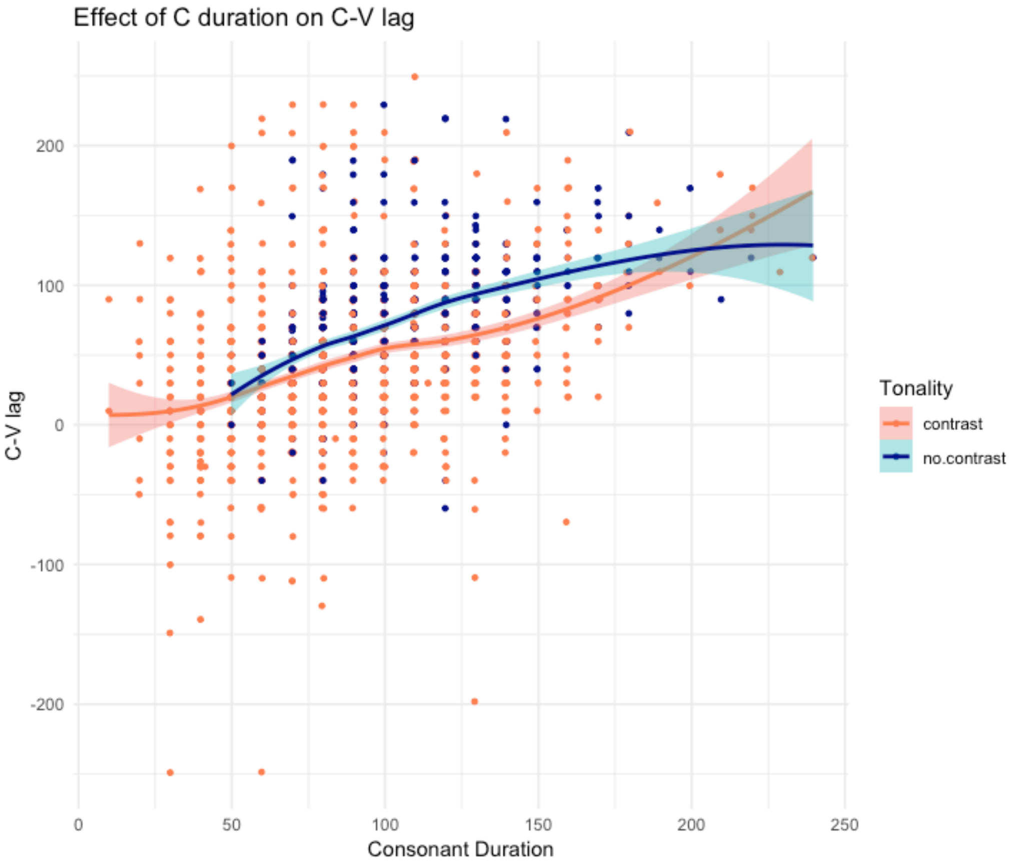


Fig. 4.8: Effect of consonant duration on C-V lag. Consonant duration calculated as beginning of gesture to attainment of target (both with 20% velocity thresholds); C-V lag calculated as the time of the beginning of vowel gesture minus time of beginning of consonant gesture. Trendlines calculated using Loess smoothing. Note that the alignment of data points at intervals of 10 ms reflects the EMA sampling rate.

We performed a linear mixed-effects analysis of the relationship between consonant gesture duration and C-V lag. A baseline model included fixed effect of onset consonant [m p p^h] and random effects of speaker and lexical item. Building on how speakers were shown to vary in the realization of aspiration in section 4.3.2, lexical items listed as containing [p^h] onsets with low tone were recoded as [p] for three-category speakers, and the [p] ~ [p^h] contrast was removed for speaker F04 because that speaker did not produce these items with consistent VOT. This was compared to a model that also included a fixed effect of consonant duration, as well as to a third model that included a fixed effect of tonality: whether or not a speaker produced a tone contrast (see section 4.3.1) As shown in Table 4.3, the model that included a fixed effect of consonant duration represented a better fit than the baseline model.

model	AIC	BIC	log likelihood	X^2	p-value
baseline	31776	31810	-15882		
baseline + C duration	31764	31804	-15875	13.9403	0.0001887
baseline + C duration + tonality	31764	31810	-15874	1.7267	0.1888314

Table 4.3: Comparison of linear mixed-effects models predicting raw C-V lag. The model including consonant duration shows improved fit over the baseline model, but including tonality does not further improve the model.

These results indicate that consonant gesture duration is positively correlated with C-V lag. This supports the hypothesis of an anti-phase coordination between C and V gestures, as the duration of the C-V lag increases with the C gesture duration. Crucially, however, there is no difference between speakers who contrast tone and those who do not. This indicates that whatever relationship is present, it is not affected by the difference in tone contrast status.

4.4 Discussion

This chapter investigated three factors that were proposed to affect C-V lag in Tibetan: the speaker's tone contrast status, the aspiration of the prevocalic consonant, and the duration of the prevocalic consonant. The first two, tone contrast and aspiration, did not affect C-V lag. We interpret this result as evidence that C-V lag is phonologized in Tibetan, rather than the result of competitive gestural coupling relations. The third factor, consonant duration, was positively correlated with C-V lag. The remainder of this section discusses each in turn.

For tone and C-V lag, the competitive-coupling model of tone of Gao (2008) predicted that speakers with a tone contrast would exhibit a longer C-V lag than speakers without a tone contrast. However, this was not the case: speakers with and without a tone contrast exhibited similar C-V lag. The stability of C-V timing across speakers with and without a tone contrast challenges this gestural coupling model. However, the observed values of C-V lag are relatively long, as to be likewise inconsistent with the C-V synchrony predicted for CV syllables. Approximately 50 ms of C-V lag resembles the values reported for tonal languages, but similar values are here found to hold for both tonal and non-tonal speakers.

The second question investigated the potential role of perceptual recoverability through the relationship between aspiration and C-V lag. Here, adjusting C-V lag to aid perceptual recovery of tone predicted longer C-V lag for aspirated stops than for unaspirated stops. This was also not borne out in the data. The fact that aspiration does not affect C-V lag suggests that C-V lag does not vary across specific segments, but is determined more globally. Additionally, since C-V lag is consistent across tonal and non-tonal speakers, tone does not affect C-V lag in Tibetan. Some speakers do use tone, so it is possible that the presence of tone in the Tibetan-speaking community may have influenced all speakers to arrive at similar values of C-V lag. Thus, recoverability of tone for some speakers may still play a role in determining C-V timing.

With competitive coupling and perceptual recoverability failing to fully explain the C-V lag, the third question investigated the covariation of consonant gesture duration and C-V lag. In-phase coupling predicts no relation between consonant gesture duration and C-V lag, while anti-phase coupling predicts C-V lag would increase as consonant gesture duration increased. Observing a positive correlation between consonant duration and C-V lag, we find support for anti-phase rather than in-phase coupling. How might this result be interpreted?

Firstly, the competitive-coupling model of tone gestures may be fundamentally correct, but require significant adjustments to derive the Tibetan data observed in this study. Specifically, those Tibetan speakers who do not contrast tone would need to be able to use competitively-coupled gestural coordination like that in Fig. 4.1(b)-(c) (reproduced in Fig. 4.6(a)) with neither a second consonant gesture nor a contrastive tone gesture. We could imagine these speakers using a single non-contrastive gesture, possibly a non-contrastive tone gesture, in all words. Doing so would allow them to use competitive coupling to derive the same C-V timing as speakers with a tone contrast. The positive correlation between consonant duration and C-V lag would be driven by the anti-phase coupling of consonant and (noncontrastive) tone, as mediated by the in-phase coupling of (noncontrastive) tone and vowel gestures. However, this explanation would present a significant departure from previous work in Articulatory Phonology, which has treated gestures as units of contrast; non-contrastive gestures do not have precedent in the theory.

Alternatively, the Tibetan speakers may be using a different coordination pattern, namely one with anti-phase coupling between consonant and vowel. This would avoid competitive coupling altogether, while still explaining the long C-V lag and the positive correlation between consonant duration and C-V lag. Speakers who contrast tone would have a tone gesture timed in-phase to the vowel gesture (as for other languages in Katsika et al. 2014, or Zsiga 2020), while speakers without a tone contrast would simply lack a tone gesture. For tone-contrasting speakers, the anti-phase C-V coupling would reduce overlap of consonant and tone gestures, thereby improving the perceptual recoverability of tone (see Section 4.1.2). For speakers lacking the tone contrast, these coupling patterns would rely on a different motivation, which we identify as a pressure for members of a speech community to converge in the domain of gestural timing. The lack of a direct coupling relation between consonant and tone gestures differs from work on Mandarin, Serbian, and Thai in Gao (2008) and Karlin (2014, 2018), but is consistent with the interpretation of Igbo data by Zsiga (2020). The differences between these two analyses are shown in Fig. 4.9, with competitive coupling in Fig. 4.9(a) and the anti-phase C-V coupling in Fig. 4.9(b).

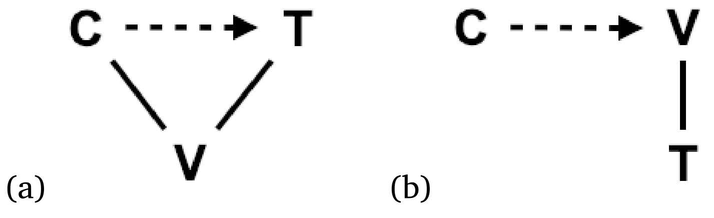


Figure 4.9. Revised coupling graph. (a) Original gestural model of tone with competitive coupling (b) Revised model with anti-phase C-V coupling.

If the coupled oscillator model of planning is to be maintained, these two coupling graphs offer the best accounts for the data. Either Fig. 4.9(a) or Fig. 4.9(b) is consistent with the results obtained for tone-contrasting speakers. Modifications of each can account for the speakers without a tone contrast: Fig. 4.9(a) can apply if the “T” refers to a non-contrastive gesture, while Fig 4.9(b) requires only omitting the “T” gesture. Both accounts require amending the theory, either by expanding the typology of coupling relations to include anti-phase C-V timing, or by allowing for non-contrastive gestures. The former would undermine the motivation of in-phase C-V timing as an explanation for the unmarked status of C-V syllables (Nam et al. 2009; section 1.2.3), and the latter would demand reevaluation of the gesture as a unit of contrast (see 1.2.1).

Regardless of which account is correct, or whether the data would be more convincingly explained by some yet-to-be-developed mechanism, the same core finding remains. Any model or theory seeking to explain the Tibetan data requires a mechanism for dynamically scaling C-V lag with consonant duration. Such a mechanism must be able to apply in the absence of lexical tone, and must be present in some languages but not others.

4.5 Chapter summary

This chapter presents the results of an experiment testing the relationship between C-V lag, aspiration, and tone in Tibetan. The discussion of results was divided into three questions:

> Is C-V lag different for speakers with and without tone? (Section 4.3.1)
> Does aspiration affect C-V timing? (Section 4.3.2)
> How are consonant, vowel, and tone gestures coordinated? (Section 4.3.3)

Speakers were found to vary in whether or not they produced a tone contrast and in the alignment of their aspiration and voicing contrasts. However, neither of these factors was found to affect C-V lag, leading to negative answers for the first two questions. As for the third question, it was found that C-V lag covaried with consonant duration for all speakers. This was taken as evidence of anti-phase coupling—either between C and V gestures directly, or between gestures that these are in-phase coupled to. This is consistent with the prediction of the competitive-coupling model of Gao (2008) for tone-contrasting speakers, but not for speakers lacking a tone contrast. Thus, speakers who have lost the tone contrast in their own production still maintain the same C-V timing as tone-contrasting speakers.

4.6 Chapter bibliography

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Boersma, Paul & David Weenink. 2018. Praat: Doing phonetics by computer [Computer software]. Version 6.0. 43.

Bombien, Lasse. 2011. Segmental and prosodic aspects in the production of consonant clusters–On the goodness of clusters. *München: Ludwig-Maximilians-Universität doctoral dissertation.*

Browman, Catherine P. & Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45(2–4). 140–155.

Browman, Catherine P. & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49(3–4). 155–180.

Browman, Catherine P. & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée* (5). 25–34.

Browman, Catherine P. & Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

Byrd, Dani. 1995. C-Centers Revisited. *Phonetica* 52(4). 285–306. https://doi.org/10.1159/000262183.

Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of phonetics* 24(2). 209–244.

Chitoran, Ioana. 1999. Accounting for sonority violations: the case of Georgian consonant sequencing. In Proceedings of the 14th International Congress of Phonetic Sciences. Berkeley: Department of Linguistics, University of California, Berkeley, 101–104.

Chitoran, Ioana, Louis Goldstein & Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. *Gussenhoven, Warner, Papers in Laboratory Phonology* 7. 419–447.

DiCanio, Christian. 2011. *Duration Script for Praat.*

DiCanio, Christian. 2018. Pitch Dynamics Script for Praat, version 6.

Fowler, Carol A. 1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General* 112(3). 386.

Gafos, Adamantios I., Jens Roeser, Stavroula Sotiropoulou, Philip Hoole & Chakir Zeroual. 2019. Structure in mind, structure in vocal tract. *Natural Language & Linguistic Theory* 1–33.

Gao, Man. 2008. Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics), Yale University, CT.*

Geissler, Christopher. 2018. Phonological Koinéization in Kathmandu Tibetan. In *Proceedings of the Annual Meetings on Phonology*, vol. 5.

Gibson, Mark, Stavroula Sotiropoulou, Stephen Tobin & Adamantios Gafos. 2017. On some temporal properties of Spanish consonant-liquid and consonant-rhotic clusters. (Ed.) Malte Beltz, Susanne Fuchs, Stefanie Jannedy, Christine Mooshammer, Oxana Rasskazova & Marzena Zygis. *Proceedings of the 13th Tagung Phonetik und Phonologie im deutschsprachigen Raum (PP13)* 73–76.

Goldstein, Louis, Ioana Chitoran & Elisabeth Selkirk. 2007. Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVIth international congress of phonetic sciences*, 241–244.

Goldstein, Louis, Hosung Nam, Elliot Saltzman & Ioana Chitoran. 2009. Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in phonetics and speech science* 239–250.

Haken, Hermann, JA Scott Kelso & Heinz Bunz. 1985. A theoretical model of phase transitions in human hand movements. *Biological cybernetics* 51(5). 347–356.

Haller, Felix. 1999. A brief comparison of register tone in Central Tibetan and Kham Tibetan. *Linguistics of the Tibeto-Burman Area* 22(2). 77–97.

Hermes, Anne, Martine Grice, Doris Mücke & Henrik Niemann. 2008. Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. *Proceedings of the 8th International Seminar on Speech Production* 433–436.

Hermes, Anne, Doris Mücke & Bastian Auris. 2017. The variability of syllable patterns in Tashlhiyt Berber and Polish. *Journal of Phonetics* 64. 127–144.

Hermes, Anne, Doris Mücke & Henrik Niemann. 2012. Articulatory coordination and hte syllabification of word-initial consonant clusters in Italian. In Phil Hoole, Marianne Pouplier, Lasse Bombien, Christine Mooshammer & Barbara Kühnert (eds.), *Consonant Clusters and Structural Complexity* (Interface Explorations), 157–176. Mouton de Gruyter.

Hermes, Anne, Rachid Ridouane, Doris Mucke & Martine Grice. 2011. Kinematics of syllable structure in Tashlhiyt Berber: The case of vocalic and consonantal nuclei. In *9th International Seminar on Speech production.*, 1–6.

Hu, Fang. 2016. Tones are not abstract autosegmentals. In *Speech Prosody*, 302–306.

Karlin, Robin. 2014. The articulatory TBU: gestural coordination of tone in Thai. In *Cornell Working Papers in Linguistics*.

Katsika, Argyro, Jelena Krivokapić, Christine Mooshammer, Mark Tiede & Louis Goldstein. 2014. The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics* 44. 62–82.

Kim, Kong-On. 1975. The nature of temporal relationship between adjacent segments in spoken Korean. *Phonetica* 31(3–4). 259–273.

Kozhevnikov, Valeriĭ Aleksandrovich & Li͡udmila Andreevna Chistovich. 1965. Speech: Articulation and perception.

Kühnert, Barbara, Phil Hoole & Christine Mooshammer. 2006. Gestural overlap and C-center in selected French consonant clusters. In *7th International Seminar on Speech Production (ISSP)*, 327–334.

Löfqvist, Anders & Vincent L. Gracco. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *The Journal of the Acoustical Society of America* 105(3). 1864–1876.

Marin, Stefania. 2013. The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics* 41(3–4). 211–227.

Mücke, Doris, Martine Grice & Taehong Cho. 2014. More than a magic moment–Paving the way for dynamics of articulation and prosodic structure. *Journal of Phonetics* 44. 1–7.

Mücke, Doris, Hosung Nam, Anne Hermes & Louis Goldstein. 2012. Coupling of tone and constriction gestures in pitch accents. In Philip Hoole, Bombien, Lasse, Pouplier, Marianne, Mooshammer, Christine & Barbara Kühnert (eds.), *Consonant clusters and structural complexity* (Interface Explorations), vol. 26, 205. De Gruyter Mouton.

Nam, Hosung & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of the Phonetic Sciences*.

Nam, Hosung, Louis Goldstein & Elliot Saltzman. 2009. Self-organization of Syllable Structure: A Coupled Oscillator Model. In François Pellegrino, Egidio Marsico, Ioana Chitoran & Christophe Coupé (eds.), *Approaches to Phonological Complexity*. Berlin, New York: Walter de Gruyter.

Niemann, Henrik, Doris Mücke, Hosung Nam, Louis Goldstein & Martine Grice. 2011. Tones as Gestures: The Case of Italian and German. In *Proceedings of the 17th International Congress of the Phonetic Sciences*, 1486–1489.

Öhman, Sven EG. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39(1). 151–168.

Pastätter, Manfred & Marianne Pouplier. 2017. Articulatory mechanisms underlying onset-vowel organization. *Journal of Phonetics* 65. 1–14.

Prom-on, Santitham, Yi Xu & Bundit Thipakorn. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America* 125(1). 405–424. https://doi.org/10.1121/1.3037222.

R Core Team. 2013. R: A language and environment for statistical computing.

Rij, Jacolien van, Martijn Wieling, R. Harald Baayen, Hedderik van Rijn & Maintainer Jacolien van Rij. 2016. *Package 'itsadug.'*

Rose, Phil. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech communication* 6(4). 343–352.

Saltzman, Elliot & Dani Byrd. 2000. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19(4). 499–526.

Shaw, Jason A., Karthik Durvasula & Alexei Kochetov. 2019. The temporal basis of complex segments. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 676–680.

Shaw, Jason A. & Wei-rong Chen. 2019. Spatially Conditioned Speech Timing: Evidence and Implications. *Frontiers in Psychology* 10(2726).

Shaw, Jason A. & Shigeto Kawahara. 2018. The lingual articulation of devoiced/ u/in Tokyo Japanese. *Journal of Phonetics* 66. 100–119.

Shaw, Jason, Adamantios I. Gafos, Philip Hoole & Chakir Zeroual. 2009. Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology* 26(1). 187–215.

Silverman, Daniel. 1997. *Phasing and Recoverability* (Outstanding Dissertations in Linguistics). New York: Garland.

Sóskuthy, Márton. 2017a. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv preprint arXiv: 1703.05339*.

Sóskuthy, Márton. 2017b. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv:1703.05339 [stat]*. http://arxiv.org/abs/1703.05339 (5 December, 2019).

Teo, Amos, Lauren Gawne & Melissa Baese-Berk. 2015. A case study of tone and intonation in two Tibetic language varieties. In *Proceedings of the 18th International Congress of the Phonetic Sciences*.

Tiede, Mark. 2005. Mview: software for visualization and analysis of concurrently recorded movement data.

Tilsen, Sam. 2016. Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics* 55. 53–77.

Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70. 86–116.

Wood, S. N. 2017. mgcv: mixed GAM computation vehicle with automatic smoothness. R package version 1.8-22.

Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(1). 3–36.

Wright, Richard Albert. 1996. *Consonant clusters and cue preservation in Tsou*. Department of Linguistics, University of California, Los Angeles.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of phonetics* 25(1). 61–83.

Xu, Yi. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55(4). 179–203.

Xu, Yi. 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58(1–2). 26–52.

Xu, Yi. 2009. Timing and coordination in tone and intonation—An articulatory-functional perspective. *Lingua* 119(6). 906–927.

Xu, Yi & Fang Liu. 2006. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics* 18(1). 125.

Xu, Yi & Q. Emily Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech communication* 33(4). 319–337.

Yi, Hao & Sam Tilsen. 2014. Gestural timing in Mandarin tone sandhi. In *Proceedings of Meetings on Acoustics 168ASA*, vol. 22, 060003. ASA.

Zhang, Muye, Christopher Geissler & Jason Shaw. Gestural representations of tone in Mandarin: Evidence from timing alternations. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Melbourne, Australia.

5 Discussion

The overall aim of this dissertation was to investigate the sources of systematically in the intergestural coordination of Tibetan as spoken in diaspora. Following a review of the historical changes in the phonological history of Tibetan (chapter 2), empirical data on the coordination of laryngeal, supralaryngeal, and tonal gestures was presented from an acoustic corpus study (chapter 3) and an EMA experiment (chapter 4). The sections that follow review the empirical results (section 5.1) and interpret these results in terms of target uniformity (5.2) and gestural coupling (5.3). This is followed by a discussion of the relationship between these findings and diachronic sound change (5.4), and a general summary (5.5).

5.1 Review of results

The corpus study reported in Chapter 3 investigated the phonetic basis of the laryngeal contrasts in Common Tibetan as spoken in diaspora. As discussed in Chapter 2, the laryngeal and tonal contrasts vary substantially across Tibetan dialects, meaning that speakers raised in diaspora have been exposed to speakers with different systems of contrasts. Therefore, the first empirical goal was to establish which categories were present for the diaspora speakers.

The phonetic parameters of VOT and F0, as well as their covariation, were investigated as cues to laryngeal and tonal contrasts. The stop contrast was largely one of aspiration rather than voicing, but varied across tone categories. In low-tone words, unaspirated stops were variably prevoiced in 12 of 19 speakers in the corpus study and 2 of 6 participants in the EMA experiment (Fig. 3.4, Fig. 4.5); prevoicing was not observed in the remaining speakers. The same

2 of 6 EMA participants with variable prevoicing also produced low-tone aspirated stops with long VOT, That is, the same speakers who produced *sdom* [tòm ~ dòm] ‘spider’ with variable prevoicing also produced *dom* [tòm ~ t^hòm] ‘bear’ with aspiration. Other participants produced all low-tone words with short VOT, i.e. [tòm] for both ‘spider’ and ‘bear’. However, all speakers used consistent aspirated and unaspirated stops with high tone, such as in the high-tone words *rta.mag* [tá.mák] ‘cavalry’ and *tha.mag* [t^há.mák] ‘cigarette’. This finding is consistent with aggregated corpus data showing low-tone aspirated stops with a range of VOT values between that of unaspirated and high-tone aspirated stops (Fig. 3.5). Only 11 of 19 speakers in the corpus study (Table 3.3) and 4 of 6 participants in the EMA study (Fig. 4.2-4.3) were found to produce a tone contrast. Importantly, the VOT patterns were independent of the actual tone contrast itself: for example, even speakers who have merged the tones only produce prevoicing in those unaspirated stops that occur with low-tone for tonal speakers. VOT and F0 at the onset of voicing covaried unevenly across categories: a positive correlation was only observed in high-tone aspirated stops (Fig. 3.6). The patterning of prevoicing and aspiration indicate that speakers with the tone merger maintain distinct representations of consonants corresponding to the tone categories.

The EMA study (Chapter 4) expanded on the acoustic results by investigating the timing of supra-laryngeal gestures across laryngeal and tonal categories. All speakers showed a similar, positive C-V lag: consonant gestures began before vowel gestures for all speakers, both those who produced a tone contrast and those who did not (Fig. 4.8). This contradicts the competitive coupling of tone hypothesis, which predicted different C-V lag in the presence vs. absence of a tone gesture. Neither tone nor consonant category affected C-V lag (Fig. 4.4, 4.6). However, the duration of the consonant gesture was found to positively correlate with C-V lag, suggesting cluster-like or eccentric timing (Fig. 4.8).

The following sections discuss implications of these findings for different areas of phonology and phonetics: the uniformity of timing relations (5.2), the place of coupling modes in phonology (5.3), and connections to sound change (5.4).

5.2 Community-level temporal target uniformity

A key result of the EMA experiment is that C-V lag patterns similarly in speakers with and without a tone production contrast. In this section, I relate this result to the concept of phonetic target uniformity, and argue that this principle should apply not just to articulatory targets but to timing relations as well.

Target uniformity refers to the tendency of articulations to be produced with maximal similarity across phonological categories. In a typical case, segments that share a feature would be produced with similar articulation. As surveyed in Section 1.2.4, target uniformity, also known as ‘gestural economy’, has been invoked to explain similarities between singly- and doubly-articulated consonants in Ewe (Maddieson 1995), consistent VOT in some English speakers by prosodic context (Keating 2003), and speaker-specific consistency in VOT across place of articulation in English (Chodroff & Wilson 2017), and English and Czech sibilant fricatives (Chodroff 2017). In Chodroff (2017), target uniformity was one of three constraints used to account for structured variation in phonetics, along with “contrast uniformity” (requiring similar acoustic productions across a phonological category) and “pattern uniformity” (requiring equivalent distances between targets across speakers). Target uniformity can help explain why languages' sound inventories tend to be structured into series that are similar rather than maximally dispersed. In several studies of uniformity (Keating 2003, Faytak 2018), some speakers exhibited greater uniformity in articulation, while others permitted some variability in articulation in order to maintain greater consistency in acoustics. These differences may mirror individual differences in other aspects of phonology and cognition (e.g. Yu 2016; see Yu & Zellou 2019 for a review on individual differences in phonology). In the EMA study presented in Chapter 4, it was found that speakers varied in their tone and consonant contrasts, but not in their C-V lag.

From a slightly different perspective, target uniformity is less related to contrast maintenance than to articulatory re-use. Faytak (2018) found target uniformity in Suzhou Chinese fricative vowels: speakers produce rounded and unrounded fricative vowels with very similar tongue position at the cost of acoustic variability induced by rounding. The explanation offered is one of articulatory reuse: over the course of L1 acquisition, speakers tend to use the same, familiar articulations where possible, rather than learning new articulations for each segment in the inventory. This return to familiar articulations resembles the substitution of reliable articulations by young children early in acquisition (McAllister Byun et al 2016). Both the Suzhou Chinese adults and the child acquirers prioritize consistent articulation even when doing so compromises phonological contrasts.

Any constraints favoring uniformity must be violable, since there are cases where speakers achieve consistent acoustic output through varying articulatory mechanisms. One notable counter-example to uniform articulation is the set of articulatory differences between French nasal and oral vowels. Carignan (2014) argues that speakers use a range of different articulatory means to attain similar acoustic results. The shared effect is to enhance the contrast within each oral-nasal pair. Whereas Suzhou fricatives involve acoustic variability with articulatory uniformity, French nasal vowels involve acoustic dispersion through multiple articulatory strategies. Any constraint enforcing articulatory uniformity within French oral-nasal pairs is apparently overridden by a competing drive for acoustic dispersion. The French example thus demonstrates the violability of uniformity in opposition with the also-violable dispersion. The variable articulations employed by different speakers shows that speakers do not always resemble each other. The English rhotic is another example of a similar acoustic target for which speakers use a range of different articulations (e.g. Smith et al. 2019, Harper et al. 2020).

As in the case of Suzhou fricatives, the Tibetan speakers in this study also seem to prioritize a consistent articulation. However, this uniformity is not in space, but in time, and represents consistent behavior across members of the speech community. Competitive coupling between onset consonant and tone gestures may have been the original cause of C-V lag in tonal Tibetan speakers. In diaspora, some Tibetan-acquiring children developed a non-tonal system, which would not be predicted to cause C-V lag based on the competitive coupling hypothesis. Instead, these non-tonal speakers produce the same C-V lag as their tonal counterparts. This is evidence for a constraint enforcing speakers to use similar articulation as their fellow community members—a different kind of target uniformity. This constraint must be violable, as the above-cited examples from French nasals and English rhotics show that speakers can sometimes differ in their articulatory strategies.

5.3 Coupling relations

5.3.1 Eccentric coupling

The results of the EMA experiment reported in Chapter 4 found unusual patterns suggesting eccentric C-V timing. The scale of the C-V lag and its co-variation with consonant duration are consistent with both competitive coupling and anti-phase C-V coupling depicted in Fig. 4.6 (reproduced as Fig. 5.1, below). However, both rely on the presence of a tone gesture, which would not be present for speakers lacking a tone contrast. Eccentric C-V coupling offers a compelling alternative to anti-phase coupling, since both are consistent with the observation that C-V lag covaries with consonant duration. This is because the dynamical systems approach to gestures (Nam and Saltzman 2003, Iskarous 2017) treats each gesture as a cycle. Any phasing other than strict in-phase coupling (or in-phase coupling with a fixed delay) predicts the observed covariation between duration and lag.

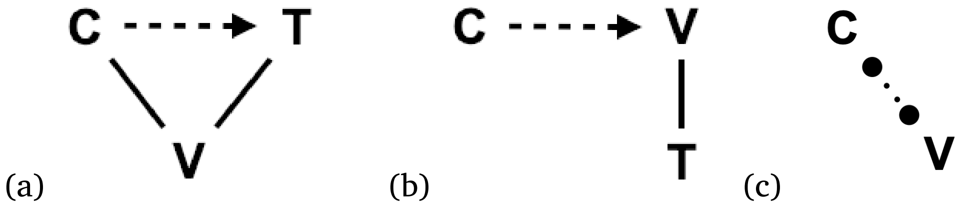


Figure 5.1. Revised coupling graph. (a) Original gestural model of tone with competitive coupling (b) Anti-phase C-V coupling. (c) Eccentric C-V coupling

Moreover, a unified account at the level of the coupling graph reflects the fact that "lexical tone" and "intonational tone" are not rigid categories: languages can have both, and some systems appear intermediate between the two (e.g. Serbian, Karlin 2018). This allows the tone gesture framework in Katsika et al. (2014) to be applied to any language, and mirrors the autosegmental-metrical (e.g. Pierrehumber 1980, Beckman & Pierrehumbert 1986) view of tones as being associated with vowels more than with (most) consonants.

Assuming that competitive coupling motivated C-V timing among tone-contrasting speakers, how might the non-tonal speakers have acquired such a system? Perhaps, all speakers passed through a stage of acquisition that featured a longer C-V lag. This could either have come from competitive coupling with a tone gesture (i.e. being tonal speakers), or through eccentric timing to match tonal speakers around them. When some of these speakers later settled on a non-tonal system, they would have maintained the gestural timing from this earlier stage. While these speakers had concrete evidence of VOT contrast and tonal- and non-tonal prosodic systems, speakers may not have been exposed to evidence contradicting the C-V timing patterns they already had. Following target uniformity, they could simply continue with the same C-V lag as they went on to acquire their adult phonological systems. The maintenance of this earlier stage of temporal coordination may have been facilitated by exposure: with a multidialectal and multilingual input, the diaspora-raised Tibetan speakers might not accumulate sufficient evidence to acquire an alternative system of coupling. As such, they maintained the relative C-V timing as adults.

The account just sketched is speculative, but follows from an extension of the principle of target uniformity into the temporal realm. While the competitive coupling model is predicated on in-phase and anti-phase coordination being most readily learned, the possibility of “eccentric timing” is not excluded. By this account, the C-V lag would have resulted from competitive coupling during a tonal phase, but remain as eccentric timing later in life. The account predicts that other cases of eccentric timing can be traced to timing patterns established under a different set of control systems at an earlier stage of acquisition.

Whatever the cause of eccentric timing, it requires an explanation for why speakers do not revert to the unmarked synchronous timing. In the absence of a tonal gesture, the only remaining coordination relation in a CV syllable is that between the consonant and vowel gestures. C-V coupling is generally predicted to be in-phase, including in cases of competitive coupling, and therefore should result in synchronous gestural start times. C-V synchrony has been shown to emerge experimentally under repetition and constrained speech rate (e.g. Gleason et al. 1996), but also in perception of repeated speech (de Jonge et al 2004). The emergence of C-V synchrony is not purely biomechanical, however, as language-specific phonotactics also play a role in the way speakers reorganize gestures in rate-limited repetition tasks (Chitoran & Tiede 2013). The appearance of language-specific patterns indicates that participants are using learned linguistic systems rather than purely general principles of motor organization. If linguistic experience can condition the ways in which C-V synchrony emerges, it stands to reason that linguistic experience can also limit the emergence of C-V synchrony. If Tibetan speakers continue to produce eccentric timing in rate-limited repetition tasks, this would indicate that language-specific coordination can overcome an innate bias toward C-V synchrony.

5.3.2 Alternatives to eccentric coupling

One interpretation of the C-V timing results would posit a “gesture” whose only articulatory consequence is its effects on the relative timing of other articulators. This would explain the fact that C-V timing in non-tonal speakers is consistent with competitive coupling in the absence of evidence for a gesture to be anti-phase coupled to the consonant. Such a gesture would lack an articulatory target of its own, but could be indirectly observed through its effects on the timing of other gestures. While such “targetless" gestures have not been proposed before, the concept resembles Gradient Symbolic Representations that have been invoked to explain segmental phenomena such as French liaison (Smolensky & Goldrick 2016) and Japanese rendaku (Rosen 2016), as well as other phenomena that are only weakly active in the phonology of a language (Zimmerman 2019). These segments are characterized by a higher activation threshold, with the result that they only appear (or disappear) under specific circumstances. If certain segments can appear or disappear only in certain environments, it is conceivable that a phonological unit—in this case, a tone—could have no realization but still exist in the representation. Such an account would require a number of assumptions beyond that of Smolensky & Goldrick (2016), Rosen (2016), or Zimmerman (2019): gradient representations would need to apply to tone gestures as well as segments, and would need to affect intergestural timing even while not controlling a movement specific to the gesture itself.

The sound changes in the history of Tibetan have involved extensive changes in gestural coordination, with the deletion of consonants in clusters (and codas) and the addition of tone, so it is reasonable to ask if the timing patterns might have been preserved even while the gestures changed. However, this is not consistent with the patterns of change both for cluster simplification and tonogenesis.

In historical clusters, tone gestures could not have replaced lost consonant gestures for two reasons. First, some words in Old Tibetan did have simplex onsets; among stops, these developed into contemporary high-tone aspirated and low-tone aspirated (or voiceless). If the timing of these clusters had been preserved, a systematic difference in C-V lag would be seen across onsets, where historically-simplex words would have simultaneous C-V timing and historical clusters would have C-V lag. These differences are not observed. Second, the preservation of cluster timing predicts different timing patterns than what is observed. This is because the segment retained from most clusters is the prevocalic one, not the initial. If the first consonant in a cluster were replaced with a tone gesture (or targetless gesture), the retained second consonant would start after the vowel gesture rather than before. There would be a C-V lag, but in the opposite direction of what is observed[3]. These differences are schematized in Fig. 5.2, below.

The gestural reanalysis involved in tonogenesis would also not predict the modern Tibetan C-V lag. In tonogenesis, the historical laryngeal contrast was replaced by a tone contrast: voiced onsets were reanalyzed as low tone and voiceless onsets (aspirated and unaspirated) were reanalyzed as high tone (see Section 2.9, Table 2.6). As these pitch perturbations were phonologized, tone gestures were added. However, the laryngeal gestures did not disappear, and indeed voicing, aspiration, and voiceless-unaspirated stops all persist in the contemporary tonal varieties. Thus, the tonogenesis process involved the addition of a tone gesture, rather than just a reorganization of existing gestures,

[3] The exception to this is prevocalic glides: *w has disappeared, while *j and *r have coalesced with the preceding stop to produce palatal or retroflex stops. For example, *bsgrubs* ‘accomplish’, has become [ɖùp] or [ʈùp] in Central Tibetan dialects after the loss of *b* and *s* and the coalescence of *gr > [ɖ].

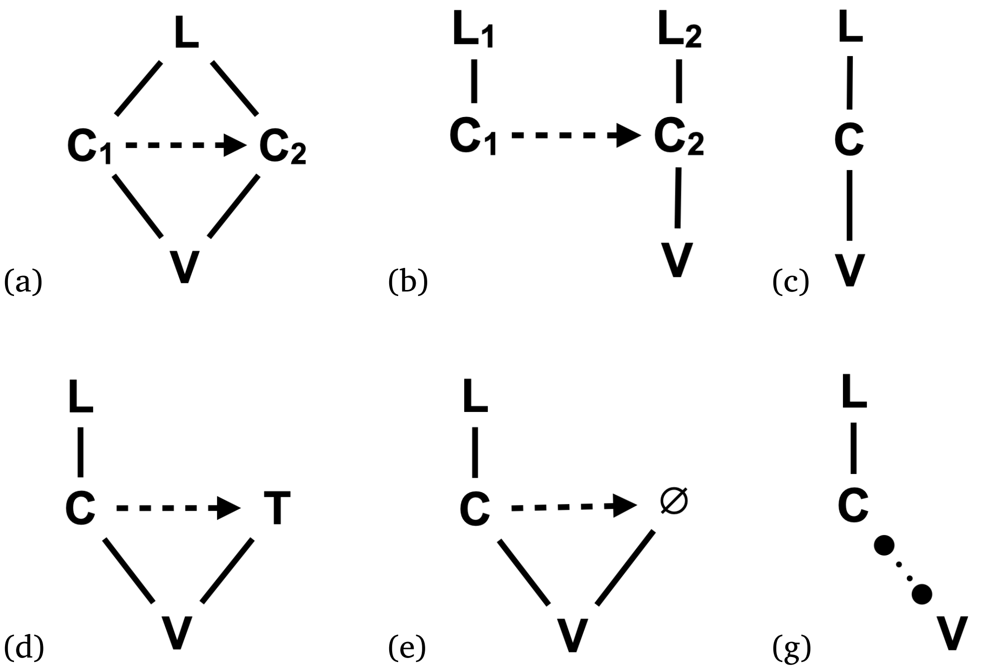


Figure 5.2. Hypothesized coupling diagrams with clusters, tones, and laryngeal (L) gestures. (a-c) Tibetan syllables before tonogenesis: (a) CCV syllable with competitive coupling and shared laryngeal gesture; (b) C.CV syllable with anti-phase C-C coupling; (c) CV syllable. (d-g) Tibetan syllables after tonogenesis: (d) CV syllable with tone gesture; (e) CV syllable with targetless gesture; (g) CV syllable with eccentric timing.

The hypothesized coupling diagrams shown in Fig. 5.2 summarize the possible shifts in gestural coordination before and after tonogenesis. Whether Old Tibetan complex onsets had competitive or anti-phase coupling (see discussion in Section 2.4), in Fig. 5.2(a-b), cluster simplification meant the loss of C_1, not C_2, and so the phasing of the remaining consonant to the vowel changed. In tonogenesis, the complex onsets (a-b) and simplex onsets (c) acquired an additional tone gesture, resulting in (d). Those speakers who have subsequently lost tone but retained tone-like C-V coordination now have either

(e) a targetless gesture or (f) eccentric C-V coupling. These processes of cluster simplification, tonogenesis, and tone loss thus retain C-V timing, but through different coupling graphs rather than substitution of gestures.

5.3.3 Coupling as phonology

The investigation of gestural timing in the preceding chapters has shown that coordination relations are language-specific structural elements in the phonological system. Coordination relations play a crucial role in the interface between discrete and continuous aspects of speech production and perception. On the one hand, phonological analysis rests upon dividing speech into discrete units, often in linear order. In Articulatory Phonology, gestural specifications for constriction location and constriction degree are discrete, and continuous coupling options are often discretized into in-phase and anti-phase coupling. An important challenge is to explain the emergence of discretized behavior of gestures in time, which will involve the interplay of contrastivity and uniformity.

A lexical contrast can be instantiated by changing underlying units of phonological representation, such as phonemes, features, gestures, or the units of temporal coordination. Changing the timing of gestures can also express phonological contrasts. For example, the contrast between the English words *tack* and *cat* can be described as different orders of the phonemes [t æ k], or as changing the relative timing of alveolar and velar closures. Gestures can also capture subsegmental structure such as the contrast between Russian palatalized stops and stop + glide sequences, which have been analyzed as consisting of the same gestures coordinated in different ways (Shaw et al 2019). Tone contrasts can also manifest in time, either through association with different syllables (as in many African languages), or within a single syllable, as has been proposed for Serbian (Karlin 2018), Shilluk (Remijsen & Ayoker 2014; Barnes et al 2019) and Luganda (Myers et al 2019). A mechanism for the emergence of discrete patterns in the spatial aspect of gestures has been offered by Quantal Theory (e.g. Stevens 1989); analogous mechanisms for the temporal aspect may be rooted in agent-based modeling (Browman & Goldstein 2000) and/or syllable structure (Shaw & Gafos 2015). The examples just cited indicate that coupled oscillators are a promising tool for for bridging continuous and discrete patterns in the temporal dimension.

However, in order to treat gestural coupling as phonology, more is needed beyond the possibility of discrete behavior. Optimality Theory constraints have referenced the alignment of gestural landmarks (e.g. Gafos 2002) or coupling relations and moraic structure (Walker and Proctor 2019). A nimplementation that accounts for the Tibetan facts would need to account for target uniformity. This could be done by supplementing an Optimality-Theoretic grammar such as that of Walker and Proctor (2019) with versions of the constraints from Chodroff (2017), generalized to intergestural timing. While Chodroff (2017) defines constraints in terms of features and contrasts, the Tibetan case differs in that C-V lag is consistent across contexts.

What would these constraints look like? Faithfulness constraints assess the difference between underlying and surface forms; as such, they reference characteristics that are potentially contrastive. Coupling relations certainly can be contrastive, in terms of which gestures are coupled (e.g. *cat* vs. *tack*) and whether a given coupling is in-phase or anti-phase (Shaw et al. 2019). Data from this dissertation is not sufficient to determine whether eccentric coupling can be contrastive, however, since the eccentric coupling proposed in the analysis of Tibetan C-V timing is consistent across productions and across speakers. Instead, this resembles markedness constraints, which assess the well-formedness of surface representations.

Markedness constraints for target uniformity in Tibetan could enforce a consistent C-V coupling, but the VOT results remain to be explained. Despite the variation across speakers (whether two or three categories of VOT) and the variable aspiration and voicing (among three-category speakers), uniformity can still be found in the consistency of individual VOT productions. That is, once a particular surface form is selected, the observed VOT does form clusters—tokens are either prevoiced or not, aspirated or unaspirated, not spread across a continuum of intermediate forms (see 4.3.2). This indicates variability across and within speakers in the selection of output forms, but uniformity in the forms themselves. Temporal target uniformity thus appears active among the set of possible markedness constraints.

5.4 Contrast maintenance

5.4.1 Tibetan diachrony

The sociolinguistic circumstances of Tibetan speakers in diaspora provide a unique window on language change. Speakers raised in this environment belong to an interconnected network of Tibetan enclaves embedded within larger communities speaking other languages. Children are exposed to Tibetan speakers of diverse backgrounds while simultaneously acquiring one or more other languages. Tibetan-acquiring children develop their phonological systems informed by a microcosm of the extensive restructuring of the tonal and laryngeal contrasts that has been ongoing across the Tibetan-speaking world.

The results of this process provide valuable insight about how language change unfolds in an increasingly mobile and multilingual world.

This dissertation has investigated two aspects of variation present in speakers of common Tibetan in diaspora: tones that are merged or unmerged and VOT that falls into two or three categories. As demonstrated in sections 3.3.3 and 4.3.2 and replicated below in Table 5.1, even speakers with merged tones still follow the tone-based categories in their VOT: speakers with two VOT categories only produce aspiration in historically high-tone words, and speakers with three VOT categories only produce variable prevoicing and aspiration in the appropriate low-tone words.

Orthography	པ་	ཕ་	བ་	རྦ་	མ་	རྨ་
Old Tibetan	*pa	*p^{h}a	*ba	*rba	*ma	*rma
Central Tibetan: Lhasa	pá	p^{h}á	p^{h}à	pà	mà	má
Central Tibetan: Shigatse	pá	p^{h}á	p^{h}à	pà	mà	má
Eastern Tibetan: Dege	pá	p^{h}á	pà	bà	mà	má
Northeastern Tibetan: Golok	pa	p^{h}a	ba > wa	ʁba	ma	ʁma
Diaspora: 3 VOT categories, tone contrast	pá	p^{h}á	pà	pà	mà	má
Diaspora: 2 VOT categories, tone contrast	pá	p^{h}á	p^{h}à	pà ~ bà	mà	má
Diaspora: 3 VOT categories, tone merger	pa	p^{h}a	p^{h}a ~ pa	pa ~ ba	ma	ma
Diaspora: 2 VOT categories, tone merger	pa	p^{h}a	pa	pa	ma	ma

Table 5.1. VOT and tone contrasts in some Tibetan varieties and diaspora speakers

The structure of the variation among the diaspora speakers is notable for its points of difference and points of consistency. The variants present—tonal and non-tonal, two and three VOT categories—shown in Table 5.1 reflect common patterns across dialects in diaspora-raised speakers' linguistic input. Diaspora speakers have down-selected from many possible systems for VOT to just two, while not (or not yet) settling on a single variant. As surveyed in Geissler (2018), the variants present resemble the demographically- and socially-dominant dialects among those entering diaspora, while avoiding socially- and structurally-marked forms. The process just described could be understood as a step in the development of a new dialect unique to diaspora speakers. For situations where a new dialect develops from the combination of several existing dialects, Trudgill (1986) identifies two key processes. The first, "simplification," occurs when a second generation acquires a subset of the forms in their input, and only the most structurally unmarked forms. The second, "focusing" occurs when a subsequent generation chooses one variant from among those remaining. In this framework, Tibetan as spoken in diaspora has undergone "simplification" to a small number of variants, but not yet "focusing," since speakers are not homogenous.

In spite of variation in voicing/aspiration and tonality, the Tibetan speakers in this study demonstrated remarkable consistency in C-V coordination. The timing of their speech gestures was uniform despite differences in the presence and nature of their tonal and laryngeal gestures. While we do not yet know the mechanism by which speakers settle on similar C-V coordination, it is clear that speakers have converged in this domain.

5.4.2 Multiple cues

Finally, the variable voicing and aspiration among non-tonal speakers with three VOT categories highlights the importance of contrasts as specifying sets of lexical items rather than smaller units such as phonemes. Tonal three-category speakers' knowledge of Tibetan includes the fact that variably-voiced onsets occur only in certain low-tone words such as *sdom* /tòm ~ dòm/ 'spider,' while variably-aspirated onsets only occur in other low-tone words such as *dom* /tòm ~ t^hóm/ 'bear.' Non-tonal three-category speakers also produce these VOT lengths in a variable manner, and with the same lexical items. This indicates that all speakers, whether or not they contrast tone, represent the same sets of lexical items in order to allow them to pattern together. Other languages of the Himalayan region and both Sino-Tibetan and non-Sino-Tibetan languages of Mainland Southeast Asia exhibit contrasts cued by interactions of voicing, aspiration, phonation type, pitch, duration, and other phonetic parameters. The term *register* is used to describe these contrastive sets characterized by multiple, often diachronically-unstable cues (Huffman 1976). One particularly interesting comparison comes from Chru, which Brunelle (2019) and Brunelle & Kirby (2020) describe as primarily contrasting in F0. Some Chru speakers also produce variable prevoicing in one register. Other speakers may not produce prevoicing, but it still affects their perception of the register contrast.

The Chru and Tibetan examples are similar in that both feature a group of speakers who seem to maintain awareness of a contrast they do not produce. In Chru, speakers who do not produce prevoicing still use it in perception. In Tibetan, some speakers do not produce a tone contrast, but the appearance of aspiration and variable prevoicing is still conditioned by the tone categories. Both languages are undergoing change in progress, but they show that, at least for a time, a contrastive representation may be maintained among all members of a speech community as long as some members continue to produce the contrast. That both cases involve VOT and F0 may suggest that listeners are particularly tolerant of interspeaker variability in these cues.

There are other ways the non-tonal Tibetan speakers could have acquired this contrast: for example the stops may exhibit different behavior in other morphological contexts (i.e. intervocalic deaspiration for one set, voicing for the other). If this is the case, the analysis of the voicing contrast would need to be revised to accommodate a neutralization in word-initial position. Future work, including perceptual studies and examination of more morphological positions are needed to learn more about the nature of these categories. The observation remains, though, that non-tonal speakers maintain sets of lexical items that, in other speakers, correspond to tonal categories.

5.5 Summary

The findings of this dissertation support the integration of continuous and discrete aspects of phonetics and phonology. Emphasis is placed on the balance of contrast with phonetic uniformity. Tibetan speakers raised in diaspora use combinations of phonological characteristics from established dialects, including one of two sets of laryngeal contrasts. Some speakers do not produce a tone contrast, but the effects of tone categories remain in all speakers. Specifically, details of VOT depend on lexical tone categories, and all speakers produce consistent C-V lag that varies dynamically with consonant duration. This provides evidence for articulatory target uniformity in the temporal domain, within and across speakers. Extending uniformity to gestural coupling complements previous work on the emergence of articulatory gestures: discrete patterns arise for both spatial and temporal aspects of articulation.

5.6 Chapter bibliography

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology*. Cambridge University Press 3. 255–309.

Browman, Catherine P. & Louis Goldstein. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée* (5). 25–34.

Brunelle, Marc, Thành Tấn Tạ, James Kirby & Lư Giang Đinh. 2019. Obstruent Devoicing and Registrogenesis in Chru. In *Proceedings of the 19th International Congress of Phonetic Sciences*, 5. Melbourne.

Brunnelle, Marc & James Kirby. 2020. Relative cue weighting in the production and perception of register. In. Vancouver.

Byun, Tara McAllister, Sharon Inkelas & Yvan Rose. 2016. The A-map model: Articulatory reliability in child-specific phonology. *Language* 92(1). 141–178. https://doi.org/10.1353/lan.2016.0000.

Carignan, Christopher. 2014. An acoustic and articulatory examination of the "oral" in "nasal": The oral articulations of French nasal vowels are not arbitrary. *Journal of Phonetics* 46. 23–33. https://doi.org/10.1016/j.wocn.2014.05.001.

Chitoran, Ioana & Mark Tiede. 2013. Gestural reorganization under rate pressure interacts with learned language-specific phonotactics. In, 060199–060199. https://doi.org/10.1121/1.4799024.

Chodroff, Eleanor R. 2017. *Structured variation in obstruent production and perception*. Johns Hopkins University Dissertation.

Chodroff, Eleanor & Colin Wilson. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61. 30–47. https://doi.org/10.1016/j.wocn.2017.01.001.

de Jong, Kenneth J., Byung-jin Lim & Kyoko Nagao. 2004. The Perception of Syllable Affiliation of Singleton Stops in Repetitive Speech. *Language and Speech* 47(3). 241–266. https://doi.org/10.1177/00238309040470030201.

Faytak, Matthew. 2020. Articulatory, but not acoustic, target uniformity in Suzhou Chinese. Poster presentation presented at the 2020 Annual Meeting of the Linguistics Society of America, New Orleans.

Faytak, Matthew Donald. Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese. 168.

Gafos, Adamantios I. A Grammar of Gestural Coordination. *Natural Language & Linguistic Theory* 20(2). 269–337.

Gafos, Adamantios I. & Stefan Benus. 2006. Dynamics of Phonological Cognition. *Cognitive Science* 30(5). 905–943. https://doi.org/10.1207/s15516709cog0000_80.

Gleason, P., B. Tuller & J.A.S. Kelso. 1996. Syllable affiliation of final consonant clusters undergoes a phase transition over speaking rates. In, vol. 1, 276–278. IEEE. https://doi.org/10.1109/ICSLP.1996.607099. http://ieeexplore.ieee.org/document/607099/.

Harper, Sarah, Louis Goldstein & Shrikanth Narayanan. 2020. Variability in individual constriction contributions to third formant values in American English /ɹ/. *The Journal of the Acoustical Society of America* 147(6). 3905–3916. https://doi.org/10.1121/10.0001413.

Huffman, Franklin E. 1976. The Register Problem in Fifteen Mon-Khmer Languages. *Oceanic Linguistics Special Publications* (Austroasiatic Studies Part I) 13. 575–589.

Iskarous, Khalil. 2017. The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics* 64. 8–20. https://doi.org/10.1016/j.wocn.2017.05.003.

Katsika, Argyro, Jelena Krivokapić, Christine Mooshammer, Mark Tiede & Louis Goldstein. 2014. The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics* 44. 62–82.

Keating, Patricia. 2003. Phonetic and other influences on voicing contrasts. In *Proceedings of the 15th international congress of phonetic sciences*, 375–378.

Maddieson, Ian. 1996. Gestural economy. *UCLA Working Papers in Phonetics* 94. 1–6.

Nam, Hosung & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. In *Proceedings of the 15th International Congress of the Phonetic Sciences*.

Pierrehumbert, Janet Breckenridge. 1980. The phonology and phonetics of English intonation. Massachusetts Institute of Technology.

Rosen, Eric. 2016. Predicting the unpredictable: Capturing the apparent semi-regularity of rendaku voicing in Japanese through harmonic grammar. In *Proceedings of BLS*, vol. 42, 235–249.

Shaw, Jason A., Karthik Durvasula & Alexei Kochetov. 2019. The temporal basis of complex segments. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, 676–680.

Smith, Bridget J., Jeff Mielke, Lyra Magloughlin & Eric Wilbanks. 2019. Sound change and coarticulatory variability involving English /ɹ/. *Glossa: a journal of general linguistics* 4(1). 63. https://doi.org/10.5334/gjgl.650.

Smolensky, Paul, Matthew Goldrick & Donald Mathis. 2014. Optimization and Quantization in Gradient Symbol Systems: A Framework for Integrating the Continuous and the Discrete in Cognition. *Cognitive Science* 38(6). 1102–1138. https://doi.org/10.1111/cogs.12047.

Smolensky, Paul & Matthew Goldrick. 2016. Gradient symbolic representations in grammar: The case of French liaison. *Rutgers Optimality Archive* 1552.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics*. LONDON: Elsevier Ltd 17(1–2). 3–45. https://doi.org/10.1016/S0095-4470(19)31520-7.

Stevens, Kenneth Noble & Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38(1). 10–19. https://doi.org/10.1016/j.wocn.2008.10.004.

Walker, Rachel & Michael Proctor. 2019. The organisation and structure of rhotics in American English rhymes. *Phonology* 36(3). 457–495. https://doi.org/10.1017/S0952675719000228.

Yu, Alan C. L. 2016. Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *The Journal of the Acoustical Society of America* 139(4). 1672–1690. https://doi.org/10.1121/1.4944992.

Yu, Alan C.L. & Georgia Zellou. 2019. Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics* 5(1). 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815.

Zimmermann, Eva. 2019. Gradient Symbolic Representations and the Typology of Ghost Segments. *Proceedings of the Annual Meetings on Phonology* 7. https://doi.org/10.3765/amp.v7i0.4576. http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4576 (4 February, 2021).