

Dear Search Committee:

Please consider my application for the position of Postdoctoral Fellow in Linguistics. After receiving my Ph.D. from Yale University in 2021, I have been a postdoctoral researcher in the Speech, Lexicon, and Modeling Lab at Heinrich Heine University Düsseldorf (HHU).

In my research, I use acoustic and articulatory phonetics, to understand phonological representation, with a particular interest in timing. My dissertation and related work investigates the relative timing of speech gestures, using data from Tibetan as spoken in diaspora. I conducted sociolinguistic interviews in Nepal to gather an audio corpus of spoken Tibetan, as well as laboratory experiments in the US with electromagnetic articulography (EMA). Key results include variation across speakers in laryngeal phonology and loss of the tone contrast among diaspora speakers, as well as remarkable consistency in intergestural timing. For this work I was awarded a Doctoral Dissertation Research Improvement Grant from the National Science Foundation (USA), allowing me to run more EMA subjects and improve the corpus by having it fully transcribed, forced-aligned, and made publicly available for use by researchers and community members.

At HHU, I am currently working with corpora and simulation while organizing the set-up of our new phonetics lab. Prof. Kevin Tang and I are studying the relationship between phonetic reduction and the contextual predictability of segments, syllables, and words. Specifically, we are interested in the interaction across levels (e.g. predictable segments in unpredictable words) and in how reduction of acoustic duration manifests in articulation. We are in the process of applying for external funding from the German Research Foundation as well as grants to develop innovative teaching materials. Through the use of articulatory simulation, in TADA and VocalTractLab, I have been able to add new data and perspectives both to the probabilistic reduction research and to the study of articulatory timing. This position has allowed me to develop project-oriented courses that bring students into the laboratory. This integration of research and teaching has been a mutually beneficial experience that I would seek to expand in the future.

As of the coming semester, I will have taught ten courses and served as a teaching fellow for seven more. These courses draw from all aspects of my linguistics experience: training in generative linguistics, dissertation research in Articulatory Phonology and historical linguistics, and current work in lexical-usage-based morphophonetics. To complement my research background with training in pedagogy, I undertook the Certificate of College Teaching Preparation offered by Yale's Poorvu Center for Teaching and Learning, and I continue my professional development through participation in the Linguistic Society of America's Scholarship of Teaching Special Interest Group and co-organizing a program of classroom observations in my current department. As part of my ongoing interest in pedagogy, I am deeply interested in the scholarship of teaching, and began contributing with a poster at the recent Annual Meeting on Phonology. I was also coauthor of two papers on gender bias in linguistics example sentences, which have been published in *Language* and the language-pedagogy journal *Babylonia*. Conducting and disseminating research on teaching practices is an area I hope to focus on in the future alongside my work in phonology and phonetics.

I have listed as references my postdoctoral supervisor Kevin Tang, my dissertation advisor Jason Shaw, and graduate school mentor Claire Bowern. Please feel free to contact me with any questions.

Sincerely,  
Christopher Geissler

# Christopher Geissler

Benzenbergstraße 7  
40219 Düsseldorf  
Germany  
[cageissler.github.io](https://cageissler.github.io)

christopher.geissler@hhu.de  
+49 157 5347 4429

## Positions and Education

2021-present	<b>Postdoctoral researcher in English Linguistics</b> Heinrich Heine University Düsseldorf
2021	<b>Ph.D in Linguistics</b> Yale University
2021	<b>Certificate in College Teaching Preparation</b> Yale University Poorvu Center for Teaching and Learning
2016, 2017	<b>M.A., M.Phil in Linguistics</b> Yale University
2013	<b>B.A. in Linguistics, Religion</b> Swarthmore College

## Selected Publications

Weber, Julika & **Christopher Geissler**. [submitted]. Accommodation to passive exposure in the L2. *Proceedings of the 19<sup>th</sup> International Congress of Phonetic Sciences*. Prague, Czechia.

Stein, Anna Sophia, Kevin Tang & **Christopher Geissler**. [submitted]. Modeling probabilistic reduction across domains with Naïve Discriminative Learning. *Proceedings of the 19<sup>th</sup> International Congress of Phonetic Sciences*. Prague, Czechia.

Kotek, Hadas, Rikker Dockum, Sarah Babinski & **Christopher Geissler**. 2021. “Gender bias and stereotypes in linguistic example sentences.” *Language*.

**Geissler, Christopher**, Jason Shaw, Fang Hu & Mark Tiede. 2021. Eccentric C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts. In Tiede, Mark, Doug Whalen, & Vincent Gracco, Eds. *Proceedings of the 12th International Seminar on Speech Production*. Providence, RI.

**Geissler, Christopher**. 2019. Tonal and laryngeal contrasts in Diaspora Tibetan. *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences*. Melbourne, Australia.

Zhang, Muye, **Christopher Geissler** & Jason Shaw. 2019. Gestural representations of tone in Mandarin: Evidence from timing alternations. *Proceedings of the 18<sup>th</sup> International Congress of Phonetic Sciences*. Melbourne, Australia.

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

## Other Publications

Kotek, Hadas, Rikker Dockum, Sarah Babinski & **Christopher Geissler**. 2021. "Gender stereotypes and inclusion in language teaching." *Babylonia*.

**Geissler, Christopher**. 2020. Review of Hill, Nathan W. 2019. *The Historical Phonology of Tibetan, Burmese, and Chinese*. Cambridge University Press. *Phonology* 37(3), p. 495-500.

DeMille, Mellissa MC, Kevin Tang, Chintan M. Mehta, **Christopher Geissler**, Jeffrey G. Malins, Natalie R. Powers, Beatrice M. Bowen, Andrew K. Adams, Dongnuh T. Truong & Jan C. Frijters. 2018. Worldwide distribution of the DCDC2 READ1 regulatory element and its relationship with phoneme variation across languages. *Proceedings of the National Academy of Sciences* 115(19). 4951–4956.

## Conference Presentations

Weber, Julika, Lara Rüter, **Christopher Geissler** & Kevin Tang. [submitted]. Reviewing introductory phonetics and phonology with e-learning.

O'Reilly, Eoin, **Christopher Geissler** & Kevin Tang. 2023. Uncover articulatory correlates of acoustic duration with analysis-by-synthesis: the case of diphthongs. *Phonetica und Phonologie in Europa* 2023. Nijmegen, Netherlands.

**Geissler, Christopher**. 2022. From gestural landmarks to analysis-by-synthesis: tone-driven timing in Diaspora Tibetan. Poster presented at 18th Phonetik und Phonologie im deutschsprachigen Raum. Bielefeld, Germany.

**Geissler, Christopher**. 2021. Backward-designing a graduate phonology seminar. Poster presented at Workshop on Teaching Phonology, *Annual Meeting on Phonology 2021*.

**Geissler, Christopher** and Jason Shaw. 2020. Eccentric C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts. *12th International Seminar on Speech Perception*. Online.

**Geissler, Christopher**, Jason Shaw, Fang Hu & Mark Tiede. 2020. Consistent C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts. *94th Annual Meeting of the Linguistic Society of America*. New Orleans, LA.

Kotek, Hadas, Rikker Dockum, Sarah Babinski & **Christopher Geissler**. 2020. Gender bias in linguistic example sentences. *94th Annual Meeting of the Linguistic Society of America*. New Orleans, LA.

**Geissler, Christopher A.** 2019. Phonetic timing and Tibetan consonant, vowel, and tone contrasts. *Phonology of the Northeast 2019*. New Haven, CT.

**Geissler, Christopher A.** 2018. VOT in Tibetan is conditioned by tone. *The Journal of the Acoustical Society of America* 143(3). 1756–1756. Minneapolis, MN.

**Geissler, Christopher**. 2018. Deriving three surface VOT categories from a two-way phonological contrast in Tibetan. *26th Manchester Phonology Meeting*. Manchester, UK.

Whalen, Doug H., Christian DiCanio, **Christopher Geissler** & Hannah King. 2016. Acoustic realization of a distinctive, frequent glottal stop: The Arapaho example. *The Journal of the Acoustical Society of America* 139(4). 2212–2213.

**Geissler, Christopher.** 2016. Semi-Creole Language Typology in Tibeto-Burman. *4th Himalayan Studies Conference*. Austin, TX.

**Geissler, Christopher.** 2016. Identifying semi-creoles in Tibeto-Burman. *Society for Pidgin and Creole Linguistics*. Washington, D.C.

Whalen, Doug H., Christian DiCanio, **Christopher Geissler** & Hannah King. 2016. Acoustic realization of a distinctive, frequent glottal stop: The Arapaho example. *Society for the Study of the Indigenous Languages of the Americas*. Washington, D.C.

## Courses Taught

### As Instructor at Heinrich Heine University Düsseldorf:

- Introduction to English Linguistics, Part II. Summer 2023.
- Laboratory Phonology (advanced seminar). Summer 2023.
- Introduction to English Linguistics, Part I. Winter 2022/2023.
- Laboratory Phonology (method module). Winter 2022/2023.
- Experimental Phonetics (M.A. seminar). Summer 2022.
- Languages of the United States (advanced seminar). Summer 2022.
- The Phonetics-Phonology Interface (advanced seminar). Winter 2021/2022.
- Morphology/Word-Formation (intermediate seminar). Winter 2021/2022.

### As Instructor at Yale University:

- Topics in Phonology: The Phonetics-Phonology Interface. Spring 2021.
- Introduction to Linguistics. Yale Summer Session 2020

### As Teaching Assistant at Yale University:

- Language and Mind. Fall 2020. Instructor: Maria Piñango
- Grammatical Diversity in U.S. English. Spring 2020. Instructor: Raffaella Zanuttini
- Introduction to Linguistics, Fall 2019. Instructor: Jason Shaw
- General Phonetics, Spring 2018. Instructor: Jonathan Manker
- Introduction to Linguistics, Fall 2017. Instructors: Jim Wood, Matt Barros
- General Phonetics, Spring 2017. Instructor: Jason Shaw
- Introduction to Phonological Analysis, Fall 2016. Instructor: Ryan Bennett

## Supervisions

Weber, Julika. 2023 (anticipated). M.A. Thesis: *Phonological accommodation in the L2*. Heinrich Heine University.

Stein, Anna Sophia. 2023 (anticipated). B.A. Thesis, *Modeling reduction with Naïve Discriminative Learning*. Heinrich Heine University. (co-advised with Kevin Tang)

Nguyen, Lena. 2023. B.A. Thesis: *The Stylistics of Morrison and Hurston - A comparative and corpus-based style analysis of African-American fiction*. Heinrich Heine University. (second reader)

Smith, Slater. 2021. B.A. Thesis: *Sounds in flux: modeling post-vocalic laryngeals in Mayan languages*. Yale University. (second reader)

## Grants and Awards

Heinrich Heine University “E-Learning Föderfonds”

Title: Choose Your Own Adventure: Teleporting to a Virtual Phonetics Laboratory

Role: Co-applicant (with Kevin Tang and Eoin O'Reilly)

Awarded February 2023

2023-2024

9,727.78 EUR

Heinrich Heine University “E-Learning Föderfonds”

Title: Interactive web-based review units for phonetics and phonology

Role: Co-applicant (with Kevin Tang and Akhilesh Kakolu Ramarao)

Awarded February 2022

2022-2023

9,423.69 EUR

National Science Foundation

Title: *Doctoral Dissertation Research: Articulatory and Corpus Study of Relative Timing among Consonants, Vowels, and Tones in Tibetan.*

Role: Co-Principal Investigator

Principal Investigator: Jason Shaw

Awarded July 2019

2019-2022

\$9,903.55

**University Fellowship.** Yale University. 2014-2021

**South Asia Council Field Research Award.** Yale University. July 2014

## Membership and Affiliations

LSA - Linguistic Society of America

LSA Special Interest Group: Scholarship of Teaching and Learning

ASA - Acoustical Society of America

LabPhon - Association for Laboratory Phonology

ANHS - Association for Nepal and Himalayan Studies

## Service

**Co-organizer**, “AI for Social Good”, presentation at Night of Science 2022, Düsseldorf

**Graduate Affiliate**, Branford College, Yale University. 2015-2020

**Student Co-organizer**, Phonology of the Northeast conference (PhoNE), 2015 and 2019

**Organizer**, Yale Phonology Group, 2015-2018

**Website Manager**, Department of Linguistics, Yale University. 2015-2017

**Representative**, Graduate Student Assembly, Yale University. 2014-2016

## **Research Statement**

### **Tone and articulatory timing**

In my dissertation and related work, I test the empirical predictions made by theoretical proposals, particularly in terms of subsegmental representation as framed in Articulatory Phonology. I am drawn to this paradigm for its ability to make falsifiable predictions about the temporal coordination between subsegmental units (gestures). I used electromagnetic articulography (EMA) to measure C-V lag, the difference in time between the start of a word-initial consonantal gesture and the start of the following vowel. Results of experiments with Tibetan as spoken in Diaspora, alongside fieldwork conducted in Nepal, expand the typological sample of languages studied in this way, and shed light on the representation of tone and on the systems of coordination among articulatory gestures. As part of this study, I gained experience in every aspect of laboratory research, participated in the initial setup of the laboratory, trained and mentored other students. I also learned the value of drawing from multiple types of data, such as highly-controlled experimental data from few speakers, and less-controlled, corpus-style data from many speakers. More recently, I have begun using articulatory simulation to test predictions of the tone gesture hypothesis, thereby using a different line of evidence to approach the problem.

As an early test of the use of articulatory simulation, I am currently working with a Ph.D. student in my department (Eoin O'Reilly) on simulating phonetic reduction in English diphthongs. In a semi-automated process, we make a number of articulatory simulations, each varying a particular parameter, then see which best matches real-world data. This work has been accepted for presentation at the upcoming Phonetics and Phonology in Europe 2023 conference, and we eagerly continue working through this complex data. In the long term, we hope to extend this system to develop a method to rapidly automate this process in order to make gestural annotations of articulatory data. I fully intend to continue collaborating with this project at my next position.

### **Phonetic detail in probabilistic reduction**

At Heinrich Heine University, I am currently working with corpora and simulations while we set up our new phonetics lab. Kevin Tang and I are studying the relationship between phonetic reduction and the contextual predictability of segments, syllables, and words. Specifically, we are interested in the interaction across levels (e.g. predictable segments in unpredictable words) and in how reduction of acoustic duration manifests in articulation. This research is motivated in the context of lexical-usage-based morphology, and is in coordination with several other projects in a multi-site research collaboration. We hope to use the probabilistic reduction effect to shed light on the nature of lexical representation and the effect of morphology on phonetics.

Kevin Tang, two students, and I are working on two papers that compare different ways to model probabilistic reduction of acoustic duration: a traditional n-gram language model and Naïve Discriminative Learning. In the coming months and years, we intend to complement this project with articulatory data, using both existing articulatory corpora and articulatory modeling to study how probabilistic reduction takes place in articulation.

## **Scholarship of teaching and other projects**

Alongside phonetics and phonology, I am increasingly interested in conducting research that seeks to improve the field of linguistics. In papers published in *Language* and *Babylonia* in 2021, my co-authors and I report our findings on the distribution of gender in the example sentences of linguistics research articles. For example, we have found about twice as many male-gendered arguments as female-gendered arguments, with female arguments proportionately less likely to be grammatical subjects or agents, among other stereotypically gendered trends.

In the future, I hope to continue turning the tools of research toward scholarly communication and effective, equitable teaching practices. As an example, my colleagues and I at Heinrich Heine University received a grant to develop self-paced online review units for phonetics and phonology. Having tested them in our courses, we have submitted our findings for presentation at the upcoming 2023 Conference on Scholarly Teaching in Linguistics.

## Teaching Philosophy

As a teacher, I am particularly drawn to teaching linguistics in the context of a broader education that includes not only linguistics but a range of other fields. I consider phonetics, phonology, and historical linguistics to be my core areas of expertise, but I have also taught morphology, sociolinguistics, and a number of introductory-level courses from different perspectives. The many areas of linguistics naturally interface with the study of the cognitive sciences, languages, social sciences, and computer science, among others. I cherish opportunities to work closely with students as they develop not only as linguists, but also as scholars who bring the tools and findings of linguistics to their other interests.

At every level of a course, I try to structure my teaching around defined overarching goals, following the principles of Backward Design. In this framework, instructors start by articulating learning goals and outcomes, then designing assessments to measure how well students have achieved those goals, then plan the activities and assignments to prepare students accordingly. For example, in a recent seminar, I identified reading academic journal articles as a core skill. Therefore, I began the course with explicit instruction in how to read an article and regularly revisit the topic in discussion. Short weekly posts to a discussion forum required students to engage with articles, but to do so in a way that reflected their own needs and interests.

This focus on learning goals also centers the diversity of students who come to linguistics classes. Most students come to linguistics along with other interests, and I greatly enjoy helping students use methods and findings from linguistics to inform their other studies. Self-directed assignments can be a valuable tool: an opera singer in a phonetics class wrote a final paper investigating the acoustics of vowels sung at high pitches, while a computer science student in the Grammatical Diversity in US English class investigated the kinds of errors made by YouTube's automatic captioning on the English of African-American speakers. In both cases, the courses were small enough that the instructor and I could spend sufficient time mentoring these students, and I hope to encourage future students to develop these creative connections.

Questions of "JEDI" (justice, equity, diversity, and inclusion) are especially important in linguistics due to the nature of our subject matter. We impress on our students that there is no such thing as 'good' or 'bad' language, but this is not sufficient to address the colonial, imperialist history of linguistics or the ways language is used as a tool of power and oppression. This is especially salient for indigenous, endangered, and minority language varieties. In my work with Tibetan, I strive to maintain the attitude of a guest to the language and its speaker community. Native speakers are experts whose knowledge we linguists seek, and those of us who are not community members are visitors.

In the future, I hope to implement Skills-Based Grading<sup>1</sup> in my teaching. This approach replaces traditional grading rubrics with a list of skills, and presents assignments and assessments as opportunities to demonstrate proficiency in those skills. A student's grade is determined by the number of skills in which they have demonstrated proficiency, and they have multiple opportunities to do so. I admire the transparency of this approach for how it invites students into their own learning process and how it shifts the psychology of grades from (self-)judgement to progress. This approach has the potential to benefit all students, but as with Backward Design, the habit of clearly communicating expectations is key to making the classroom a more equitable place.

---

<sup>1</sup> Zuraw, Kie, Ann M. Aly, Isabelle Lin & Adam J. Royer. 2019. Gotta catch 'em all: Skills grading in undergraduate linguistics. *Language* 95(4). e406–e427. <https://doi.org/10.1353/lan.2019.0081>.

## **Statement on Diversity, Equity, and Inclusion**

As a linguist, I have found that questions of diversity, equity, and inclusion (DEI) invariably arise due to the nature of our subject matter. We impress on our students that there is no such thing as ‘good’ or ‘bad’ language, but this alone is not sufficient to address the colonial, imperialist history of linguistics or the ways language is used as a tool of power and oppression. This is especially salient for indigenous, endangered, and minority language varieties. In my work with Tibetan, I strive to maintain the attitude of a guest to the language and its speaker community, the language belongs to its speakers and the rest of us are visitors. Outside the specific topic areas of linguistics, strategies to address DEI injustices can benefit all students and participants in academic communities.

The clear communication embodied by teaching frameworks such as Backward Design and skills-based grading has the potential to benefit all students, but especially those who are struggling or vulnerable. For example, first-generation students and those from underrepresented backgrounds often find it difficult to navigate the complex network of implicit assumptions embedded in college life and the cultures of particular disciplines. One prominent example has been that many students I teach have never been taught how to read a research article. Those from certain backgrounds are more likely to be able to handle this task, but all students will benefit from explicit instruction. So, for many classes, I include explicit instruction on how to approach a research article by leveraging its structure. The more that expectations and skills are communicated transparently, the more equitable the classroom will be.

Focusing on skills is especially valuable in the context of an undergraduate education, which particularly encourages students to develop connections between disciplines and systems of knowledge. I love most of all when my students pursue associations that would never occur to me, like the opera singer who wrote about the acoustics of vowels sung at high pitches, or the computer science student who investigated the errors made by YouTube’s automatic captioning on the English of African-American speakers. In this light, the responsibility of the instructor is to support students as they develop their own trajectories.

In order to better understand the role of implicit bias in the academy, I have been engaged in collaborative research on the distribution of gendered arguments (i.e. characters) in the example sentences of contemporary linguistics journal articles. For example, we have found about twice as many male-gendered arguments as female-gendered arguments, with female arguments proportionately less likely to be grammatical subjects or agents, among other stereotypically gendered trends. We were able to hire a team of undergraduate research assistants, for whom this served as an entry point for engaging in research. We have been able to disseminate this through presentations, a webinar, and publications in the linguistics journal *Language* and in the language-pedagogy journal *Babylonia*. In the future, I hope to continue turning the tools of research toward scholarly communication and effective, equitable teaching practices.

# Eccentric C-V timing across speakers of diaspora Tibetan with and without lexical tone contrasts

Christopher Geissler<sup>1</sup>, Jason A. Shaw<sup>1</sup>, Mark Tiede<sup>2</sup>, Fang Hu<sup>3</sup>

<sup>1</sup>*Yale University, USA*

<sup>2</sup>*Haskins Laboratories, USA*

<sup>3</sup>*Chinese Academy of Social Sciences, China*

christopher.geissler@yale.edu, jason.shaw@yale.edu, tiede@haskins.yale.edu,  
hufang@cass.org.cn

## Abstract

This paper investigates the relative timing of onset consonant and vowel gestures in Tibetan as spoken in the Tibetan diaspora. According to the coupled oscillator model of articulatory timing (Browman & Goldstein 2000, Nam & Saltzman 2003), the most readily-available coupling modes among gestures are in-phase (synchronous) or anti-phase (sequential) timing, with competition among these modes also giving rise to a stable timing pattern. The model predicts that other timing relations, i.e. “eccentric timing”, are possible but not as readily available. Data gathered using electromagnetic articulography (EMA) shows relative C-V timing consistent with either competitive coupling or eccentric timing. Competitive coupling is a plausible explanation for CV syllables in a tone language (Gao 2008), but acoustic analysis showed that some speakers do not produce a pitch contrast corresponding to tone. In the apparent absence of a tone gesture, we conclude that these speakers exhibit eccentric C-V timing.

**Keywords:** speech prosody, articulation, tone, Tibetan, Articulatory Phonology, Electromagnetic Articulography

## 1. Introduction

Languages differ in how articulatory gestures are coordinated in time, including the relative timing between consonants and vowels, i.e., C-V timing. Previous work has observed that in lexical tone languages, such as Mandarin (Gao 2008), Thai (Karlin 2014), and Lhasa Tibetan (Hu 2016), the lag between consonant and vowel gestures (C-V lag) is longer than in non-tonal languages, such as English (Löfqvist & Gracco 1999). Additionally, C-V lag is longer for tonal syllables than non-tonal syllables in Mandarin (Zhang et al. 2019). These observations support treating tone as an articulatory gesture (Gao 2008; Niemann et al. 2011) which can be organized with other gestures as competitively-coupled oscillators (Browman & Goldstein 2000, Nam & Saltzman 2003).

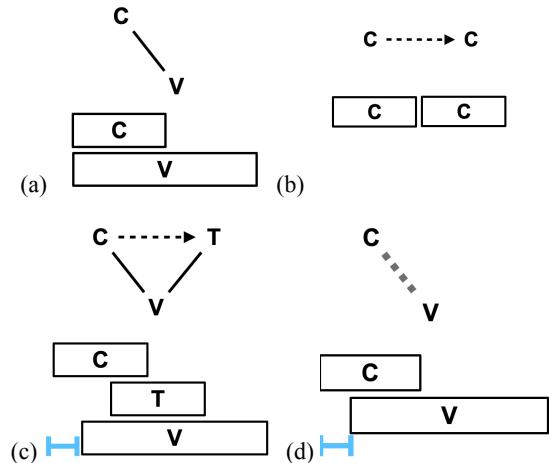
In the present study, we focus on Tibetan as spoken in the diaspora enclaves in South Asia and around the world, outside of traditional Tibetan-speaking regions. Speakers raised in these communities are exposed to a mixed linguistic input comprising speakers of tonal and non-tonal Tibetan dialects, along with extensive contact with other languages. Of these diaspora-raised speakers, some produce a tone contrast while others do not. This fact makes diaspora Tibetan speakers a unique case in which to test the effect of tone on articulatory timing.

### 1.1. Coupling relations

Pairwise coupling relations are a prominent way to account for the timing of gestures in Articulatory Phonology. By hypothesis, any relative phasing of gestures is possible, but two cou-

pling modes are considered to be intrinsically stable: *in-phase* (synchronous) and *anti-phase* (sequential) (Saltzman & Byrd 2000). Language-specific timing patterns can arise from the interaction of these two coupling modes, as in the case of *competitive coupling*. Competitive coupling has been invoked to explain the partial overlap in onset consonant clusters (Nam & Saltzman 2003, Goldstein et al. 2000) as well as the difference in C-V timing between languages with and without lexical tone (Gao 2008). A schematic depiction of coupling graphs and their associated gestural scores is shown in **Figure 1**. The first two diagrams depict (a) in-phase coupling and (b) anti-phase coupling using common examples of a CV syllable and a CC sequence, respectively. In **Figure 1(c)**, competitive coupling is illustrated with anti-phase coupling of consonant and tone gestures and in-phase coupling of both with a vowel gesture.

While in-phase, anti-phase, and competitive coupling should be most readily available to speakers, cases of other timing relations are possible as well. Other timing relations are called *eccentric timing* (Goldstein 2011) and represent any timing values besides those already mentioned. Importantly, eccentric timing allows a pairwise timing between two gestures similar to that found in cases of competitive coupling, as illustrated in **Figure 1(d)**.



**Figure 1:** (a) In-phase and (b) anti-phase coupling of articulatory gestures. (c) Competitive coupling of consonant, vowel, and tone gestures can result in similar C-V timing as (d) eccentric timing.

### 1.2. Predictions

As a language with lexical tone, Tibetan may be expected to show similar C-V lag to other tone languages. Following Gao (2008), a tone gesture with competitive coupling would predict a C-V lag value of ~50 ms. However, Tibetan as spoken in

diaspora presents a high degree of diversity, with some speakers lacking a tone contrast. Speakers who do not produce a tone contrast would be expected to exhibit in-phase C-V timing, and therefore ~0 ms C-V lag. Alternatively, with eccentric C-V timing, the C-V lag of non-tonal speakers could take on other values, including resembling that of tonal speakers.

## 2. Methods

Six native speakers (four female, two male) of Tibetan participated in this study. All were raised in Tibetan diaspora communities in India and Nepal and were living in the United States.

Target items ( $N = 71$ ) were one- and two-syllable words that varied in word-initial consonant (/m/, /p/, /p<sup>h</sup>/) and tone (high-level and low-rising); first-syllable vowels drew from the set of back vowels in Tibetan: /u/, /o/, /a/. Speakers produced each item 4–10 times. Target words were presented using Tibetan orthography in the carrier phrase /tsʰik tʃ \_\_\_\_ tük/ ‘This word is \_\_\_\_.’ This phrase provided a front vowel preceding the target item, allowing the back vowel of the target item to be identified with tongue dorsum retraction.

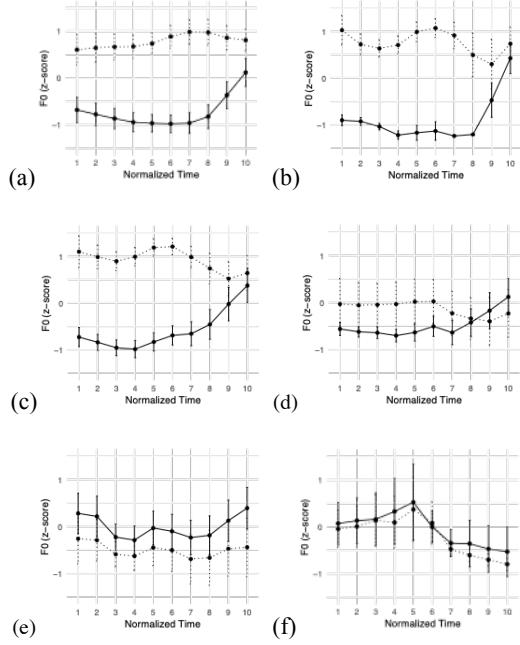
Recordings were made using the NDI Wave EMA system using sensors aligned with each speaker’s midsagittal plane. Consonant gestures were identified using lip aperture, the Euclidean distance between sensors on the upper and lower lips, and vowel gestures using movement in the longitudinal (anterior-posterior) dimension of a sensor placed on the tongue dorsum. Articulatory gestures were identified using *Mview* (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. C-V lag was calculated as the difference in start time between consonant and vowel gestures.

In order to identify which speakers produced a tone contrast, time-normalized F0 trajectories were analyzed for systematic differences by lexical tone category. Data for this analysis came from monosyllabic nasal-initial tokens (60 per speaker). The target syllables were identified as beginning with the acoustic start of the nasal consonant and ending with the end of a visible second formant at the onset of the following stop closure. F0 was measured at ten evenly-spaced intervals using a script (DiCanio 2012) in Praat (Boersma & Weenink 2018)

## 3. Results

### 3.1. Tone contrast status

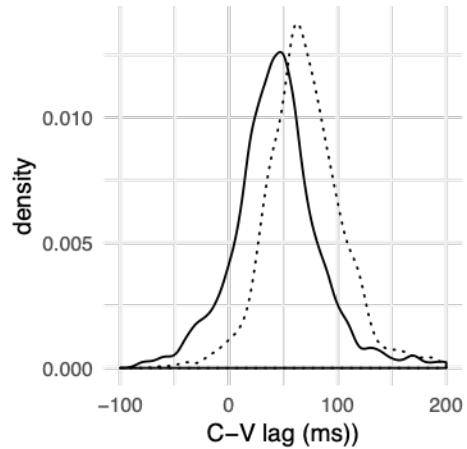
Four of the six speakers were found to produce a significant difference in F0 trajectories on nasal-initial syllables, which we interpret as evidence for a tone contrast in production. Generalized Additive Mixed Models (*mgcv*, Wood 2011; in R: R Core Team 2011) were fit to the data of each speaker, predicting F0 based on terms for lexical tone category, a smooth for normalized timestep, a difference smooth by lexical tone category, and random smooths by word. The difference smooth for two participants (one female, one male) was not significant, indicating they did not produce a lexical tone contrast. The difference smooth for the other four participants was significant. Raw pitch trajectories are presented in **Figure 2**.



**Figure 2:** Time-normalized F0 trajectories on /mV/ syllables. F0 values are z-scored by speaker. Model comparison with GAMMs found significant differences by tone in (a)–(d) but not for (e)–(f)

### 3.2. C-V lag

Overall C-V lag values were similar to those observed in other languages with lexical tone, centering around 50 ms, and these values did not differ by tone. However, C-V lag values were similar for speakers with and without tone contrasts, as established in 3.1. C-V lag values according to tone contrast status are presented in **Figure 3**.



**Figure 3:** C-V lag among speakers with tone contrast (solid line) and without tone contrast (dotted line).

Furthermore, we found that C-V lag is positively correlated with the duration of the onset consonant gesture. This held true for both speakers with and without the tone production contrast, as shown in **Figure 4**. This pattern is not expected if the gestures are in-phase (Shaw, Durvasula, Kochetov, 2019).



**Figure 4:** Positive correlation of consonant gesture duration on C-V lag.

These observations—the lack of difference across speakers by tonality, and the correlation of C duration on C-V lag—were substantiated using linear mixed-effects models (lme4; Bates et al. 2015). We performed a linear mixed-effects model analysis of the relationship between consonant gesture duration and C-V lag. A baseline model included fixed effect of onset consonant [m p p<sup>h</sup>] and random effects of speaker and lexical item. This was compared to a second model that also included a fixed effect of consonant duration. Comparison was also made to a third model that included a fixed effect of tonality: whether or not a speaker produced a tone contrast. As shown in **Table 1**, the inclusion of consonant duration led to an improvement over the baseline model, but the further inclusion of tonality did not further improve the model.

model	AIC	BIC	loglik	Chi-sq	p-value
baseline	31776	31810	-15882		
baseline + C dur	31764	31804	-15875	13.940 3	0.0002
baseline + C dur + tonality	31764	31810	-15874	1.7267	0.1888

**Table 1:** Comparison of linear mixed-effects models shows improvement with consonant duration, but no improvement with tone contrast status of speakers.

#### 4. Discussion and conclusion

Observed C-V lag values were around 50 ms and were positively correlated with consonant gesture duration, consistent with both competitive coupling and eccentric timing. However, speakers lacking a tone contrast exhibited similar C-V lag to speakers with a contrast, which we interpret as evidence that non-tonal speakers, at least, use eccentric timing. This is because these speakers lack a tone gesture, and so do not have another gesture with which to make competitive coupling possible. We further rule out the possibility of C-V timing being anti-phase because the C-V lag values are substantially less than consonant duration (even defined as just the closing phase of the consonant gesture).

We thus conclude that while tone-contrasting Tibetan speakers may use competitive coupling, the non-tonal speakers in our sample used eccentric timing. All the speakers in this study learned a pattern of C-V timing common to their com-

munity, and so some speakers must have accomplished this using eccentric timing (Marin & Pouplier 2010). Such eccentric coupling is predicted to be possible (e.g. Goldstein 2011), but less likely than in-phase and anti-phase coupling, and the present study demonstrates an example of eccentric timing that mimics competitive coupling.

#### 5. Acknowledgements

We would like to thank the speakers for participating in this study, Yale University and the National Science Foundation Graduate Research Fellowships Program for funding support, and Muye Zhang for assistance with EMA recording. We are also grateful for feedback from audiences at Yale Linguistics, LSA 2020, and ISSP 2020.

#### 6. References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer software]. Version 6.0. 43.
- Browman, C. P., & Goldstein, L. (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.
- DiCanio, C. (2012). *Pitch Dynamics Script for Praat, version 6*. [https://www.acsu.buffalo.edu/~cdicano/scripts/Pitch\\_Dynamics\\_6.praat](https://www.acsu.buffalo.edu/~cdicano/scripts/Pitch_Dynamics_6.praat)
- Gao, M. (2008). Tonal alignment in Mandarin Chinese: An articulatory phonology account. *Unpublished Doctoral Dissertation (Linguistics)*, Yale University, CT.
- Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In Proceedings of the XVIth international congress of phonetic sciences (pp. 241–244).
- Goldstein, L. (2011). *Back to the Past Tense in English* (R. Gutiérrez-Bravo, L. Mikkelsen, & E. Potsdam, Eds.; pp. 69–88). Linguistics Research Center, UC-Santa Cruz Department of Linguistics.
- Hu, F. (2016). Tones are not abstract autosegmentals. *Speech Prosody*, 302–306.
- Karlin, R. (2014). The articulatory TBU: gestural coordination of tone in Thai. *Cornell Working Papers in Linguistics*.
- Löfqvist, A., & Gracco, V. L. (1999). Interarticulator programming in VCV sequences: Lip and tongue movements. *The Journal of the Acoustical Society of America*, 105(3), 1864–1876.
- Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, 14(3), 380–407.
- Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of Syllable Structure: A Coupled Oscillator Model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to Phonological Complexity*. Walter de Gruyter. <https://doi.org/10.1515/9783110223958>
- Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. *Proceedings of the 15th International Congress of the Phonetic Sciences*.
- Niemann, H., Mücke, D., Nam, H., Goldstein, L., & Grice, M. (2011). Tones as Gestures: The Case of Italian and German. *Proceedings of the 17th International Congress of the Phonetic Sciences*, 1486–1489.
- R Core Team. (2013). R: A language and environment for statistical computing.

- Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science*, 19(4), 499–526.
- Shaw, J. A., Durvasula, K., & Kochetov, A. 2019. The temporal basis of complex segments. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019 (pp. 676-680). Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- Tiede, Mark. (2005). Mview: Software for visualization and analysis of concurrently recorded movement data.
- Wood, S. 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.
- Zhang, M., Geissler, C., & Shaw, J. (2019). Gestural representations of tone in Mandarin: Evidence from timing alternations. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019 (pp. 1803-1807). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

## 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	8
Acknowledgement	10
Dedication	12
Note on nomenclature and conventions	13
1 Introduction	15
1.1 General introduction	15
1.2 Target Uniformity	17
1.3 Tibetan	19
1.4 Theoretical motivation: Articulatory Phonology	21
1.5 Dissertation roadmap	24
1.6 Chapter bibliography	25
2 Diachrony	30
2.1 Chapter overview	30
2.2 Sino-Tibetan	30
2.3 Tibetan language history	33
2.4 Tone and laryngeal contrasts in Tibetan varieties	36
2.5 Tibetan in diaspora	38
2.6 Common Tibetan Consonants	39
2.7 Common Tibetan Vowels	42
2.8 Common Tibetan tone	44
2.9 Laryngeal contrasts in Tibetan language history	47
2.10 Chapter summary	55
2.11 Chapter bibliography	56
3 Corpus study	64

3.1 Introduction	64
3.1.1 Toward phonetic predictions	64
3.1.1 Aspiration and VOT	64
3.1.2 Tone, F0, and VOT	66
3.2 Corpus Methods	67
3.2.1 Participants	67
3.2.2 Procedures	68
3.2.3 Stimuli	69
3.2.4 Data analysis	70
3.3 Results	72
3.3.1 F0 and Tonality	72
3.3.2 VOT	74
3.3.3 VOT and F0	78
3.4 Discussion	79
3.4.1 Enhancement account	80
3.4.2 Gestural scores	83
3.5 Ongoing corpus development	93
3.6 Chapter summary	93
3.7 Chapter bibliography	94
4 EMA study	97
4.1 EMA study introduction	97
4.1.1 C-V lag	97
4.1.2 Perceptual factors in C-V timing	101
4.1.3 Research Questions	104
4.2 EMA methods	105
4.3 EMA results	107

4.3.1 Question 1: Is C-V timing different for speakers with and without tone?	107
4.3.2 Question 2: Does aspiration affect C-V timing?	113
4.3.3 Question 3: How are consonant, vowel, and tone gestures coordinated?	
117	
4.4 Discussion	122
4.5 Chapter summary	126
4.6 Chapter bibliography	127
5 Discussion	134
5.1 Review of results	134
5.2 Community-level temporal target uniformity	136
5.3 Coupling relations	139
5.3.1 Eccentric coupling	139
5.3.2 Alternatives to eccentric coupling	141
5.3.3 Coupling as phonology	145
5.4 Contrast maintenance	147
5.4.1 Tibetan diachrony	147
5.4.2 Multiple cues	149
5.5 Summary	151
5.6 Chapter bibliography	152

## **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-occurrence 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)		Two-way contrast (Dutch)	
	[ + voice]		[ + voice]
	[ -voice]		[ -voice]
(a) [-SG]		aspirated	[ + SG]
		plain	
			(b) [-SG]
			prevoiced
			plain

Three-way contrast (Thai)		Four-way contrast (Hindi)	
	[ + voice]		[ + voice]
	[ -voice]		[ -voice]
		aspirated	[ + SG]
			breathy
(c) [-SG]	prevoiced	plain	aspirated
			[ -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, Tsing 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.*

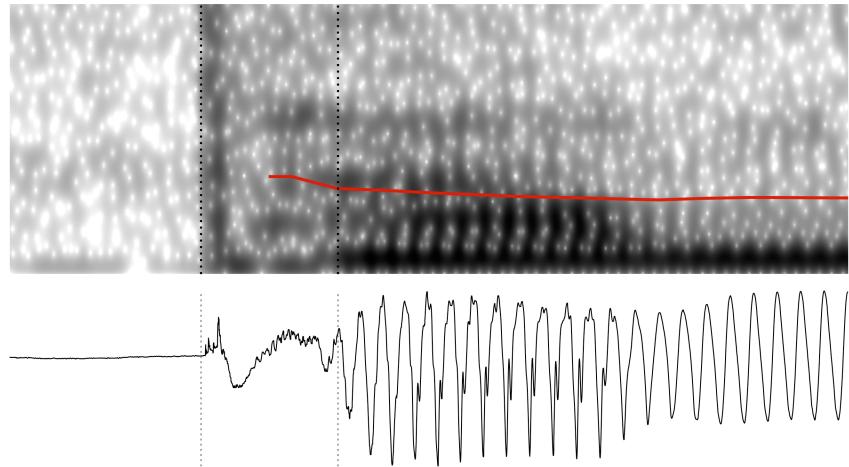
Each place of articulation was represented by four or five items, though the full four-way aspiration/tone contrast was only represented in the dental series, which comprised two minimal pairs for aspiration: the high-tone pair /tá.mák/ *rta dmag* ‘cavalry’ and /t<sup>h</sup>á.mák/ *tha mag* ‘cigarette,’ and the low-tone pair /tōm/ *dom* ‘bear’ and /dōm/ *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 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 [tʰá.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.*

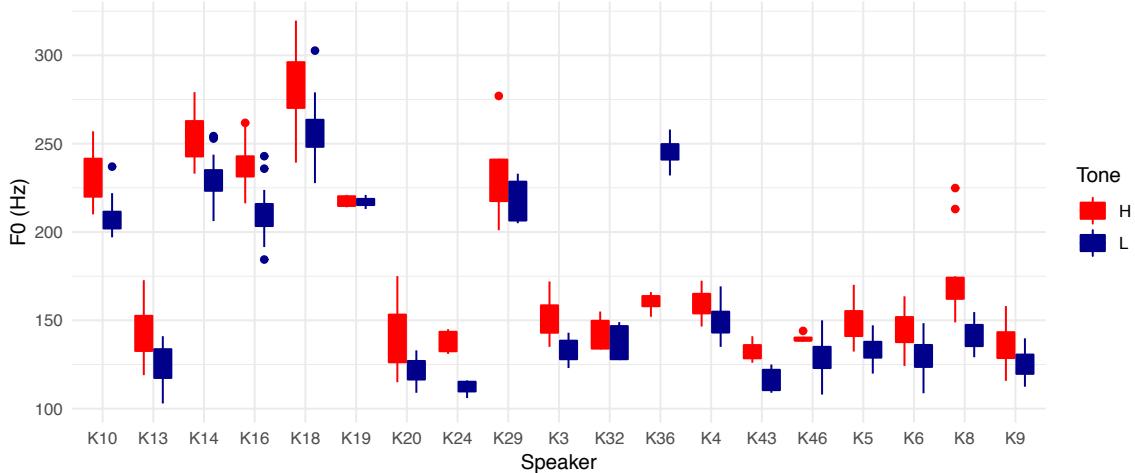
F0 was measured using Praat’s built-in pitch tracker, with a time step of 0.01 seconds and a pitch range of 75-500 Hz for all speakers. Whenever possible, the pitch value recorded was that interpolated at the first period following the release burst; in most cases, this coincided with the endpoint of 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	<i>p</i> -value	Speaker	Coefficient	Chisq	<i>p</i> -value
<b>*K10</b>	23.144	28.5104	1.305E-06	K32	13.900	3.2099	0.1463923
<b>*K13</b>	19.086	19.6039	0.0001143	K36	-76.100	96.2155	< 2.2e-16
<b>*K14</b>	23.302	29.4728	8.506E-07	K4	10.931	6.3785	0.0412732
<b>*K16</b>	28.916	45.0427	3.47E-10	K43	24.872	9.0594	0.0179229
<b>*K18</b>	27.517	40.4372	3.249E-09	K46	19.900	6.5791	0.0412732
K19	9.400	1.4679	0.2256740	<b>*K5</b>	17.356	16.0929	0.0006031
<b>*K20</b>	28.900	13.8759	0.0017576	<b>*K6</b>	14.513	10.6203	0.0089482
<b>*K24</b>	34.900	20.2356	8.901E-05	<b>*K8</b>	34.085	42.7435	1.061E-09
K29	23.400	9.0969	0.0179229	K9	12.730	8.5759	0.0179229
<b>*K3</b>	18.768	19.3094	0.0001223				

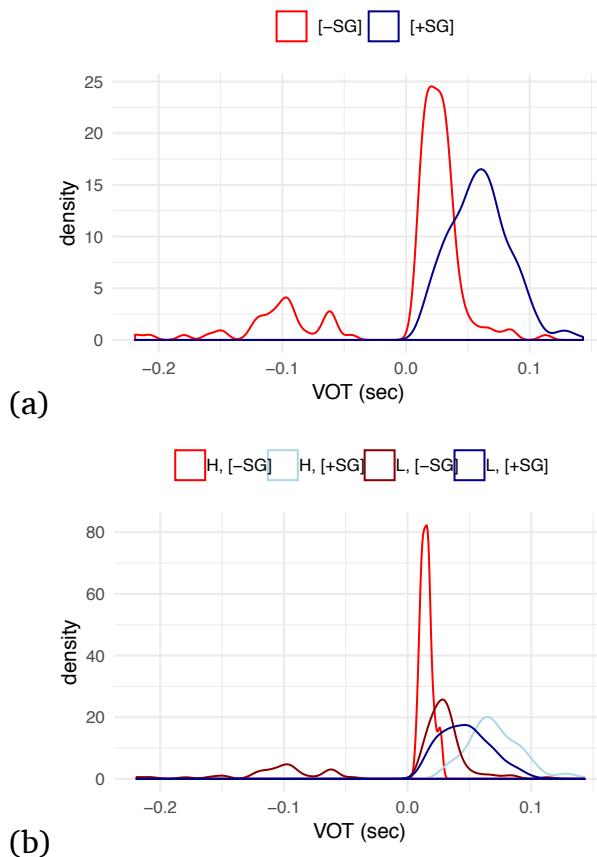
*Table 3.3. Post-hoc analysis of speaker\*tone interactions. Chi-square values vary substantially, with larger values corresponding to larger differences between high and low tones. 11 of 19 speakers, indicated with asterisks (\*) and in bold, have *p* > .01 and positive coefficients, indicating higher 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:



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

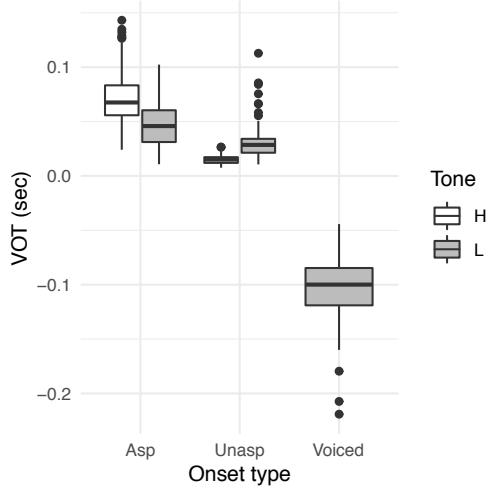
As expected, most items exhibit positive VOT values, and [+SG] stops show longer VOT than [-SG] stops. Prevoicing (negative VOT) is observed in a

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.tʃá/ ‘pheasant’; 4/23 tokens of /tóm/ ‘spider’; 12/28 tokens of /tú/ ‘barley’; and 10/24 tokens of /tük/ ‘dragon’, all of which were non-[SG], as well as two [SG] tokens: 1 token of /pʰà.tʃúk/ ‘cow’ and 1 token of /tʰóm/ ‘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)  $\text{VOT} \sim (1|\text{Place}) + (1|\text{Speaker}) + (1|\text{Word})$
- (2)  $\text{VOT} \sim \text{SG} + (1|\text{Place}) + (1|\text{Speaker}) + (1|\text{Word})$
- (3)  $\text{VOT} \sim \text{Tone} + \text{SG} + (1|\text{Place}) + (1|\text{Speaker}) + (1|\text{Word})$
- (4)  $\text{VOT} \sim \text{SG} * \text{Tone} + (1|\text{Place}) + (1|\text{Speaker}) + (1|\text{Word})$

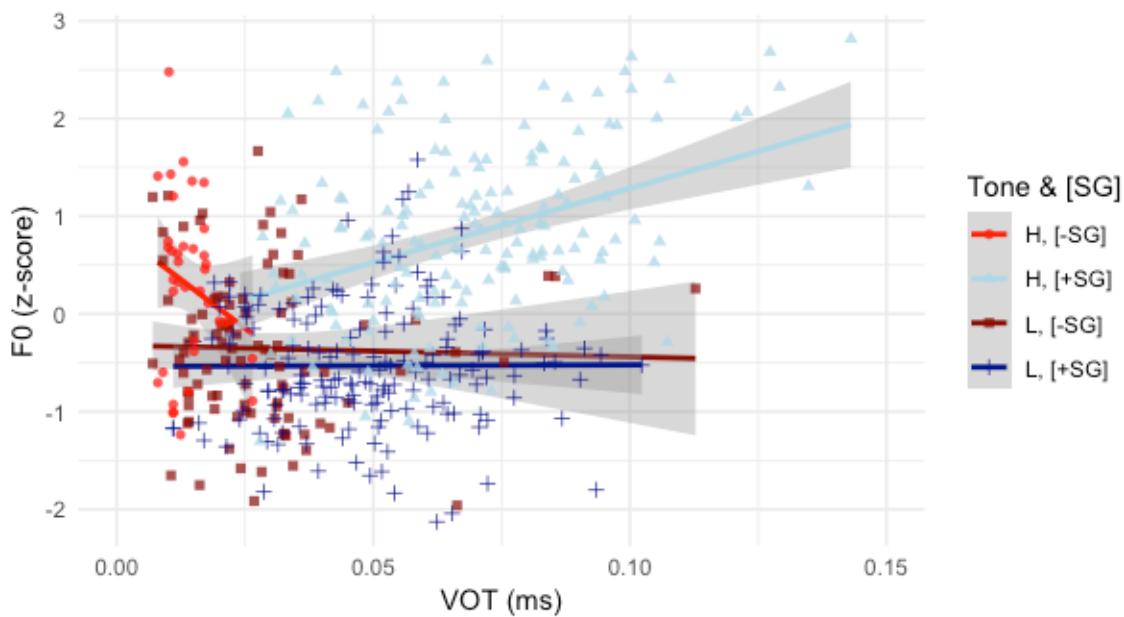
Model	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	p-value
Baseline	5	-2,844.69	-2,823.37	1,427.35	-2,854.69	NA	NA	NA
SG	6	-2,859.83	-2,834.24	1,435.92	-2,871.83	17.14	1	3.472E-05
SG + Tone	7	-2,863.84	-2,833.99	1,438.92	-2,877.84	6.01	1	0.0142
SG*Tone	8	-2,865.87	-2,831.75	1,440.94	-2,881.87	4.03	1	0.0447

*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 unspirates 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 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).

	[pá] and [pà]	
labial	Cclosure	
glottal	Copen	
(a)		
	[bà]	
labial	Cclosure	
glottal	Ccritical	
(b)		
	[p <sup>h</sup> á] (duration)	
labial	Cclosure	
glottal	Cspread :	
(c)		
	[p <sup>h</sup> à] (duration)	
labial	Cclosure	
glottal	Cspread	
(d)		
	[p <sup>h</sup> á] (timing)	
labial	Cclosure	
glottal	Cspread	
(e)		
	[p <sup>h</sup> à] (timing)	
labial	Cclosure	
glottal	Cspread	
(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 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 Tsiring 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

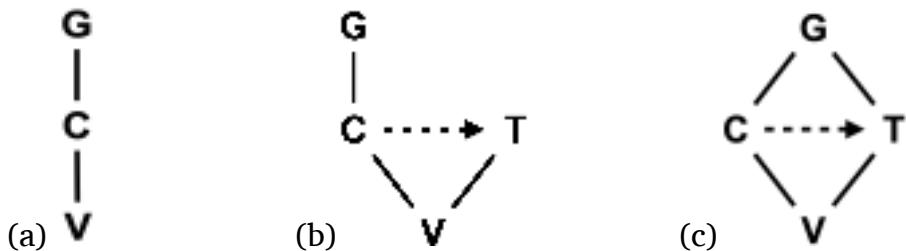
& Abramson 1964, Cho & Ladefoged 1999), but a language with three positive VOT contrasts remains unattested. It is thus more consistent with the typological literature to derive the intermediate-VOT stops of Tibetan from another category of stops, as in the gestural timing account. Again, the gesture 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 coordination. 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<sup>2</sup> 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*

---

<sup>2</sup> 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.

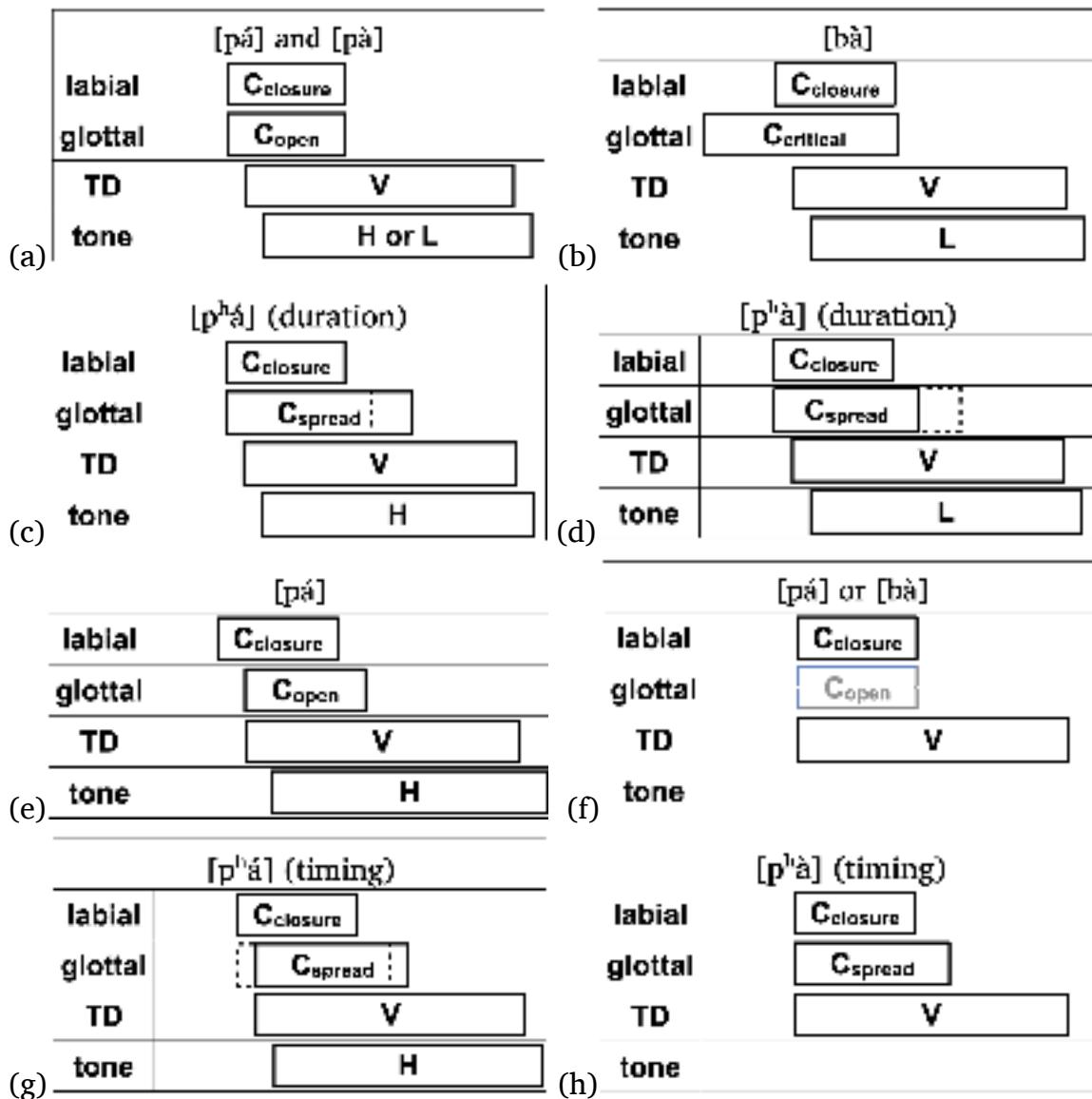


Figure 3.9. Gestural scores for Tibetan onset stops. (a-d) Gestural scores according to gesture duration account: (a) Short-VOT stop with C-G synchrony and C-V lag; (b) Negative-VOT stop with C-V lag; (c) High-tone long-VOT stop with long glottal spreading gesture; (d) Low-tone intermediate-VOT stop 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
<b>Articulatory prediction</b>	<b>C-V timing not different (simultaneous) by tone</b>	<b>C-V lag longer with H tone than with L tone</b>

*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-Ø analysis may still require a high tonal target in the “Ø” 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> (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>.

## **Teaching Experience**

### **Courses taught at Heinrich Heine University Düsseldorf:**

- Laboratory Phonology (“advanced” course), Summer 2023
- Introduction to English Linguistics, Part 2, Summer 2023
- Laboratory Phonology (“method” course), Winter 2022/2023
- Introduction to English Linguistics, Part 1, Winter 2022/2023
- Experimental Phonetics, Summer 2022
- Language in the United States, Summer 2022
- The Phonetics-Phonology Interface, Winter 2021/2022
- Morphology / Word-Formation, Winter 2021/2022

### **Courses taught at Yale University:**

- Topics in Phonology: Phonetics-Phonology Interface, Spring 2021.
- Introduction to Linguistics, Yale Summer Session 2019.

### **Courses taught as Teaching Fellow (teaching assistant):**

- Grammatical Diversity in US English, Spring 2020. Instructor: Raffaella Zanuttini
- Language and Mind, Fall 2020. Instructor: Maria Piñango
- Introduction to Linguistics, Fall 2019. Instructor: Jason Shaw
- General Phonetics, Spring 2018. Instructor: Jonathan Manker
- Introduction to Linguistics, Fall 2017. Instructors: Jim Wood, Matt Barros
- General Phonetics, Spring 2017. Instructor: Jason Shaw
- Introduction to Phonological Analysis (Phonology I), Fall 2016. Instructor: Ryan Bennett

### **Certificate of College Teaching Preparation (awarded Spring 2021)**

- This program, run by the Poorvu Center for Teaching and Learning at Yale University, prepares graduate students and postdoctoral associates for teaching at the college and university level. Completion requires attending a Fundamentals of Teaching course, at least eight teaching workshops, two instances of being observed teaching, two instances of observing another teach, assembling a teaching portfolio, and a narrative reflection. More information is available here:  
<https://poorvucenter.yale.edu/teaching/professional-development/certificate-college-teaching-preparation/>
- Workshops and short courses were completed on the following topics:
  - Fundamentals of Teaching, Gender in the Classroom, Teaching with Technology, Classroom Observations, Mentoring as a Teaching Fellow, Writing Across Disciplines, Preparing and Delivering an Effective Lecture, Course Design, “Let’s talk about it”: Difficult Conversations in Higher Education, Contemporary Controversies in Pedagogy

### **Teaching-related grants:**

- 2022-2023. “Interactive web-based review units for phonetics and phonology”  
Co-applicant with Kevin Tang and Akhilesh Kakolu Ramarao  
Heinrich Heine University E-Learning Förderfonds.
- 2023-2024. “Choose Your Own Adventure: Teleporting to a Virtual Phonetics Laboratory”  
Co-applicant with Eoin O'Reilly and Kevin Tang  
Heinrich Heine University E-Learning Förderfonds.

## Teaching Evaluations

A complete record of teaching evaluations is available [here](#). Below are evaluations from two recent courses:

### Student Evaluations: Laboratory Phonology, HHU Düsseldorf, Winter 2022/2023

- 4 responded

*What is something that helped your learning in this class?*

- Ilias Quiz [note: this refers to self-paced online review units we developed for phonetics and phonology topics from the introductory class] (verbatim from two students)
- learning how to read papers and what the most important parts are
- reading papers, going through them during class, being given an order in which to read the papers' sections
- giving personal advice, that is honest and shows some “behind the scenes” of academia and in general a very open teaching style
- a focus on learning skills (papers reading/write), Praat...
- intake interview
- reading papers together in class and discussing them
- analysing data step by step in class before assignments (Praat, R)

*What is something that did not help your learning in this class?*

- some papers were quite difficult; easier literature would have helped
- for me personally, group work isn't always the best way to go because most of the time it's awkward if you don't know the person
- some papers were too difficult to read

*What is something that could be done differently next time?*

- having a fixed and clear syllabus
- focus more of the class on actually writing the term paper or at least parts of it
- a little bit more structure regarding the topic of the AP paper
- different structure of the course, e.g.: deciding on topics/questions for the short paper at an earlier point in the course; starting the writing process in class/during semester (rather than the lecture-free period) for more guidance and support in the writing process

## **Student Evaluations: Introduction to English Linguistics, Part 1: Structure, HHU**

**Düsseldorf, Winter 2022/2023**

- 54 participated, 15 responded

*What is something that helped your learning in this class?*

- Being generally more involved and having the opportunity to hear questions from other students
- Doing the quizzes helped me in finding out what I should focus on and what I had to study more.
- Your excitement for the topics in class makes them less boring
- Getting [sic] fun with linguistics
- You've answered all my questions I had :)
- That linguistics is more than just learning a language
- You did a very great job by the way!! Keep teaching that way :)
- All of the examples and things we discussed regarding the differences between having English as your first language vs. as a second language → it was nice that you tried to think about how we as mostly German students learn English
- Textbook oriented
- Prof. Geissler's motivation :)
- The handouts we did with the exercises on them (so basically the practical use of the class's content)
- Your way of presenting & your enthusiasm for linguistics
- Reading the textbook before class
- The warm-up exercise (→Syntax II)
- Quizzes have helped me because while I was studying, they served me as a review
- Practical exercises to apply the theory and the quizzes to know the most important information
- The textbook, warm-up exercises, humor
- Textbook, slides, humor, warm-up exercises
- This class helped me to have more knowledge about the English language in phonetics, morphology and syntax. I think it was very helpful.
- Reading the chapters before class and then at class doing some practical activities.

*What is something that did not help your learning in this class?*

- My lack of motivation :(
- The slides are kinda difficult to understand after class when looking over them again
- I would like you to answer your emails
- nothing :)
- In the beginning the class had so many students in it, so I kind of had the feeling I didn't need to learn as much because it was not that necessary to participate → but I like how that changed & with less students I learned so much more
- Class exercises sometimes confuse me because they were very complicated
- Sometimes the exercises were from a higher level than the theory and I got confused.
- There were some activities that were difficult
- Some activities were a little bit confusing, and it was difficult for me to understand them.

*What is something that could be done differently next time?*

- I think everything was great :)
- Maybe more quizzes to check the stand [sic] of my knowledge
- To my mind, I really appreciated that you had taken our (kinda) feedback from the study week and made clear the learning goal. That was very nice :)
- I could've asked more questions
- As you can read above, the class size was a bit too much for me personally
- I really like when you compare our topics with things in Tibetan but I can't really understand it because I don't know Tibetan ;)
- I always like to access presentations before the lecture to prepare! (or during the lecture) [Note: lecture slides were only posted after class, not before]
- Not much because I have learnt how to take advantage of the resources you gave us. I would have liked to have more exercises to practice
- Not so many difficult exercises that may confuse us and make the theory seem to be harder. I would like to have more classes for reviewing the previous contents.
- More group work to practice
- Warm-up exercises for all topics (not only syntax)
- I liked the way this class is done and the teacher really likes the subject. I think the activities and powerpoints are a very good idea to understand the topics.
- I think that it was a very good idea to print some activities and work [on] them in class, so I would use them more.

## **References**

Dear Search Committee:

References may be requested from the individuals listed below. Kevin Tang is my current supervisor, and we collaborate closely in all areas of my current work. Jason Shaw was my Ph.D. advisor; we taught together, and he and I continue to collaborate in research. Claire Bowern was a significant mentor throughout my graduate training, and in particular supervised my field research.

- Kevin Tang (Professor, Heinrich Heine University Düsseldorf): [kevin.tang@hhu.de](mailto:kevin.tang@hhu.de)
- Jason Shaw (Associate Professor, Yale University): [jason.shaw@yale.edu](mailto:jason.shaw@yale.edu)
- Claire Bowern (Professor, Yale University): [claire.bowern@yale.edu](mailto:claire.bowern@yale.edu)

Please feel free to contact me with any questions, or for contact information from additional sources (e.g. students, advisees, collaborators).

Sincerely,

Christopher Geissler