UNIVERSITY OF CALIFORNIA

Los Angeles

Optimal stop inventories: Laryngeal settings

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Linguistics

by

Jahnavi Narkar

2026

ABSTRACT OF THE DISSERTATION

Optimal stop inventories: Laryngeal settings

by

Jahnavi Narkar
Doctor of Philosophy in Linguistics
University of California, Los Angeles, 2026
Professor Patricia Keating, Co-Chair
Professor Megha Sundara, Co-Chair

Although a large body of work has been dedicated to understanding the contents of vowel inventories, consonant inventories have received far less attention. In this dissertation, I investigate the contents of stop inventories, with special reference to their laryngeal features, to determine the factors that affect their composition.

Chapter 3 presents a probabilistic model of stop inventory typology, showing that modeling functional pressures representing maximizing minimum distance, minimizing articulatory complexity and maximizing the number of contrasts accurately predicts the contents of stop inventories. This suggests that the interaction of these phonetically defined constraints can explain much of the structure found in stop inventories, without the need for explicit structural constraints that enforce the maximal use of distinctive features. I also present a case study demonstrating how such a model can be used to inform models testing hypotheses about the ancestral states shared by a set of inventories.

A major obstacle to the study of stop inventories is the lack of understanding of the basic phonetic space of stop contrasts. Even though aspects of stop articulation and acoustics are well understood, the primacy of different acoustic correlates across languages is poorly understood, as is their relationship to stop articulation. To enable the acoustic modeling of stop

inventories, Chapter 5 develops a cross-linguistic acoustic space of stops. I show that thirteen acoustic parameters can be collapsed into a four-dimensional space whose individual dimensions are phonetically interpretable. Broadly speaking, two of these dimensions can be interpreted as indexing stop place contrasts, and the other two as indexing stop laryngeal contrasts.

The stop acoustic space thus constructed is used to examine the contents of stop inventories using two different approaches. Chapter 6 investigates the contents of the same 3020 stop inventories in terms of their acoustic properties. It evaluates two different aspects of inventory attestedness – what determines if a given inventory is attested at all, and if it is, how frequently it is attested among the world's languages. Like the MaxEnt model in Chapter 3, this model finds that stop inventories are shaped by the same phonetically-defined predictors – maximizing dispersion between the closest pair of stops, minimizing articulatory effort and maximizing the size of the inventory. This model also finds that acoustic symmetry is not required as a predictor to capture the frequency distribution of inventories found in the world's languages.

In Chapter 7, the problem of mapping the continuous acoustic space onto a discrete number of categories to form an inventory is conceptualized as a search problem, using a model inspired by the principles of biological evolution. This model shows that the goal of increasing perceptual distance between stop categories can be achieved by maximizing the size and annexing more of the acoustic space after articulatorily simpler options have been exhausted, in addition to the phonetically based factors found to influence the featural and acoustic makeup of stop inventories.

In sum, although there is evidence for many of the factors proposed in previous work, I find that some effects may be derivable from the others. Minimizing articulatory complexity, maximizing minimum dispersion and the number of stops are likely directly implicated, but the effects of feature economy and symmetry can be chalked up to the mechanisms by which inventories strike a balance between the phonetically-based factors.

The dissertation of Jahnavi Narkar is approved.

Kie Ross Zuraw

Bruce P. Hayes

David Michael Goldstein

Megha Sundara, Committee Co-Chair

Patricia Keating, Committee Co-Chair

University of California, Los Angeles

2026

For Aba

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The list of people I would like to extend my thanks to is long, but I will try to keep this short.

I would never have finished this dissertation without the guidance and encouragement of my co-chairs, Pat Keating and Megha Sundara. Pat was extremely generous with her time, and read millions of drafts I bombarded her with. I benefited tremendously from her insight into every aspect of my work, from acoustic minutiae to big picture questions. Pat helped me believe in the value of my work and offered a broad perspective on it in a way only she could have done. Megha's unrelenting focus on clarity and details made my work infinitely more legible. I am convinced she could advise a dissertation in any field with her ability to sniff out the essence of any question. I have gone to Megha for advice since my first year at UCLA and I cannot imagine what my research trajectory would have looked like without her. With their powers combined, Pat and Megha made completing this dissertation a truly enjoyable and, dare I say, smooth experience!

Other members of my committee have their fingerprints all over this dissertation. I was lucky to have Kie Zuraw as a committee member and teacher. I have learned so much from Kie, not just about phonology, but also about teaching and learning. I hope to be half the linguist she is! I could count on Bruce Hayes' wisdom for getting unstuck any time I felt bogged down. Thanks, Bruce, for your friendly skepticism, historical anecdotes and fun analogies! I still think of parts of this dissertation as the Goddess' Grammar! David Goldstein graciously entertained this amateur historical linguist's efforts to weigh in on Proto-Indo-European. I enjoyed every chat I had with him about language and biological evolution!

Thanks to my undergraduate research assistants, Atlas Boulom, Kaitlyn Khoe, Hannah Lin, Johanna Liu, Max Meszaros and Arthur Zhao, for poring over dozens of acoustic studies. The chapter on acoustic typology was made possible with their assistance.

This work has benefited from discussions with a number of other people. Canaan Breiss and Connor Mayer responded to my many questions about MaxEnt. Participants in the UCLA phonetics, phonology and PIES seminars provided invaluable feedback on my work, as did audiences at *Issues in Phonological Typology* in Tromsø, *The Edinburgh Symposium on Historical*

Phonology and the 2026 meeting of the *Linguistic Society of America*. Colleagues and faculty from the Program in Indo-European Studies, especially John Clayton, Muhammad Rehan and Tony Yates, took all my naïve questions about Proto-Indo-European seriously. The good folks over at UCLA OARC, especially Siavash Jalal and Andy Lin, provided valuable insight into statistical methods.

The last six years at UCLA Linguistics have been extremely enriching, thanks to the faculty I had the privilege of learning from. Ben Eischens, Sun-Ah Jun, Jonah Katz and Claire Moore-Cantwell have made me a better phonetician and phonologist. A shout-out also to Dylan Bumford, Anoop Mahajan and Ethan Poole. Despite their best efforts, my life has neither structure nor meaning, but that's on me! Current and former members of the p-lab – Jake Aziz, Coralie Cram, Hironori Katsuda, Jennifer Kuo, John McGahay, Jian-Leat Siah, Yang Wang, Z.L. Zhou, and many others – made my time at UCLA fulfilling and fun! Special thanks to Arthur, Coralie, Corrina, Hannah, Hashmita and Zach for your camaraderie. Outside Linguistics, I was lucky to learn from Oona Paredes, Hongyin Tao and Bu Juliana Wijaya during my time at UCLA. Writing groups held by the Graduate Writing Center helped me write regularly and actually finish this dissertation. The UCLA linguistics staff made navigating administrative matters a breeze. Thank you all!

All my teachers before coming to UCLA influenced me in immeasurable ways. Linguists at Wayne State, especially Haiyong Liu, Ljiljana Progovac, Natasha Rakhlin, Martha Ratliff and Peter Staroverov, helped me lay a strong foundation and taught me how fun the field was. My very first linguistics teacher, Olga Dmitrieva, spurred my love for phonetics and phonology. Stuart Bolton's acoustics classes at Purdue quite literally changed my life! Peter Cole, Gaby Hermon and Yanti introduced me to the thrill of collecting primary data. I was lucky to share that experience with my fieldwork buddies from Indonesia – Tyler Lemon, Tamisha Tan and Grace Wivell. I feel like the prodigal child for not having written my dissertation on my fieldwork language like you all did, but hey, someone's gotta play that role!

Thanks also to my family who supported me in my meandering. Above all, my grandfather, a meanderer himself, whose curiosity, kindness and gregariousness never ceased to amaze me. I miss him dearly and dedicate this dissertation to his memory. My Aji, my wonderful

Aji, who said to me, “I don’t know what it is you do, but whatever it is, I’m proud of you!” My parents, who never fail to feign interest in my work, despite finding it quite dull, really; although, I do think I may have finally gotten through to my dad with the genetic algorithm. My sister, who shares my love of language, but has decided to use it for evil (I kid)! Thanks for letting me talk your ear off! And the entire cast of characters that make up ragtag family. I know I can just turn up in Detroit, Goa, Kolhapur, Louisville or Toronto, as I have done so many times, and be fed and entertained and loved!

The last six years were made pleasant by my little family here in LA. My amazing partner, Derek, without whom I would still be sitting at a computer in a sea of grey cubicles – thank you for your infinite love and encouragement. And for the endless supply of coffee, video games and beautiful stationery! My non-human relatives, Murderface and Toki, who taught me to trust in the divine power of snacks and naps! I can’t wait to spend more time playing with you instead of shooing you off! To all my friends and family who lent me a shoulder to cry on – especially Anisha, Corrina, John, Madhura, Olive and Wreck – your friendship means a lot to me!

This work used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Office of Advanced Research Computings Research Technology Group and was supported by a Dissertation Year Award.

VITA

2024	C.Phil. (Linguistics), University of California, Los Angeles
2020	M.A. (Linguistics), Wayne State University, Detroit
2016	M.S. (Mechanical Engineering), Purdue University, West Lafayette

PUBLICATIONS

Narkar, Jahnavi. "Reconceptualizing VOT: Further contributions to marking 50 years of research on voice onset time." Journal of Phonetics (2025).

Narkar, Jahnavi. "Acoustic Properties of the Four-Way Laryngeal Contrast in Bengali Infant Directed Speech." Proceedings of the 20th International Congress of Phonetic Sciences (2023).

Narkar, Jahnavi and Peter Staroverov. "An acoustic study of voiceless stops in Indian English." Journal of South Asian Linguistics (2022).

Narkar, Jahnavi. "The Curious Absence of Aspiration in Indian English: The Role of Phonetics in Adaptation." Proceedings of the 37th West Coast Conference on Formal Linguistics (2021).

CHAPTER 1

Introduction

Tiger got to hunt, bird got to fly;

Man got to sit and wonder 'Why, why, why?'

– The Books of Bokonon

KURT VONNEGUT, Cat's Cradle

Why do languages have the sounds they do? The idealized set of contrastive sounds of a language are collectively called its *inventory*. A *phoneme inventory* lists the consonants and vowels used in a language in terms of the gross phonetic dimensions along which meaningful linguistic contrasts are organized. The phonetic space, which represents the acoustic and articulatory properties of speech, is continuous and languages carve it up differently to make meaningful distinctions, which are categorical. Take voice onset time (VOT), for example, the primary acoustic correlate of obstruent voicing contrasts in a wide range of languages. VOT of stops is the time interval between the release of the stop burst and the onset of voicing (Lisker and Abramson, 1964). Thus, we can think of the possible phonetic space of stop voicing contrasts as a time scale that goes from $-\infty$ to $+\infty$, as voicing can either precede the burst (as in voiced stops, yielding negative VOT) or follow it (as in voiceless stops, yielding positive VOT). Different languages can carve this continuous phonetic space into categories in different ways – some contrast negative values of VOT with positive values, as in Spanish, or small positive values with large positive vales, as in English (Lisker and Abramson, 1964).

Even within a particular type, languages can vary in terms of where on the VOT continuum they draw a boundary. Figure 1.1 shows the VOT of voiceless stops in a handful of languages from Cho and Ladefoged (1999). All the stops in this figure are voiceless, which means they all have positive values of VOT. Figure 1.1 shows that the same stop, in this case the coronal aspirated stop /tʰ/, has a VOT of 58 ms in Apache on the lower end, but in the rightmost bar, in Navajo it has a VOT of 130 ms. So the /tʰ/ of Navajo is more than twice as long as the /tʰ/ of Apache, and yet they belong to the same category. By contrast, the Aleut coronal stop, which is virtually identical to the Apache one, is described as being unaspirated, that is /t/ (Cho and Ladefoged, 1999). Therefore, two stops that are very similar to each other can be categorically different, as in the case of Aleut and Apache. And conversely, two stops that are very different from each other phonetically can still belong to the same category, as in the case of Apache and Navajo.

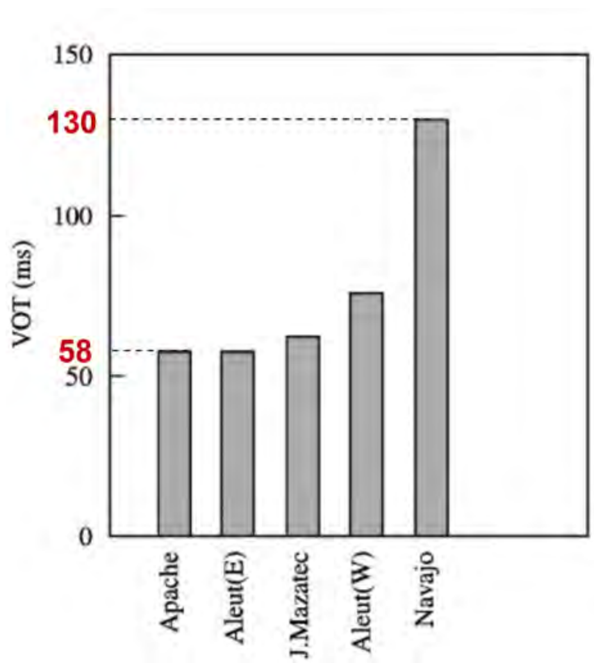


Figure 1.1: *Variation in VOT of voiceless stops across languages from Cho and Ladefoged (1999).*

Despite this wide range of possible variation, no language uses the entire phonetic space to create contrasts. In the case of stop voicing, for example, no language makes more that three distinctions in terms of VOT (Lisker and Abramson, 1964). In fact, there is no language that uses the entire IPA chart either, which is a gross categorical representation of the continuous phonetic spaces of various speech sounds.

Languages also vary greatly in the number of sounds in their inventories. For example, Rotokas (Papuan, Papua New Guinea), famously, has one of the smallest inventories in the world with only eleven segments (Maddieson, 1984). By contrast, !Xu (Khoisan, southern Africa) has the most segments of any documented language – 141 consonants and vowels (Maddieson, 1984).

Systematic trends have been noted across inventories despite the variation in the composition and size of inventories. For example, small inventories of comparable sizes are usually similar in composition, and differences start to emerge as inventory size increases (Rice and Avery, 1993). Another near-universal is that if a language has three vowels, they are most likely /i/, /u/ and /a/ (Liljencrants and Lindblom, 1972). For consonants, it has been shown that across a wide range of languages, around 70% of all consonants in an inventory are obstruents and the other approximately 30% are sonorants (Lindblom and Maddieson, 1988).

That such similarities are found between distant and unrelated languages suggests that there are general factors affecting the contents of inventories. Some observations regarding inventory trends have served as evidence for new theoretical claims and frameworks. For example, the observation that front vowels are usually unrounded while non-low back vowels are usually rounded has been used as evidence for *Dispersion Theory*, which claims that rounding reinforces the front-back contrast to maximize the perceptual contrast between these categories (Liljencrants and Lindblom, 1972, *et seq.*). Some influential proposals that aim to explain inventory content are described below.

1.1 Background

1.1.1 Vowel inventories

Numerous studies have attempted to describe and explain the structure and size of vowel inventories. A large body of work has been conducted on them, partly because the acoustic space of vowels is well-understood, as is its relationship to articulatory properties. The acoustic space of vowels is largely reducible to two dimensions, F_1 and F_2', which can be determined by the first three formant frequencies of each individual vowel. Moreover, these neatly map on the articulatory properties of tongue height, backness and roundness. The parameterization of this space, which relates acoustic and articulatory properties of vowels, has enabled work investigating the properties that govern vowel inventories.

Dispersion Theory (Liljencrants and Lindblom, 1972) claims that languages tend to have sounds that are maximally dispersed in the available phonetic space. Based on this principle of dispersion, whereby the sounds in an inventory are maximally dispersed in the phonetic space, they propose a computational model of vowel inventories, which simulates the dispersive behavior of individual vowels as electrically charged particles repelling each another in the continuous $F_1 - F_2'$ space. The results of their simulation match the contents of attested vowel inventories fairly well, especially for small-sized vowel inventories. For example, their model predicts that the five-vowel inventory that achieves the greatest total dispersion is [i ɛ a/æ ɑ u], which is similar to the most frequent inventory of size 5, [i e a o u].

A rather different approach to explaining the cross-linguistic frequency of some segments is Stevens's (1972, 1989) *Quantal Theory*. This theory claims that languages exploit regions of stability where variations in articulations do not have significant acoustic consequences, and avoid regions where small articulatory differences have relatively large acoustic consequences. For example, as noted before, the F2 of a vowel is an acoustic correlate of its backness. However, the relationship between F2 and tongue position is non-linear, such that some differences in tongue backness do not result in proportionally large differences in F2. There are some regions in the F2-backness relationship that are more stable than others. Stevens suggests that the regions of stability are exploited by languages to create contrasts and are thus associated with sounds that are frequently found in the world's languages.

Schwartz et al. (1997) marry ideas from Dispersion Theory and Quantal Theory to propose the principle of *Dispersion-focalization* according to which vowel inventories prefer perceptually dispersed vowels and vowels that have an enhanced percept, specifically those that are made salient when two of their formants are close in frequency, an effect they call "focalization." The focalization aspect of their model lends more acoustic stability to quantal vowels, making them easier to perceive. Their model performs better than Liljencrants and Lindblom (1972) at matching the contents of attested vowel inventories.

Both dispersion theory and dispersion-focalization have found empirical support. In a large study of 555 vowel inventories, Becker-Kristal (2010) finds a correlation between the number of vowels in an inventory and the acoustic space it occupies, thus supporting the principles of dispersion theory. He also finds evidence supporting focalization, which boosts vowels whose F2 and F3 are close together.

1.1.2 Consonant inventories

Compared to vowels, a relatively modest body of work has focused on the contents of consonant inventories. In response to dispersion theory, Ohala (1980) points out that if consonant inventories similarly maximized perceptual distance, we would expect the optimal consonant inventory to be something like the set in (1.1). However, such an inventory is wildly pathological. Instead, Ohala (1980) claims that consonant and vowel inventories are structured by different organizing principles, with consonant inventories being organized to maximally use distinctive features, and not to be maximally dispersed.

(1.1) [ɗ, k', ts, ɬ, m, r, ɹ]

In support of dispersion theory as an explanation for the contents of consonant inventories, Lindblom and Maddieson (1988) contend that inventories of both consonants and vowels are not "maximally" but "adequately" dispersed, following Lindblom (1986). They claim that "consonant inventories tend to evolve so as to achieve maximal perceptual distinctiveness at minimum articulatory cost" (Lindblom and Maddieson, 1988: 72). According to them, the inventory in (1.1) is unattested because the articulatory cost involved in producing it is too great. The true vowel equivalent of (1.1) is shown in (1.2) (Lindblom and Maddieson, 1988: 74), with a modal /i/, nasal /e/, breathy /a/, creaky /o/ and pharyngealized /u/, which is just as pathological. Therefore, Lindblom and Maddieson (1988) argued that the optimal interplay between articulatory complexity and perceptual distinctiveness is the organizing principle shaping both consonant and vowel inventories.

(1.2) /i ẽ a̤ o̰ uˤ/

This tension between dispersion and articulatory cost has been modeled in Optimality Theory (OT) (Prince and Smolensky, 1994) by Flemming (2013). In addition to maximizing dispersion and minimizing articulatory cost, Flemming (2013) suggests that maximizing the number of segments in an inventory also plays a role in determining the contents of optimal inventories. This proposal is discussed in detail in Chapter 3. Dispersion has also been modeled in agent-based simulations (De Boer, 2000) and in Exemplar Theory (Wedel, 2006).

Lindblom and Maddieson (1988) sketch out a mechanism by which articulatory complexity and dispersion may be balanced in inventories. They propose that as inventory size increases, a basic phonetic space must first be expanded and saturated. Beyond this, the space itself must become more complex by incorporating additional dimensions of contrast. The details of this proposal are discussed in Chapter 7.

Although dispersion effects are well established for vowel inventories, there is also evidence for its role in determining the contents of consonant inventories. Boersma and Hamann (2008) modeled the emergence of dispersion in sibilant inventories in an OT-based framework[1]. Kokkelmans (2021) did the same with larger sibilant inventories in a similar bidirectional model based on OT and neural nets. Finally, Schwartz et al. (2012) showed that the labial-coronal-velar series, which is nearly universal in stop inventories, is perceptually optimal in terms of dispersion, although an ad-hoc restriction on pharyngeals is required to exclude them from the possible set of place contrasts.

The lack of research on consonant inventory structure is even more acute when we consider whether and how laryngeal settings are treated. For instance, no voicing contrasts were included in Schwartz et al.'s (2012) study of stops and Boersma and Hamann's (2008) study of sibilants, and only a two-way voiced-voiceless distinction was included in the larger study of sibilants Kokkelmans (2021).

An influential phonological proposal, known as *Feature Economy*, has attempted to ex-

[1]Boersma and Hamann (2008) andKokkelmans (2021) argue that the effects of dispersion can be emergent, and that specific constraints that refer to it, like those proposed by Flemming (2013), are not required. This point is irrelevant to the present discussion as this dissertation presents typological models and not models of the grammar. The point here is simply that attested sibilant inventories exhibit a tendency to be perceptually dispersed.

plain inventory structure in a way that includes laryngeal settings. This proposal claims that inventories maximize the ratio of segments over features (Clements, 2003). Since distinctive features are usually based on articulatory (and acoustic) properties, maximizing feature economy involves reusing the same gestures to yield more segments. The idea that languages may reuse gestures at the phonetic level has also been proposed (Maddieson, 1995). Feature economy thus refines Lindblom's (1986) principle of "articulatory cost" by combining articulations while maximizing perceptual distinctiveness.

Feature economy is a subset of the older and broader notion of *symmetry* according to which inventories maximize symmetry (Hockett, 1955), but, presumably, not necessarily economy. The difference between economy and symmetry is demonstrated by the inventories in (1.3) and (1.4) (Clements, 2003:92).

(1.3) p t k
b d g
f s x

(1.4) p t k
b d g
f s x
v z ɣ

For every phoneme in (1.3) that uses place, manner and voicing features, a symmetrical phoneme is present, since /p/ is to /t/ is to /k/ as /b/ is to /d/ is to /g/ as /f/ is to /s/ is to /x/. However, the features employed in this inventory are not combined to produce the maximal number of contrasts, since the fricatives are not combined with voicing even though they are combined with place. Therefore, (1.3) is symmetrical but not economical. (1.4), on the other hand, combines fricatives and voicing to produce the maximal number of segments possible from the features already present in the inventory. Thus, (1.4) is both symmetrical and economical.

Even though the notions of symmetry and economy have a long history (de Groot, 1931; Martinet, 1939; Jakobson et al., 1952), their precise definitions are not agreed upon (Clements, 2003; Hall, 2007; Mackie and Mielke, 2011; Dunbar and Dupoux, 2016). For example, Dunbar and Dupoux (2016) point out that the inventory in (1.3) can be considered to be symmetric or asymmetric under different assumptions. If voicing and manner are considered to be independent dimensions of contrast, (1.3) is not symmetrical, since there are only half as many voiced as voiceless consonants, and only half as many fricatives as stops. By contrast, if voiceless stops, voiced stops and voiceless fricatives are assumed to be ordered along a continuum, the inventory in 1.3) is both symmetrical and economical. When defined in a way that is able to tease these two structural properties apart, evidence has been found for both symmetry and economy independently influencing inventory content (Dunbar and Dupoux, 2016).

The acoustic equivalent of feature economy is acoustic symmetry. Empirical, cross-linguistic work on the effect of acoustic symmetry on inventory contents is sparse, compared to the work on featural symmetry and economy. However, its effects have been found at the level of individual speakers (Keating, 2003; Chodroff and Wilson, 2017, 2022), although in most of this literature the term *uniformity* is used instead. These studies have found that the acoustic realization of a particular contrast is *uniform* within speakers such that segments sharing a distinctive feature exhibit similar acoustic properties. For example, speakers whose realizations of $[\mathrm{p^h}]$ have relatively long VOT also produce $[\mathrm{k^h}]$ with proportionally longer VOT (Chodroff and Wilson, 2017). In other words, $[\mathrm{p^h}]$ and $[\mathrm{k^h}]$ exhibit symmetry in acoustic space.

Statistical studies of inventory databases have found support for the effect of structural constraints on inventory composition, in the form of feature economy (Mackie and Mielke, 2011; Dunbar and Dupoux, 2016; Nikolaev and Grossman, 2020; Wang, 2020) and feature symmetry (Dunbar and Dupoux, 2016) in both consonant and vowel inventories. Economy has been found to influence the contents of vowel inventories as well, with larger vowel inventories reflecting feature economy, but smaller vowel inventories reflecting maximal dispersion (Mukherjee et al., 2007). A large-scale acoustic investigation of vowels also found evidence for adaptive dispersion, with a minor effect of acoustic symmetry (Becker-Kristal, 2010).

Note that other influential proposals, such as Quantal Theory (Stevens, 1972, 1989) and Evolutionary Phonology (Blevins, 2004), which explain inventory content and are supported by similar studies[2], are not considered here as they do not make explicit system-level predictions regarding the makeup of optimal inventories.

Finally, *Evolutionary Phonology* claims that inventories are shaped by the factors involved in language change (Blevins, 2004; Brown, 2006). However, since change itself can have a phonetic basis, Evolutionary Phonology does not necessarily involve an explanation that is unrelated to phonetic factors. It is, therefore, not incompatible with the theories mentioned so far, even though it makes different predictions about the nature and contents of the synchronic grammar of speakers.

1.1.3 Factors affecting inventory content

To summarize, there is evidence for at least four factors affecting inventory content. Dispersion encourages the segments in an inventory to be perceptually distinct. Articulatory cost acts as an opposing force, encouraging the use of simpler segments. Striking a balance between these conflicting pressures, known as *Adaptive Dispersion* (Lindblom, 1986), along with maximizing the number of contrasting segments, has also been successful in the modeling inventory content (Flemming, 2013). Finally, there is also evidence for the role of a structural factor in the form of feature economy, the tendency of inventories to use the features they employ to create the maximum number of contrasts (Clements, 2003). I take the acoustic equivalent of feature economy to be acoustic symmetry, as detailed above.

1.2 This dissertation

This dissertation aims to explain the contents of stop inventories with particular reference to their laryngeal settings. It evaluates how the various factors that have been shown to affect inventory content shape stop inventories with respect to their discrete categorical and continuous acoustic properties. I present three models investigating the factors that influ-

[2]Becker-Kristal (2010) finds evidence for the influence of so-called focalization (Schwartz et al., 1997), which combines aspects of Dispersion Theory and Quantal Theory, in the acoustics of vowel inventories.

ence the contents of stop inventories. The first model, presented in Chapter 3, evaluates stop categories in featural terms and presents a formal model, couched in Dispersion Theory (Liljencrants and Lindblom, 1972; Flemming, 2013) and Maximum Entropy Harmonic Grammar (Goldwater and Johnson, 2003; Hayes and Wilson, 2008). This model probes how dispersion in the featural space, articulatory complexity, inventory size and featural economy influence stop inventory contents. The second model evaluates stop categories in terms of their acoustic, rather than featural, properties. This model, detailed in Chapter 6, evaluates how acoustic dispersion, articulatory complexity, inventory size and symmetry in the acoustic space influence inventory contents. Finally, Chapter 7 presents a model that searches the acoustic space to devise stop inventories while balancing acoustic dispersion, symmetry, articulatory complexity and inventory size.

1.2.1 Why stops

As noted before, most work on inventory content has focused on the organization of vowel inventories, and consonant inventories have received far less attention by comparison. Stops, in particular, have more complicated articulatory and acoustic properties compared to vowels and even other consonants, such as fricatives. Yet languages tend to employ stops prolifically. There are no languages without stops. The smallest consonant inventory, Rotokas (Papua New Guinea), has five consonants, /p, t, k, g, β/, four of which are oral stops. Moreover, one pair, /k-g/, contrasts a laryngeal feature. Languages also tend to have maximal laryngeal contrasts in stops rather than in non-stops (Rice and Avery, 1993). Stops also feature in the modal inventory ([p t k b d g m n ŋ s l]), a set of basic consonants that occur in nearly all inventories (Maddieson, 1984).

Maddieson (1984) and Ladefoged and Maddieson (1996) describe laryngeal settings in stop inventories comprehensively, but there has been no work on explaining laryngeal settings in stop inventories. By contrast, some work has attempted to explain the near-universality of the labial-coronal-velar places in stop inventories (Schwartz et al., 2012). Most previous work on inventories abstracts away from laryngeal distinctions completely (Schwartz et al., 2012;

Bybee and Easterday, 2022). In work that has attempted to explain laryngeal distinctions in stops (e.g., Kingston, 1993), voicing is considered the only dimension along which laryngeal distinctions differ. However, stop inventories are known to be more complex with four, five and even six-way laryngeal contrasts (Maddieson, 1984; Ladefoged and Maddieson, 1996). Characterizing the laryngeal contrast in stops is, therefore, crucial and laryngeal contrasts that exist elsewhere can be understood as a subclass of stop laryngeal contrasts.

1.2.2 Methodology

Throughout this dissertation, I assume that possible laryngeal contrasts are produced by combining two basic types of laryngeal gestures – glottal vibration and glottal aperture, as proposed by Narkar (2025). This is represented schematically in Figure 1.2. In this figure, "T" represents plain voiceless stops, "Th" represents voiceless aspirated stops, "Tc" voiceless constricted stops, "D" plain voiced stops, "Dh" voiced aspirated stops and "Dc" voiced constricted stops. This notation is used throughout the dissertation.

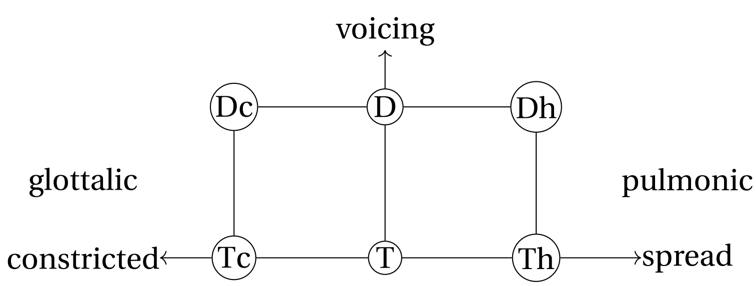


Figure 1.2: *The complete phonetic space of laryngeal contrasts*

The space in Figure 1.2 is defined in articulatory terms – the x-axis represents the state of the glottis, with glottal constriction on the negative x-axis and glottal spread on the positive x-axis. The y-axis represents glottal vibration, with categories with a positive value on the y-axis representing stops produced with glottal vibration, and categories with a zero value on the y-axis representing stops that do not involve any glottal vibration. The stops on the negative x-axis – Tc and Dc – are glottalic stops, produced with glottal constriction, and the stops on the zero and positive x-axis – T, D, Th and Dh – are pulmonic stops. All the stops that appear in the analyses that follow are classified into one of these six laryngeal categories.

This organization abstracts away the phonetic detail and considers stops that may differ in phonetic implementation as belonging to the same gross category, extending Keating's (1984a) proposal for two-way VOT contrasts to the six-way laryngeal contrast shown in Figure 1.2. For example, Indic languages like Marathi, Bamileke languages like Yemba and Armenian languages like Yerevan Armenian all have voiced aspirates (Seyfarth and Garellek, 2018; Dmitrieva and Dutta, 2020; Faytak and Steffman, 2024), although their phonetic implementation differs in important ways. Marathi voiced aspirated stops stops have prevoicing followed a period of breathy release, Yemba voiced aspirates have prevoicing followed by a period of voiceless aspiration, and these same stops in Yerevan Armenian have prevoicing followed by breathiness on the following vowel without any prevocalic aspiration or breathiness. Under the model in 1.2, these differences are assumed to be differences in language-specific phonetic implementation, largely captured by differences in the magnitude and phasing of glottal gestures, and not differences in phonological specification or categories.

Along similar lines, both the constricted categories are quite heterogeneous, with the Tc category consisting of ejectives, pre- and post-glottalized, (voiceless) creaky and devoiced implosive stops, and the Dc category consisting of implosives, pre- and post-glottalized, creaky voiced and voiced ejective stops. We could add a third dimension of laryngeal movement to the space in Figure 1.2 to separate the ejectives and implosives from the other constricted stops. However, there are two issues to consider. The first is that there is a lot of variation in how ejectives and implosives are implemented within and across languages, such that laryngeal movement is not necessarily present in their production (Lindau, 1984; Kingston, 1985; Nihalani, 1986; Clements and Osu, 2002; Ladefoged and Loeb, 2002; Brandt and Simpson, 2021). Second, there is no language that contrasts any of the stop types within a category (e.g., implosives and creaky voiced stops). Since we are modeling *phonemic* inventories, incorporating laryngeal movement introduces added complexity to the space and somewhat muddies it.

Thus, the task here is to determine what explains the categories in Figure 1.2 that co-occur as a set in the most common stop inventories found in the world's languages. This is done by first addressing what separates inventories attested in the world's languages from those that are possible but never attested. Then, among those that are attested, what separates the most commonly attested inventories from unattested ones?

Attested inventories are taken to be those included in PHOIBLE (Moran and McCloy, 2019), a comprehensive, openly accessible cross-linguistic repository of phonological inventories based on data from individual language descriptions and from other databases. Data from PHOIBLE have been used in recent studies of inventories (Nikolaev and Grossman, 2020; Wang, 2020). Unlike many databases, PHOIBLE has multiple entries for the same language, corresponding to distinct sources that may disagree about the number and/or identity of the phonemes in that language. I included all the inventories from PHOIBLE so as to yield the largest possible dataset. Differences between different inventories of the same language are retained as these are assumed to be either reflections of genuine dialectal variation, or of disagreement in transcription and/or analysis. The former is variation that we want to include in our dataset, but the latter is not. However, in the absence of an objective metric to determine which inventory is closest to the real inventory, and assuming that the likelihood of transcriber error is small in this large dataset, I did not exclude any inventories with the assumption that the noise from transcriber disagreement would not have a large effect on the overall pattern.

Table 1.1 shows how the bilabial phonemes in PHOIBLE were classified into the six laryngeal categories shown in Figure 1.2. Only bilabial phonemes are shown as an illustration. Phonemes at the other places of articulation were classified into laryngeal categories in the same manner.

Category	**Phoneme**
T	p, b̥
D	b
Th	pʰ, ʰp, ʰpʰ, b̤̥, b̥ʰ
Tc	p', pˀ, ˀp, p̰ , ɓ̥
Dh	b̤, bʰ, bʱ ʰb̃
Dc	ɓ, ˀb, bˀ, b̰, 'b

Table 1.1: *Classification of bilabial phonemes in PHOIBLE.*

Note that I use non-IPA symbols when describing the contents of inventories to represent this abstraction. For example, /ph th kh b d g/ is used instead of /pʰ tʰ kʰ b d g/. The use of IPA symbols is reserved for describing the inventories of specific languages. For example, /pʰ tʰ kʰ b d g/ is used to refer to the stop inventory of Swedish. Thus, when non-IPA symbols are used the intended meaning is the gross category described by the classification scheme in 1.1, whereas the use of IPA symbols signifies the specific phoneme within a category. Thus, /ph/ represents a voiceless-aspirated stop, which could be any phoneme of the set [pʰ, ʰp, ʰpʰ, b̤̥, b̥ʰ], while /pʰ/ specifically refers to a post-aspirated voiceless bilabial stop. I also use slanted brackets rather than square brackets as we are dealing with *phonemic* rather than phonetic inventories.

Figure 1.3 shows the distribution of phonemic stop inventory sizes for all the inventories in PHOIBLE, including stops at all possible places of articulation and one of the six laryngeal categories in Figure 1.2. Chapter 2 investigates the contents of these stop inventories in detail and compares them with Maddieson's findings on UPSID.

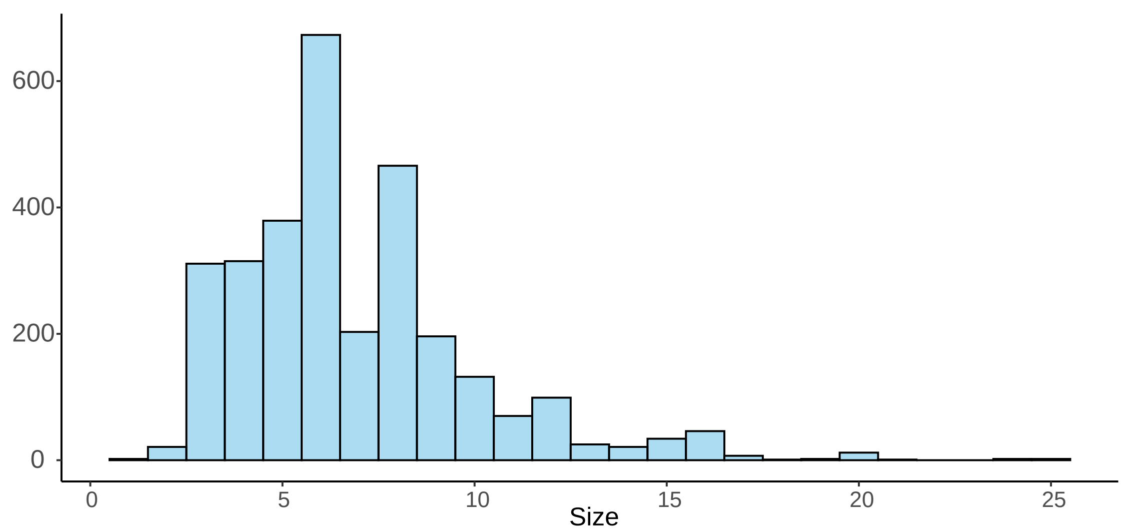


Figure 1.3: *The distribution of stop inventory sizes in PHOIBLE.*

As seen in Figure 1.3, stop inventory sizes range between 1 to 25 stops, with the most common size being 6. The spike at size 3 on the x-axis represents the high frequency of the /p t k/ inventory. This figure also shows peaks for even-numbered inventory sizes, relative to the adjacent sizes, at 6, 8, 12 and 16 on the x-axis, suggesting a preference for symmetrical inventories.

Investigating the contents of these inventories, we find that laryngeal distinctions start to emerge at size 4, although the most common strategy at this size is to add another place of articulation to the basic /p t k/ series. The vast majority of these languages (158 out of 206) add a second coronal stop, like in many Australian languages. However, some also add a laryngeal contrast in one of the places. For example, 28 inventories of this size consist of /p, b, t, k/. There are 50 other inventories that exhibit this pattern of having a laryngeal contrast at one place in addition to the /p t k/ series.

Inventories of size 5 are approximately evenly split between inventories with two laryngeal contrasts at two out of three places of articulation (e.g., /p t k b d/) and two places of articulation in addition to the /ptk/ series (e.g., /p t ʈ c k/). In inventories of size 4 and 5 that employ voicing distinctions, the voiced velar stop, /g/ is disfavored. In inventories of size 4, /g/ appears in 28 languages, compared to 52 occurrences of /b/ and 35 occurrences of /d/. In inventories of size 5, /g/ appears in 102 languages compared to 163 and 172 for /b/ and /d/ respectively. This dispreference for /g/ is expected given its aerodynamic disadvantage, since it is difficult to maintain voicing when the space between the glottis and the velar closure is small (Keating, 1984b; Westbury and Keating, 1986).

The next spike at 6 in Figure 1.3 represents the addition of voicing contrasts to the basic inventory, yielding /p t k b d g/. This is the most common makeup of inventories of any size – three places of articulation that contrast two laryngeal settings. The vast majority of inventories of size 6 employ voicing as opposed to aspiration or constriction, suggesting the voicing dimension is far more preferred than the aspiration-constriction dimension.

Inventories of size 7, 8 and 9 also exploit laryngeal contrasts. Size 7 inventories are mostly comprised of two laryngeal contrasts at three places with an additional place (e.g., /p t k b d g q/ or /p t ʈ k b d g/). A large proportion of size 8 inventories have a two-way laryngeal contrast in four places (e.g., /p t c k b d g ɟ/), with a considerable number of languages with a three-way laryngeal contrast at two places and a two-way contrast at one place, with the gap usually at /g/ (e.g., /p t k p^h t^h k^h b d/) or /p/ (e.g., /t k p^h t^h k^h b d g/). Finally, the most common pattern in inventories of size 9 is a fully symmetric 3-way place and 3-way laryngeal contrast (/p t k p^h t^h k^h b d g/).

Larger stop inventories that seem to be more common compared to others of comparable sizes have three laryngeal contrasts at four places (e.g., /p t ʈ k pʰ tʰ ʈʰ kʰ b d ɖ g/) or four laryngeal contrasts at three places (e.g., /p t k pʰ tʰ kʰ p’ t’ k’ b d g/); and four laryngeal contrasts at four places (e.g., /p t ʈ k pʰ tʰ ʈʰ kʰ b d ɖ g bʰ dʰ ɖʰ gʰ/). The largest stop inventory in PHOIBLE is Sindhi with the 25 stops /p t ʈ c k q pʰ tʰ ʈʰ cʰ kʰ b d ɖ ɟg bʰ dʰ ɖʰ ɟʰ gʰ ɓ ɗ ʄ ɠ/. The largest laryngeal inventory is Santali, which uses all six laryngeal categories at four places, with some gaps (/p t ʈ k pʰ tʰ ʈʰ kʰ p̰ t̰ k̰ b d ɖ g b̤ d̤ ɖ̤ g̤ b̰/).

Thus, the labial-alveolar-velar series is nearly universal among stop inventories, occurring in small and large inventories alike. As inventories get larger, beyond size 3, they combine this basic place contrast with laryngeal contrasts and/or expand to other places, namely, retroflex, palatal and uvular. Since this dissertation is concerned with developing an understanding of how laryngeal contrasts work in stops, I limit the place contrasts to the labial-alveolar-velar series. In all the models presented in this dissertation, the possible stops are limited to three places and to the six laryngeal categories described above.

The combination of three place and six laryngeal configurations yields a total of 18 possible segments. Assuming that each segment is either present or absent in an inventory, this yields a total of 2^{18} possible inventories. Figure 1.4 shows the distribution of inventories of different sizes.

The blue curve represents the actual distribution of inventories (the same data as shown in Figure 1.3, but limited to the 18 stops under consideration) and the orange curve represents the possible distribution of inventory sizes under the assumption that inventories can be created by randomly sampling from the set of 18 stops. This yields 18 possible unique inventories of size 1, all the way to a single possible inventory of size 18. The inventory sizes between 1 and 18 are normally distributed with size 9 showing the greatest number of possibilities. Notice that the possible inventories (shown on the right y-axis) far outnumber the attested inventories (shown on the left y-axis). If there were no regularities affecting the content of stop inventories, the shape of the blue curve in Figure 1.4 would resemble that of the orange curve. However, we see clear spikes in the blue curve, indicating that some sizes, and indeed some segments, are preferred over others. The task is to determine what separates the

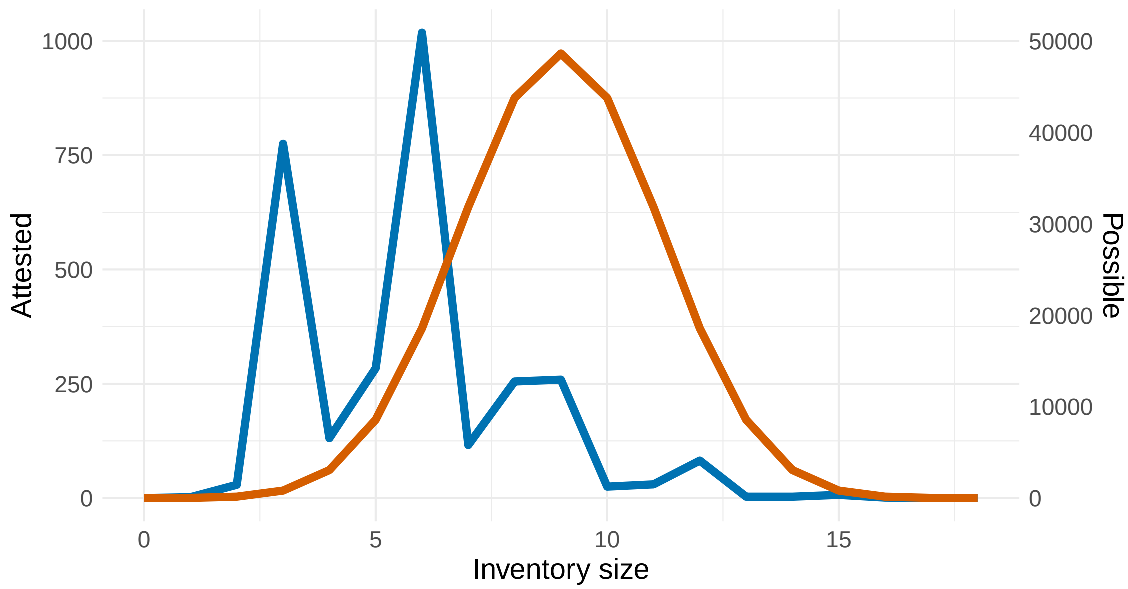


Figure 1.4: *Attested versus possible stop inventories.*

blue curve from the orange one.

Note that the stops considered here were limited to place and laryngeal categories. Other attributes of stops, such as double and secondary articulations and geminates, were ignored, since we are primarily concerned with stop laryngeal contrasts.

1.2.3 Caveats

Before moving on, it is necessary to acknowledge that any phoneme inventory is an idealized abstraction. The use of inventories in linguistic typological research has been criticized for conflating phonetic and phonological details, since what is claimed to represent a phoneme in these phonemic inventories really is an allophone (Simpson, 1999; Kiparsky, 2018), usually, the most frequent one (Maddieson and Precoda, 1989). Despite the abstraction involved in the formulation of inventories, it has long been recognized that they represent real cognitive objects (Sapir, 1933; Jaeger, 1980). This is evidenced by the perception of native phonemes, especially consonants and stops even more so, as it differs from the perception of non-native phonemes, broadly speaking. Therefore, even though phoneme inventories are idealized abstractions, they represent important generalizations about the phonological systems they represent. Moreover, to make meaningful, explanatory generalizations regarding languages, some level of abstraction is essential.

A second problem concerns using data specifically from PHOIBLE, which is not genetically and areally balanced like some other databases are. Therefore, it is possible that the patterns present in the data from PHOIBLE are representative of a certain genetic or areal feature, and not more generally representative of the organization of inventories. However, there is value in probing patterns without correcting for areal and/or genetic affiliation. If a certain pattern is a prominent feature of related languages or geographical areas, this pattern must be preferred in some way as it shows diachronic fitness, persisting over time, and spreading to unrelated languages (Greenberg, 1978; Blevins, 2018). Note that such patterns do not preclude other explanations. Take the example of voiced aspirates in South Asia, for instance. These have persisted in the inventories of Indic languages for centuries and have been borrowed by other unrelated languages, possibly due to their economic organization whereby the features [voice] and [spread] are employed maximally. Therefore, in the analyses that follow, I do not separate out areal and genetic factors, as done in more recent work on inventories using PHOIBLE data (Nikolaev and Grossman, 2020; Wang, 2020). However, it must be noted that this approach treats some individual languages as independent data points when they might not, in fact, be independent.

Finally, it is important to note that the models presented in this dissertation are not meant to be models of the speaker or learner. It is not claimed to be a grammar, like some work using dispersion theory (e.g., Padgett, 2003). Instead, like dispersion theory (Liljencrants and Lindblom, 1972; Flemming, 2013), it is a deductive typological model of languages.

1.3 Overview of chapters

The dissertation is structured as follows. Chapter 2 presents the results of a large-scale study of 3020 stop inventories recorded in PHOIBLE (Moran and McCloy, 2019) as an update to Maddieson's (1984) seminal work on patterns of sounds. It shows that his conclusions regarding stop inventories in the areally and genetically balanced sample of 317 languages in

UPSID are largely replicated in the large, imbalanced sample from PHOIBLE. This suggests that the lack of genetic and areal factors may not influence the contents of stop inventories to a great extent. Chapter 3 presents a probabilistic model of stop inventory typology, couched in Dispersion Theory, showing that phonetically-defined constraints suffice to explain stop inventory content, and that specific constraints that refer to structural constraints like symmetry and economy are not required. Chapter 4 sketches an outline of how the phonological model presented in Chapter 3 can be used to test hypotheses regarding ancestral states, by testing the probabilities of different proposed Proto-Indo-European stop inventories producing the stop inventory typology attested in the modern Indo-European languages. Chapter 5 constructs a cross-linguistic acoustic space for stop contrasts. This space provides a low-dimensional parameterization of the complex acoustic space associated with stop place and laryngeal contrasts. Chapter 6 uses this space to investigate which factors explain the contents of stop inventories. Comparing attested and unattested stop inventories to vowel inventories, it shows that, broadly speaking, languages prefer stop inventories that are articulatorily simpler, dispersed and larger in size. Mirroring results from the MaxEnt model, this model finds that symmetry in the acoustic space is not required to sufficiently explain the higher frequencies of the most common stop inventories. Chapter 7 presents a model which simulates the mapping of the continuous acoustic space onto discrete categories, subject to various objectives using Genetic Algorithms. This produces a menu of stop inventories, each of which is optimal when a different goal is prioritized. The success of this method in producing a menu that resembles the most commonly attested inventories is limited, but it may hold promise for future work on inventories.

To summarize, this dissertation evaluates the contents of stop inventories with reference to their laryngeal settings. It evaluates how different phonetic and structural properties influence the categorical and continuous characteristics of stop inventories.

CHAPTER 2

Patterns of stops

There is no blue without yellow and without orange, and if you put in the blue, then you must put in the yellow and orange too, mustn't you?

– Vincent Van Gogh

This chapter presents an update to Maddieson's (1984) findings on stops in his seminal work *Patterns of Sounds.* I find that his conclusions regarding stop inventories in the areally and genetically balanced sample of 317 languages in UPSID are largely replicated in the large, imbalanced sample from PHOIBLE. This suggests that stop inventories across geographies and families look remarkably similar and may be explained by the same underlying organizational principles regardless of areal or genetic affiliation.

I present results from statistical tests of co-occurrence patterns of different place and laryngeal configurations of stop consonants. Determining these co-occurrence patterns has two purposes. The first is to evaluate implicational statements, which are commonplace in linguistics but are seldom substantiated by rigorous statistical tests (cf. Maddieson, 1984). For example, it is often claimed that the presence of a voiced stop in a language implies the presence of a voiceless stop. Voicing in stops is widely claimed to be a marked feature, so much so that the claim that this laryngeal articulation can only occur when it contrasts with the unmarked one is taken to be conventional wisdom. I evaluate the implicational relationship between voiced aspirates and plain, voiceless, voiceless, aspirated and plain, voiced stops, as predicted by the principle of Feature Economy (Clements, 2003) or maximal use of features (Ohala, 1980), discussed in Chapter 1. Extending this analogy to the glottalic stops, I also evaluate the statement that the presence of voiced constricted stops implies the presence of plain voiceless, voiceless constricted and plain voiced stops.

Parallel to the laryngeal patterns, I also evaluate implicational and co-occurrence patterns among places of articulation. For example, it is well known that languages that have voiced stops are more likely to have a gap at /g/ than at /b/ or /d/, since maintaining voicing with a velar closure is more difficult as the space between the glottis and the stop closure is small (Keating, 1984b; Westbury and Keating, 1986). Therefore, we can say that the presence of /g/ implies the presence of /b/ and /d/. Along similar lines, the presence of /p'/ may imply the presence of /t'/ and /k'/. In voiceless aspirated stop series, by contrast, there are no obvious gaps, meaning the presence of any stop out of /p^h/, /t^h/, /k^h/ may imply the presence of the others. In other words, while /b/ and /d/ may occur with or without /g/, /p^h/, /t^h/ and /k^h/ tend to occur together or not at all (Maddieson, 1984; Kingston, 1993).

The second motivation for evaluating co-occurrence and implicational relationships is to determine how place and laryngeal features pattern in stop inventories in order to formulate reasonable constraints to be used in the models in Chapters 6, 3 and 7 to model the typology of stop inventories.

2.1 Data

To assess if implicational relationships of the type discussed above are substantiated by statistical analyses, I tested these relationships based on segment data from PHOIBLE. For every segment, PHOIBLE also provides an associated matrix of phonological features. The specification of place as coded by the values of these place features was used to assign a value of place to each phoneme. For example, if a segment was coded as having a positive value of the feature [labial] in PHOIBLE, it was recorded as a labial, and so on.

The case of coronals, however, was found to be somewhat complicated. First, many descriptions that PHOIBLE uses were inconsistent in their treatment of dental stops. These were very often transcribed with plain /t/, and not with a dental diacritic. Moreover, some coronal stops are coded as "dental or alveolar" in PHOIBLE. To avoid excluding any of these segments, I combined them into a single group called "denti-alveolar." This group contained segments described as “dental,” “alveolar” or “dental or alveolar” in PHOIBLE. For languages that had both dental and alveolar stops, e.g., Arrernte, I coded only one category, namely, denti-alveolar, collapsing such distinctions. Moreover, I only coded for a two-way coronal distinction – denti-alveolar and retroflex – as this was the most common place contrast in languages that had more than one coronal place of articulation. This led to collapsing coronal contrasts in the 52 inventories that have more than two coronal stops, all of which are Australian languages. As a result, there is a fair degree of abstraction in place and the results presented in this chapter do not account for finer place distinctions. The coding for place done in this manner resulted in a total of six possible place features – labial, denti-alveolar, retroflex, palatal, velar and uvular.

Recall that there is also some abstraction in the coding of laryngeal settings. I assume that the organization in Figure 1.2 also accounts for more substantial differences, traditionally treated as separate categories. My assumption is that pre-aspirated stops are language- or context-specific instantiations of aspirated stops, just like post-aspirated stops. This is also assumed for pre- and post-glottalized stops (Keating et al., 2021). See Table 1.1 for the classification scheme used for laryngeal categories.

The resulting six place and six laryngeal features were combined to form 36 segments. Every language was specified as either having or not having each of these 36 segments with 1s and 0s (representing presence and absence respectively).

2.2 The distribution of place and laryngeal features in PHOIBLE

Figure 2.1 shows the distribution of laryngeal contrasts at every place of articulation in all the inventories in PHOIBLE (Moran and McCloy, 2019). It shows that the labial, denti-alveolar and velar places are far more common than retroflex, palatal and uvular places of articulation. This figure also shows that voicing is dispreferred for places of articulation farther back in the mouth, and that this dispreference is more acute in the voiced glottalic stops than in voiced pulmonic stops. Conversely, posterior places of articulation are more preferred by the voiceless glottalic stops than by stops produced with any other laryngeal configuration.

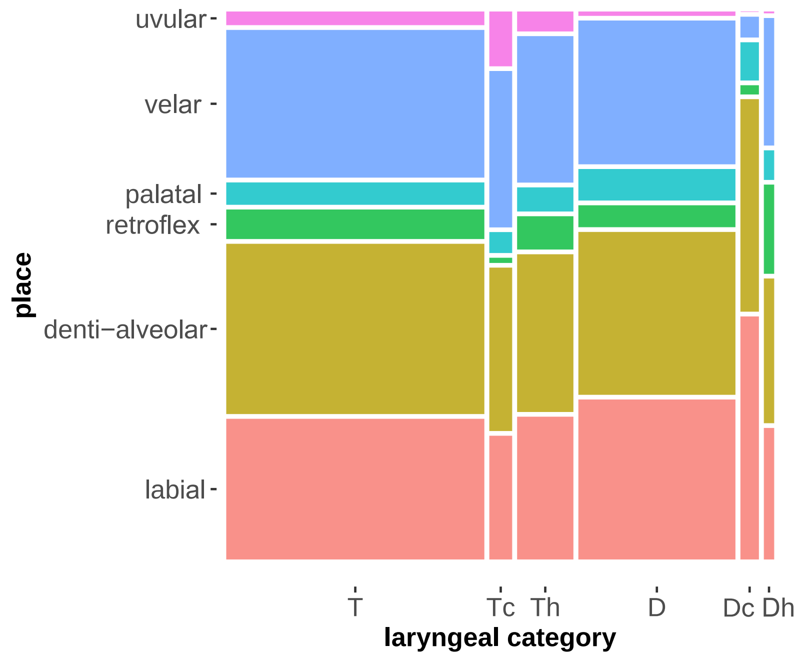


Figure 2.1: *Laryngeal contrasts at each place of articulation*

Table 2.1 lists the ratios of the frequencies of stops at the three major places of articulation – labial, denti-alveolar and velar. This table shows that the dispreference for voicing at more posterior places is more exaggerated in the glottalic stops than in the pulmonic stops – the ratio D/T at the velar place is slightly lower compared to the labial and denti-alveolar places, but the ratio Dc/Tc is much lower at the velar than at the two other places. In fact, only a very small number of languages have voiced velar constricted stops like the implosive /ɠ/. This trend is reversed in the voiceless constricted stops. Such stops become less common the farther front they are articulated, as indicated by the lower value of Tc/T at the labial place of articulation. Moreover, these ratios show that aspirated stops, both voiced and voiceless, do not seem to be influenced by place of articulation – their ratios, indicated by Th/T and Dh/D, remain fairly constant across all three major places.

The dispreference for voicing at posterior places is likely due to its aerodynamic disadvantage, since it is difficult to maintain voicing when the space between the glottis and the closure is small (Keating, 1984b). The gap at bilabial place in the plain voiceless (p) and voiceless constricted stops (pʼ etc.) may also have a phonetic explanation, for the exact opposite

	labial	**alveolar**	**velar**
D/T	0.735	0.667	0.627
Th/T	0.22	0.2	0.215
Tc/T	0.076	0.083	0.092
Dc/Tc	1.582	1.047	0.105
Dc/D	0.175	0.15	0.016
Dh/Th	0.163	0.162	0.154
Dh/ D	0.052	0.056	0.056

Table 2.1: *Ratios of number of languages with different laryngeal settings at the three major places of articulation.*

reason. The closure at the bilabial place requires more intra-oral pressure than at other places since the lips are both active articulators and more volume must be filled with air. On the perceptual side, it is also possible that the more diffuse spectrum and lower intensity of bilabial stops makes them more difficult to be perceived faithfully. The exaggerated gap at the glottalic velar stops is likely due to the fact that the articulatory maneuvers required to distinguish such a stop from the plain, voiced stop, /g/, are more limited with this posterior closure, as compared to other places. Therefore, there is considerable reason to believe that the contents of stop inventories can be explained phonetically.

Table 2.2 compares the results from PHOIBLE with Maddieson's (1984) results from UPSID_{317}, an earlier version of the database that had 317 languages as opposed to the latest version which has 451 languages. Maddieson (1984) lists the generalizations shown in Table 2.2 regarding the makeup of stop inventories. There is a high degree of similarity between these two sets of results, suggesting that the results from PHOIBLE are likely reliable and are not skewed in favor of or against a language family or geographical area, despite the lack of areal and genealogical balance in this database. Note that Maddieson (1984) uses the term "series" to refer to laryngeal contrasts. Therefore, the statement, "A language is most likely to have two series of stops" can be interpreted as "A language is most likely to have two series of stops that contrast a laryngeal state," using the terminology used in this dissertation.

Despite broad similarities between the patterns shown in Table 2.2, there are a couple of differences between the UPSID and PHOIBLE samples. First, there are proportionally more languages in UPSID (51.1%) with two stop series than in PHOIBLE (45.8%). This difference

Generalization	Maddieson (1984)	This chapter
All languages have stops.	100%	100%
A language is most likely to have two series of stops.	51.5%	45.8%
A language is highly likely to have a series of plain voiceless stops.	91.8%	95.0%
If a language has only one stop series, that series is plain voiceless.	98.0%	98.5%
If a language has two stop series, it has a VOT contrast between them.	88.9%	90.2%
If a language has three stop series, it is most likely to have two series with contrasting VOT and one glottalic series.	65.8%	57.4%
A language is most likely to have stops at 3 places of articulation.	53.9%	57.1%
A language most typically includes stops at bilabial, dental or alveolar, and velar places of articulation.	98.4%	98.0%
Stops at other than these three places do not occur unless the language contrasts stops at 4 or more places.	98.1%	98.5%

Table 2.2: *Comparison with findings from Maddieson (1984).*

is driven by a greater percentage of languages with a single series which do not contrast any laryngeal state in PHOIBLE (25.9%) than in UPSID (15.8%). The second difference concerns the makeup of languages with three stop series – 65.8% of such languages in UPSID have a VOT contrast (i.e., a contrast between prevoiced and voiceless stops, D vs T, or unaspirated and aspirated stops, T vs Th) and one glottalic series (Tc or Dc) – compared to 57.4% in PHOIBLE. This difference is driven by the greater proportion of languages in PHOIBLE that have a three-way VOT contrast and glottalic series, D-T-Th, compared to UPSID – 33.7% vs 25.0%. All other generalizations are very similar in the two databases, as seen in Table 2.2.

In addition to these, Maddieson (1984) also reports the generalizations in 2.1-2.6. Note that Maddieson uses /t/ and /d/ to represent the combined categories of dentals and alveolars. Comparable patterns are found in PHOIBLE.

(2.1) If a language has /p/ then it has /k/, and if it has /k/ then it has /t/ (4 counterexamples in UPSID).

In PHOIBLE, out of a total 2866 languages that have at least one voiceless unaspirated stop, only 47 have /p/ but no /k/, and 17 have /k/ but no /t/, as opposed to 186 languages that have /t/ but no /p/, and 187 that have /k/ but no /p/.

(2.2) If a language has /g/ then it has /d/, and if it has /d/ then it has / b/ (3 counterexamples in UPSID).

Out of the 2002 languages that have at least one voiced unaspirated stop in PHOIBLE, 73 have /g/ but no /d/, and 69 have /g/ but no /b/. By contrast, 233 languages have /d/ and 262 have /b/ to the exclusion of /g/. Additionally, 57 languages have /d/ but no /b/, and 90 have /b/ but no /d/.

(2.3) If a language has /p'/ it also has /t'/.

(2.4) If a language has /t'/ it also has /k'/.

(2.5) If a language has only one ejective stop, it is /k'/.

Of the 299 PHOIBLE languages that have at least one ejective stop, only 13 have /p'/ but no /t'/ as opposed to 68 that have /t'/ but no /p'/ and 197 that have both. This supports Maddieson (1984)'s generalization in (2.3). Additionally, there are 28 languages that have /t'/ but no /k'/ and 28 others that have /k'/ but no /t'/. By contrast, there are 237 languages that have both /t'/ and /k'/. Therefore, /t'/ and /k'/ are likely to co-occur and the presence of one predicts the presence of the other. It is the case that the presence of /t'/ implies the presence of /k'/, but the reverse is also true – the presence of /k'/ also implies the presence of /t'/. However, if we consider the makeup of the 36 languages that have exactly one ejective stop, 21 languages have /k'/, while only 6 and 9 languages respectively have /p'/ and /t'/ only. This directly supports Maddieson's generalization in (2.5). It also suggests that ejectives prefer the velar place of articulation, and therefore the implicational relationship in (2.4) is supported.

(2.6) A language with any implosives or laryngealized voiced stops has /ˀb/ and /ˀd/.

Of the 363 languages with at least one implosive in PHOIBLE, 264 have both /ˀb/ and /ˀd/, and only 30 have /ˀg/. Moreover, there are no languages with only /ˀg/, but there are 67 with only /ˀb/ and 26 with only /ˀd/. This supports the generalization in (2.6).

These generalizations are further evaluated with more statistical rigor in Sections 2.3 and 2.4.

2.3 Co-occurrence of places of articulation

Since the three major places of articulation – labial, denti-alveolar and velar – are far more common as shown in Figure 2.1, only these places were included in the statistical analyses. The results largely confirmed results from previous studies. Inventories are likely to have pulmonic stops that share the same laryngeal articulation at all the major places, with inventories with gaps at textipa/p/ and /g/ just as likely as inventories without any gaps (see Sections 2.3.2.1 and 2.3.2.2). This is also the case for voiceless constricted stops, which are just likely to occur at all three major places as they are to have a gap at textipa/p'/ (Section 2.3.2.4). Both voiceless and voiced aspirated stops tend to occur at all major places (see Sections 2.3.2.3 and 2.3.2.5). Finally, the voiced constricted stops are the only series that are less likely to occur at all three places. While inventories are likely to have both bilabial and denti-alveolar voiced constricted stops, these stops are likely to occur without the velar stop (Section 2.3.2.6).

2.3.1 Method

The method for assessing the co-occurrence of place features was largely inspired by Kingston (1993), who investigates place co-occurrences in the stop inventories of UPSID 451 (Maddieson and Precoda, 1989). However, unlike Kingston's frequentist models, Bayesian log-linear models were built and implemented with the `brms` package (Bürkner, 2017) in R (R Core Team, 2021). The Bayesian implementation was chosen for several reasons. First, Bayesian models allow the incorporation of prior knowledge through the use of prior distributions. Thus, we can specify what we know about co-occurrence patterns in stop inventories and evaluate the data against these prior beliefs. Bayesian models are more flexible and are better equipped for regularization through informative priors, which helps prevent over-fitting the data. Finally, results from Bayesian models are more intuitive and easier to interpret as they offer direct probability statements about parameters and predictions. Bayesian models estimate a range of probable values for parameters, known as the 95% Credible Interval (95% CI). An effect is considered robust if this interval does not include zero. Conversely, if the interval does include zero, it suggests there is reason to believe that the parameter under consideration has no effect on the dependent variable.

Separate Bayesian log-linear models were built to assess the co-occurrence of place features within a laryngeal category. The dependent variable was the number of languages and the independent variables were the (binary) presence or absence of a phoneme. The fully specified model evaluated the main effects of each category and all two-way interactions. Interactions were subsequently dropped and these models were compared to the fully specified model. Better model fit with the inclusion of interaction terms suggests co-occurrence of the categories included in the interaction. If dropping an interaction term, say $x_1 * x_2$, does not degrade the model fit, we can conclude that the presence of x_1 is not associated with the presence of x_2. On the other hand, if dropping this interaction does worsen the model, we can conclude that this interaction is crucial and, therefore, x_1 and x_2 must co-occur. If a model containing no interactions is better than the models with interactions, we can conclude that the presence of any category is independent of the presence or absence of another category.

The different models thus built were compared using LOO (Leave One Out) cross-validation in the R package `loo` (Vehtari et al., 2024) which performs model comparison by estimating the expected log predictive density (`elpd`) for each model. This method helps evaluate how well a model generalizes to new data by assessing its performance when one observation is left out at a time and the model is refitted without that observation (hence the name "*Leave One Out*"). The model with the higher `elpd` is preferred, provided the added complexity is justified by better predictive performance. LOO cross-validation was chosen over Bayes factor, another method of comparing Bayesian models, since LOO is less sensitive to choice of priors. That is, if the amount of data is small compared to the number of predictive parameters, Bayes factor is more sensitive to the priors, risking spitting out what was put into the model. Performing model comparison with LOO, however, allows the data to overwhelm the prior in such cases. Additionally, if the models are not very close to the true model, Bayes factor can be more unstable compared to LOO cross-validation.

The first set of models evaluated whether bilabial, denti-alveolar and velar stops co-occur with the same laryngeal setting in stop inventories. The dependent variable, `count`, represented the number of languages that have a certain inventory. The independent variables were all binary and took the value 0 if that phoneme was absent and 1 if it was present in the inventory. As an illustration, there are 801 inventories that have /p, t, k, b, d, g/. These inventories are represented as the variables `p, t, k, b, d, g` having the value 1, all other independent variables having the value 0 and the dependent variable, `count`, having a value of 801. The models used for voiceless unaspirated (T) stops are shown below as an illustration. Note that terms like `pt` represent interaction terms. A full list of models and their specifications is included in A.1 in the Appendix.

(2.7) Model 1 (all two-way interactions): `count ~ p + t + k + pt + tk + pk`

(2.8) Model 2 (no labial x velar): `count ~ p + t + k + pt + tk`

(2.9) Model 3 (no labial x denti-alveolar): `count ~ p + t + k + pk + tk`

(2.10) Model 4 (no denti-alveolar x velar): `count ~ p + t + k + pt + pk`

(2.11) Model 5 (main effects only): `count ~ p + t + k`

Negative binomial and Poisson distributions were both used in separate models and compared, as both these distributions are appropriate for modeling count data (Winter and Bürkner, 2021). The data were found to be credibly over-dispersed, prompting the choice of the negative binomial distribution over the Poisson distribution. The expectations shown in Table 2.3 were specified in the priors based on previous studies (Gordon, 2016; Maddieson, 1984). Note that the segments in this table are a shorthand for different segment types, as discussed in

Laryngeal category	**Expectation**
Plain voiceless (T)	gap at p
Plain voiced (D)	gap at g
Voiceless aspirated (Th)	no gaps
Voiceless constricted (Tc)	gap at p’
Voiced aspirated (Dh)	gap at g^h
Voiced constricted (Dc)	gap at g’

Table 2.3: *Expectations that informed the choice of priors.*

Chapter 1. For example, "p" in Table 2.3 stands for both /p/ and /b̥/.

2.3.2 Results

2.3.2.1 Plain voiceless stops

Figure 2.2 shows the composition of inventories with plain voiceless (T) stops. The x-axis shows various co-occurrence patterns and the y-axis shows the number of languages that have the corresponding pattern. As seen in this figure, inventories with all three stops – /p/, /t/ and /k/ – far outnumber inventories with gaps at one of the places. However, a gap at /p/ is more common than a gap at /k/ or /t/.

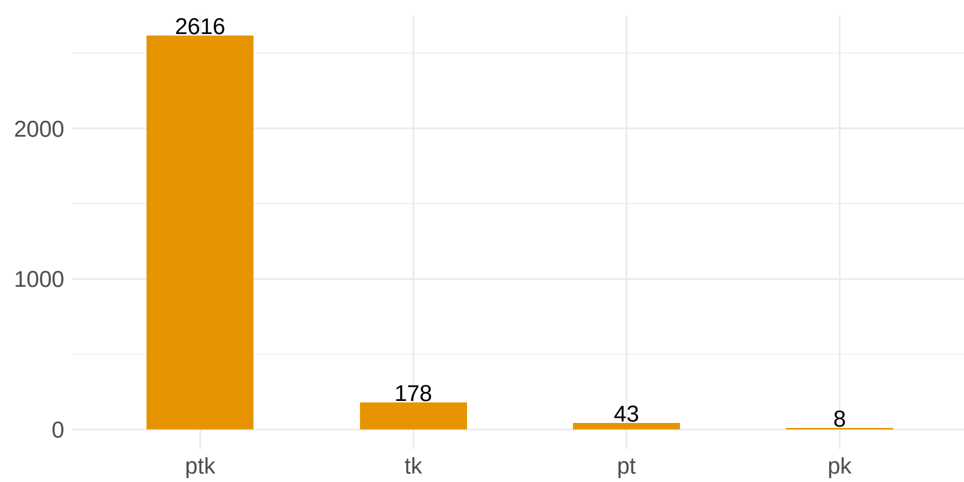


Figure 2.2: *The composition of inventories with plain voiceless stops.*

The results from model comparisons of the T stops are shown in Table 2.4. There is a three-way tie between the model with all two-way interactions, the model without an interaction between /p/ and /k/, and the model without an interaction between /p/ and /t/. The performance of these three models is not credibly different, although Model T1 outperforms Model T2, which, in turn, outperforms Model T3. However, dropping the interaction between /t/ and /k/ leads to a credible drop in model performance. This means that the interaction between the denti-alveolar and velar stops is crucial, and it must be present in the model to achieve the best fit. Dropping interactions with /p/, on the other hand, as in Models T2 and

T3, does not credibly worsen the fit. These results suggest that inventories are more likely to have /p t k/ or /t k/, than to have /p t/ or /p k/. This provides evidence towards the likely gap at /p/ in stop inventories. Results from the best performing models are included in the Appendix.

	Model	Result	Inventory
T1	all 2-way interactions	top performing	pt, tk, pk co-occur
T2	no labial x velar		pt, tk co-occur
T3	no labial x denti-alveolar		pk, tk co-occur
T4	no denti-alveolar x velar	credibly worse than T1, T2,	pt, pk co-occur
T5	main effects only	T3	no co-occurrence of ptk

Table 2.4: *Model comparisons for voiceless, unaspirated (T) stops*

2.3.2.2 Plain voiced stops

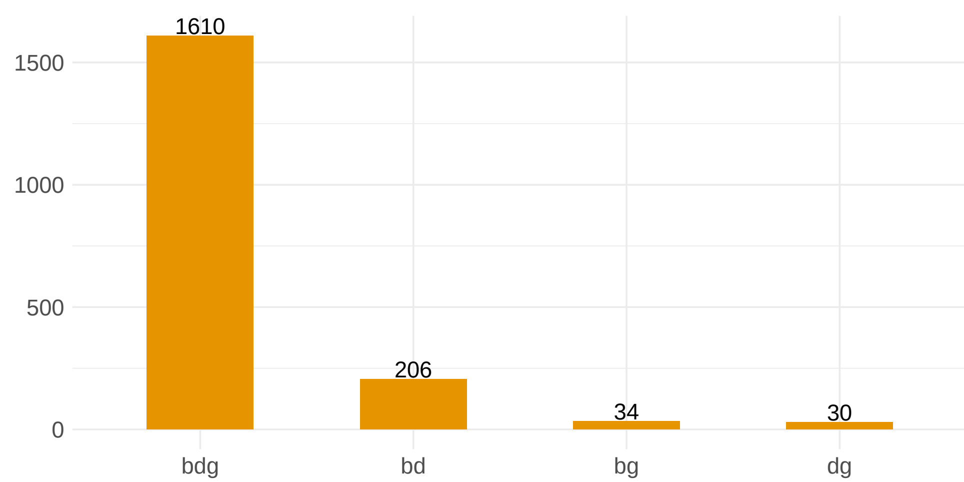


Figure 2.3: *The composition of inventories with plain voiced stops.*

Figure 2.3 shows inventories with plain voiced (D) stops. This figure shows that inventories with /b/, /d/ and /g/ are the most common. Additionally, a gap at /g/ is more common than a gap at /b/ or /d/, in line with expectations.

The results from model comparisons, shown in Table 2.5, are consistent with this pattern. The model with all two-way interactions is comparable to the models that includes the interaction between /b/ and /d/. These three models are better than the model which does not include the interaction between /b/ and /d/ but does include their interactions with /g/, suggesting that the co-occurrence of voiced bilabial and denti-alveolar stops is likely in more inventories.

	Model	Result	Inventory
D1	all 2-way interactions	top performing	bd, dg, bg co-occur
D2	no labial x velar		bd, dg co-occur
D3	no denti-alveolar x velar		bg, dg co-occur
D4	no labial x denti-alveolar	credibly worse than D1, D2, D3	dg, bg co-occur
D5	main effects only	worst-performing	no co-occurrence of bdg

Table 2.5: *Model comparisons for voiced, unaspirated (D) stops*

2.3.2.3 Voiceless-aspirated stops

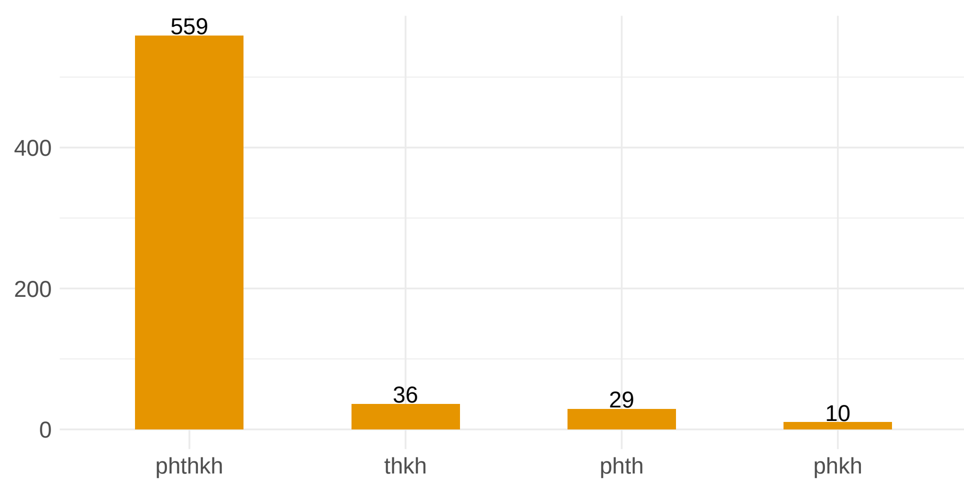


Figure 2.4: *The composition of inventories with voiceless-aspirated stops.*

Figure 2.4 shows the composition of inventories with voiceless-aspirated (Th) stops. This figure shows that inventories with stops at all three places are the most common and even though inventories with a gap at /p^h/ slightly outnumber those with gaps at /t^h/ or /k^h/, the difference is much smaller than in the case of plain voiceless and voiced stops (see Figures 2.2 and 2.3).

The results from model comparisons, shown in Table 2.6, show the model with all two-way interactions outperforms all other models. That is, dropping any two-way interaction,

	Model	Result	Inventory
Th1	all 2-way interactions	top performing	p^ht^h, t^hk^h, p^hk^h co-occur
Th2	no labial x denti-alveolar	credibly worse than Th1	p^hk^h, t^hk^h co-occur
Th3	no denti-alveolar x velar		p^ht^h, p^hk^h co-occur
Th4	no labial x velar	credibly worse than Th1, Th2, Th3	p^ht^h, t^hk^h co-occur
Th5	main effects only	worst-performing	no co-occurrence of $p^ht^hk^h$

Table 2.6: *Model comparisons for voiceless, aspirated (Th) stops*

including those with the bilabial /p^h/, credibly worsens model fit. This suggests that voiceless aspirated stops are likely to co-occur at all three basic places of articulation in more inventories, and inventories with gaps are less likely.

2.3.2.4 Voiceless-constricted stops

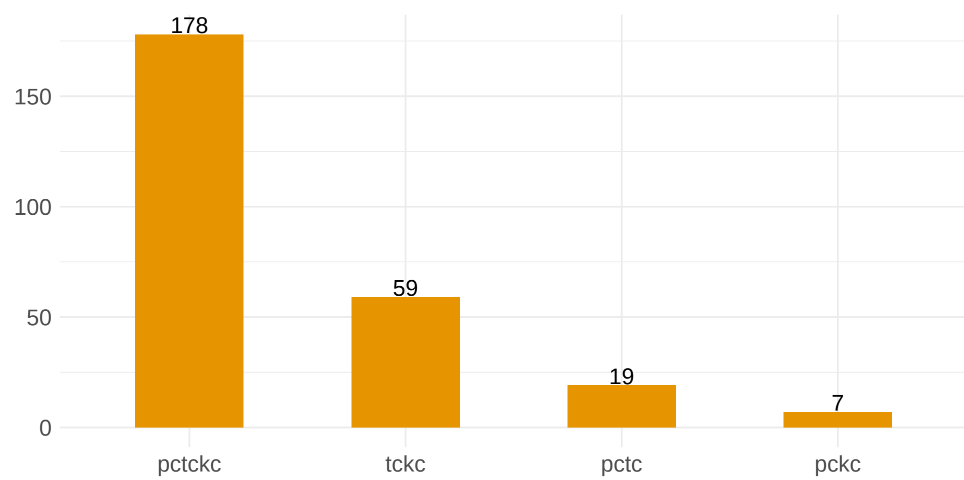


Figure 2.5: *The composition of inventories with voiceless-constricted stops.*

Figure 2.5 shows the composition of inventories with voiceless-constricted (Tc) stops. This figure shows that inventories with stops at all three places are the most common, followed by those with a gap at /p'. By contrasts, inventories with gaps at /t'/ and /k'/ are rarer.

The results from model comparisons, shown in Table 2.7, confirmed this pattern. As in the case of plain voiceless stops, the model with all three interactions was just as good as the model that excluded interactions with /p'/. By contrast, dropping the interaction between

	Model	Result	Inventory
Tc1	all 2-way interactions	top performing	p't', t'k', p'k' co-occur
Tc2	no labial x denti-alveolar	2^{nd} best	p'k', t'k' co-occur
Tc3	no labial x velar	3^{rd} best	p't', t'k' co-occur
Tc4	no denti-alveolar x velar	4^{th} best	p't', p'k' co-occur
Tc5	main effects only	worst-performing	no co-occurrence of p't'k'

Table 2.7: *Model comparisons for voiceless, constricted (Tc) stops.*

/t'/ and /k'/ credibly worsened the model.

2.3.2.5 Voiced-aspirated stops

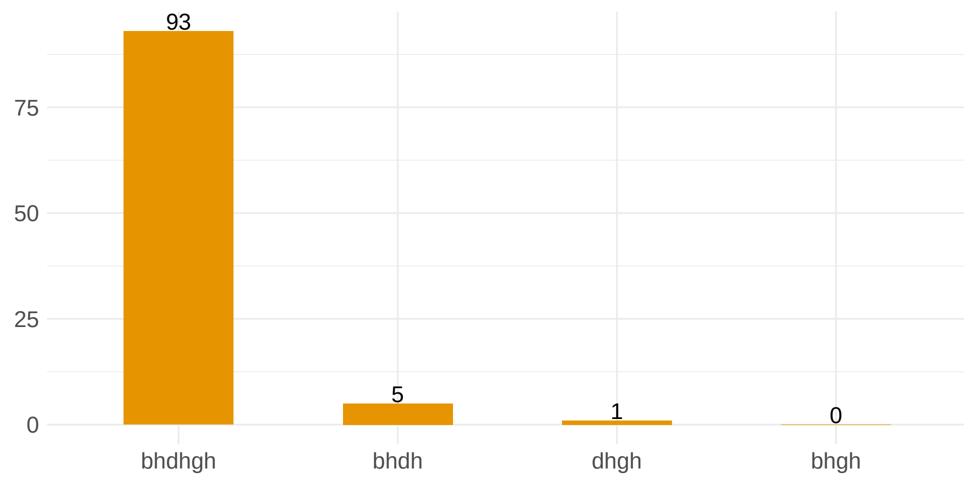


Figure 2.6: *The composition of inventories with voiced-aspirated stops.*

Figure 2.6 shows the composition of inventories with voiced-aspirated (Tc) stops. This figure shows that inventories with stops at all three places are far more common, and inventories with gaps are vanishingly rare.

The results from model comparisons, in Table 2.8, confirm this pattern as the model with all two-way interactions outperforms all other models. That is, dropping any two-way interaction, including those with the velar voiced aspirate, /g^{h}/, credibly worsens model fit. This suggests that voiced aspirated stops are likely to co-occur at all three basic places of articulation in more inventories, and inventories with gaps are less likely, contradicting the prior and the pattern suggested by the conditional probabilities in Figure 2.6.

	Model	Result	Inventory
Dh1	all 2-way interactions	top performing	b^hd^h, d^hg^h, b^hg^h co-occur
Dh2	no labial x denti-alveolar	credibly worse than Dh1	b^hg^h, d^hg^h co-occur
Dh3	no denti-alveolar x velar		b^hd^h, b^hg^h co-occur
Dh4	no labial x velar		b^hd^h, d^hg^h co-occur
Dh5	main effects only	worst-performing	no co-occurrence of $b^hd^hg^h$

Table 2.8: *Model comparisons for voiced-aspirated (Dh) stops.*

2.3.2.6 Voiced-constricted stops

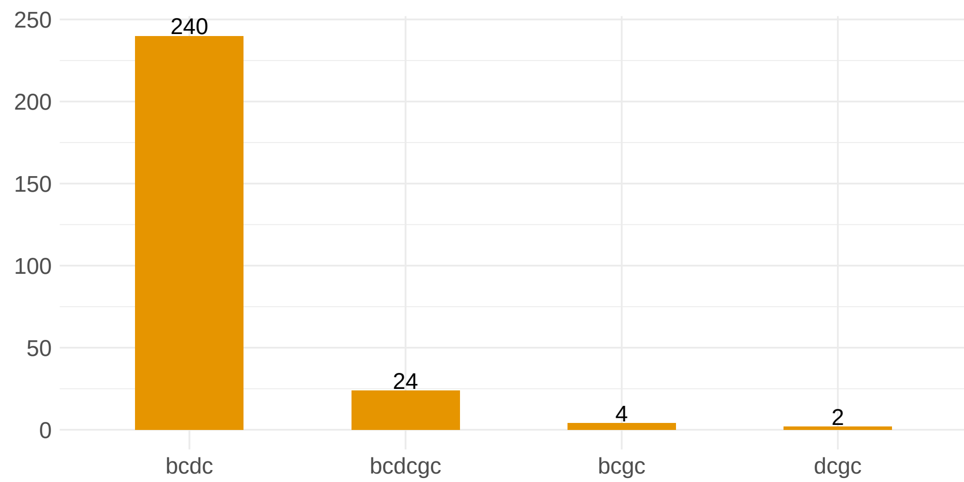


Figure 2.7: *The composition of inventories with voiced-constricted stops.*

Figure 2.7 shows the composition of inventories with voiced-constricted (Tc) stops. In contrast to all the other co-occurrence patterns considered so far, the most common inventory type is not the one with stops at all three places of articulation. Inventories with /ɓ/ and /ɗ/ to the exclusion of /ɠ/ are far more common.

Consistent with this, the results from model comparisons, in Table 2.9, show that the model with all two-way interactions does not best fit the data. Instead, the best performing model is the one in which the interaction between /ɗ/ and /ɠ/ is not included.

Since this pattern is different from all the others, results from the best performing model (Dc1) are shown in Table 2.10. The other model summaries can be found in the Appendix. Table 2.10 shows results from the model with two-way interactions between /ɓ/ and /ɗ/, and /ɓ/ and /ɠ/. The main effect of /ɠ/ and its interaction with /ɓ/ are not credible since the

	Model	Result	Inventory
Dc1	no denti-alveolar x velar	top performing	ɓɗ co-occur; ɓɠ do not
Dc2	no labial x velar	2^{nd} best	ɓɗ co-occur; ɗɠ do not
Dc3	all 2-way interactions	3^{rd} best	ɓɗ, ɗɠ, ɓɠ co-occur
Dc4	main effects only	4^{th} best	no co-occurrence of ɓɗɠ
Dc5	no labial x denti-alveolar	worst-performing	ɓɠ, ɗɠ do not co-occur

Table 2.9: *Model comparisons for voiced, constricted (Dc) stops.*

	Estimate	**Error**	**95% CI**
Intercept	2.09	0.08	[1.94, 2.24]
ɓ	-1.10	0.24	[-1.58, -0.62]
ɗ	-1.63	0.29	[-2.19, -1.04]
ɠ	-0.6	0.38	[-1.35, 0.17]
ɓɗ	**1.99**	**0.39**	**[1.22, 2.74]**
ɓɠ	**-0.65**	**0.38**	**[-1.39, 0.09]**

Table 2.10: *Dc1 result summary (all reference levels are 0, i.e., absence).*

95% Credible Intervals (CI) contains zero. Note, however, that the confidence interval is very skewed, just barely including zero. In such cases where the 95% CI is very skewed, we can quantify the probability that the parameter under consideration is strictly positive or negative by calculating its Probability of Direction (pd). The p-direction of the interaction term ɓɠ is 92%, meaning we can be 92% confident in the size of its effect. With this caveat in mind, notice that the estimates of the two interaction terms have opposite signs. This suggests that we can say with 95% confidence that inventories are more likely to have both /ɓ/ and /ɗ/. And, we can say, with 92% confidence, that inventories are *less* likely to have both /ɓ/ and /ɠ/. Thus, languages with /ɓ/ and /ɗ/, to the exclusion of /ɠ/ are more common, while languages that have all three voiced-constricted stops are much rarer. These patterns suggest the implicational hierarchy for voiced constricted stops is: the presence of /ɓ/ and /ɗ/ implies the *absence* of /ɠ/.

2.3.3 Summary

In summary, this section has shown that inventories are more likely to have voiceless-aspirated and voiced-aspirated stops at three major places of articulation at once or at none of them, than to have them at just one or two places. Co-occurrence of all three places is also mostly likely for plain voiceless, plain voiced and voiceless-aspirated stops. However, inventories without /p/, /g/ and /p'/ are just as plausible as inventories with /p, t, k/, /b, d, g/ and /p', t', k'/ respectively. By contrast, a voiced constricted velar stop is only likely to occur in a language which also has such stops at both the bilabial and denti-alveolar places. Bilabial and denti-alveolar voiced constricted stops themselves are likely to co-occur in inventories. These results are summarized in Table 2.11.

Laryngeal Category	**Results**
Plain voiceless	ptk co-occur OR gap at p – both equally likely
Plain voiced	bdg co-occur OR gap at g – both equally likely
Voiceless-aspirated	pʰtʰkʰ co-occur
Voiceless-constricted	p't'k' co-occur OR gap at p' – both equally likely
Voiced-aspirated	bʰdʰgʰ co-occur
Voiced-constricted	ɓɗ co-occur to the exclusion of ɠ

Table 2.11: *Summary of place co-occurrence results.*

These results are largely in line with Maddieson's findings. Like him, I find that a gap at /p/ is more likely than gaps at /t/ or /k/. However, I also find that inventories without gaps are just as likely. Maddieson (1984) does not make this explicit comparison, since he is more immediately concerned with describing gaps rather than statistically comparing co-occurrence patterns. The gap at /g/ also matches Maddieson's findings from UPSID, as does the gap at /p'/. In these cases, as well, the co-occurrence of the full series is just as likely. Maddieson does not directly address gaps by place in the two aspirated series, but I find that these tend to occur as a full set rather than with gaps. Finally, like Maddieson, I find that the bilabial and denti-alveolar voiced-constricted stops are more likely to co-occur to the exclusion of the velar, than they are likely to co-occur as a full series.

2.4 Co-occurrence of laryngeal categories

2.4.1 Method

Two implicational relationships between laryngeal categories were evaluated – (i) the presence of voiced-aspirated stops (Dh) implies the presence of voiced-unaspirated stops (D) and voiceless-aspirated stops (Th); (ii) the presence of voiced-constricted stops (Dc) implies the presence of plain voiced stops (D) and voiceless-constricted stops (Tc). Both these relationships are expected to hold if stop inventories make efficient use of distinctive features.

Implicational relationships between the laryngeal categories were evaluated in the following manner[1]. To test if the presence of a category implies the presence of other categories, I used a Bayesian implementation of logistic regressions in R with the laryngeal category whose presence implies the presence of other categories as the predictor variable, and the other categories whose presence is implied as separate dependent variables in different models. For example, to check if the presence of a voiced velar stop implies the presence of a voiceless velar stop, I ran a logistic regression with the presence or absence of the voiced velar stop as the predictor and the presence or absence of the voiceless velar stop as the dependent variable. Along the same lines, to check if the presence of a voiced aspirated velar stop (g^{h}) implies the presence of a plain voiced velar stop (g), a voiceless aspirated velar stop (k^{h}), and a plain voiceless stop (k), I ran three separate logistic regressions. These are shown in 2.12, 2.13 and 2.14. In each of these, the presence or absence of the voiced aspirated velar stop was the predictor variable. The dependent model in the first regression was the presence or absence of /g/, in the second, it was the presence or absence of /k^{h}/, and in the third, the dependent variable was the presence or absence of /k/. In all these models the Bernoulli distribution, a discrete probability distribution, was used, since it models the probability of a binary outcome, in this

[1]A different method was used here since plain voiceless stops are found in virtually every inventory, meaning the dependent variable, `count`, was largely predictable by the inclusion of this category. Dropping this category led to a considerable drop in model performance, with no credible differences between individual models that included other categories and excluded the plain voiceless category. This was to be expected as, typically, the presence of this neutral category does not predict the presence of any other category. The model comparison works in the models that evaluate the co-occurrence of place contrasts because the three major places are more or less equally frequent in languages and none occurs far more frequently than the others.

case presence (1) or absence (0) of any laryngeal category.

(2.12) The presence of /dʰ/ implies the presence of /d/: d ~ dʰ

(2.13) The presence of /dʰ/ implies the presence of /tʰ/: tʰ ~ dʰ

(2.14) The presence of /dʰ/ implies the presence of /t/: t ~ dʰ

2.4.2 Results

Figure 2.8 shows the dependencies between /dʰ/ the one hand and /t/, /tʰ/ and /d/ on the other. Each pair of bars shows the number of PHOIBLE inventories with and without /dʰ/, broken up by whether or not they also have /t/ (left panel), /tʰ/ (middle panel) and /d/ (right panel). In the stacked bars, blue represents the lack of the corresponding less marked stop, i.e. /t/ in the left panel, /tʰ/ in the middle panel and /d/ in the right panel.

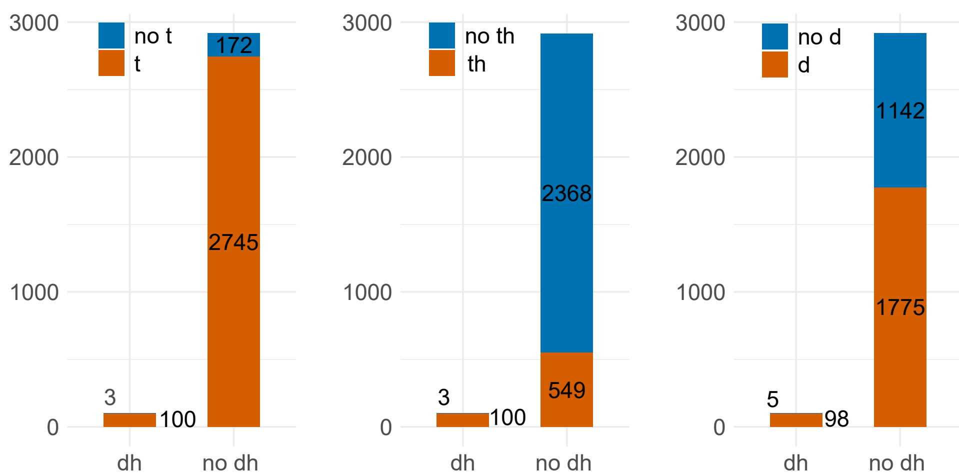


Figure 2.8: *Dependence of the number of inventories with and without /dʱ/.*

This figure shows that across the three laryngeal categories, inventories that have /dʰ/ are proportionally more likely to have /t/, /tʰ/ and /d/. Take the leftmost panel, for example. The probability of an inventory having /t/ when it also has /dʰ/ is given by

(2.15) $P(t = 1 \mid dh = 1) = \frac{100}{3+100} \approx 0.971$

By contrast, the probability of an inventory having /t/ when it does not have /dʰ/ is given by

$$P(t=1 \mid dh \neq 1) = \frac{2745}{172+2745} \approx 0.941 \qquad (2.16)$$

This difference is even more acute for the Th and D stops.

The logistic regression confirmed this pattern across all three places. The presence of /d^{h}/ led to an increase in the log-odds of /t/ by a factor of 0.91 (CI = [0.19, 1.70]); of /t^{h}/ by a factor of 4.49 (CI = [3.85, 5.19]); and of /d/ by a factor of 2.34 (CI = [1.76, 2.96]). In the labials, the presence of /b^{h}/ led to an increase in the log-odds of /p/ by a factor of 1.62 (CI = [0.53, 3.02]); of /p^{h}/ by a factor of 4.19 (CI = [3.45, 5.06]); and of /b/ by a factor of 3.14 (CI = [2.06, 4.63]). Finally, the presence of /g^{h}/ led to an increase in the log-odds of /k/ by a factor of 1.01 (CI = [0.05, 2.41]); of /k^{h}/ by a factor of 6.68 (CI = [4.78, 9.88]); and of /g/ by a factor of 3.04 (CI = [2.09, 4.25]).

The results for all three places suggest that presence of voiced-aspirated stops in an inventory increases the probability of plain voiceless, voiced and voiceless-aspirated stops also being present, compared to a case where voiced-aspirates are absent. Thus, there is credible evidence for the implicational relationship between voiced aspirates and all other pulmonic stops within a given place.

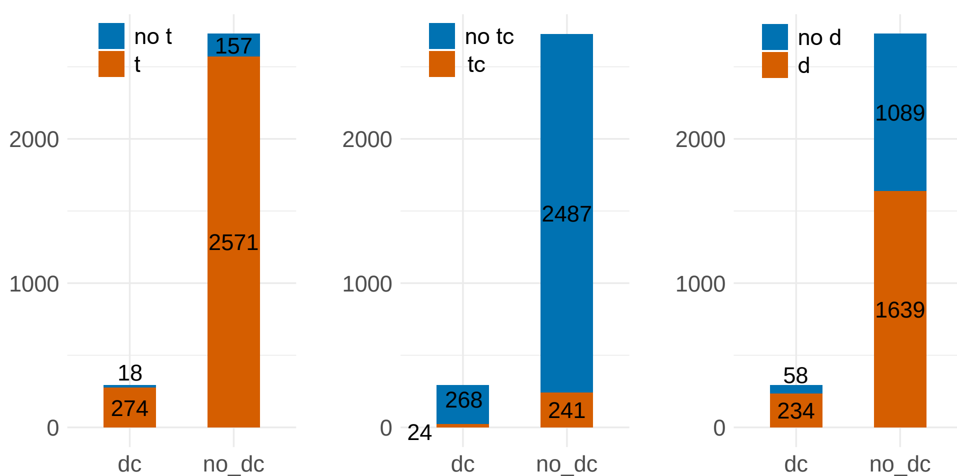


Figure 2.9: *Dependence of the number of inventories with and without* /ɗ/.

Figure 2.9 an equivalent plot for the denti-alveolar voiced-constricted stops. The pattern here is not as clear as it is in the case of voiced-aspirated stops. The conditional probability of plain voiceless stops in the presence of Dc stops is about 0.939 (274/(18 + 274)), which is comparable to the conditional probability of the presence of T in the absence of Dh, 0.942 (2571/(157 + 2571)). The dependency between Dc and Tc stops (shown in the middle panel) is similar ($P(Tc = 1 \mid Dc = 1) \approx 0.082$; $P(Tc = 1 \mid Dc \neq 1) \approx 0.088$), although the baseline probabilities of both Tc and Dc stops are low. By contrast, the presence of voiced constricted stops seems to increase the conditional probabilities of plain voiced stops, as in the case of Dh and D stops.

The logistic regression confirmed that this pattern is credible for all three places. The presence of /ɓ/ increased the log-odds of /b/ by a factor of 0.98 (CI = [0.70, 1.28]); the presence of /ɗ/ increased the log-odds of /d/ by a factor of 0.92 (CI = [0.66, 1.18]); and the presence of /ɠ/ increased the log-odds of /g/ by a factor of 1.09 (CI = [0.23, 2.10]). There were no credible effects of Dc stops on the presence of T or Tc stops. Therefore, the implicational relationship predicted by feature economy for the glottalic stops does not find support in the statistical analyses.

Together, these results are largely in line with Maddieson (1984). Although he does not explicitly set up an implicational hierarchy between the Dh, D, Th and T stops, he observes that inventories with these four stop types exhibit "the most completely filled out 4-series patterns" (Maddieson, 1984:30), thus gesturing towards the economy displayed by such stop systems. The finding that voiced-constricted stops imply the presence of plain voiced stops mirrors Maddieson's finding the the presence of implosives and voiced stops tend to co-occur. However, in a departure from his result showing that ejectives and implosives co-occur at a rate greater than expected from their individual rates of occurrence, I find that the presence of voiced-constricted stops is unrelated to voiceless-constricted stops.

2.5 Summary

The statistical analyses showed that place features tend to occur within a laryngeal category, except in the case of voiced constricted (Dc) stops. In the case of voiceless-constricted (Tc) and plain voiced (D) and voiceless (T) stops, inventories are just as likely to have gaps at /p'/, /g/ and /p/ respectively. Similar gaps are not likely in the voiceless aspirated (Th) and voiced aspirated (Dh) stops. The analysis of the voiced constricted (Dc) stops showed that it is more likely for these stops to occur at the labial and denti-alveolar places together, rather than at just one of these places. Moreover, the voiced constricted velar stop (like /ɠ/ or /g̰/) is likely to be absent in such inventories. Therefore, except for these voiced constricted stops, all other laryngeal articulations are more likely to occur at all three major places of articulation at once or at none of them, than to occur at just one or two places. In other words, languages tend to use these laryngeal articulations maximally by employing them at all three major places.

Statistical tests of the implicational relationships within laryngeal categories found that the presence of voiced aspirates (Dh) implies the presence of all other pulmonic laryngeal categories (T, Th and D). However, the same relationship did not hold for the glottalic stops. The presence of voiced constricted stops implies the presence of voiced stops only. This is the case for all three places of articulation. Moreover, the presence of the voiced constricted stops does not depend on the presence or absence of plain voiceless (T) and voiceless constricted (Tc) stops.

Setting aside the voiced constricted stops, the co-occurrence patterns of places of articulation with different laryngeal configurations showed that languages tend to make maximal use of place features, in line with Ohala (1980) and Clements (2003). This pattern also showed that gaps in inventories were likely a result of phonetic effects, with gaps most likely at /p/, /p'/ and /g/, as these segments are aerodynamically challenged. Inventories that employ Dh stops were also more likely to use place features maximally, even though /g^{h}/ is also similarly challenged. In these cases, however, the requirement to use distinctive features maximally is satisfied at the expense of tolerating these aerodynamic challenges.

The co-occurrence patterns also showed evidence for economy for the features corresponding to voicing and glottal spread. The presence of voiced aspirates (Dh) implied the presence of plain voiceless (T), voiceless aspirated (Th) and plain voiced (D) stops at all three major places. Therefore, these inventories prioritize the maximal use of place and the features, [voice] and [spread glottis], at the expense of articulatory complexity.

The voiced constricted type patterned differently from all other stop types, in terms of both place and laryngeal distinctions. These stops were likely to co-occur at the bilabial and denti-alveolar places, to the exclusion of the velar. This is a reflection of the phonetic dispreference against these types of segments that arises in the difficulty of differentiating it from a plain voiced velar stop. The presence of Dc stops predicted the presence of D stops, but not T or Tc stops. Thus, voiced-constricted stops do not pattern like voiced-aspirated stops, as inventories with Dc stops do not use their features maximally like inventories with Dh stops do.

In summary, pressure against articulatory complexity and the preference for using features maximally were found to conflict and resolve in different ways. There is no single principle that can explain the organization of stop inventories. Given the variation in which some objectives are prioritized over others, these inventories are ideal candidates for being modeled using approaches that can model the interaction of such conflicting objectives. Three such models are presented in Chapters 6, 3 and 7, each of which model different aspects of stop inventories.

CHAPTER 3

A dispersion-theoretic model of stop inventory typology

And now that you don't have to be perfect, you can be good.

– John Steinbeck, East of Eden

In this chapter I present a formal model of stop inventory typology, couched in Dispersion Theory (Liljencrants and Lindblom, 1972; Flemming, 2013) and Maximum Entropy Harmonic Grammar (Goldwater and Johnson, 2003; Hayes and Wilson, 2008). Although the contents of stop inventories are influenced by dispersion, markedness and feature economy, in this chapter I argue that these factors do not have equal status. Specifically, I show that structural constraints like economy are not needed in a model that explicitly models only dispersion-theoretic functional pressures. This chapter suggests that the interaction of functional phonological constraints in a probabilistic grammar can explain much of the structure found in the world's stop inventories.

3.1 Background

3.1.1 Maximum Entropy Harmonic Grammars

The model is couched in MaxEnt (Goldwater and Johnson, 2003; Hayes and Wilson, 2008), a probabilistic variant of Optimality Theory (Prince and Smolensky, 1994). An advantage of using a probabilistic model over classical OT for present purposes is that the former generates not an individual winner, but probabilities over candidate inventories, which can then be compared to stop inventories attested in the world's languages. This is appropriate for the purposes of this chapter, whose goal is not to determine which stop inventory is the best one (i.e., the winning candidate), but to capture how the interaction of the organizing principles in Chapter 6 could explain the distribution of attested inventories.

Like other types of harmonic grammar, MaxEnt grammars (Goldwater and Johnson, 2003) produce constraint weights rather than a ranking. For the input-output pair (x, y), the harmony is given by

(3.1)

$$\mathscr{H} = \sum_{i=1}^{n} w_i C_i(x, y)$$

where n represents the number of constraints, w_i represents the vector of constraint weights and $C_i(x, y)$ represents the vector of constraint violations incurred by the mapping of x to y. MaxEnt grammars work by mapping this harmony score to the probability of the output y given the input x. This probability is given by

(3.2)

$$P(y \mid x) = \frac{e^{-\mathscr{H}(x,y)}}{\sum_{y'} \mathscr{H}(x, y')}$$

The goal of the MaxEnt model is to find the set of weights that maximizes the probability of the observed data. A measure of how well a set of weights captures the observed data is given by

(3.3)

$$L(D) = \prod_{i=1}^{n} P(y_i \mid x_i; w)$$

where the likelihood L of N points the dataset D is given by the product of the conditional probabilities assigned to each input-output pair (x_i, y_i) under the weights w. A higher value of the likelihood indicates that a set of weights fits the data better than a set that yields a lower likelihood. Thus, when finding weights for a given dataset, the objective of the MaxEnt algorithm is to find a set that maximizes the likelihood of the data.

Since multiplying several probabilities together can produce very small values, likelihood maximization is implemented as maximizing the log-likelihood of the data, given by

(3.4)

$$LL(D) = \sum_{i=1}^{n} ln P(y_i \mid x_i; w)$$

This returns less unwieldy values compared to (3.3), which enables the implementation of this algorithm in a computational software. In sum, MaxEnt models work by finding a set of weights that maximizes the log-likelihood of the observed data.

Note that other probabilistic models, including but not only alternative probabilistic implementations of OT, are appropriate for this task. I employ a MaxEnt model here as it lends itself to the application of standard statistical inference procedures (Flemming, 2021; Mayer et al., 2024).

3.1.2 A Formal Model of Dispersion

The second component of the model presented in this chapter is based on Flemming's (2013) dispersion theory of contrast, which formalizes Adaptive Dispersion Theory (Lindblom, 1986, 1990). According to this theory, the contents of inventories are determined by three functional goals that represent the communicative function of language.

(3.5) Functional goals:

a. Maximize the distinctiveness of contrasts.

b. Maximize the number of contrasts.

c. Minimize articulatory effort.

The first goal (3.5a) enforces maximal auditory distinctiveness to aid the listener by minimizing confusion. The second goal (3.5b) promotes maximizing the number of contrasting segments to generate a rich lexicon while avoiding excessively long words. Finally, the third goal (3.5c) promotes effort minimization, a tendency exhibited by both linguistic and non-linguistic behavior. Crucially, these goals can conflict – maximizing distinctiveness may require more effortful articulation, for example. Languages resolve this conflict in differing ways, which is held to be the source of cross-linguistic variation.

Flemming (2013) formalizes these conflicting goals as constraints in classical OT (Prince and Smolensky, 1994). For illustrative purposes, consider the following toy example. Assume that the complete set of stops possible in a language is the one shown in Figure 3.1. In this figure, "p" denotes voiceless bilabial stops, "ph" denotes voiceless-aspirated bilabials stops and "pc" denotes voiceless-constricted stops[1]. Under Flemming's approach the functional goals in 3.5 can be formalized as the following constraints.

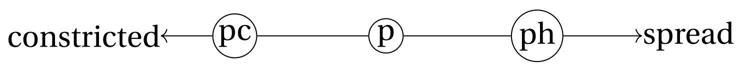


Figure 3.1: *Toy example: Possible stops.*

(3.6) Constraints in Flemming's model

a. MINDIST=2[2]: Assign a violation for every pair of contrasts that does not differ by at least 2 units along the contrasting dimension.

b. MAXIMIZECONTRASTS: Assign a positive point for every stop in the inventory.

c. *MARKED - Assign a violation for every marked articulation.

Note that (3.6b) is a positive scalar constraint that awards candidates rather than penalizing them like the other two constraints. *MARKED is a broadly defined markedness constraint that penalizes marked articulations – pc and ph in this example. Thus, we can generate the tableau in Table 3.1, where different rankings generate different winners. A language that ranks MAXIMIZE CONTRASTS >> MINDIST=2, *MARKED selects candidate (a) as its inventory. A language that ranks ranks *MARKED >> MINDIST=2, MAXIMIZE CONTRASTS selects (b), and one that ranks MINDIST=2 >> MAXIMIZE CONTRASTS, *MARKED selects (c).

[1]See Table 1.1 in Chapter 1 for details on how stop types were coded.

[2]Note that Flemming also formulates MINDIST=1, which is excluded here since none of the candidates violate it.

	MINDIST=2	MAXIMIZE CONTRASTS	*MARKED
a. pc-p-ph	**	✓✓✓	**
b. p-ph	*	✓✓	*
c. pc-ph		✓✓	**

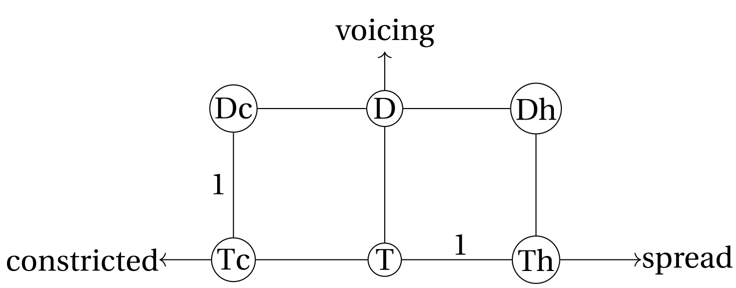
Table 3.1: *Toy example: OT Tableau under Flemming's approach*

Flemming's is a generative model of languages and it is not claimed to be the model of a speaker or learner. The model introduced in this chapter is, similarly, a typological model of languages, and not a grammar of a speaker of a language.

3.2 A probabilistic dispersion-theoretic model

We can now build a model that evaluates candidate inventories and predicts their probabilities given a set of phonetically-motivated constraints, by combining elements from MaxEnt grammars and Dispersion Theory. With this model, instead of generating constraint rankings corresponding to different languages, we can find constraint weights that best fit the stop inventories attested in the languages of the world. A successful model is one that assigns high probabilities to stop inventories that are common in languages, and low probabilities to those that are unattested.

Possible stops are assumed to be those shown in Figure 3.2. In this figure, "T" represents neutral stops, "Th" voiceless-aspirated stops, "Tc" voiceless-constricted stops, "D" plain voiced stops, "Dh" voiced-aspirated stops and "Dc" voiced-constricted stops. Recall that this configuration assumes some abstraction to comprehensively account for all possible laryngeal contrasts.



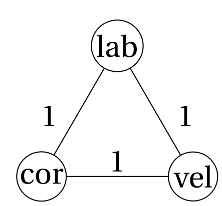


Figure 3.2: *The phonetic space of stop contrasts: laryngeal (left) and place (right).*

The laryngeal categories are assumed to have the organization shown in the left panel of Figure 3.2 (Narkar, 2025)[3]. As in the other chapters, the possible places of articulation are limited to the near-universal labial, denti-alveolar and velar series. These are assumed to have the two-dimensional triangular organization shown in the right panel of Figure 3.2. This configuration assumes that stops at all three places of articulation are perceptually equidistant from each other – /p/ is as distinct from /t/ as it is from /k/, and so on.

Candidates to the MaxEnt tableau are inventories generated by all possible combinations of the 18 stops in Figure 3.2 (6 laryngeal categories × 3 places), yielding 2^{18} inventories in total. Attested inventories are assigned frequencies from PHOIBLE (Moran and McCloy, 2019), while unattested inventories are assigned a frequency of zero. No inputs are specified, following Flemming (2013). Table 3.2 shows an example part-tableau with various candidates and their corresponding frequencies. The full tableau has 262,144 rows, where each row corresponds to a possible stop inventory.

	FREQ
a. p-t-k-b-d-g	801
b. p-t-k	752
c. p-t-k-b-d-g-bc-dc	149
d. p-t-k-b-d-g-ph-th-kh	138
e. p-t-k-b-d	92
f. p-t-k-b-d-g-ph-th-kh-bh-dh-gh	72
g. p-t-k-ph-th-kh-pc-tc-kc	33
h. p-k-g-bh-bc-gc	0
i. p-k-kh-g-bh-bc-gc	0
j. ...	...

Table 3.2: *Example part-tableau showing output candidates and their associated frequencies.*

Every candidate inventory is evaluated for its violation of the specified constraints, described in detail in Sections 3.4-3.6. Thus, a tableau is populated with no input, 2^{18} candidate outputs and their associated frequencies and constraint violations. The resulting tableau is fit using the R package `maxent-ot` (Mayer et al., 2024) to find the optimal constraint weights

[3]As shown later in Chapter 5, the y-axis in this figure is equivalent to the second dimension of the stop acoustic space and the x-axis is equivalent to the third dimension.

that maximize the objective function (see Chapter 1 for details). Gaussian priors are specified for each constraint with a mean of zero and a standard deviation of 100. The zero mean, commonly used in MaxEnt modeling (Goldwater and Johnson, 2003; Martin, 2011; Mayer et al., 2024), represents an agnostic default and avoids overfitting. The high standard deviation allows the constraint weights to deviate from zero without a large penalty. Thus, all constraints start on equal footing but are allowed to deviate from the prior with a very low penalty. The constraint set that that best fits the distribution of attested inventories is taken to represent the factors governing the makeup of stop inventories.

3.3 The dimensions of stop contrasts

Before describing the constraints used in the model, I establish the phonetic space of stop contrasts. Stops are assumed to differ in terms of place of articulation, voicing and glottal spread, as shown in Figure 3.2. It should be noted that this space is defined in articulatory, rather than auditory parameters, even though the latter is more relevant to constraints that refer to perceptual dissimilarity.

The first relevant dimension that distinguishes stops is their place of articulation. The distances between the three places are assumed to be as below:

(3.7) Place dimension:

1	labial – denti-alveolar
1	denti-alveolar – velar
1	labial – velar

Thus, the three places of articulation are assumed to be equidistant from each other, as shown in Figure 3.2.

The second dimension relevant to stop contrasts is glottal width, represented by the x-axis in Figure 3.2. Glottal width is the smallest in constricted stops and largest in aspirated stops. The distances along this dimension are therefore:

(3.8) Glottal width dimension:

1	constricted
2	neutral
3	aspirated

Finally, glottal vibration or voicing, shown on the y-axis in Figure 3.2, has two possible settings – voiceless or voiced. The distances are therefore:

(3.9) Voice dimension:

1	voiceless
2	voiced

We can now formulate constraints on stop contrasts based on the phonetic space thus defined.

3.4 Minimize articulatory effort

From the contents of Table 3.2, it is apparent that the most common inventories have articulatorily simple stops and more complex stops start appearing only in the presence of less marked stops. Therefore, the functional goal of minimizing articulatory effort (3.5c) must be a crucial pressure affecting the contents of stop inventories. We can thus define markedness constraints in the following manner.

The markedness constraints must represent the penalty for the use of effortful articulations in an inventory. This includes the use of any laryngeal articulation – glottal vibration (i.e. voicing), glottal spreading (i.e. aspiration) and glottal constriction (i.e. constriction). This yields the following corresponding constraints.

(3.10) *VOICE: Assign a violation if an inventory employs voicing.

(3.11) *ASPIRATION: Assign a violation if an inventory employs aspiration.

(3.12) *CONSTRICTION: Assign a violation if an inventory employs glottal constriction.

I assume that the concurrent articulation of two laryngeal maneuvers requires an articulatory effort greater than the effort involved in the articulation of each laryngeal maneuver separately. For instance, the effort required to produce a voiced-aspirated stop is greater than the sum of the effort involved in producing a voiced stop and a voiceless-aspirated stop. Therefore, voiced-aspirated and voiced-constricted stops are associated with specific constraints beyond those associated with voicing, aspiration and constriction. This yields two additional constraints.

(3.13) *VOICED-ASPIRATION: Assign a violation if an inventory employs voiced aspiration.

(3.14) *VOICED-CONSTRICTION: Assign a violation if an inventory employs voiced constriction.

It is assumed that the use of each place is associated with a markedness constraint, but this does not differ by location of the closure, yielding the markedness constraints below.

(3.15) *LABIAL: Assign a violation if an inventory employs a labial stop.

(3.16) *CORONAL: Assign a violation if an inventory employs a denti-alveolar stop.

(3.17) *VELAR: Assign a violation if an inventory employs a velar stop.

Finally, combinations of place and laryngeal features that are known to be effortful are assumed to incur an additional penalty. There is a dispreference for voicing at posterior places due to its aerodynamic disadvantage, since it is difficult to maintain voicing when the space between the glottis and the closure is small (Keating, 1984a). Voiceless-constricted stops, like ejectives, at the bilabial place of articulation are also articulatorily marked, for the exact opposite reason. The closure at the bilabial place requires more intra-oral pressure than at other places since the lips are both active articulators and more volume must be filled with air.

(3.18) *G: Assign a violation if an inventory employs any voiced velar stop.

(3.19) *PC: Assign a violation if an inventory employs a bilabial voiceless constricted stop.

(3.20) *GC: Assign a violation if an inventory employs a velar voiced constricted stop.

Note that these constraints are binary – candidate inventories are penalized for the use of effortful articulations, not for how many effortful articulations they employ[4]. For example, the inventory /p t k b d/ violates *VOICE only once, not twice. It is thus comparable to the inventory /p t k b d g/ in terms of *VOICE violations. Note that /p t k b d g/ is still more marked than /p t k b d/ as the former also violates *G. Conceptually, this binary formulation of constraints suggests that it is the use of effortful articulations that is relevant for quantifying the markedness of a stop inventory, not *how many* effortful articulations it employs. That is, the markedness penalty an inventory incurs for employing, say, voicing, does not increase with the number of voiced stops it employs. Once an inventory employs an effortful articulation like voicing, it is marked with respect to the corresponding constraint regardless how many stops use the marked articulation. An example tableau illustrating markedness violations is shown in 3.3.

	*VOICE	*ASP	*CONS	*V-ASP	*V-CONS	*LAB	*COR	*VEL	*G	*PC	*GC	FREQ
a. p-t-k-b-d-g	1					1	1	1	1			801
b. p-t-k						1	1	1				752
c. p-t-k-b-d-g-ph-th-kh	1	1				1	1	1	1			138
d. p-t-k-b-d	1					1	1	1				92
e. ...	...	...	...	...	...	...	...	...	...	...	...	...

Table 3.3: *Example part-tableau showing markedness violations.*

We can test the effectiveness of markedness constraints in capturing stop inventory typology by fitting a model that includes only these constraints. Since this model contains over a quarter of a million candidates, only a subset of candidates are considered here. Here, a smaller tableau containing all 239 attested unique stop inventories and 2000 unattested inventories, randomly sampled from the remaining (2^{18} - 239) candidates. Weights are fit using the `maxent-ot` function `optimze_weights`. Figure 3.3 shows parity plots with observed frequencies of different stop inventories on the x-axis and frequencies predicted by the model on the y-axis. Each circle represents a candidate stop inventory and the number next to it shows its attested frequency in PHOIBLE. A perfectly fit model would have points lying exactly on the diagonal. The left panel shows all the inventories in the model. The right one zooms in on the inventories that cluster in the bottom left, which are harder to visually separate in the left panel.

[4]Comparison of this model with one that assigned non-binary markedness violations showed that the former model fits attested inventories better than the latter.

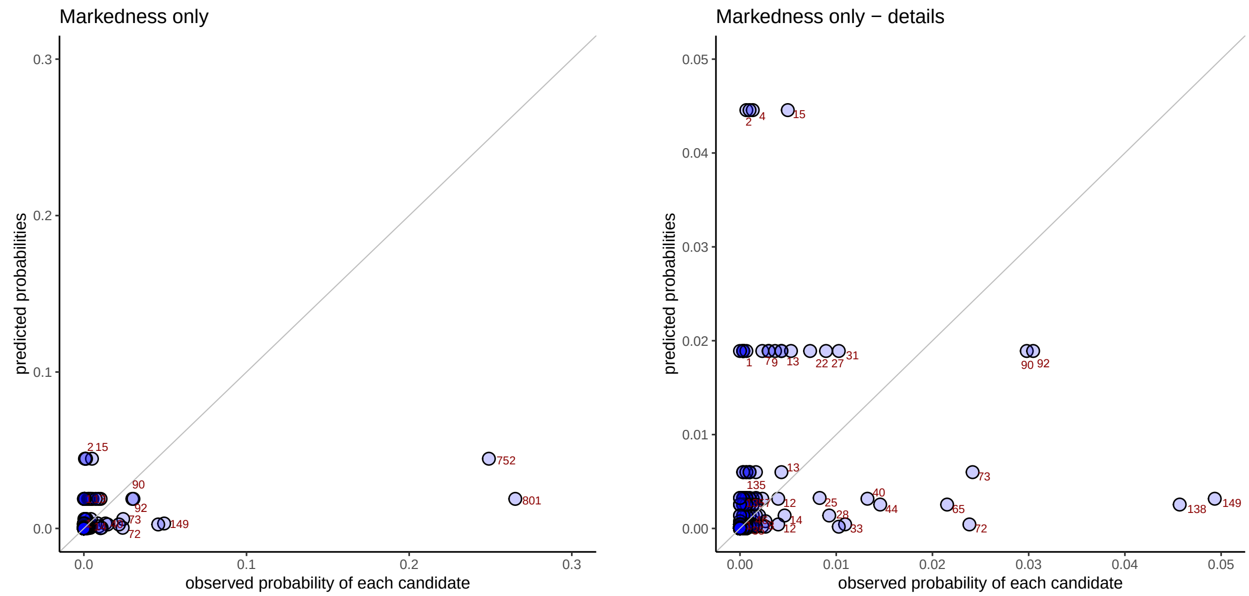


Figure 3.3: *Observed versus predicted probabilities with markedness constraints.*

Figure 3.3 shows that the MaxEnt model overestimates the frequency of simpler inventories. The two outliers which have very high observed probabilities represent /p t k/ (752) and /p t k b d g/ (801). The cluster of inventories with the highest predicted frequencies above the diagonal consists of simple inventories of size 2 – /t k/, /p t/ and /p k/ – which are actually quite uncommon (with frequencies of 15, 4 and 2 respectively). This is expected since all the constraints in the model penalize marked articulations, suggesting markedness alone cannot explain the contents of stop inventories. Weights assigned to the various markedness

Constraint	**Weight**
*VOICE	0.86
*ASPIRATION	2.00
*CONSTRICTION	2.62
*VOICED-ASPIRATION	1.78
*VOICED-CONSTRICTION	1.79
*GC	1.41
*LABIAL	0
*CORONAL	0
*VELAR	0
*G	0
*PC	0

Table 3.4: *Constraint weights in the model with only markedness constraints.*

constraints in this model are shown in Table 3.4. As shown in this table, the model assigns non-zero weights only to the constraints penalizing the use of laryngeal articulations (and the velar voiced-constricted stop). The weights of all the other constrains that penalize the use of place are zero. Thus, this markedness-only model is unable to predict inventory gaps that known to occur (e.g., at /p/, /g/ etc.).

3.5 Maximize contrasts

We can enrich the model in Section 3.4 with constraints reflecting a further functional goal of Dispersion Theory, namely, maximizing the number of contrasts (3.5b). The constraint MAXIMIZECONTRASTS is defined as:

(3.21) MAXIMIZECONTRASTS: An inventory with n stops incurs $(18-n)$ violations of this constraint.

Note that this definition of MAXIMIZECONTRASTS differs from Flemming's – whereas he implements MAXIMIZECONTRASTS as a positive constraint rewarding more contrasts, I implement it as a negative constraint, in line with all the other constraints in the model, for reasons of parsimony. Due to the MaxEnt math, there is no functional difference between Flemming's positive constraints and the negative ones used here.

In addition to maximizing the total number of stops, as enforced by MAXIMIZECONTRASTS, constraints promoting more contrasts along each dimension in Figure 3.2 are also specified as:

(3.22) MAXIMIZEWIDTHCONTRASTS: A stop inventory that employs x glottal widths incurs $(3-x)$ violations of this constraint.

(3.23) MAXIMIZEVOICECONTRASTS: A stop inventory that employs y distinctions of voicing incurs $(2-y)$ violations of this constraint.

(3.24) MAXIMIZEPLACECONTRASTS: A stop inventory that employs z places of articulation incurs $(3-z)$ violations of this constraint.

An illustrative example of MAXIMIZECONTRASTS violations is shown in Tableau 3.5.

	MAXCONT	MAXWIDTHCONT	MAXVOICECONT	MAXPLACECONT	FREQ
a. p-t-k-b-d-g	12	2			801
b. p-t-k	15	2	1		752
c. p-t-k-b-d-g-ph-th-kh	9	1			138
d. p-t-k-b-d	13	2			92
e. p-k-kh-g-bh-bc-gc	11			1	
f. ...	...	...	...	...	...

Table 3.5: *Example part-tableau with* MAXCONTRASTS *violations.*

We can now add MAXIMIZECONTRAST constraints to the model in Section 3.4. An illustrative part-tableau with MAXIMIZECONTRASTS and markedness violations is shown in Table 3.6. Note that violations within each constraint type are summed for the purposes of illustration.

	ΣMAXCONTRASTS	ΣMARKEDNESS	FREQ
a. p-t-k-b-d-g	14	5	801
b. p-t-k	18	3	752
c. p-t-k-b-d-g-bc-dc	11	6	149
d. p-t-k-b-d-g-ph-th-kh	7	9	138
e. p-t-k-b-d	5	5	92
f. p-t-k-b-d-g-ph-th-kh-bh-dh-gh	7	7	72
g. p-t-k-ph-th-kh-pc-tc-kc	7	11	33
h. p-k-g-bh-bc-gc	12	7	0
i. p-k-kh-g-bh-bc-gc	13	7	0
j. ...	...	...	...

Table 3.6: *Example part-tableau with* MAXIMIZECONTRASTS *and markedness constraints.*

Results from a model with all attested inventories and 2000 randomly sampled inventories are shown in Figure 3.4. Compared to the model with only markedness constraints, the fit of this model, which also has MAXIMIZECONTRASTS constraints, is greatly improved. Table 3.7 shows the results of a model comparison in terms of the Akaike information criterion (AIC). In general, lower values of AIC indicate better model fit while accounting for the number of constraints. As seen from this table the added complexity of the model with both types of constraints is justified as it has a much lower value of AIC, when compared to the model with only markedness constraints.

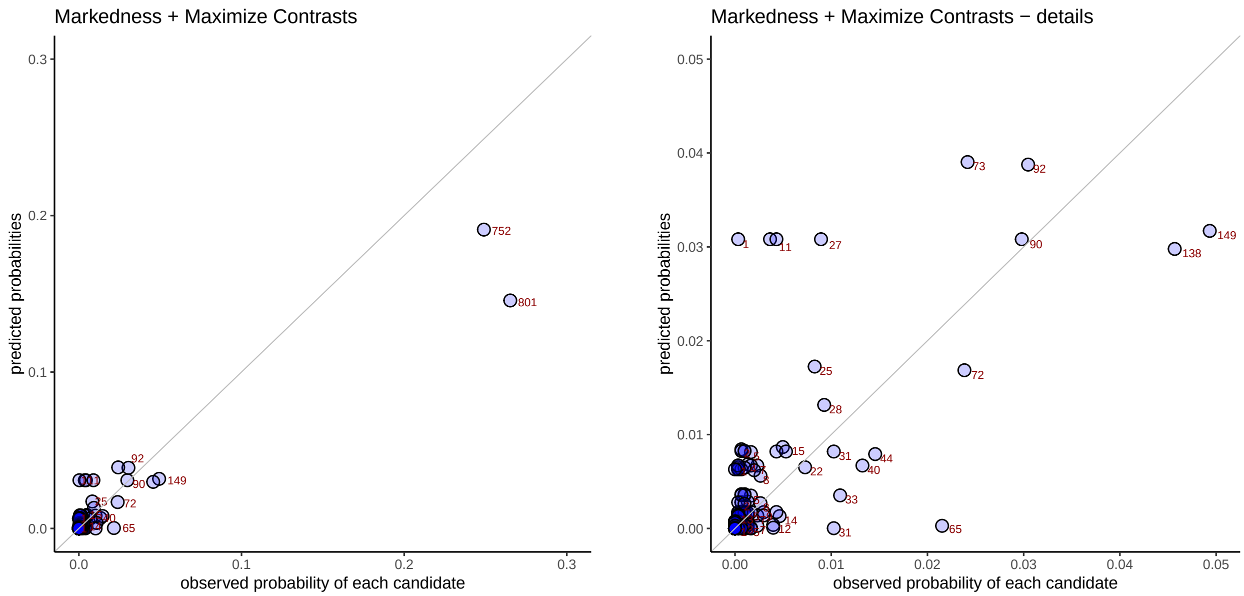


Figure 3.4: *Observed versus predicted probabilities with* MAXIMIZECONTRASTS *and markedness constraints.*

Model	**AIC**	**ΔAIC**
Markedness + MAXIMIZECONTRASTS	21754.88	0
Markedness only	28390.83	6635.96

Table 3.7: *Model comparison with and without* MAXIMIZECONTRASTS *constraints.*

The frequencies of inventories predicted by this model more closely match their observed frequencies as evidenced by the shift of the circles closer to the diagonal in Figure 3.4, compared to Figure 3.3. Moreover, this improved fit can be entirely attributed to the inclusion of two constraints – MAXIMIZEPLACECONTRASTS and MAXIMIZECONTRASTS – the only MAXIMIZECONTRASTS constraints to receive non-zero weights in the model, as shown by the model constraint weights in Table 3.8. The higher weights of these constraints allow the larger inventories /p t k/ (752 observed inventories), /p t k b d g/ (801 observed inventories), /p t k ph th kh b d g/ (128 observed inventories) and /p t k ph th kh/ (73 observed inventories) to become more favored over the previously favored, less marked inventories of size 2.

Comparison with constraint weights in the markedness-only model, shown in Table 3.4, reveals that the addition of the MAXIMIZECONTRASTS family also changes the weights assigned to the markedness constraints. The relative weights of the constraints penalizing the use of la-

Constraint	Weight
MAXIMIZECONTRASTS	1.55
MAXIMIZEPLACECONTRASTS	5.46
*VOICE	4.70
*ASPIRATION	6.25
*CONSTRICTION	6.26
*VOICED-ASPIRATION	5.23
*VOICED-CONSTRICTION	4.63
*G	0.23
*PC	0.80
*GC	3.29
*LABIAL	1.19
*VELAR	0.95
MAXIMIZEVOICECONTRASTS	0
MAXIMIZEWIDTHCONTRASTS	0
*CORONAL	0

Table 3.8: *Constraint weights in the model with markedness and* MAXIMIZECONTRASTS *constraints.*

ryngeal maneuvers are comparable in both models – *ASPIRATION, *CONSTRICTION > *VOICED-ASPIRATION, *VOICED-CONSTRICTION > *VOICE. In contrast to the previous model, however, now the constraints *PC and *G have non-zero weights, in addition to *GC. Finally, two constraints penalizing place, *LABIAL and *VELAR, also receive considerable weights. As a result, inventories with gaps at these places, such as /p t k b d/ (92 attested inventories), are favored by the model. Note also that the frequency of occurrence of the inventory with a gap at /p/, /t k b d g/ (90 such attested inventories), is predicted by the model almost perfectly. The considerable weight of *LABIAL alone drives this pattern and there is no need for a constraint like *P.

Thus, combining maximizing the place of articulation of stops and preferring larger stop inventories, while being subject to minimizing articulatory difficulty, predicts a typology of stop inventories that more closely matches the observed typology.

3.6 Maximize dispersion

We can enrich the MaxEnt model even further by adding constraints that reflect the goal of maximizing the distinctiveness of stop contrasts (3.5a). To this end, a set of constraints, MINDIST (Flemming, 2013), is defined as

(3.25) MINDIST: Maximize the distinctiveness of the contrasts in Figure 3.2.

MINDIST is a family of constraints of the form MINDIST=*Dimension:distance*, where *Dimension* refers to the perceptual dimension under consideration, and *distance* refers to the minimum distance along that dimension that the corresponding constraint enforces. The complete set of MINDIST constraints, therefore, includes:

(3.26) MINDIST=PLACE:1: Assign one violation for every pair of stops that does not differ in place by at least 1 unit.

(3.27) MINDIST=WIDTH:1: Assign one violation for every pair of stops that does not differ in glottal width by at least 1 unit.

(3.28) MINDIST=WIDTH:2: Assign one violation for every pair of stops that does not differ in glottal width by at least 2 units.

(3.29) MINDIST=VOICE:1: Assign one violation for every pair of stops that does not differ in voicing.

Note that under Flemming's approach, violations of MINDIST constraints are assessed over pairs of contrasts. That is, it is not simply the presence of a pair of stops differing along a certain dimension with a distance smaller the the one prescribed by the constraint that incurs a violation – the minimum distances between all individual pairs of stops that make up an inventory count towards violations. MINDIST=WIDTH:1, for instance, enforces a distinction in glottal spread of at least one unit. Therefore, the stop inventory /k kh kc/ does not violate this constraint, since all pairs of contrasts differ in at least 1 unit in glottal width. However, /p t k/ incurs 3 violations MINDIST=WIDTH:1, since none of the pairs, /p-t/, /t-k/ and /p-k/, have a minimal difference of 1 in glottal width. The same inventory, /p t k/, violates

MINDIST=WIDTH:2 – enforcing a glottal width difference of at least two units – three times, since the all the pairs have the same glottal width. An example of the violation profiles of a handful of illustrative inventories is shown in the tableau in 3.9. Note that "MINDIST=" is omitted from the name of each constraint here. Although there are three possible places of articulation a stop may have, these places are all assumed to be equidistant from each other, as shown in Figure 3.2. Therefore, there is no MINDIST=PLACE:2 as there are no pairs that differ in place by two units.

	PLACE:1	WIDTH:1	WIDTH:2	VOICE:1	FREQ
a. p-t-k-b-d-g	3	15	15	6	801
b. p-t-k		3	3	3	752
c. p-t-k-b-d-g-ph-th-kh	9	18	36	18	138
d. p-t-k-b-d	2	10	10	4	92
e. p-k-kh-g-bh-bc-gc	9	5	17	9	0
f. ...	...	...	...	...	...

Table 3.9: *Example part-tableau with pairwise* MINDIST *violations.*

These constraint violations can now be added to the model introduced in Section 3.5, yielding the part-tableau in 3.10. Note that, once again, violations within each constraint family are summed for the purposes of illustration.

	ΣMINDIST	ΣMAXCONTRASTS	ΣMARKEDNESS	FREQ
a. p-t-k-b-d-g	39	14	5	801
b. p-t-k	9	18	3	752
c. p-t-k-b-d-g-bc-dc	64	11	6	149
d. p-t-k-b-d-g-ph-th-kh	81	7	9	138
e. p-t-k-b-d	26	5	5	92
f. p-t-k-b-d-g-ph-th-kh-bh-dh-gh	144	7	7	72
g. p-t-k-ph-th-kh-pc-tc-kc	81	7	9	33
h. p-k-g-bh-bc-gc	30	12	7	0
i. p-k-kh-g-bh-bc-gc	40	13	7	0
j. ...	...	...	...	...

Table 3.10: *Example with* MINDIST, MAXIMIZECONTRASTS *and markedness violations.*

Results from the model with all three types of constraints fitting attested inventories and 2000 randomly sampled inventories are shown in Figure 3.5. This figure is comparable to Fig-

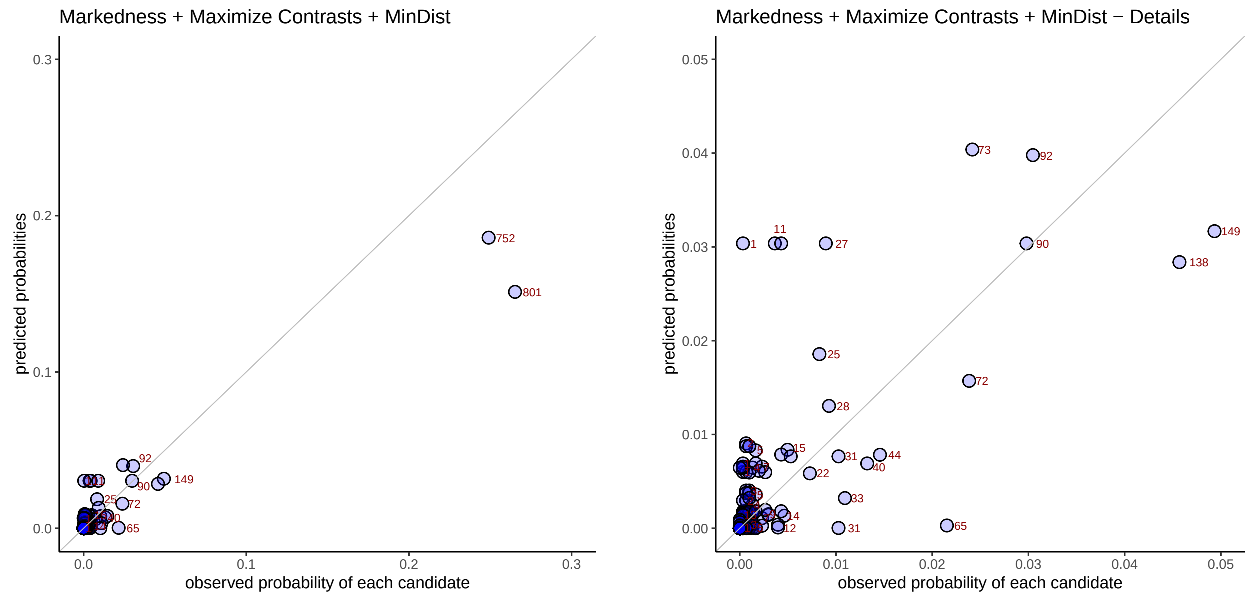


Figure 3.5: *Observed versus predicted probabilities with* MINDIST, MAXIMIZECONTRASTS *and markedness constraints.*

ure 3.4, which included only MAXIMIZECONTRASTS and markedness constraints. The added complexity of the model with all three types of dispersion constraints results in a worse (i.e. higher) value of AIC, as shown by the model comparison in Table 3.11. This table shows that the lowest AIC is associated with the model with only MAXIMIZECONTRASTS and markedness constraints. This suggests that including the added complexity of MINDIST constraints in the model is not worth it in terms of improving the fit to the attested data[5].

Model	**AIC**	**Δ(AIC)**
MAXIMIZECONTRASTS + Markedness	21754.88	0
MINDIST + MAXIMIZECONTRASTS + Markedness	21759.53	4.66
Markedness only	28390.83	6635.96

Table 3.11: *Model comparison.*

Table 3.12 shows the constraint weights in the model with all three types of constraints. Two dispersion constraints, MINDIST=VOICE:1 and MINDIST=WIDTH:2 receive non-zero weights.

[5] Note that the combinations of the three models in Table 3.11 are used for the sake of exposition. The model with MAXIMIZECONTRASTS and markedness constraints is the best performing model when compared to other combinations as well. Additionally, results from k-fold cross-validation showed that the best model does not overfit the data and generalizes well.

Constraint	Weight
MINDIST=VOICE:1	1.81
MINDIST=WIDTH:2	1.91
MAXIMIZECONTRASTS	1.74
MAXIMIZEPLACECONTRASTS	4.80
*VOICE	4.88
*ASPIRATION	6.24
*CONSTRICTION	6.29
*VOICED-ASPIRATION	5.01
*VOICED-CONSTRICTION	4.65
*G	0.27
*PC	0.73
*GC	3.15
*LABIAL	0.95
*VELAR	0.69
MINDIST=PLACE:1	0
MINDIST=WIDTH:1	0
MAXIMIZEVOICECONTRASTS	0
MAXIMIZEWIDTHCONTRASTS	0
*CORONAL	0

Table 3.12: *Constraint weights in the model with markedness and* MAXIMIZECONTRASTS *constraints.*

All other constraint weights are comparable to the weights in Section 3.5. Since some constraint pairs, like MINDIST=VOICE:1 and MAXIMIZECONTRASTS, have the same effect, the frequencies of some inventories are over-predicted by this model, compared to the simpler model which fits the data better. The non-zero weight of MINDIST=WIDTH:2 does improve fit to some inventories that utilize the extreme ends of the glottal-state dimension (e.g., /pc tc kc p t k ph th kh/), but the overall effect of inclusion of the MINDIST constraints worsens the model fit.

3.7 The nature of MINDIST constraints

While the model presented in Section 3.6 presents one way of measuring distinctiveness, namely, pairwise distances (Flemming, 2013), it is not the only one. In this section, I present two other ways of measuring distances and present results from models that evaluate violations of MINDIST based on each of these methods of measuring the distinctiveness of the stop categories in Figure 3.2. The formulation of the MINDIST constraints in the best performing model can be interpreted as most closely capturing the distinctiveness of the stop categories under consideration.

3.7.1 Euclidean distance

A different way of evaluating how dispersed a stop inventory is by measuring the sum of all pairwise Euclidean distances between its stops. This is comparable to Liljencrants and Lindblom's (1972) approach to measuring dispersion in vowel systems. In the case of stop inventories, given a set of stops $P = \{p_1, p_2, \ldots, p_n\}$ where each stop p_i has coordinates (x_i, y_i, z_i) in the three-dimensional space, the total pairwise distance is calculated as

(3.30)
$$r = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

In this case there is only one dispersion constraint, MINDIST, as opposed to a set of constraints as in Section 3.6, which prefers larger values of r. Violations of MINDIST are therefore operationalized as the reciprocal of r so that inventories with smaller global distances are penalized more and those will larger distances are penalized less. Note that the violations of all the constraints are normalized so that their values fall between 0 and 1.

(3.31) MINDIST: Assign $\frac{1}{r}$ violations to an inventory whose total pairwise distance is r.

3.7.2 The bottleneck approach

Instead of measuring minimum pairwise distances or global pairwise distance, dispersion can be formalized based on the minimum distance between any pair of contrasts in an inventory (Flemming, 2005; Becker-Kristal, 2010). Under this conception, it is not the distinctiveness of *every* pair of contrasts that is relevant to the dispersion of its inventory, but only the most confusable one. That is, an inventory "is only as good as its worst contrast" (Flemming, 2005: 9). Becker-Kristal (2010) refers to this conception of dispersion where only the shortest between-category distances are maximized as the 'bottleneck approach' and I adopt his term here.

The following constraints are defined to evaluate dispersion of stop inventories based on the bottleneck approach.

(3.32) MINDIST=PLACE: Assign a violation if the closest stop pair based on place of articulation does not differ by at least one unit.

(3.33) MINDIST=VOICE: Assign a violation if the closest stop pair based on voicing does not differ by at least one unit.

(3.34) MINDIST=WIDTH: Assign a violation if the closest stop pair based on glottal spread does not differ by at least one unit.

Notice that there are two fewer MINDIST constraints under the bottleneck approach compared to the pairwise approach as there is only one constraint each defined along the place and width dimensions, compared to two separate constraints enforcing minimum distances of 1 and 2 units along these dimensions in Section 3.6. Note also that this way of implementing the bottleneck approach differs from both Becker-Kristal's (2010) and Flemming's (2005) approaches which measure acoustic distances between vowels in formant space.

3.7.3 Comparison of distance measures

We can now compare the fit to the attested inventory content of models with MINDIST constraints that reflect the three different approaches to measuring dispersion. Table 3.13 shows the results of this model comparison. The model that uses dispersion constraints based on the bottleneck approach has the lowest AIC, suggesting this approach to measuring the dispersion of stop inventories best captures their attested typology. Therefore, as is the case for vowels (Flemming, 2005; Becker-Kristal, 2010), stop inventories are generally organized to minimize the distinctiveness of the most confusable pair while distances from stops farther away are ignored.

As shown in this table, the model with dispersion constraints based on the bottleneck approach outperforms the previously optimal model with only MAXIMIZECONTRASTS and markedness constraints. Thus, there is evidence for all three functional goals of Dispersion Theory –

Model	AIC	Δ(AIC)
Bottleneck approach	21579.54	0
No dispersion constraints	21754.88	175.34
Total pairwise Euclidean distance	21756.88	177.34
Pairwise MINDIST	21759.53	178.00

Table 3.13: *Model comparison of distance measures.*

maximizing the number of contrasts, maximizing their distinctiveness and minimizing articulatory effort – in the contents of stop inventories when dispersion is based on the most confusable pair[6].

3.8 The optimal model

The optimal model is, thus, one that includes dispersion constraints based on the bottleneck approach, in addition to markedness and MAXIMIZECONTRASTS constraints. Figure 3.6 shows the model fit of the optimal model to the attested data. Constraint weights from this model are shown in Table 3.14. Constraints that do not appear in this table received weights of zero so that dropping them from the model did not deteriorate model fit.

A number of patterns are apparent in this table. In general, place and laryngeal features pattern differently. First, markedness constraints that penalize laryngeal features – *VOICE, *ASPIRATION, *CONSTRICTION, *VOICED-ASPIRATION and *VOICED-CONSTRICTION – are far more influential in shaping inventories than are those that penalize place features – *LABIAL and *VELAR. Contrasts are maximized in terms of place but not in terms of laryngeal configuration to the same extent. Moreover, distinctiveness in terms of laryngeal features is maximized, but this is not true for place. Both these patterns are likely a result of the fact that only the most common places of articulation – labial, denti-alveolar and velar – were included in the dataset, and less common places – retroflex, palatal and uvular – were excluded. By contrast, all possible laryngeal configurations were included. The different behavior of MINDIST and MAX-

[6] I also ran a model with a bottleneck approach to Euclidean distance with $r_{\min} = \min_{i,j}\sqrt{(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2}$ and a corresponding MINDIST constraint that incurred $\frac{1}{r_{\min}}$ violations. The result from this model was almost identical to the model with pairwise Euclidean distances summed together. In both models, the dispersion constraint received zero weight.

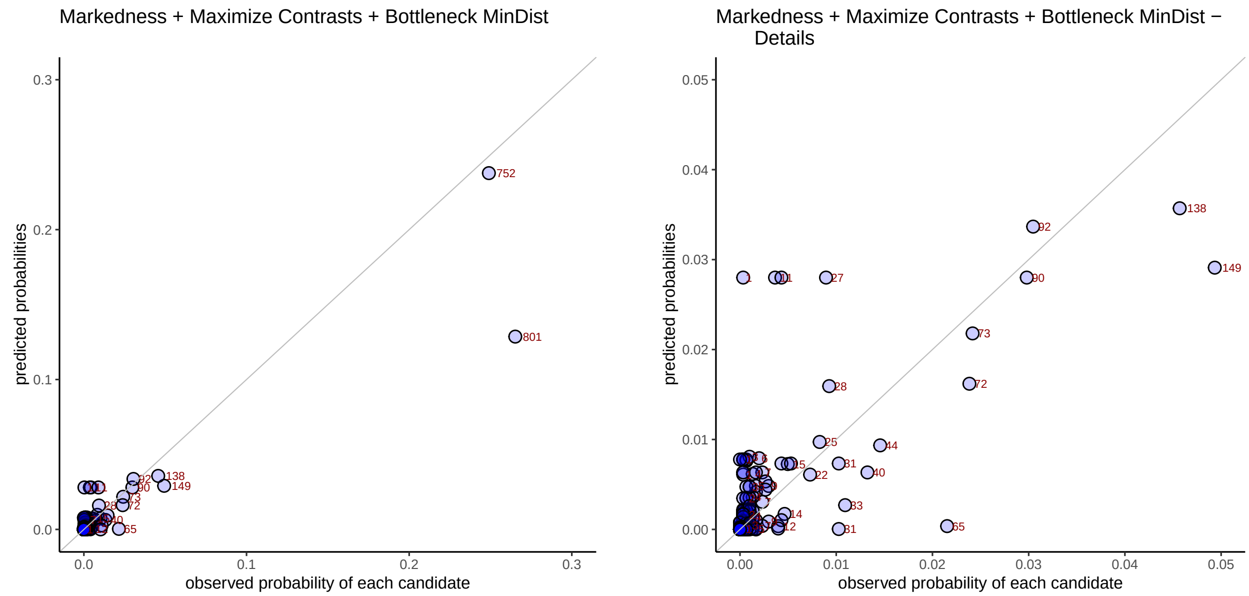


Figure 3.6: *Observed versus predicted probabilities with bottleneck dispersion constraints.*

Constraint	**Weight**
MAXIMIZECONTRASTS	1.53
MAXIMIZEPLACECONTRASTS	5.98
MAXIMIZEVOICECONTRASTS	0.27
MINDIST=PLACE	1.11
MINDIST=VOICE	1.54
MINDIST=WIDTH	0.61
*VOICE	4.17
*ASPIRATION	5.86
*CONSTRICTION	5.84
*VOICED-ASPIRATION	5.36
*VOICED-CONSTRICTION	4.54
*G	0.18
*PC	0.83
*GC	3.22
*LABIAL	1.03
*VELAR	0.86
MAXIMIZEWIDTHCONTRASTS	0
*CORONAL	0

Table 3.14: *Constraint weights in the optimal model.*

IMIZECONTRASTS constraints referring to place versus laryngeal features is, therefore, likely simply a reflection of the composition of the dataset and do not necessarily suggest that place and laryngeal features generally pattern differently.

To better understand the kinds of inventories the model prefers and disprefers, consider the top and bottom 10 inventories shown in Tables 3.15 and 3.16. Table 3.15 shows the top 10 inventories according to the model, that is those inventories that are assigned the greatest probabilities of occurrence. Note that all the inventories labeled "6" in this table are tied in that position. The column "observed frequency" shows the number of inventories of the corresponding composition and "predicted frequency" shows the number of inventories predicted by the model. Note that IPA symbols are not used to describe these inventories since there is a fair degree of abstraction in what each symbol represents, as detailed in Chapter 1. For example, /gh/ represents any stop out of the set [g̤, $\text{g}^{\text{ɦ}}$, $^{\text{ɦ}}\text{g}$].

	inventory	**observed frequency**	**predicted frequency**
1	ptk	752	718
2	ptk bdg	801	389
3	ptk bdg phthkh	138	108
4	ptk bd	92	102
5	ptk bdg bcdc	149	88
6	tk bdg	90	85
	ptk bg	27	85
	ptk dg	13	85
	pt bdg	11	85
	pk bdg	1	85

Table 3.15: *Top 10 inventories predicted by the optimal model.*

This table shows that all the top inventories preferred by the model are also preferred by real languages. The inventory /ptk/ is preferred over /ptk bdg/ since the former only violates MaximizeContrasts and MaximizeVoiceContrasts while the latter does not. Moreover, /ptk bdg/ violates *g, Mindist:place and *voice, which has a very high weight. By contrast, /ptk/ does not violate any of these constraints. The preference for /ptk bdg phthkh/ over /ptk bdg bcdc/ is driven by the preference for larger inventories, as the former inventory has fewer violations of MaximizeContrasts, despite being slightly more marked, as indicated by the slightly higher weight of *aspiration compared to *voiced-constriction. By contrast, the preference for /ptk bd/ over /tk bdg/ is driven by the small weight of the markedness constraint *g. The model is unable to distinguish between all the inventories in position 6, which all have 5 stops and the same violation profiles for markedness and dispersion. If non-binary markedness constraints were used, inventories with fewer voiced stops would be preferred, but this way of evaluating markedness violations degrades overall fit, since it penalizes /tk bdg/, which is fairly popular, more than /ptk bg/ and /ptk dg/.

Another way of interpreting the comparable status of the fairly popular inventory with a gap at /p/ with the other inventories that are just as favored is that the model predicts that these inventories are comparable in terms of the functional factors considered here. This is in line with the claim that the lack of /g/ in inventories is a true reflection of markedness, while the lack of /p/ is a genetic or areal artefact (Maddieson, 1984). Moreover, a common diachronic pathway that leads to the loss of /p/ from inventories is via sound change to /f/. Therefore, under the functional constraints defined here, all the inventories in the 6^{th} position are in fact comparable. More broadly, it is possible that those inventories whose frequencies the model predicts poorly may be explained by mechanisms of sound change and by expanding the analysis to include fricatives and other classes of sounds.

Finally, it should be noted that the one language that has /p k b d g/ does have the retroflex stop /ʈ/. Its stop inventory, /p ʈ k b d ɖ g/, is therefore not accurately represented by the classification scheme used for coronal stops here.

Table 3.16 shows the bottom 10 inventories that are attested in the languages of the world. The model assigns the lowest probabilities to inventories whose observed probabilities are zero, which are not shown here. That is, the languages predicted by the model to be the least probable are those that do not occur in the world's languages. Table 3.16, thus, shows the lowest probabilities assigned by the model to real inventories that are attested among real languages. Notice that in contrast to Table 3.15, this table uses IPA symbols, since the inventories occur in specific languages. The languages which have the associated inventories are shown in the fourth column. Note that for Maidu, the corresponding inventory appears in two separate descriptions.

As seen in this table, the model succeeds at capturing the rarity of these inventories, which only occur in one-off languages. Several inventories in this table are missing the neutral se-

	inventory	observed frequency	Language
1	/p^h t^h k^h p' t' k' b̤ d̤ g̤/	1	Xhosa
2	/t^h k^h k' b d d'/	1	Jibbali
3	/p^h t^h k^h t' k' ɓ/	1	Jacaltec
4	/t t^h t'/	1	Hupa
5	/p^h t^h k^h p' t' k' ɓ ɗ/	2	Maidu (2)
6	/t^h k^h t' k' d g/	1	Eyak
7	/p t k^h p' t' k' b g ^{ɦ}d/	1	Günün Yajich
8	/p^h t^h k^h p' k' b d g ɗ/	1	Gamo
9	/t k t^h ɓ ɗ/	1	Southern Vietnamese
10	/p t k t^h ɗ/	1	Hanoi Vietnamese

Table 3.16: *Bottom 10 real inventories predicted by the optimal model.*

ries (#1, 2, 3, 5, 6, 8), meaning they have more markedness violations compared to inventories of comparable sizes, which share their violations of MAXIMIZECONTRASTS. All except #1 have gaps, either in the place or laryngeal dimensions, incurring violations of the specific MAXIMIZECONTRASTS constraints. Thus, through the dispersion-theoretic constraints and their interactions, the model is able to capture the intuition that the inventories in Table 3.16 are less probable.

3.9 Implications

3.9.1 Structural properties

In addition to improving fit to the attested data, the optimal model has another desirable characteristic – it predicts the geometric properties of stop inventories. As evidenced by the top and bottom 10 inventories, the interactions of the dispersion-theoretic constraints predict larger inventories balanced in terms of markedness. Thus, it predicts symmetry and economy to be properties of optimal inventories. Symmetry quantifies the degree to which the feature combinations in a stop inventory are balanced rather than gapped. Thus, both /p t k ph th kh b d g bh dh gh/ and /p t k ph th kh b d g/ are perfectly symmetrical, although only the former, as we will later see, is perfectly economical. Symmetry is calculated as the ratio of the number of actually occurring combinations of place and laryngeal setting to the total number of logically possible combinations given the set of unique places and laryngeal settings present in the inventory. A perfectly symmetrical inventory is such that if a feature like voicing occurs at one place of articulation, it occurs at all places of articulation present in that inventory. Therefore, both /p ph b bh k kh g gh/ and /p ph b k kh g/ have symmetry scores of 1, but an inventory with a gap, say /t k b d g/, has a symmetry score less than 1 (5/6 to be exact, since it has 5 stops when the combinations of already present place and laryngeal features predicts 6 possibilities) since it is missing /p/.

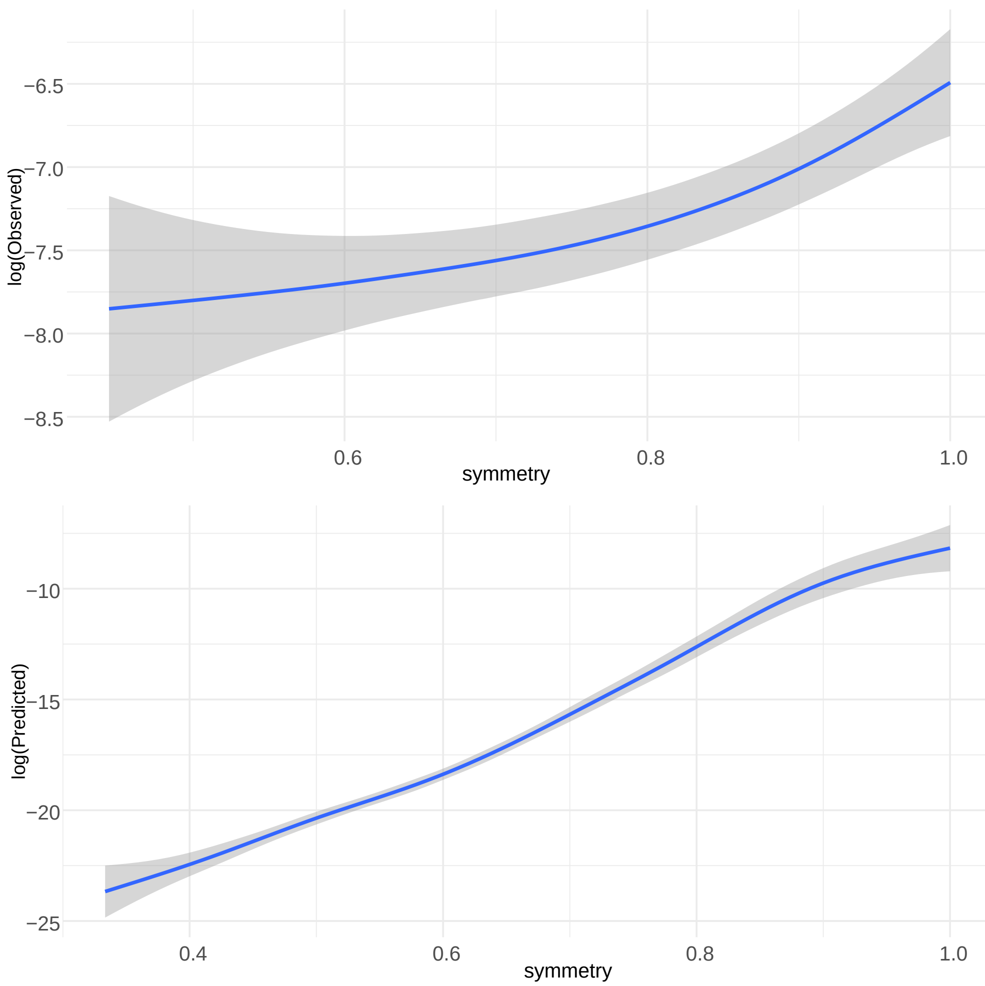


Figure 3.7: *Symmetry versus observed (top) and predicted (bottom) inventory frequencies. Note the different scales.*

Figure 3.7 shows the relationship between symmetry and observed (top panel) and predicted (bottom panel) inventory frequency. Symmetry is a moderately correlated with observed frequency ($\rho = 0.41, p << 0.001$) and strongly correlated with predicted frequency

($\rho = 0.58, p << 0.001$). This pattern suggests that the model expects symmetric inventories to occur more frequently than they actually do, suggesting it overestimates the strength of the correlation between symmetry and inventory frequency.

A similar pattern is observed with inventory economy. The plots in Figure 3.8 show the relationship between economy and observed (top panel) and predicted (bottom panel) inventory frequency. Here economy is calculated as the ratio of the number of stops found in

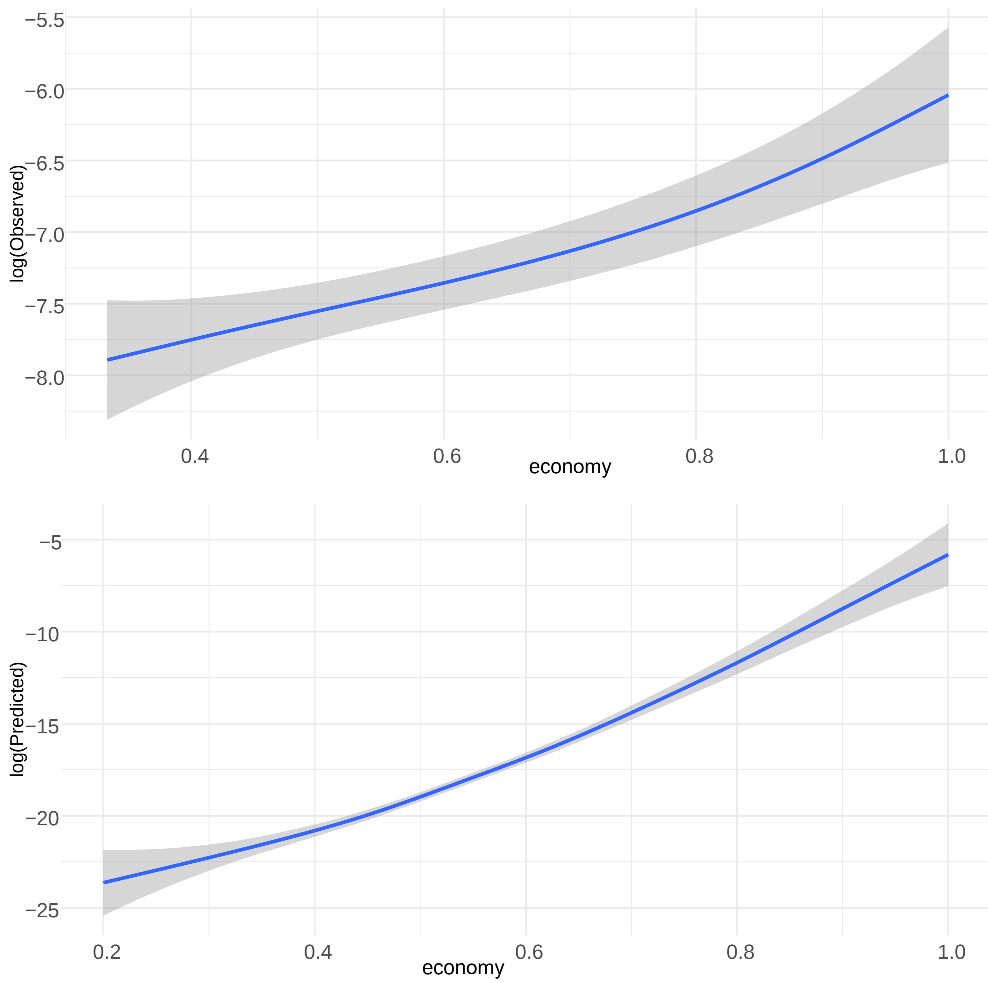


Figure 3.8: *Economy versus observed (top) and predicted (bottom) inventory frequencies. Note the different scales.*

an inventory and the maximum number of stops possible based on the features it employs. Thus, it quantifies how many distinct stops an inventory has relative to the maximum number that could be generated from its available features. This definition is based on Clements (2003), who defines the economy index (E) of an inventory with S speech sounds characterized by F features as

(3.35) $E = S/F$

The features relevant to stop contrasts include some whose combinations are incompatible, like [spread] and [constricted], for instance. Therefore, to avoid penalizing inventories that have more incompatible features just as much as those with compatible features, the denominator in 3.35 was changed to the maximum number of stops that can possibly be derived from the available features. In contrast to Clements' definition, this reformulation of economy yields a number between 1 and 0, making its interpretation more intuitive. Note that this way of calculating economy predicts exactly one perfectly economical inventory for a set of features. For instance, assuming privative features, for the features [voice], [spread] and [labial], the only perfectly economical inventory is /p ph b bh/. The perfectly symmetrical inventory /p ph b/ is uneconomical under this formulation of economy since it does not maximally employ [voice] and [spread] to produce voiced aspirated stops.

As seen in Figure 3.8, both observed and predicted inventories generally tend to be more economical. There is a weak correlation between observed frequency and economy ($\rho = 0.17$) and a stronger correlation between predicted frequency and economy ($\rho = 0.53$). This indicates that the model overpredicts the frequency of economical inventories. Therefore, although economy, as defined in 3.35, is predicted to emerge via the interaction of markedness, MAXIMIZECONTRASTS and MINDIST constraints, real inventories exhibit a weaker tendency towards economy thus defined.

Although the model overestimates the frequency of symmetrical and economical inventories compared to the attested frequencies, it accurately predicts these structural properties to hold for the optimal stop inventories at each size. Table 3.17 shows the optimal inventories predicted by the model at each size, along with the most attested inventory for that size[7].

This table shows that the model makes incorrect predictions for only four out of the sixteen

[7]The two languages in PHOIBLE with only one stop, /t/, are Kwaio (Austronesian, Solomon Islands) and Mwotlap (Austronesian, Vanuatu), whose names, ironically, contain /k/ and /p/. These languages also have pre-nasalized and doubly articulated stops, which may be represented by orthographic $< k >$ and $< p >$ in their names.

size	predicted optimal	most attested	frequency of predicted optimal
1	t	t	2
2	tk	tk	15
3	ptk	ptk	752
4†	ptk b	ptk b	31
5	ptk bd	ptk bd	92
6	ptk bdg	ptk bdg	801
7*	ptk bd bcdc	ptk bdg bc (40)	2
8	ptk bdg bcdc	ptk bdg bcdc	149
9	ptk bdg phthkh	ptk bdg phthkh	138
10*	ptk phthkh bd bcdc	ptk pctc bdg bcdc (3)	2
11*	ptk phthkh bdg bcdc	ptk pctckc bdg bcdc (4)	3
12	ptk phthkh bdg bhdhgh	ptk phthkh bdg bhdhgh	72
13	ptk thkh bdg bhdhgh bcdc	ptk thkh bdg bhdhgh bcdc	1
14*	ptk phthkh pctckc bdg bcdc	ptk phthkh bdg bhdhgh bcdc (2)	1
15	ptk phthkh bdg bhdhgh bcdcgc	ptk phthkh bdg bhdhgh bcdcgc	2
16	ptk phthkh pctckc bdg bhdhgh bc	ptk phthkh pctckc bdg bhdhgh bc	1

Table 3.17: *Best inventory predicted correctly by the model and the most commonly attested inventories at each size. The numbers in parentheses represent the attested frequencies where predictions do not match the attested inventories.*

inventories. These sizes are all marked with * in the first column in Table 3.17. Note also that the model predicts three different inventories to be equally probable at size 4, one of which is the most commonly attested inventory at this size.[8]. In these cases, where the model makes incorrect predictions, the attested frequency of the most common inventory at each size is shown in parentheses in the third column. Notice that other than size 7, the sizes where the model makes the wrong predictions are rarely attested and model predictions are off by only 1 inventory. Moreover, sizes 4 and 7 are sizes which exhibit a wide range of variation in their composition as shown in Chapter 1.

Consider the distribution of inventory sizes from Chapter 1, shown in Figure 3.9 below. In this figure, the orange line shows the number of possible inventories, the blue line shows the number of attested inventories and the green line shows the relative diversity of inventories at each size. For instance, at size 6, the blue curve shows a high peak since over 1000 inventories have 6 stops, 801 of which have /p t k b d g/. The green curve shows that there are only 36

[8]The two other inventories that are predicted to be optimal by the model are /ptk d/ (16 attested inventories) and /tk bd/ (13 attested inventories).

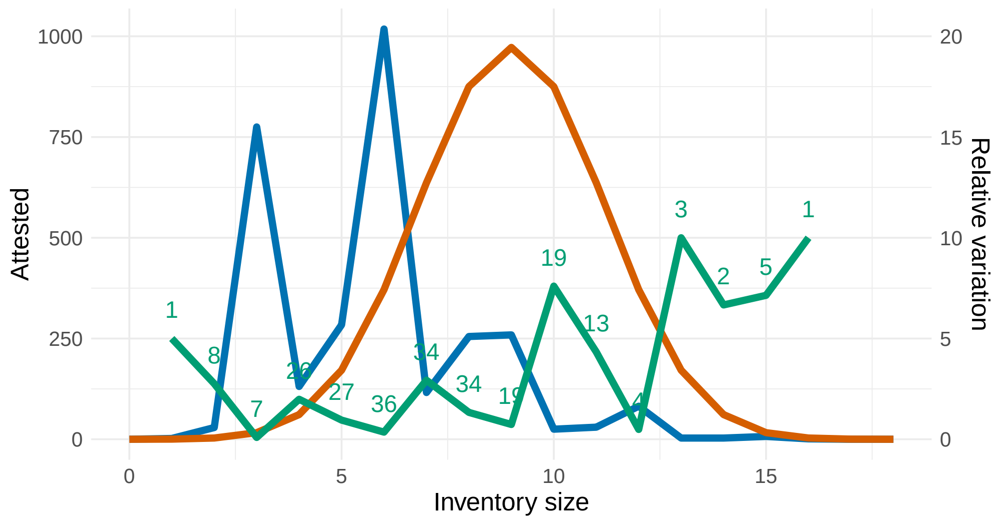


Figure 3.9: *The distribution of stop inventory sizes.*

unique inventories at this size, which is relatively low compared to how frequent inventories of this size are. Therefore, despite being a popular size, the makeup of inventories of size 6 is relatively uniform. This suggests that the high frequency of size 6 inventories is driven mainly by the high frequency of inventories that contain /p t k b d g/. The same is true of /p t k/. This pattern of relative diversity is reversed at sizes 4, 7 and 10, which are relatively rare but have varied composition. For instance, there are only 116 inventories of size 7, but there are 34 unique inventories of this size, almost just as many as those of size 6. Therefore, the model makes incorrect predictions in exactly those cases where attested inventories show a great deal of variation and there is no clear optimal candidate inventory. Moreover, the inventories predicted to be optimal by the model are attested and typologically plausible.

Investigating the contents of the predicted optimal inventories at sizes 3, 6, 9 and 12 reveals that the optimal inventories which can employ their features maximally indeed do so. Thus, the property of being economical, which is a prominent feature of stop inventories, is emergent in this model. The interaction of phonetically defined constraints predicts the optimality of a system that employs its features economically. Therefore, economy need not be modeled directly in the MaxEnt model, as its effects are emergent from the interaction of the dispersion-theoretic factors.

The preference for economical inventories is further illustrated by the simplified example tableaux shown in 3.18 and 3.19. The first tableau shows inventories of size 3 and their violations of the various dispersion-theoretic constraints. The second shows inventories of size 12. Constraints whose violation profiles do not differ between the candidate inventories are not shown for reasons of space. Note that the math shown here does not exactly match the math in the model, which has many more rows. Nevertheless, the relative magnitudes of the predicted probabilities remain comparable between these tableaux and the model.

	weight	MAXIMIZEPLACE 5.98	MAXIMIZEVOICE 0.27	MINDIST=PLACE 1.11	MINDIST=WIDTH 0.61	*VOICE 4.17	*ASPIRATION 5.86	*CONSTRICTION 5.84	*VELAR 0.86	$\mathcal{H}$	eH	Z	P
a.	p-t-k		1		1				1	1.74	0.176	0.179	0.98
b.	b-t-k				1	1			1	5.64	0.004	0.179	0.02
c.	p-ph-b	2		1	1	1	1			23.71	5.05×10^{-11}	0.179	≈ 0
d.	p-tc-kh		1				1	1	1	12.83	2.7×10^{-6}	0.179	≈ 0

Table 3.18: *Example tableau showing stop inventories of size 3.*

First consider the inventories of size 3. Candidate (a) /p t k/ violates MAXIMIZEVOICE-CONTRASTS since it does not utilize the voicing dimension to create contrasts; it also violates MINDIST=WIDTH since none of its stops are separated along the glottal width dimension; and it violates *VELAR on account of having /k/. All three of these constraints have relatively low weights and candidate (a) ends up with a low harmony score, $\mathcal{H}$ of 1.74.

In contrast to candidate (a), (b) does not violate MAXIMIZEVOICECONTRASTS as it uses the voicing dimension, but to do that it has to violate *VOICE, which has a much larger weight. This results is a greater $\mathcal{H}$ of 5.64. Candidate (c) avoids incurring a violation of *VELAR but to preserve the total number of stops in the inventory, it has to violate a number of other markedness constraints. Thus, it ends up with a much higher $\mathcal{H}$ of 23.71. Finally, candidate (d) does not violate MINDIST=WIDTH like (a) does, but its violations of the markedness constraints penalizing aspiration and constriction yield a harmony of 12.83.

The probabilities for these candidates calculated by the MaxEnt algorithm show that candidate (a) is vastly preferred over the other candidates. Candidate (b) is a distant second, and candidates (c) and (d) are predicted to be almost impossible. This matches the typology well – there are 752 inventories in PHOIBLE that have /p t k/, 9 that have /b t k/ and none that have /p ph b/ or /p tc kh/.

Thus, the least marked inventory which does not have redundant features is the most probable, since to maintain the number of stops at three, any changes made to /p t k/ result in violations of higher-weighted constraints.

Next consider inventories of size 12, shown in 3.19. This tableau uses a shorthand to denote the five candidates to save space. "T" represents the complete neutral series, "T^h" represents the complete voiceless-unaspirated series etc. Thus, candidate (a) represents the fully economical inventory /p t k ph th kh b d g bh dh gh/. Likewise, candidate (b) is fully economical as it uses the features [voice] and [constriction] maximally. Candidates (c) and (d), which contain redundant features, are not fully economical, and candidate (e), which represents /p t k pc tc kc b d g bc dc kh/, is also uneconomical. It avoids the highly marked /gc/ and instead has the relatively less marked /k^h/. Once again, constraint violations that do not differ between the candidates are not shown.

		*ASPIRATION	*CONSTRICTION	*V-ASPIRATION	*V-CONSTRICTION	*PC	*GC	$\mathcal{H}$	eH	Z	P
	weight	5.86	5.84	5.36	4.54	0.83	3.22				
a.	T-Th-D-Dh	1		1				11.22	1.34×10^{-5}	1.77×10^{-5}	0.75
b.	T-Tc-D-Dc		1		1	1	1	14.43	5.41×10^{-7}	1.77×10^{-5}	0.03
c.	T-Th-Tc-D	1	1			1		12.51	3.69×10^{-6}	1.77×10^{-5}	0.21
d.	T-D-Dh-Dc	1		1	1	1	1	18.98	5.72×10^{-9}	1.77×10^{-5}	≈ 0
e.	T-Tc-D-bc-dc-kh	1	1		1	1		16.24	8.85×10^{-8}	1.77×10^{-5}	0.01

Table 3.19: *Example tableau showing stop inventories of size 12.*

First notice that the candidates differ only in terms of violations of the markedness constraints. As inventories get larger, they have to utilize more of the stop space. Large inventories must exhaustively use the place dimension, since MAXIMIZEPLACECONTRASTS is the highest weighted constraint in the model. However, since the MINDIST constraints are based on the bottleneck as formulated in (3.32)-(3.34), a glottal width separation of two units does not provide any advantage over a separation of one unit in terms of dispersion. Thus, inventories can exhaustively use either the glottalic or the pulmonic dimension instead of having to use both to be maximally dispersed.

As seen in the tableau above, the pulmonic inventory in candidate (a), which exhaustively uses place, voicing and aspiration, has the lowest harmony score as it only violates two constraints. Its glottalic equivalent, candidate (b) has a higher $\mathscr{H}$ since it violates two markedness constraints that penalize specific glottalic stops, *PC and *GC, in addition to having comparable violations of the markedness constraints that penalize voicing and a glottal articulation. This is also true of candidate (c), which has a similar $\mathscr{H}$ to (a) plus a violation of *PC. This leads to a slightly greater harmony score for candidate (c). Candidate (d), which avoids the use of any glottal articulations, is harmonically bound by candidate (a). Finally, candidate (e), which avoids a violation of the highly weighted *GC, is still forced to violate other markedness constraints and thus incurs a higher value of $\mathscr{H}$.

The probabilities predicted by the model for these inventories match the typological data well. There are 72 inventories that have /p t k ph th kh b d g bh dh gh/, 8 that have /p t k ph th kh pc tc kc b d g/ and one that has /p t k pc tc kc b d g bc dc gc/.

The emergence of economy is also facilitated by the nature of the markedness constraints. As discussed in 3.4, markedness violations are formulated as binary, such that an inventory that uses a marked articulation twice incurs the same number of violations as one that uses the same marked articulation only once. Thus, the use of the same feature is not penalized proportionally to the number of distinct segments that employ it, promoting economy.

Thus, the structural properties, symmetry and economy, which are prominent features of the most commonly attested stop inventories (Dunbar and Dupoux, 2016; Nikolaev and Grossman, 2020), need not be modeled directly into the grammar at the same level as the phonetically-defined constraints. These properties can instead emerge via the interaction of the dispersion theoretic constraints, particularly the tendency towards more contrasts and less effort, combined with the bottleneck approach to distinctiveness.

3.9.2 Implicational relationships

The model architecture presented in the previous sections can be used to assess implicational relationships between different laryngeal categories. Specifically, we can test if the presence of voiced aspirates implies the presence of voiceless aspirates by assessing violations of markedness constraints that follow from assuming this implicational relationship. The fit of this model can then be compared to one which does not assume an implicational relationship between voiced aspirated and voiceless aspirated stops. Similarly, we can test if a similar relationship holds for the glottalic stops. That is, we can test if the presence of voiced constricted stops implies the presence of voiceless constricted stops.

All the models presented so far assume implicational relationships we found support for in Chapter 2. Here, I will consider the following possibilities[9]:

(3.36) Voiced aspirated (Dh) stops violate *VOICE on account of being voiced, *ASPIRATION on account of being aspirated and *VOICED-ASPIRATION on account of being voiced-aspirated.

(3.37) Dh stops violate *VOICED-ASPIRATION and *VOICE, but not *ASPIRATION.

(3.38) Voiced constricted (Dc) stops violate *VOICE on account of being voiced, *CONSTRICTION on account of being constricted and *VOICED-CONSTRICTION on account of being voiced-constricted.

(3.39) Dc stops violate *VOICED-CONSTRICTION and *VOICE, but not *CONSTRICTION.

In terms of the implicational relationships evaluated in Chapter 2, (3.36) represents the implicational relationship between voiced-aspirated and voiceless-aspirated stops and (3.39) represents the lack of such an implicational relationship between voiced-constricted and voiceless-constricted stops. (3.37) and (3.38) represent the other two possibilities respectively – that the

[9]All other possible combinations were also evaluated and none were found to fit the data better than the ones reported here.

presence of Dh stops does not depend on the presence of Th stops, and the presence of Dc stops does depend on the presence of Tc stops. We will therefore evaluate the four models in Table 3.20. Model 1 is consistent with the findings in Chapter 2 while Models 2-4 consider alternate possibilities.

Model	**Dh-Th relationship**	**Dc-Tc relationship**
Model 1	Dh depends on Th	Dc independent of Tc
Model 2	Dh depends on Th	Dc depends on Tc
Model 3	Dh independent of Th	Dc independent of Tc
Model 4	Dh independent of Th	Dc dependent on Tc

Table 3.20: *Summary of models tested.*

The constraint violations of a few illustrative candidate inventories are shown in Tableaux 3.21-3.24. Note that only those markedness constraints relevant to laryngeal settings are shown. Since Model 1 assumes the independence of Dc from Tc, the presence of /bc/ and /dc/ in candidate (a) violates *VOICED-CONSTRICTION but not *CONSTRICTION in Tableau 3.21. By contrast, since Model assumes the dependence of Dh on Th, the presence of /bh/ in candidate (d) violates both *VOICED-ASPIRATION and *ASPIRATION.

	*VOICE	*ASP	*CONS	*V-ASP	*V-CONS	FREQ
a. p-t-k-b-d-g-bc-dc	1				1	149
b. p-t-k-b-d-g-ph-th-kh-bh-dh-gh	1	1		1		72
c. p-t-k-ph-th-kh-pc-tc-kc		1	1			33
d. p-k-g-bh-bc-gc	1	1		1	1	
e. ...	...	...	...	...	...	...

Table 3.21: *Example part-tableau showing markedness violations in Model 1.*

Since Model 2 assumes that Dh and Dc depend on Th and Tc respectively, as seen in Tableau 3.22, (a) violates both *VOICED-CONSTRICTION and *CONSTRICTION just as (d) violates *VOICED-ASPIRATION and *ASPIRATION.

Since Model 3 assumes the independence of both Dh and Dc, candidates (a) and (d) no longer violate *CONSTRICTION and *ASPIRATION respectively, but only *VOICED-CONSTRICTION and *VOICED-ASPIRATION, as seen in Tableau 3.23.

	*VOICE	*ASP	*CONS	*V-ASP	*V-CONS	FREQ
a. p-t-k-b-d-g-bc-dc	1		1		1	149
b. p-t-k-b-d-g-ph-th-kh-bh-dh-gh	1	1		1		72
c. p-t-k-ph-th-kh-pc-tc-kc		1	1			33
d. p-k-g-bh-bc-gc	1	1	1	1	1	
e. ...	...	...	...	...	...	...

Table 3.22: *Example part-tableau showing markedness violations in Model 2.*

	*VOICE	*ASP	*CONS	*V-ASP	*V-CONS	FREQ
a. p-t-k-b-d-g-bc-dc	1				1	149
b. p-t-k-b-d-g-ph-th-kh-bh-dh-gh	1	1		1		72
c. p-t-k-ph-th-kh-pc-tc-kc		1	1			33
d. p-k-g-bh-bc-gc	1			1	1	
e. ...	...	...	...	...	...	...

Table 3.23: *Example part-tableau showing markedness violations in Model 3.*

Finally, Model 4 assumes the opposite pattern from the findings in Chapter 2, with Dc depending on Tc, but Dh being independent from Th, thus generating the constraint violations shown in Tableau 3.24.

	*VOICE	*ASP	*CONS	*V-ASP	*V-CONS	FREQ
a. p-t-k-b-d-g-bc-dc	1				1	149
b. p-t-k-b-d-g-ph-th-kh-bh-dh-gh	1	1		1		72
c. p-t-k-ph-th-kh-pc-tc-kc		1	1			33
d. p-k-g-bh-bc-gc	1			1	1	
e. ...	...	...	...	...	...	...

Table 3.24: *Example part-tableau showing markedness violations in Model 4.*

Model	**AIC**	**Δ(AIC)**	**Log-likelihood**
Model 1	21579.54	0	-10770.77
Model 3	21609.92	30.38	-10785.96
Model 2	22872.65	1293.11	-11417.32
Model 4	22964.62	1385.08	-11463.31

Table 3.25: *Model comparison testing implicational relationships. Models are ordered from best to worst.*

Table 3.25 shows results from the model comparison. The AIC and log-likelihood of each model are reported. As seen from this table, Model 1 outperforms all the other models. That is, the model that incorporates the findings from Chapter 2 best fits the distribution of attested and unattested stop inventories. Therefore, in terms of markedness violations, inventories that contain voiced-aspirated stops are harmonically bounded by those that contain only voiceless-aspirated stops. However, inventories that contain voiced-constricted stops are not similarly bounded by those with voiceless-constricted stops.

3.9.3 The status of implosives

We can similarly employ model comparison to probe why the implicational relationship does not hold between glottalic stops. In particular, we can probe the validity of classifying implosives as voiced-constricted stops. Implosives have been shown to phonologically pattern like sonorants in some languages and like stops in others and even show mixed behavior in some (Sande and Oakley, 2023). Based on the phonetic and phonological properties of Ikwere implosives, Clements and Osu (2002) propose that implosives are non-obstruents and do not bear the feature [constricted glottis]. Following this proposal, we can remove implosives from the voiced constricted (Dc) category, leaving only glottalized and creaky voiced stops in this category. We can subsequently implement a MaxEnt model with this categorization of Dc stops and compare it to the model from previous sections which includes implosives. Moreover, we can change the violations of *CONSTRICTED to either assume an implicational relationship between voiced constricted and voiceless constricted stops, or assume that no such relationship holds.

Results from such a model comparison are shown in Table 3.26. As seen in this table, when

Model	**AIC**	**Δ(AIC)**	**Log-likelihood**
Without implosives; Dc implies Tc	17023.24	0	-8490.62
Without implosives; Dc does not imply Tc	17680.98	657.74	-8819.49

Table 3.26: *Model comparison with and without implosives.*

implosives are removed from the model, assuming an implicational relationship between Dc stops (which do not include implosives) and Tc stops improves fit. This is the opposite pattern to the findings in Section 3.9.2, which showed that assuming the lack of an implicational relationship improved the model fit when Dc stops included implosives. As the removal of implosives leads to the removal of inventories that contain them, the datasets being modeled in the previous section and this one are different. Therefore, they cannot be directly compared to each other for a direct test of the claim that implosives are non-stops (Clements and Osu, 2002). However, since they are the only stops that do not follow the implicational relationships evaluated here, the model comparison finds support for the broad claim that implosives are unlike other stops. This suggests that glottalic stops pattern just like pulmonic stops in terms of feature economy, when implosives are excluded from the set of stops.

3.10 Conclusion

This chapter has presented a probabilistic OT model of stop inventories, showing that a Maximum Entropy Harmonic Grammar accurately predicts attested inventories. Moreover, the structural properties SYMMETRY and ECONOMY are emergent under this approach, despite constraints modeling only functional pressures. Taken together these results suggest that the interaction of phonetically-based constraints in a probabilistic grammar can explain much of the structure found in the world's stop inventories. The explanation for cases in which model predictions differ from the attested typology must be found in a diachronic typological approach to phonology (cf. Easterday and Bybee, 2023). Additionally, we found broad typological evidence that implosives do not pattern like other stops, consistent with the proposal in Clements and Osu (2002). We also found evidence for the bottleneck approach over other methods of measuring dispersion in stop inventories.

Further investigation of the contents of inventories predicted by the model presented in this chapter can inform theories of sound change under the view that static inventories are the product of diachronic change. In Chapter 4, I demonstrate how the model presented in this chapter, when enriched with constraints assessing the changes to the contents of an inventory, can be used for hypothesis testing of ancestral states.

CHAPTER 4

Hypothesis testing of ancestral states

We make our own history; we do not make it under circumstances of our own choosing, but under circumstances existing already, given and transmitted from the past.

– KARL MARX, The Eighteenth Brumaire of Louis Bonaparte

4.1 Introduction

This chapter demonstrates how the typological model developed in Chapter 3 can be used for testing hypotheses regarding the ancestral state of a set of languages assumed to have descended from a common ancestor. I present a case study illustrating the use of this model to test the stop inventory of Proto-Indo-European (PIE). The PIE stop inventory has been the subject of much debate as the most commonly accepted result of reconstruction yields the synchronic inventory *T-D-Dh, which is typologically exceedingly rare, if not entirely unattested. In this chapter I adopt a typological approach to this question that goes beyond determining whether or not the proposed PIE inventory is attested. Instead, I test how likely this inventory is to have been the initial state that generated the observed stop typology of the modern Indo-European languages. This idea that attestation is not the same as likely or common is taken up again in Chapter 6.

As a preview of the results, I find that the standard reconstruction is not the most likely initial state. It is, however, more likely as an initial state than a different stop inventory that is typologically sensible. Therefore, whether or not a proto inventory is typologically attested is a poor metric by which to judge its plausibility as a synchronic inventory. Although I find that *T-D-Dh is not the most likely synchronic PIE inventory, a more nuanced approach to this question reveals that the standard reconstruction, despite its typological rarity, is a better predictor of modern IE stop typology than an inventory that is typologically attested but ill-suited as an initial state for this set of languages.

4.2 Background

Proto-Indo-European (henceforth *PIE*) is reconstructed to have voiced-aspirated stops, since the daughter languages exhibit sound correspondences between voiced and aspirated stops. Moreover, the vast majority of languages in the Indo-Aryan (or Indic) sub-branch have voiced aspirates, as did the language these languages have descended from, Sanskrit. Based on this, early reconstruction of the PIE stop inventory was the four-way system (e.g., Brugmann, 1897), shown in Table 4.1[1]. I will henceforth refer to this reconstruction as the 'four-way' reconstruction.

p	p^h	b	b^h
t	t^h	d	d^h
k	k^h	g	g^h

Table 4.1: *The classical reconstruction of PIE stops as the four-way system.*

Of the four laryngeal categories in Table 4.1, voiceless-unaspirated (T), voiced-unaspirated (D) and voiced-aspirated (Dh) stops have distinct reflexes in different branches of Indo-European. Consider the Greek, Latin and Sanskrit forms in Table 4.2 from Comrie (2014). Voiceless-unaspirated and voiced-unaspirated stops straightforwardly emerge as T and D in all three languages. Voiced-aspirates are retained only in Sanskrit; they turn into *voiceless*-aspirated stops in Greek, and in Latin as voiceless fricatives or voiced-unaspirated stops depending on place of articulation and position in the word. Such reflexes are harder to establish for

[1] Note that, as in the rest of this dissertation, I only consider the three places of articulation shown in this table, and ignore the palatals and labialized velars in PIE.

	Greek	Latin	Sanskrit	gloss
*t	*treĩs*	*trēs*	*trayas*	'three'
*d	*dómos*	*domus*	*damas*	'house'
*b^h	*phérō*	*ferō*	*bharāmi*	'I carry'

Table 4.2: *Reflexes of T, D and Dh in Greek, Latin and Sanskrit.*

the voiceless-aspirated stops. The strongest evidence for this laryngeal type comes from the Indo-Iranian languages. Voiceless-aspirated stops were, therefore, proposed to be an innovation specific to the Indo-Iranian languages, and the PIE inventory was revised to exclude them (Bally and Gautier, 1922). The resulting inventory was, thus, *T-D-Dh, which I will refer to as the 'standard' reconstruction.

It was quickly noted, however, that the standard inventory is typologically unusual, with Jakobson (1957: 23) proclaiming "theories operating with the three phonemes /t/-/d/-/dh/ in Proto-IE must reconsider the question of their phonemic essence." In response to this criticism, the PIE inventory was revised yet again. I will address two influential revisions here. The 'Glottalic Theory' (Gamkrelidze and Ivanov, 1995; Hopper, 1973) reinterpreted *T-D-Dh as *T(h)-Tc-D(h), where D stops are instead ejectives, with T (remaining unchanged) and D (coming from Dh) stops having their aspirated counterparts as allophones. Several revisions have been made to this initial proposal, but I will limit my discussion to the original inventory, *T-Tc-D, which is typologically well-formed and attested. A different approach to revising the standard reconstruction has been to return to the classical four-way reconstruction (Elbourne, 1998)[2].

In what follows I evaluate the plausibility of three reconstructions –

(4.1) PIE_1, the standard reconstruction – *T-D-Dh

(4.2) PIE_2, the four-way reconstruction – *T-Th-D-Dh

(4.3) PIE_3, the glottalic theory reconstruction – *T-Tc-D

Note that throughout the rest of the chapter I leave out '*' to mark reconstructed forms of PIE

[2]Beyond typological arguments, Elbourne (1998) reexamines voiceless-aspirates in PIE and provides independent evidence for their existence.

in order to avoid confusion with constraints whose names start with the same symbol.

4.3 Method

Typological evidence brought to bear on the plausibility of the T-D-Dh series as the PIE stop inventory consists of whether or not such inventories are attested in the world's languages (Comrie, 2014; Stuart-Smith, 2004 and references therein). In PHOIBLE, there are only two languages that have Dh stops without also having Th stops – Mfumte and Ngishe, both Atlantic-Congo languages spoken in Cameroon[3]. In the past, the mere existence of such languages has been cited as evidence for the typological well-formedness of the standard reconstruction.

I propose an alternate evaluation metric that tests how likely a candidate PIE inventory is as the initial state, given the variation observed in the stop systems of Indo-European (IE) descended from PIE. It is conceivable that the variation in the stop inventories of the later IE languages is more likely to have originated from an unstable system like T-D-Dh rather than a stable one like T-Th-D-Dh.

The following method was used to evaluate the plausibility of various proposed PIE inventories. First, a dispersion-theoretic model of the kind introduced in Chapter 3 was built. This was used to model the predicted frequencies of all the non-Indo-European stop inventories in PHOIBLE (Moran and McCloy, 2019). A list of the languages excluded from this model is shown in Table 4.3. Note that this table lists languages, not inventories, so that a language that has several separate inventories listed in PHOIBLE is shown only once. I will refer to this model as MODEL 0. The constraint weights from MODEL 0 were taken to represent general typological tendencies in stop inventories.

Next, for each potential PIE inventory, a model was built to account for the contents of IE stop inventories descended from PIE, subject to the tendencies captured by MODEL 0. These descended inventories included all the inventories in Table 4.3 plus Hittite. In the remainder

[3] Note, however, that unlike PIE, both these languages only have stray voiced aspirates – /bʰ/ in Mfumte and /gʰ/ in Ngishe – to the exclusion of the other places. These are, therefore, quite different to the fully economical systems like PIE and the Indic languages.

	Language		Language		Language		Language
1	Afrikaans	34	German	67	Marwari	100	Saurashtra
2	Albanian	35	Greek	68	Mazanderani	101	Scots
3	Angolar	36	Gujarati	69	Mewati	102	Scottish Gaelic
4	Aragonese	37	Gujari	70	Mirandese	103	Serbian
5	Armenian	38	Halbi	71	Morisyen	104	Shina
6	Assamese	39	Hindi	72	Munji	105	Sindhi
7	Asturian	40	Icelandic	73	Naga Pidgin	106	Sinhala
8	Awadhi	41	Irish	74	Nepali	107	Slovak
9	Bakhtiari	42	Ishkashimi	75	Nigerian Pidgin	108	Slovenian
10	Balochi	43	Italian	76	Northern Frisian	109	Spanish
11	Belize Kriol	44	Jamaican Creole	77	Norwegian	110	Swedish
12	Bengali	45	Jaunsari	78	Occitan	111	Swiss German
13	Bhili	46	Jerriais	79	Odia	112	Tajik
14	Bhojpuri	47	Judeo-Tat	80	Ormuri	113	Tok Pisin
15	Breton	48	Kabuverdianu	81	Ossetian	114	Torwali
16	Brokskat	49	Kangri	82	Pashto	115	Ukrainian
17	Bukharic	50	Karipúna Creole	83	Persian	116	Upper Guinea Crioulo
18	Bulgarian	51	Kashmiri	84	Phalura	117	Upper Saxon
19	Bundeli	52	Krio	85	Pichi	118	Upper Sorbian
20	Catalan	53	Kumaoni	86	Plautdietsch	119	Urdu
21	Czech	54	Kumzari	87	Polish	120	Vaagri Booli
22	Danish	55	Kurdish	88	Portuguese	121	Varli
23	Dhivehi	56	Ladino	89	Principense	122	Vlax Romani
24	Dimli	57	Lambadi	90	Punjabi	123	Wakhi
25	Domari	58	Ligurian	91	Rajbanshi	124	Waneci
26	Dutch	59	Lithuanian	92	Réunion Creole	125	Welsh
27	Eastern Pahari	60	Lower Sorbian	93	Romanian	126	Western Frisian
28	English	61	Luxembourgish	94	Romansh	127	Wymysorys
29	Faroese	62	Macedonian	95	Russian	128	Yagnobi
30	French	63	Magahi	96	Rusyn	129	Yiddish
31	Friulian	64	Mah. Konkani	97	Saraiki		
32	Gade Lohar	65	Maithili	98	Sardinian		
33	Galician	66	Marathi	99	Sãotomense		

Table 4.3: *IE languages excluded from* MODEL 0 *and included in* PIE MODELS.

of the chapter, the term "later" languages refers to the set of languages in Table 4.3 and Hittite. The inclusion of Hittite was crucial as without it the earliest state that can be simulated at the input to the model presented here is a stage past PIE. The inclusion of Hittite allows the input to the model to be pushed farther back in time to PIE. I assume the inventory of Hittite to be /p t k kʷ pː tː kː kːʷ/, as in Yates (2017). Since geminates and secondary articulations are ignored here, this inventory is simplified to /p t k/[4]. Table 4.4 shows descended IE inventories

[4]Note that this produces results that are only minimally different from a model that assumes the Hittite in-

and their associated frequencies, when the frequency of occurrence was greater than 1. I will refer to these models as PIE MODELS. The performance of all such PIE MODELS was compared and the model that best captured the attested typology of the later IE languages was taken to be most likely synchronic inventory of PIE. I describe the architecture of MODEL 0 and the PIE MODELS in more detail below.

Inventory	Frequency
p t k b d g	107
p t k ph th kh b d g bh dh gh	34
ph th kh b d g	30
p t k ph th kh b d g	11
p t k b d	10
p t k ph th kh	8
p t k	7
p t k ph th kh pc tc kc	3
p t b d g	2
p t k ph th kh b d g bh dh gh bc dc gc	2
p t k ph kh	2
p t k th kh b d g bh dh gh	2

Table 4.4: *List of Indo-European inventories and their frequencies.*

4.4 Model 0

4.4.1 Architecture

The architecture of MODEL 0 was identical to that in Chapter 3, except instead of fitting the frequencies of all the inventories in PHOIBLE, only the non-IE inventories were fit. This was done to avoid training the model on inventories whose frequencies it would later be used to predict in the PIE MODELS. Thus, in contrast to the model in Chapter 3, MODEL 0 evaluated constraint violations incurred by each possible non-IE inventory. This model was based on the best performing model from Chapter 3. Therefore, its dispersion constraints were modeled on the "bottleneck" approach. Moreover, voiced-aspirated stops were assumed to violate both

ventory to have a voicing contrast instead (Melchert, 1994).

*ASPIRATION and *VOICED-ASPIRATION, and voiced-constricted stops were assumed to violate only *VOICED-CONSTRICTION, not *CONSTRICTION.

An example part-tableau is shown in 4.5. The constraint violations are summed for the sake of illustration. The model included the full set of individual MINDIST, MAXIMIZECONTRASTS and markedness from Chapter 3. The constraint weights from this model were implemented as priors in the PIE MODELS to simulate the effect of general tendencies exhibited by stop inventories.

	ΣMINDIST	ΣMAXCONTRASTS	ΣMARKEDNESS	FREQ
a. p-t-k	2	18	4	748
b. p-t-k-b-d-g	3	14	6	723
c. p-k	2	20	2	3
d. p-k-g-bh-bc-gc	3	13	8	0
e. ...	...	...	...	...

Table 4.5: *Example part-tableau with summed* MINDIST, MAXIMIZECONTRASTS *and markedness violations in Model 0.*

4.4.2 Results

Constraint	MODEL 0	FULL MODEL
MAXIMIZECONTRASTS	1.6	1.5
MAXIMIZEPLACECONTRASTS	5.8	6.0
MAXIMIZEVOICECONTRASTS	0.3	0.3
MINDIST=PLACE	1.2	1.1
MINDIST=VOICE	1.6	1.5
MINDIST=WIDTH	0.5	0.6
*LABIAL	0.7	1.03
*VELAR	0.9	0.9
*VOICE	4.3	4.2
*ASPIRATION	6.0	5.9
*CONSTRICTION	5.9	5.9
*VOICED-ASPIRATION	5.8	5.4
*VOICED-CONSTRICTION	4.6	4.5
*G	0.3	0.2
*PC	0.9	0.8
*GC	3.4	3.2

Table 4.6: *Constraint weights in* MODEL 0 *and the full model from Chapter 3.*

Table 4.6 shows the constraint weights from MODEL 0. Constraint weights found by the optimal model in Chapter 3 are also included for reference. The constraint weights are virtually identical in the two models. The weights from MODEL 0, shown in Table 4.6, were fed as priors to the PIE MODELS as detailed in Section 4.5.1.

4.5 PIE models

4.5.1 Architecture

The architecture of the PIE MODELS differs from that of MODEL 0 in two key ways. First, as noted above, instead of using agnostic priors which allow all the constraints to start on a level playing field, the priors used in the PIE MODELS were derived from MODEL 0 so that the model was biased towards general typological tendencies rather than being agnostic. This bias was modeled as a Gaussian prior (Goldwater and Johnson, 2003) of the form (μ, σ) on constraint weights, where μ is a vector of means and σ is a vector of standard deviations for the prior on each constraint weight. Specifying such a prior maximizes the predicted probability of the IE data minus a penalty for deviations from the prior distribution that represents general typological tendencies across stop inventories. This penalty term is given by

(4.4)
$$\sum_{i=1}^{M} \frac{(w_i - \mu_i)^2}{2\sigma_i^2}$$

where M is the total number of constraints, and w, μ and σ are the weight, prior mean and standard deviation of every i^{th} constraint. Thus, the inclusion of the prior biases the PIE MODELS to exhibit general tendencies observed in stop inventories. The more a constraint weight w deviates from its prior weight μ, the larger the numerator in (4.4) and hence the greater the penalty. The value of σ further modulates this penalty. The smaller the value of σ, representing a tighter prior distribution, the smaller the denominator in (4.4), leading to a greater penalty. In the PIE MODELS, the means of each constraint weight were set to the weights in

Table 4.5 and all standard deviations were set to 5.

The second difference between MODEL 0 and the PIE MODELS is that the latter evaluated candidate outputs against inputs. The outputs were later IE stop inventories and the inputs were various proposed PIE stop inventories. Note that this way of defining inputs and outputs differs from most standard notions of "input" and "output" in Optimality Theoretic phonology. Here, "input" refers to the starting point of a cross-linguistic model of sound change, whereas in most work in OT "input" refers to the representation that undergoes a phonological computation in a language-specific grammar. Similarly, "output" refers to the end result of sound change here, while in OT phonology, it refers to the optimal result of a phonological computation. In what follows, it is important to bear in mind these non-standard notions of "input" and "output" employed here.

Since each model evaluated input inventories and output inventories, an additional set of faithfulness constraints was defined. Recall that the sense in which the terms "input" and "output" are used here is distinct from the input and output forms that undergo phonological computation. The inputs here are the surface forms in the proto-inventory and the outputs the surface forms in the later inventories. For the sake of simplicity, later IE forms containing a stop that did not occur in the proto input inventory under consideration were assumed to be added to the inventory. Similarly, the opposite case, in which a later inventory did not contain a stop that was present in the proto-input, was assumed to be deleted. The discussion of the case for this simplification is tabled here and taken up again, along with proposals for refining it in future work, in Section 4.6. Thus, the two faithfulness constraints were of the form

(4.5) DEP - assign a violation if the segment that is not present in PIE is present in a candidate IE inventory.

(4.6) MAX - assign a violation if the segment that is present in PIE is not present in a candidate IE inventory.

An example part-tableau showing constraint violations of a few illustrative inventories derived from the standard reconstruction (PIE_1) are shown in 4.7. Note the differences between the tableaux in 4.5 and 4.7. The most frequent inventories among IE and non-IE inventories dif-

PIE1	ΣMINDIST	ΣMAXCONTRASTS	Σ*MARKED	ΣDEP	ΣMAX	FREQ
a. p-t-k-b-d-g	3	14	6	0	3	107
b. b-d-g-ph-th-kh	3	13	6	3	6	30
c. p-k	2	20	2	0	7	0
d. b-d-g-dc	3	16	7	1	6	0
e. ...	...	...	...	...	...	...

Table 4.7: *Part-tableau for the standard reconstruction, PIE_1* = /p-t-k-b-d-g-b^h-d^h-g^h/.

fer – whereas /p t k b d g/ is overwhelmingly preferred by IE languages, it is inched out by /p t k/ among the non-IE languages. Note also that the frequencies assigned to each inventory depend on the group of languages under consideration. The inventory /p k/, for example, is attested in the languages of the world (in Hawaiian, for example). However, it is unattested among the IE languages and thus receives a frequency of 0 in Tableau 4.7, but not in 4.5. Finally, Tableau 4.7 contains an input and two additional constraint types that evaluate each candidate output inventory against this input. Candidate (a) does not contain any stops that are not present in the input. Therefore, it does not incur any violations of DEP. However, since it deletes the entire Dh series from the PIE input, it incurs three violations of MAX. Note that both types of faithfulness violations are summed here for ease of illustration. The complete model contains constraints that specifically refer to each stop. In the case where the input is PIE_1, for example, there are 9 relevant constraints each for DEP and MAX since PIE_1 contains 9 out of a total possible 18 stops. Therefore, the relevant constraints are

(4.7) DEP–{ph, th, kh, pc, tc, kc, bc, dc, gc}

(4.8) MAX–{p, t, k, b, d, g, bh, dh, gh}

(4.7) represents all nine constraints that penalize the addition of each of the stops in the curly brackets. For example, if we consider the model with the standard reconstruction (PIE_1), an output inventory that contains the aspirated series (/p^h, t^h, k^h/) violates three constraints – MAX–ph, MAX–th and MAX–kh. Similarly, (4.8) represents all nine constraints that penalize the deletion of each of the stops in the curly brackets from the PIE inventory.

Note that different constraints are non-trivially relevant for different inputs. For example, in the case where the input is PIE_2, four-way series, which contains 12 out of the 18 possible stops, there are 6 specific DEP constraints and 12 specific MAX constraints.

Gaussian priors on the weights of the faithfulness constraints were defined to penalize the removal or addition of more salient stops by specifying a weak prior against any changes to such stops. Strictly speaking, this implemented bias is meant to represent a bias against changes to *perceptually* salient stops in an inventory. However, since it is not possible to quantify the absolute perceptual salience of a speech sound in a language-independent manner, I took the following approach. All neutral stops in the center of the space – /p t k/ – were assumed to be the least salient so that their removal or addition would be penalized the least. Each additional laryngeal feature required to specify a stop was assumed to add to its salience. Under this conception, an additional feature corresponds to at least one additional acoustic correlate and hence to perceptual salience. For example, a voiceless-aspirated stop requires an additional feature, [spread], and an additional acoustic correlate such as VOT, that distinguishes it from neutral stops. Therefore, aspirated stops were assumed to have one additional unit of salience, compared to neutral stops. Voiced-aspirated stops require two additional features, [voice] and [spread], corresponding to positive and negative VOT (Narkar, 2025), rendering them two units more salient than neutral stops, and one unit more salient than voiceless-aspirated stops. Thus the addition or removal of stops from an inventory follows the order T < D, Tc, Th < Dc, Dh, from least to most penalized[5]. The values of μ for each potential removal and addition are shown in Table 4.8. The value of σ was kept constant at 5. This high value of

Stop	μ
T	1
D, Tc, Th	2
Dc, Dh	3

Table 4.8: *μ values for priors on changes to segments between PIE input and later IE output.*

σ relative to the values of μ allowed the constraint weights of the faithfulness constraints to deviate from the priors with minimal penalty.

[5]A different method that took into account more acoustic detail, such as burst intensities to account for place differences, returned virtually identical results. I report results from the simplest model here.

4.5.2 Results

Table 4.9 shows the constraint weights in the three PIE MODELS. Those constraints whose weights differ considerably between the three models are shown in boldface. Constraints not shown in this table received weights of zero.

Constraint	Standard	Four-way	Glottalic
MINDIST=PLACE	0.9	0.9	0.9
MINDIST=VOICE	3.4	3.4	3.4
MINDIST=WIDTH	6.0	6.0	6.0
*LABIAL	2.0	2.1	2.0
*VELAR	1.7	1.6	1.7
*VOICE	10.9	11.4	10.8
***ASPIRATION**	4.0	8.7	4.3
***CONSTRICTION**	6.9	6.1	8.8
***VOICED-ASPIRATION**	6.8	6.3	3.1
*VOICED-CONSTRICTION	5.0	4.7	5.0
*G	2.1	1.8	2.7
***PC**	0.0	0.0	1.1
MAXIMIZECONTRASTS	1.2	0.4	1.1
MAXIMIZEPLACECONTRASTS	6.6	6.8	6.5
MAXIMIZEVOICECONTRASTS	1.6	2.3	1.4
MAXIMIZEWIDTHCONTRASTS	0.1	0.9	0.1
DEP–dc	2.0	1.5	2.0
MAX–ph	–	2.0	–
MAX–th	–	2.0	–
MAX–kh	–	2.4	–
MAX–pc	–	–	2.2
MAX–tc	–	–	0.1
MAX–kc	–	–	0.7
MAX–b	2.8	3.3	2.7
MAX–d	3.9	4.4	4.0
MAX–g	3.3	3.3	4.0
MAX–bh	1.7	2.2	–
MAX–dh	0.6	0.5	–
MAX–gh	1.9	2.6	–

Table 4.9: *Constraint weights in the* PIE MODELS.

Let us consider each family of constraints. Note that the violation profiles of markedness, MINDIST and MAXIMIZECONTRASTS do not differ between the three models as the evaluation of these is independent of the input. The evaluation of the faithfulness constraints, however, does differ between the models as does the set of specific faithfulness constraints. Since the MaxEnt algorithm finds the best constraint set as a whole that fits the observed data, weights assigned to the same constraints may differ between the models even though they have the same violations.

First notice that all MINDIST constraints received identical weights in the three models.

Next consider the weights of the markedness constraints. The weights of *LABIAL and *VELAR were comparable across the models. So were the weights of *VOICE and *VOICED-CONSTRICTED. The weights of *ASPIRATION, *CONSTRICTION and *VOICED-ASPIRATION differed considerably. This is due to the fact that the inputs to the three models differed precisely in aspiration and constriction. The higher weight of *VOICED-ASPIRATION in the Standard and Four-way PIE models ensured that voiced-aspirates, which were present in the inputs of both these models, were penalized enough so as to be absent in some of the most common modern IE stop inventories, such as /p t k b d g/, /p^h t^h k^h b d g/ and /p t k p^h t^h k^h/. Recall that voiced-aspirated stops violate both *ASPIRATION and *VOICED-ASPIRATION. Therefore, the weight of *ASPIRATION differed between the Standard and Four-way models. Since voiceless aspirated stops are not present in the Standard PIE inventory, the weight of *ASPIRATION was lower in the corresponding model so as to allow voiceless-aspirates to emerge in the later inventories. For similar reasons, *CONSTRICTION received a higher weight in the Glottalic model – to ensure that the (voiceless) glottalized stops in the PIE input did not proliferate into the output inventories. Since voiced-constricted stops violate only *VOICED-ASPIRATION and are not present in the input of any of the models, the weights of this constraint were similar across all three models. *PC received a high weight in the Glottalic model for the same reason. The remaining markedness constraints – *P and *GC received constraint weights of zero. The zero weight of *P is unsurprising since this constraint does not seem to matter in the full model in Chapter 3 or in MODEL 0 either. The lack of effect of *GC can be attributed to the additive effect of *VELAR and *G (which penalizes any voiced articulation at the velar place) together, rendering the super-additive *GC constraint unnecessary.

The MAXIMIZECONTRASTS set of constraints differed minimally between the models. The Four-way model had a slightly lower weight for MAXIMIZECONTRASTS compared to the other two models, but slightly higher weights for MAXIMIZEVOICECONTRASTS and MAXIMIZEWIDTH-CONTRASTS, with the overall effect relatively consistent across the models.

Finally, consider the faithfulness constraints, whose violations were meaningfully different in the three models. The weights of all the faithfulness constraints were comparable across the three models. Of the constraints penalizing the addition of stops not present in the input, only DEP–dc received a non-zero weight. This is due to the relative rarity of voiced-constricted stops in later IE languages, combined with the zero weight of *CORONAL. The combined effect of *LABIAL and *VELAR with *VOICED-CONSTRICTED keeps /bc/ and /gc/ from proliferating in the later IE languages, but a specific constraint against /dc/ is needed since *CORONAL has no influence. Removing the neutral stops, /p/, /t/, /k/, does not seem to be penalized as none of the constraints, MAX–{p, t, k}, received non-zero weights. This must be due to the fact that these are otherwise unmarked[6]. Removing the voiceless-aspirated stops, which is only relevant in the Four-way model, was penalized similarly across all three places, and did not deviate much from its prior weight. Removing voiceless-constricted stops is tolerated, as indicated by the low weights of MAX–{tc, kc}. MAX–pc received a high weight, only to counteract the already high weight of *PC which penalizes these stops outright. Removing the plain voiced and voiced-aspirated stops is penalized as well, subject to the mediating effect of the corresponding markedness constraints, since both these types are relatively frequent in later IE stop inventories.

Thus, the three PIE MODELS differ considerably. Table 4.10 shows results from the comparison of the performance of these models. As seen in this table, the model whose input is the four-way series, T-Th-D-Dh, had the lowest AIC. This suggests that, of the three PIE inventories under consideration, the four-way series best captures the observed typology of the later IE languages. In other words, the contents of the stop inventories of later IE languages are the least surprising if they all started out as the four-way series.

This lends support to Jakobson's dictum that the standard reconstruction must be reconsidered on the grounds that it is typologically improbable as a synchronic inventory. Table

[6]The inventory, /p^h t^h k^h b d g/, which excludes the neutral series, is relatively common in the PHOIBLE data, but all three models underestimate its frequency.

Model	**PIE inventory**	**AIC**	**Δ(AIC)**
Four-way series	T-Th-D-Dh	1143.416	0
Standard reconstruction	T-D-Dh	1186.425	43.01
Glottalic theory	T-Tc-D	1206.662	63.25

Table 4.10: *Model comparison of* PIE MODELS.

4.10 shows that the standard reconstruction is unlikely to have been the PIE synchronic stop inventory. Instead, the typologically well-attested and stable system, T-Th-D-Dh, is the most plausible candidate, consistent with proposals such as Elbourne (1998). Note, however, that the model with the standard reconstruction as the PIE inventory had a lower AIC, compared to the inventory proposed by the glottalic theory. Therefore, it is more likely to have been the PIE stop inventory than the inventory proposed by the glottalic theory, despite the latter's typological attestedness.

4.6 Discussion

The previous sections have demonstrated a model that can take the effect of general typological tendencies into account when assessing the relationship between ancestral and descended states. The case study on the PIE stop inventory was able to evaluate three different proposals and determine which of these was most likely the initial state from which the later IE languages have descended.

A few caveats are in order. The model presented in this chapter conceptualizes the changes from PIE to modern IE as one set of changes. This is a simplification since, in reality, the trajectory of change included many intermediate stages. A more sophisticated model would start with priors representing general typological tendencies, as done in this chapter, and fit the observed probabilities at the first split in the IE tree. It would then trace each split to the next one, update its priors at each stage and repeat this process iteratively[7].

Such a model would also address a related shortcoming of the model presented in this chapter. Changes between the PIE stop inventory and the modern IE inventories are eval-

[7]See Kuo (2023) for a MaxEnt model that implements such an iterative learning process.

uated as addition or deletion, not as changes per se. However, the more realistic scenario is that the stops in modern IE inventories are the result of changes from one segment into another. Therefore, they should be evaluated as violations of *MAP(x, y) constraints (Zuraw, 2010, 2013), where x and y are non-zero[8]. Take, for example, voiceless-aspirated stops (Th) in Greek, which are derived from PIE voiced-aspirated stops (*Dh). With the four-way reconstruction as the input, the model presented here does not penalize the presence of Th in Greek since these stops are also present in the PIE input. However, a more accurate model would reflect the change from PIE *Dh to Th in Greek and, therefore, assign a violation of *MAP(Dh, Th) to Greek and other inventories like it. This way of evaluating changes could also model inventories that are identical in their composition but are the result of different pathways of change.

An iterative model would also account for intermediate stages, which the current model is unable to do. Since the current model has no notion of time, it is unable to predict changes within PIE depending on its precise location on the tree. What this model specifies as the input is the Proto inventory that most closely dominates all the IE inventories in the tableau (that is, the list in Table 4.3). Since no other split on the IE tree includes all these languages, the input to the tableau reflects the initial state that eventually resulted in all the IE languages included in the model. If we were to only include the Indo-Aryan languages, the input would have to be Sanskrit since that is the ancestor language that most closely dominates all the modern Indo-Aryan languages. If we were to add the modern Iranian languages to this model, however, we would have to push the node farther back in time, to Proto-Indo-Iranian, since Sanskrit does not dominate the modern Iranian languages. Thus, this model is incapable of predicting changes to the same inventory, say Proto-Indo-Iranian, over time, unless this change resulted in a split. The inclusion of Hittite allows the node to be pushed farther back to a state that includes its branch as well as all the other branches that include the modern IE languages.

While a more sophisticated iterative approach would be a more accurate representation

[8]MAX and DEP can be thought of as special variants of the more general *MAP constraints, with MAX being the equivalent of *MAP(x,0), penalizing the mapping of a segment x to zero, i.e. deletion, and DEP being the equivalent of *MAP(0, y), penalizing the mapping of zero to a segment y, i.e. insertion.

of the changes from a proto inventory, modeling the intermediate stages requires knowledge of the contents of their stop inventories. Since many of the intermediate inventories themselves are the result of reconstruction, there is often considerable uncertainty regarding their contents. Moreover, with each intermediate stage, the number of models increases at a rate that is greater than linear. For this reason, the model presented here is only meant to be a first attempt at including typologically sensible effects when modeling change. Future work could develop and test models based on the architecture presented in this chapter, enriched with mechanisms for iterative learning.

CHAPTER 5

The acoustic space of stop contrasts

Life consists of the little things;
the important matter is to see them largely.

– OTTO JESPERSEN

5.1 Introduction

This chapter addresses the structure of the acoustic space of stops across languages. It identifies the acoustic properties that are most crucial in defining this space. It thus constructs a basic, cross-linguistic acoustic space for stops, comparable to the F1-F2 space for vowels. This space is employed in Chapters 6 and 7 to model optimal stop inventories in terms of their acoustic properties.

There are a number of acoustic correlates of voicing, such as voice onset time, preceding vowel length, onset f0, F1, voice quality etc. Indeed, the list of acoustic correlates of voicing is long and likely incomplete (Lisker, 1986). The same is true for stop place of articulation – VOT, burst spectral shape, formant transitions all cue place. However, not all these cues are weighted equally in production or perception; some cues signal the identity of phonetic categories more strongly and/or robustly than others. Moreover, the acoustic correlates of laryngeal distinctions and place far outnumber the laryngeal contrasts that the world's languages make, even after accounting for cues that co-vary and might result from "automatic" consequences of physiological and aerodynamic factors governing production. We can therefore ask if there is a basic acoustic space of stop contrasts, equivalent to that of vowels. Can the dimensions of this acoustic space be interpreted phonetically? Based on acoustic data from 67 languages, this chapter presents an attempt to determine which acoustic correlates of stops are preferred to organize stop contrasts across languages.

Most modeling studies on inventory structure have focused on vowels (Liljencrants and Lindblom, 1972; Schwartz et al., 1997; De Boer, 2000), partly because the acoustic space of vowels is well-understood, as is its relationship to the articulatory properties of vowels. The vowel space is largely reducible to two dimensions, F_1 and F_2', which can be determined by the first three formant frequencies. Moreover, these neatly map on the articulatory properties of height, backness and roundness. No such equivalent space exists for stops (or other consonants, for that matter). Both stop acoustics and articulation are highly multi-dimensional and complicated and their relationship is not straightforward.

The only major study modeling stops, Schwartz et al. (2012), proposed an articulatory-acoustic model of stop place of articulation that simulates the effect of the closure location on the first three formants. They are thus able to construct a phonetic space in terms of acoustic and articulatory properties of stop place of articulation. However, there is still no model that incorporates laryngeal distinctions in stops, not even a binary voicing distinction, into the characterization of the basic phonetic space in which they operate. While it is the case that models of vowel inventories are largely limited to monophthongal oral vowels with modal voicing, considerable research has been conducted on non-modal phonation (see Garellek (2019) for a review). The most comprehensive of these is Keating et al. (2023), which shows that vocalic phonation distinctions, based on acoustic measures from 11 languages, are largely reducible to two dimensions that are phonetically interpretable. The first dimension separates modal vowels from non-modal vowels and the second makes finer distinctions within the non-modal vowels by separating various voice qualities with creaky vowel on one end and breathy on the on the other end of the spectrum. Although Keating et al. (2023) do not include vowel quality contrasts in their analysis, we can combine the well-known results on vowel quality with their results on voice quality to construct four-dimensional space of vocalic contrasts that includes both these attributes. Such a four-dimensional space consists of two dimensions that index vowel quality (F_1-F_2') and two dimensions that index voice quality (the dimensions proposed by Keating et al. (2023)).

In this chapter, I present an equivalent phonetic space for stop place and laryngeal contrasts, adopting Keating et al.'s approach. I show that the stop space, like the vowel space, is largely four-dimensional, with two dimensions corresponding to place contrasts and two corresponding to the laryngeal contrasts.

5.2 Method

The method was largely based on Keating et al.'s (2023) study of vocalic phonation types. The main difference between their study and this one is that they measured all the acoustic correlates of voice quality in 11 languages themselves, whereas in this study, I used data collected from published acoustic studies on 66 languages. Details on the languages and acoustic measures are presented below. Another difference between Keating et al. (2023) and this study is that they included languages with allophonic phonation distinctions, while this study is limited to languages with phonemic contrasts.

5.2.1 Stop contrasts

As in previous chapters, place contrasts are limited to labials, coronals (dental and alveolar) and velars. For languages that have more than one coronal, the one described as "alveolar" or "dental" was included. Other coronals were excluded. One language included in the study, Arrernte (Tabain et al., 2016), is described as having both alveolar and dental stops (in addition to bilabials, velars, palatals and retroflexes). In this case, only the alveolar stop was included. Six laryngeal configurations were included – voiceless (T), voiced (D), aspirated (Th), constricted (Tc), voiced-aspirated (Dh), and voiced-constricted (Dc). Stops with all other secondary articulations were also excluded, as were doubly articulated stops.

5.2.2 Acoustic Properties

For each stop, the acoustic correlates of place and laryngeal setting were compiled based on a survey of published acoustic studies. See the Appendix for the sources. For each language, the following acoustic correlates were compiled for each of its stops. These acoustic parameters were chosen as they have been shown to index place and/or laryngeal distinctions and are widely reported in acoustic studies. Note that there is a great deal of variation between languages in terms of if and how the following acoustic measures cue various stop place and laryngeal contrasts. The description provided below is meant to be general and broadly applicable, but not without exception. Indeed, this variation across languages is the reason why we want to establish a basic phonetic space that captures stop contrasts in the languages of the world.

Table 5.1 shows the acoustic correlates of the stop laryngeal categories. Note that this table is meant to provide a rough idea of the acoustic properties of the different laryngeal categories as there is a lot of variation across languages. Note also that for ejectives, Table 5.1 shows the properties of the so-called "stiff" rather than "slack" ejectives (Lindau, 1984; Kingston, 2005).

	T	**D**	**Th**	**Dh**	**Tc**	**Dc**
+VOT	short	–	long	long	long	–
−VOT	–	long	–	long	–	long
Cf0, ΔCf0	raised	lowered	raised	inconclusive	raised	lowered
CD	long	short	long	short	long	short
Burst intensity	strong	weak	strong	weak	strong	weak
Burst duration	long	short	long, noisy	short, noisy	very short	short, diffuse
CPP	high	high	low	low	low	low
H1–H2	0	0	+	+	−	−

Table 5.1: *Acoustic correlates of laryngeal categories.*

In Table 5.1, VOT represents Voice Onset Time, the time interval between the stop release burst and the onset of periodicity (Lisker and Abramson, 1964). Note that I treat positive VOT, the time interval between the burst and the onset of voicing when the burst precedes voicing, and negative VOT, the time interval between these events when the burst follows voicing, separately, following Narkar (2025). Although VOT is primarily associated with voicing distinctions (Lisker and Abramson, 1964), it is also a correlate of place of articulation, with VOT values increasing as the location of the closure becomes more posterior in the oral cavity (Lisker and Abramson, 1964; Volaitis and Miller, 1992).

Consonant-induced fundamental frequency (Cf0) refers to the f0 at the onset of the following vowel as influenced by the preceding stop. Roughly speaking, voiceless stops tend to raise Cf0, while voiced stops lower it (House and Fairbanks, 1953; Dmitrieva et al., 2015). This pattern holds cross-linguistically (Ting et al., 2025), but is not without exception in production (Francis et al., 2006; Chen, 2011) or perception (Narkar and Meszaros, under review). In addition to the Cf0 at the onset, another measure of the effect of the preceding consonant on pitch was included, namely the difference between the f0 at the vowel midpoint and onset, ΔCf0.

CD represents closure duration, the time interval during which the stop closure is held before release. Burst duration measures how long the noise generated at the release of the stop persists. Like closure duration, it contributes to distinguishing laryngeal categories (Van Alphen and Smits, 2004) and place (Tabain et al., 2016). Burst intensity in Table 5.1 represents the relative burst intensity, which measures the amplitude of the burst compared to that of the following vowel and serves as a correlate of place and laryngeal category. CPP, or Cepstral Peak Prominence (Hillenbrand et al., 1994), is a harmonics-to-noise ratio that measures the strength of periodicity in the speech signal by quantifying how prominent the harmonic structure is relative to noise. Higher values indicate more regular, modal phonation, while lower values reflect increased noise or non-modal phonation. Finally, H1-H2 is the difference in amplitude between the first and second harmonics of the sound. It reflects the shape of the glottal waveform, with higher values indicating breathier voice and lower values indicating more constricted voice quality. Modal voice has H1-H2 values lying between those of breathy and creaky phonations.

Table 5.2 shows the acoustic correlates of stop place of articulation. Spectral moments serve as a correlate of place and voicing to some extent (Chodroff and Wilson, 2014). The first spectral moment, Center of Gravity (CoG), represents the mean distribution of acoustic energy in the burst spectrum. The second moment, Standard Deviation (SD), captures the spread of this energy on either side of the mean, as the name suggests. The third moment, skewness, describes whether energy is concentrated toward lower (positive skewness) or higher (negative skewness) frequencies. Finally, kurtosis reflects how peaky or flat the distribution is. Together, the spectral moments provide fine-grained distinctions in spectral shape and are strong cues

	Labial	Coronal	Velar
+VOT	short	variable	long
−VOT	variable, often long	variable	variable, often short
CD	relatively long	variable	relatively short
Burst intensity	weak, diffuse	strong	strong
Burst duration	short	short	long
CoG	low	high	mid
SD	diffuse spectrum	moderate	compact spectrum
Skew	low-frequency energy	near-symmetric/negative	variable, often negative
Kurtosis	flat spectrum	moderate	peaky spectrum
F2	low	high	higher

Table 5.2: *Acoustic correlates of place categories.*

to place of articulation. In addition to the spectral moments, the second formant frequency, F2, at vowel onset also serves as a correlate of place. Note again, that Table 5.2 provides rough descriptions. This is particularly important to bear in mind for the coronals, whose spectral properties are known to vary depending on the exact place of articulation (Jongman et al., 1985; Sundara, 2005; Tabain et al., 2016). Recall that the analysis presented here does not differentiate between dentals and alveolars, which have different spectral shapes, and ignores other coronals altogether.

Many of the acoustic measures shown in Tables 5.1 and 5.2 are known to depend on speaker attributes. Perhaps the most systematic variation is exhibited by f0, which depends on the length of the vocal tract, and thus varies by biological sex. Therefore, when Cf0 measures were reported as coming from male or female speakers, they were treated separately and z-score normalized within sex across languages. However, some acoustic measures whose measurement involves f0, such as H1-H2, were not normalized within sex as none of the studies reported by-sex measures for these acoustic properties. Furthermore, although there is evidence that some of the other acoustic measures, such as VOT, may also be influenced by sex-related differences (Whiteside and Irving, 1997), there is evidence that such measures are socio-indexical in nature and may be correlated with gender expression rather than arising solely via anatomical differences between various sexes (Yu, 2023). Therefore, no other acoustic measures beyond Cf0 were normalized within sex. Moreover, since none of the studies reviewed here reported gender as separate from (binary) sex, gender expression could not be included as a conditioning factor in the analysis presented here.

5.2.3 Languages

A total of 66 languages were included in the study. Table 5.3 lists these languages. Note that for three languages – Afrikaans, Korean and Malagasy[1] – two varieties were included, since varieties of all three languages are undergoing tonogenesis to some extent. Thus, the total number of varieties included in the sample here is 69. See the Appendix for details of acoustic measures for all the languages and their sources. Table 5.4 shows the number of languages with

	Language		Language		Language
1	Afrikaans (x2)	23	Gurindji	45	Pitjantjatjara
2	Amharic	24	Hanoi Vietnamese	46	Polish
3	Anyi	25	Hindi	47	Q'anjob'al
4	Ao-Naga	26	Hul'q'umi'num'	48	Russian
5	Armenian	27	Hungarian	49	Somali
6	Arrernte	28	Igbo	50	Shimaore
7	Bengali	29	Indonesian	51	Sindhi
8	Burushaski	30	Italian	52	Siraiki
9	Cambodian	31	Japanese	53	Standard Modern Greek
10	Canadian French	32	Kalasha	54	Swahili
11	Cantonese	33	Kannada	55	Swedish
12	Catalan	34	Korean (x2)	56	Swiss German
13	Central Vietnamese	35	Makasar	57	Tamil
14	Dahalo	36	Malagasy(x2)	58	Telugu
15	Dutch	37	Mandarin Chinese	59	Tigrinya
16	English	38	Maori	60	Turkish
17	Estonian	39	Marathi	61	Tzutujil
18	Finnish	40	Mixtec	62	Uyghur
19	Georgian	41	Navajo	63	Warlpiri
20	German	42	Oromo	64	Wubuy
21	Gipuzkoan Basque	43	Peninsular Spanish	65	Yoruba
22	Gujarati	44	Persian	66	Zuberoan Basque

Table 5.3: *List of languages used in the analysis.*

[1]These languages are not shown as separate dialects in the table, like the different varieties of Basque for example, since younger and older speakers produce the contrast differently in all three languages. For example, younger speakers of Seoul Korean are increasingly producing the traditionally three-way laryngeal contrast as a tonal contrast, but the laryngeal contrast is still produced by older speakers, who only use pitch as a secondary correlate (Silva, 2006; Kang and Han, 2013). Afrikaans, Korean and Malagasy were therefore coded and treated separately in the analysis (e.g., as "Afrikaans old" and "Afrikaans young").

each type of laryngeal contrast inventory that was included in the analysis. In this table, “T” represents neutral stops, “D” represents plain voiced stops, “Th” voiceless-aspirated stops, “Tc” voiceless-constricted stops, “Dh” voiced-aspirated stops, and “Dc” voiced-constricted stops. Note that Table 5.4 shows coarse descriptions of languages into the six laryngeal categories. It includes four languages whose stop contrasts are not strictly voicing contrasts as the descriptions in Table 5.4 might imply – Afrikaans spoken by younger speakers (Coetzee et al., 2018), Central Malagasy (Howe, 2017), Swiss German (Ladd and Schmid, 2018)[2] and Seoul Korean (Kang and Han, 2013).

Contrast type	**N = 69**
$T-D$	22
T	13
$T-Th$	5
$D-Th$	4
$T-Th-D-Dh$	4
$T-Tc-D$	4
$T-Tc-Th$	4
$T-Th-Dc$	3
$T-Th-D-Dh-Dc$	2
$T-Th-D$	2
$D-Dc$	1
$T-Tc$	1
$Th-Tc$	1

Table 5.4: *Number of languages with various types of laryngeal contrasts in the study.*

The acoustic correlates of stop types in the languages in Table 5.3 were compiled from published studies (see the Appendix for details). Since both the languages in this table and the acoustic measures listed in Tables 5.1 and 5.2 differ in the extent to which they are studied and reported on in the literature, not every combination of language, stop type and acoustic measure was available. Therefore, missing acoustic measures were reconstructed using a Bayesian multivariate imputation procedure implemented in the R package `mice` (Van Buuren and Groothuis-Oudshoorn, 2011). Imputation was carried out using Multiple Imputation by Chained Equations (MICE), which iteratively estimates a series of conditional models

[2]Note that although Swiss German is shown to be increasingly cued by closure duration, geminates, which are also primarily cued by closure duration (e.g., Ridouane, 2010; Dmitrieva, 2012), were otherwise excluded.

for each variable with missing data. At each iteration, missing values are replaced with draws from the posterior predictive distribution, allowing uncertainty in the imputation to be propagated through the model.

To ensure that the reconstructed values were phonetically realistic and acoustically plausible, the imputation model was constrained by the type of acoustic parameter. The parameters were grouped into two theoretically motivated sets – (i) correlates of laryngeal contrast (shown in Table 5.1), and (ii) place-related spectral cues (shown in Table 5.2). The predictor matrix was specified such that variables were imputed only from other variables within the same set of acoustic correlates. This restriction prevented the model from exploiting spurious cross-domain correlations (e.g., predicting temporal voicing measures from spectral shape parameters), thus ensuring that the reconstructed values were phonetically interpretable.

This imputation was performed using predictive mean matching (PMM), a semi-parametric method that replaces missing values with observed values from similar cases, ensuring that imputed values remain within the empirical range of the data. Multiple imputed datasets were generated over 20 iterations to ensure convergence of the sampling procedure. Such iterative updating reduces dependence on initial starting values, which are randomly drawn from the observed values, yielding draws from the posterior predictive distribution. Categorical variables such as stop type and language were included as predictors, allowing the reconstruction of missing values to be conditioned on these variables. The complete datasets with the missing values thus obtained were used in the statistical analysis[3].

Note that this analysis used data complied from various studies, which differed in their methodologies. This meant that two types of factors could not be controlled for – (i) the number of participants, which does not directly impact the measured parameter, but may indicate differences in reliability across studies, and (ii) elicitation method. The latter included factors such as words elicited in isolation, versus in a carrier phrase, which is known to have an effect on the acoustic parameter measured (e.g., Kessinger and Blumstein, 1997). Moreover, the context in which the stop occurred also differed, sometimes within a language. For example,

[3] These data can be found at the OSF directory at `https://osf.io/2a9tm/overview?view_only=a02731809c084bb0997def9220d3aec1`.

in most studies measuring VOT, the stop occurred in initial position, whereas in those measuring closure duration, it occurred word-medially. Additionally, the vowel contexts, which can also influence the acoustic measures, could not be controlled for.

In all the cases, mean values of each language x stop combination were used, since we are interested in differences in how different languages use the acoustic space to create contrasts, and not how different individual speakers within languages do so.

5.2.4 Statistical Analysis

The method was based on Keating et al.'s (2023) study of vocalic phonation types. They formulated a multi-dimensional acoustic space of phonation distinctions in vowels by mapping the acoustic correlates of voice quality onto a low-dimensional space that best fit the acoustic data using Multi-Dimensional Scaling (Kruskal and Wish, 1978). Multi-Dimensional Scaling, or MDS, is a method for dimensionality reduction whereby many individual measures can be reduced to a smaller number of independent dimensions. MDS is particularly well-suited for multi-collinearity and missing data (Groenen and Borg, 2013), as in our case. The MDS analysis was performed using a modified version of the code used by Keating et al. (2023).

First, to determine the appropriate number of dimensions that the acoustic parameters could be reduced down to, k, *stress values* were calculated as a measure of goodness-of-fit. The smaller the stress value, the better is the fit of the low-dimensional space. The MDS solution was then run with the best value of k.

Since there are 69 languages (the 66 languages shown in Table 5.3 plus three varieties each of Afrikaans, Korean and Malagasy) in the MDS solution, centroids for each stop category were determined by aggregating the coordinates of the individual observations in the MDS solution. Specifically, the MDS output provided each token in the dataset with coordinates along each dimension of the reduced space. These coordinates were then grouped by the categorical factors `place` and `laryngeal category`, and for each grouping, the mean value along each MDS dimension was calculated. The resulting centroid for a stop category thus represents the average position of all tokens sharing a given place of articulation and laryngeal category in the MDS space, providing a summary point for each category in each dimension.

To further understand how each stop inventory type uses the MDS space, a representative language type was chosen for each type of stop inventory shown in Table 5.4. For each inventory type with more than one associated language, the selected representative language was the one whose stop category MDS coordinates were closest to the corresponding category centroids, such that the total summed distance between the language's coordinates and all centroids was minimized. Each language thus chosen was assumed to be the prototype of its inventory. Note that the inventory types shown in this table only consider laryngeal contrasts and ignore differences in inventory resulting from place differences. Thus, a language like Dutch (stop inventory: /p t k b d/) is treated the same as Polish (stop inventory: /p t k b d g/, since they are both of the type T-D, despite differences in place of articulation. The MDS analysis was repeated on this smaller, cleaner dataset to determine any effects the inventory type might have on how the phonetic space is employed to create stop contrasts.

5.3 Results

5.3.1 Number of dimensions

Figure 5.1 shows the stress values against different numbers of dimensions. The "elbow" of the curve was identified as the sufficient number of dimensions. As suggested by Figure 5.1, there is a steep drop in stress from $k = 1$ and $k = 3$. There is a noticeable drop in stress between $k = 3$ and $k = 4$, but from $k = 4$ onwards the curve is noticeably flatter. Although stress continues to decrease beyond $k = 4$, the gains are marginal, compared to the large drops earlier. Each additional dimension adds interpretive complexity and provides diminishing returns in terms of fit. Therefore, the MDS analysis was run with four dimensions, since the elbow at $k = 4$ represents the point where adding dimensions stops paying off meaningfully.

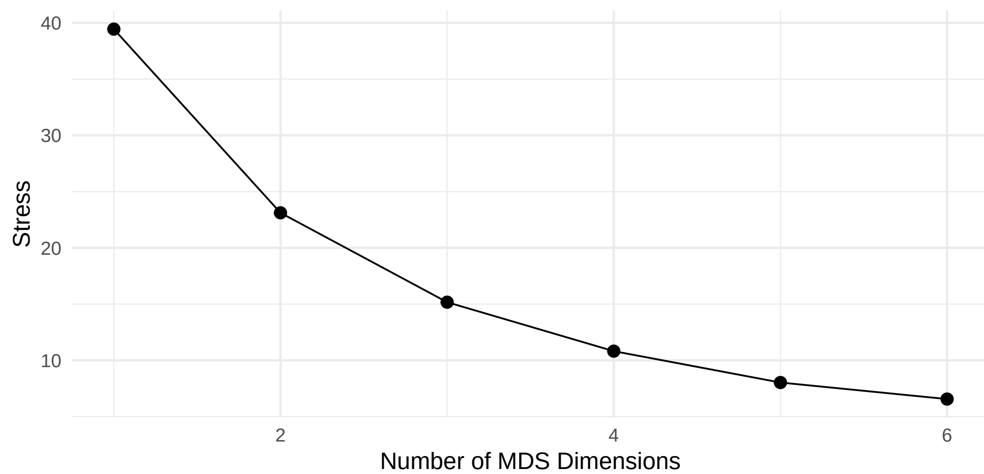


Figure 5.1: *Scree plot for the MDS solution.*

5.3.2 The acoustic space

Figures 5.2 and 5.3 show the centroids of the 18 major stop categories in three-dimensional spaces. Figure 5.2 shows the latent dimensions 1, 2 and 3 from the MDS solution, and Figure 5.3 shows the latent dimensions 2, 3 and 4. Interactive 3D plots can be found at `https://osf.io/2a9tm/overview?view_only=a02731809c084bb0997def9220d3aec1`.

Figure 5.2 shows two different views of the dimensions 1, 2 and 3 from the MDS solution. The top figure shows that MDS1 primarily separates the three places of articulation, with velar stops having positive MDS1 coordinates, coronals having negative MDS1 coordinates and labial stops in the middle. It thus separates the labials and coronals from the velars, in line with classical dimension of diffuseness from Jakobson et al. (1952). The top and bottom plots together show that the second dimension, MDS2, broadly separates the voiced and voiceless stops, resembling the VOT continuum. Along this dimension, the voiced stops (D, Dc and Dh) at all three places have negative MDS2 coordinates while the voiceless stops (T, Tc and Th) at all three places have positive MDS2 coordinates. Finally, the third dimension, MDS3, separates modal from non-modal phonation, mediated by place. The coronals stops as a set have positive coordinates of MDS3 and the labials and velars are lower on the MDS3 axis. The modally voiced stops (T and D) at all three places have negative MDS3 coordinates and the

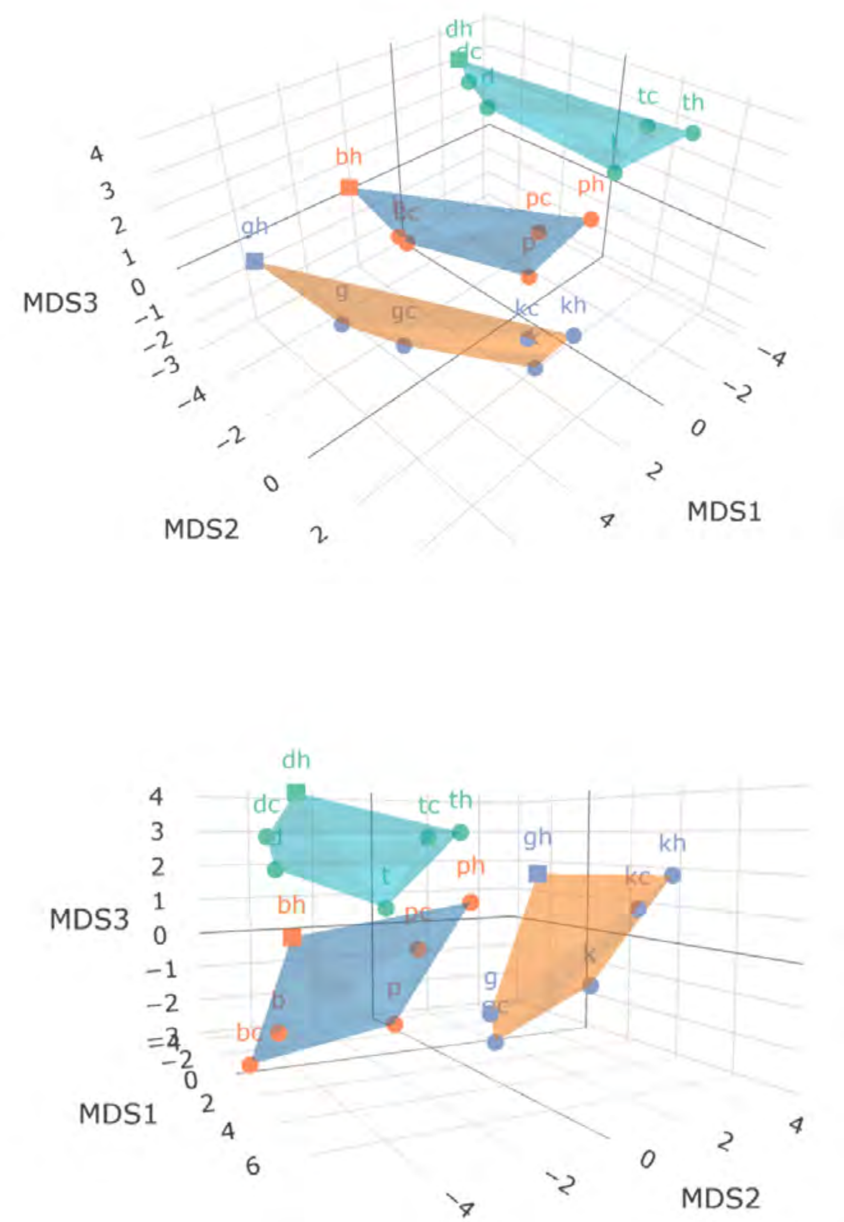


Figure 5.2: *3D plots of dimensions 1, 2 and 3 from the MDS solution.*

non-modal stops, particularly Tc, Th and Dh, have positive MDS3 coordinates. The implosives /ɓ/ and /ɠ/ seem to pattern with the modal stops. This could be because the implosives are the most underrepresented category (tied with voiced-aspirates), with only 6 languages with implosives included in the dataset (see Table 5.4). Moreover, there is a significant imbalance across place in the dataset and in languages in general, as shown in Chapter 2, such that the velar implosive is extremely rare. Out of the six languages with implosives, only two (Sindhi and Siraiki) have /ɠ/ in the dataset analyzed here. Therefore, their acoustic correlates may be less reliable. Additionally, three of the six languages, Cambodian, Central Vietnamese and Hanoi Vietnamese, have implosive stops to the exclusion of plain voiced stops, meaning their acoustic correlates may not be robustly differentiated from those of the modal-voiced, plain stops. Finally, it should be noted that there is a wide range of variation in the production of implosives. For example, implosives in Bantu languages tend to be produced with modal voicing, rather than constricted (Maddieson, 2019). Even in languages that have both implosives and modal-voiced stops, there is a fair amount of variation in terms of how the implosives are implemented (Nihalani, 1986; Clements and Osu, 2002; Ladefoged and Loeb, 2002). Therefore, it is possible that the within-category variation among the implosives might get wiped out when they are treated as a single, uniform category.

Figure 5.3 shows two views of dimensions 2, 3 and 4 from the MDS solution. This plot shows the contribution of the fourth dimension. As seen in this figure, MDS4 separates the labial stops from the coronals and velars, as well as the modal phonation from non-modal phonation within each place.

Taken together, the structure of the phonetic space of stop contrasts can be interpreted as being made up of four dimensions. The first of these (MDS1) represents place distinctions, the second (MDS2) represents broad voicing contrasts (i.e. voiced versus voiceless), the third (MDS3) represents phonation contrasts (modal versus non-modal) and the fourth represents a combination of place and phonation contrasts.

This structure is also apparent when we test which acoustic correlates are weighted on each dimension, as shown by the weights in Table 5.5. Note that the weights in this table were obtained by rotating the dimensions obtained from the MDS solution to improve interpretability. Because MDS axes are arbitrary in orientation, an orthogonal rotation was applied to the matrix of variable loadings to align each dimension with a small set of strongly contributing acoustic parameters. This redistributes variance across dimensions without chang-

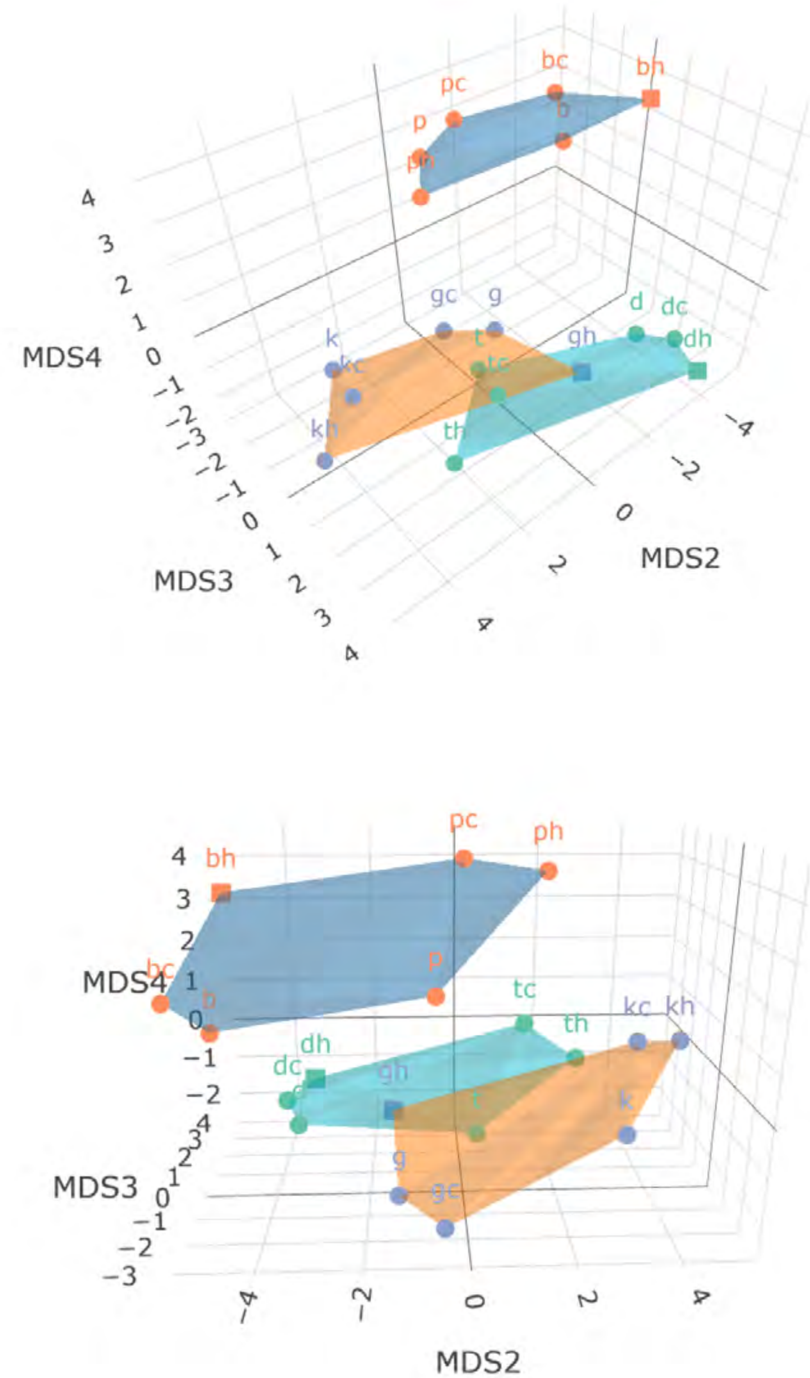


Figure 5.3: *3D plots of dimensions 2, 3 and 4 from the MDS solution.*

ing the underlying distances among observations, making it easier to interpret each dimension as reflecting a distinct phonetic property. Thus, the rotated solution provides a clearer mapping between latent dimensions and acoustic correlates while preserving the original structure of the MDS space.

	MDS1	**MDS2**	**MDS3**	**MDS4**
+VOT	0.132	0.553		-0.169
−VOT		-0.438	0.472	
Cf0		0.378		
∆Cf0			-0.388	0.134
CD				0.110
Burst intensity				-0.182
Burst duration	0.234			-0.170
CPP			0.475	
H1-H2			-0.232	-0.103
COG	-0.216			
SD	-0.267			
Skew	0.393			
Kurtosis	0.429			
F2	0.163			
Variance (%)	4.1	4.6	4.8	1.1

Table 5.5: *Weight of each acoustic measure on each dimension of the 4-D MDS solution (MDS1, MDS2, MDS3, MDS4)*

The rotated MDS loadings revealed a multidimensional structure in which different sets of acoustic measures clustered into interpretable phonetic dimensions. As seen above, the first dimension (MDS1) is primarily defined by measures of spectral shape, with strong positive loadings for skewness and kurtosis and negative loadings for COG and standard deviation. Thus, this axis captures variation in the distribution of burst energy and represents differences in place of articulation. The second dimension (MDS2) is dominated by more or less binary voicing distinctions, with opposing contributions from positive and negative VOT and a substantial contribution from Cf0, thus representing stop voicing contrasts. The third dimension (MDS3) groups together negative VOT, CPP and H1-H2, suggesting that this dimension captures contrasts related to voicing and phonation. Finally, the fourth dimension (MDS4) captures residual variation associated through differences in closure duration, burst duration, and relative burst intensity, suggesting that this dimension captures variation in place and laryngeal category. Overall, the solution separates correlates of place from correlates of laryngeal category (in terms of both voicing and phonation).

Note that the final row in Table 5.5 shows the variance accounted for by the individual MDS dimensions. The proportion of variance is relatively low, which is typical for multidimensional scaling, as the method prioritizes preservation of pairwise distances rather than maximizing the variance accounted for. Therefore, variance is distributed across multiple dimensions, and the interpretability of dimensions provides a more meaningful criterion for evaluating the solution. As we will see, the MDS structure is consistent across models, indicating that the structure discussed so far is "real."

5.3.3 The effect of inventory type

Table 5.6 shows the prototypical language, whose stop category MDS coordinates were closest to the corresponding category centroids, for each inventory type. Recall that the inventory types shown in this table only consider laryngeal contrasts and do not take into account differences in inventory resulting from place differences.

Inventory	**Language**
$T-D$	Polish
T	Arrernte
$T-Th$	Cantonese
$D-Th$	Swedish
$T-Th-D-Dh$	Marathi
$T-Tc-D$	Tz'utujil
$T-Tc-Th$	Georgian
$T-Th-Dc$	Central Vietnamese
$T-Th-D-Dh-Dc$	Siraiki
$T-Th-D$	Burushaski
$D-Dc$	Shimaore
$T-Tc$	Q'anjob'al
$Th-Tc$	Hul'q'umi'num'

Table 5.6: *Representative languages of each laryngeal inventory type.*

Results from the model with representative languages are shown in Table 5.7. The MDS solution on the smaller, representative dataset shows a clearer and more strongly differentiated structure than the full model. In the solution based on the smaller dataset, MDS1 is dominated by positive and negative VOT, compared to the baseline model, in which MDS1 was dominated by spectral shape measures. This is unsurprising since the smaller dataset was

	MDS1	MDS2	MDS3	MDS4
+VOT	-1.496	0.318	-0.192	
−VOT	1.399		0.995	
Cf0	-0.567	-0.128		
ΔCf0	-0.179		-1.097	-0.202
CD				-0.440
Burst intensity	-0.335	-0.113		0.366
Burst duration	-0.109	0.561		0.306
CPP	0.112	0.172	1.211	
H1-H2			-0.536	0.228
COG	-0.188	-0.472	-0.148	0.143
SD		-0.587		-0.201
Skew	0.122	0.679		-0.139
Kurtosis	0.107	0.835		
F2		0.346		
Variance (%)	34.0	16.7	28.7	4.4

Table 5.7: *Weight of each acoustic measure on each dimension of the 4-D MDS solution with representative languages.*

constructed based on different laryngeal inventories, ignoring some place distinctions. MDS2 in the smaller dataset is comparable to MDS1 in the baseline, with strong effects of spectral shape measures. MDS3 in both models captures a phonation-related dimension, with high CPP and negative H1H2 weights, along with a prominent role for voicing as indicated by the weight of negative VOT and and ΔCf0. Finally, MDS4 remains a minor residual dimension in both cases.

Note that a greater proportion of variance is accounted for in the smaller model, compared to the full model, as shown in the final rows of Tables 5.5 and 5.7. This likely reflects the more controlled nature of the smaller dataset, which reduces noise and thus extraneous variability. However, the overall structure of the MDS solution remains consistent across datasets, with comparable place, voicing and phonation dimensions. This consistency across datasets suggests that the MDS solution captures a robust underlying structure rather than reflecting idiosyncrasies of a particular dataset.

Another advantage of running the analysis on a smaller dataset is that now we can assess how different languages use this space, instead of overwhelming the solution with too many languages or having to resort to only investigating category centroids. Instead of showing two

3D plots as in Figures 5.2 and 5.3, here I show two 2D plots contrasting different pairs of axes in the 4D space.

Figure 5.4 shows the dimensions MDS1 and MDS2[4] from the MDS solution of the representative dataset on the x- and y-axis respectively. Different shapes in this plot represent the

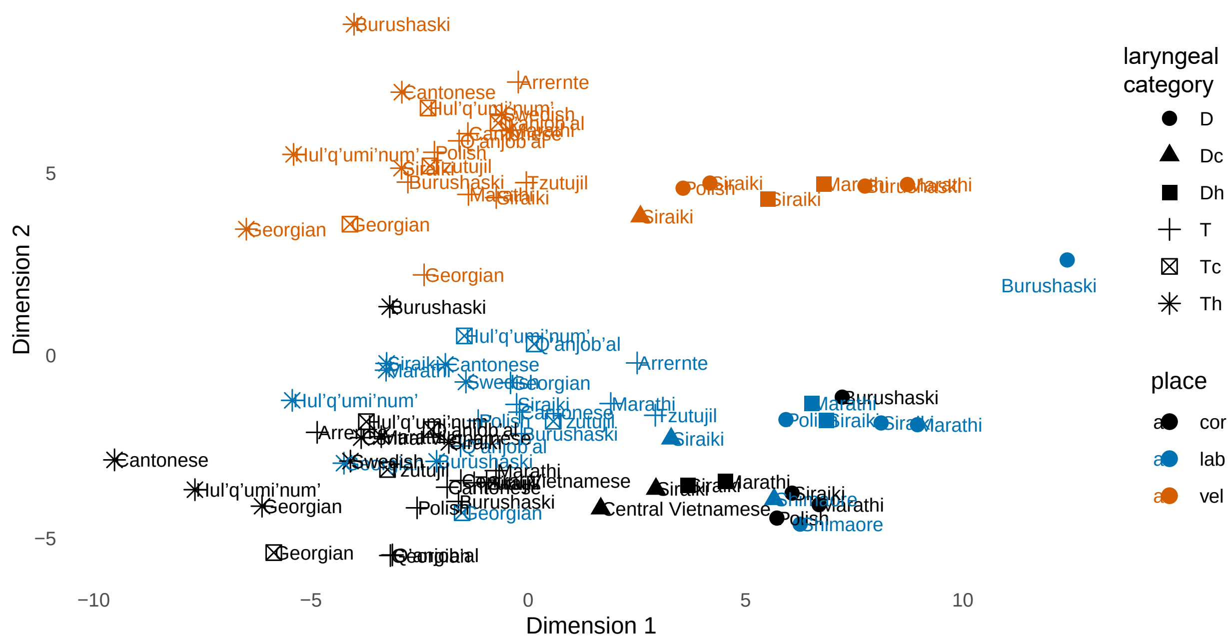


Figure 5.4: *2D plots of dimensions 1 and 2 from the MDS solution on representative languages.*

different laryngeal categories and colors represent places of articulation. Every point in this plot shows how a language x stop combination utilizes the acoustic space.

First notice that generally speaking, the same stops cluster together in parts of the MDS space across languages. Along MDS1, the voiceless stops (T, Tc, Th) are all clustered on the negative x-axis , and the voiced stops (D, Dc, Dh) on the positive x-axis. MDS1 thus serves as a voicing continuum, going from voiceless to voiced. Along MDS2, velar stops (represented by orange) all cluster on the positive y-axis, while coronals (black) and labials (blue) cluster on the negative y-axis.

Figure 5.5 shows the two other dimensions, MDS3 and MDS4. The patterns here are not

[4] The combinations of axes not shown in this chapter are included in the Appendix.

as clear as they are in MDS1 and MDS2. Plain voiced stops span the entire MDS3 range such that the /b/ of Shimaore has the lowest value of MDS3 while the /d/ of Marathi and Siraiki is very close to the opposite edge of MDS3. Generally speaking, the coronals occur on the positive end of the x-axis and the labials on the negative end, although this pattern has a few exceptions.

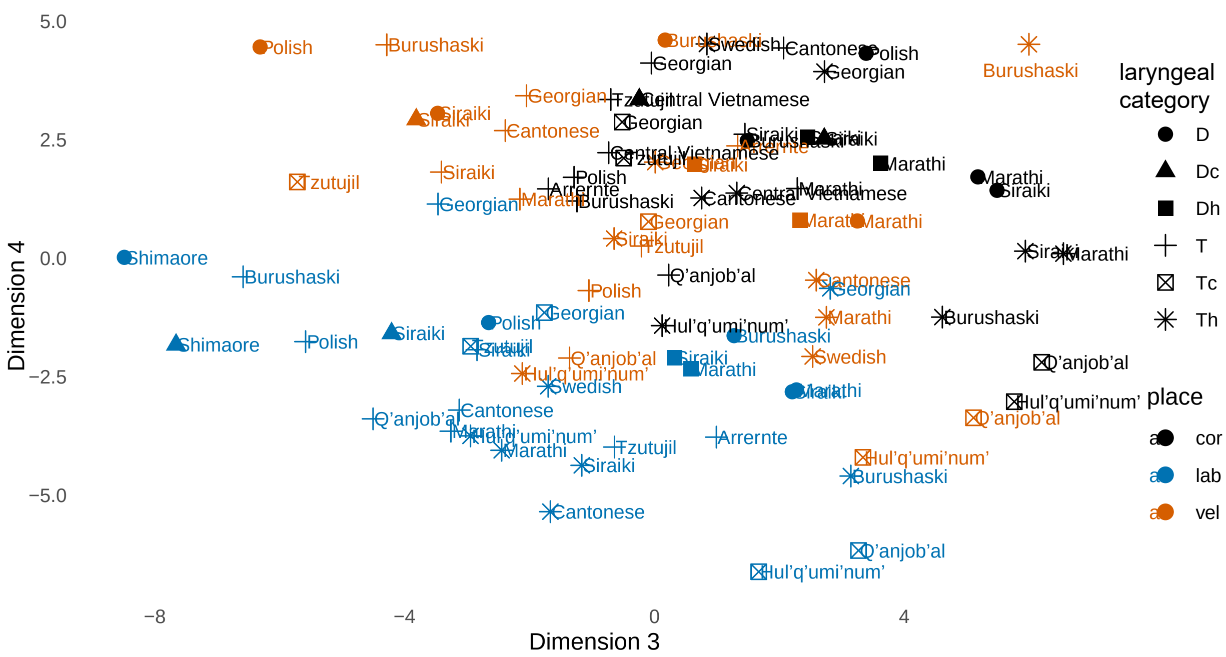


Figure 5.5: *2D plots of dimensions 3 and 4 from the MDS solution on representative languages.*

Finally, MDS4, which accounts for a smaller proportion of variance, appears to encode even more heterogeneous effects. It is primarily associated with longer closures and longer and louder bursts. One consequence of this is that MDS4 distinguishes between the stiff ejectives of Tz'utujil (Bennett, 2010) and the slack ejectives of Hulquminum and Qanjobal (Percival, 2024). Also, the labials mostly cluster along the negative y-axis, while the coronals and velars occur on either side of the origin. Overall, the patterns along MDS3 and MDS4 are more heterogeneous, compared to MDS1 and MDS2.

Figures 5.4 and 5.5 do not show any obvious patterns by language. This result is quite different from Keating et al. (2023), which found that the language with the largest inventory of phonation contrasts defined the extreme edges of the MDS space. Burushaski (stop inventory:

T-Th-D) does seem to frequently occur at the edges. Burushaski /b/ has the greatest MDS1 value and /k^h/ has the greatest MDS2 value. It also has one of the highest values of either edge of MDS3 – /p/ is very close to the negative edge of the MDS3 axis and /k^h/ is very close to the positive edge. Burushaski /k, k^h, g/ are very close to the positive edge of MDS4 and /p^h/ to its negative edge. However, this might not be the result of meaningful variation conditioned by language-specific properties and is more likely due to Burushaski having values of skewness and kurtosis that are much greater (Hussain, 2021) than for any other language in the smaller dataset. Moreover, there is no language in the dataset, the full set or the smaller one, that has all six phonation categories. Indeed such an inventory is extremely rare, with only a couple of languages in PHOIBLE, Igbo and Santali, having all six laryngeal categories. Since there are no comprehensive acoustic studies on this language it could not be included in the analysis here. Therefore, I suggest that the lack of a clear pattern may have more to do with the composition of the dataset used and not a genuine correlation between the makeup of Burushaski and its use of the MDS space.

So does this mean that larger stop inventories do not use more of the MDS space, contra the result for vowel phonation types reported by Keating et al. (2023)? To directly test this, a language-level measure of spread was computed by summing the standard deviations of each language's stops across all four MDS dimensions (MDS1-MDS4). This measure captures the overall extent to which stops are distributed in the MDS space. A linear regression model was then fitted with spread as the dependent variable and inventory size as the predictor, with each language contributing a single data point.

Results from the model are shown in Table 5.8. As seen in this table, the model showed a significant positive relationship between inventory size and spread. That is, larger inventories are associated with greater dispersion in the MDS space, such that each additional unit increase in size corresponds to an increase of ≈ 0.45 units in spread. These results suggest that larger stop inventories do tend to exhibit greater acoustic dispersion, consistent with predictions from dispersion-based theories.

Figure 5.6 shows the relationship between acoustic spread and inventory size for each language, with individual points representing language-level observations. The line represents

	Estimate	Std. Error	t value	Pr(>|t|)	Significance
Intercept	8.191	0.718	11.413	< 0.001	***
size	0.454	0.114	4.005	< 0.001	***

Table 5.8: *Results from the model testing the relationship between inventory size and acoustic dispersion.*

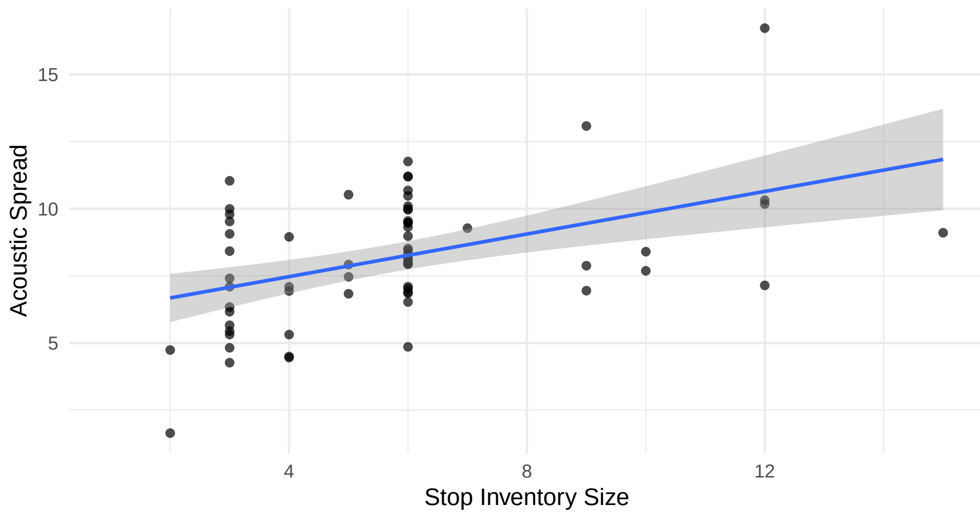


Figure 5.6: *Inventory size versus acoustic dispersion.*

the fitted linear regression line with a shaded 95% confidence interval. The plot shows a positive association, with acoustic spread increasing as inventory size increases, indicating that languages with larger stop inventories tend to exhibit greater dispersion in the MDS space.

5.3.4 Do we need +VOT and −VOT?

One result that is apparent from the MDS analysis is that the dimension most strongly associated with laryngeal category, MDS2, broadly classifies stops as voiced and voiceless. In fact, it resembles the traditional VOT continuum with prevoiced stops at one end and aspirated stops on the other, with the neutral stops lying in the middle (see Figure 5.2). However, the MDS analysis treats positive and negative VOT as separate correlates. It is conceivable that this split is not required and the traditional definition of VOT as a continuum going from negative to positive values along a single axis will produce a comparable MDS solution.

To assess whether splitting the VOT into separate positive and negative components was warranted, I ran a separate MDS analysis on the same smaller dataset. The only difference was that the VOT measure was combined into a single measure, in accordance with traditional conceptions of VOT (Lisker and Abramson, 1964).

Table 5.9 shows the rotated weights from the MDS solution with a single VOT measure. This solution revealed a different MDS structure. The single-VOT model explained 58.5% of

	MDS1	**MDS2**	**MDS3**	**MDS4**
VOT	0.180	-0.576	1.279	-0.103
Cf0	-0.110		0.511	
ΔCf0		-1.014	0.166	0.219
CD				0.418
Burst intensity	-0.102		0.298	-0.348
Burst duration	0.525			-0.284
CPP	0.152	1.103	-0.125	
H1-H2		-0.533		-0.218
COG	-0.444	-0.124	0.206	-0.128
SD	-0.558			0.196
Skew	0.644		-0.150	0.123
Kurtosis	0.792		-0.156	
F2	0.329			
Variance (%)	15.6%	22.2%	16.4%	4.3%

Table 5.9: *Weight of each acoustic measure on each dimension of the 4-D MDS solution with a single VOT measure.*

the variance across four dimensions, with MDS2 capturing the strongest signal (22.2%), primarily driven by ΔCf0 and H1-H2. MDS1 (15.6% variance) was defined by spectral moments, while MDS3 (16.4% variance) was dominated by VOT and Cf0. By contrast, the model with both positive and negative VOT explained substantially more variance overall (83.8%) and exhibited a cleaner, more phonetically interpretable structure. Here, MDS1, the voicing dimension, was the most dominant (34.0% variance). MDS2 (16.7% variance) captured spectral shape, while MDS3 (28.7% variance) was dominated by CPP and ΔCf0. The comparison demonstrates that collapsing VOT into a single measure spreads voicing-related variance across multiple dimensions. The split-VOT model explains 25.3% more of the total variance, and suggests that voicing, rather than spectral shape as in the model with a single VOT measure, is the primary organizing dimension in this dataset. This is in line with the expected result since the smaller dataset was focused on laryngeal distinctions, ignoring some place distinctions, as shown in Table 5.6.

5.4 Conclusion

This chapter has shown that the acoustic space of stop contrasts broadly consists of four dimensions – two dimensions that capture place contrasts and two that capture laryngeal contrasts. Thus, the makeup of the acoustic space of stops is comparable to that of vowels, which is made up of a two-dimensional place contrast ($F1-F2'$) and a two-dimensional voice quality contrast (the complex space of (Keating et al., 2023)). The stop space is not as separable into place and laryngeal dimensions since the analysis is based on a much larger sample of languages and models both place and laryngeal contrasts in the same solution.

This chapter has presented the first study to establish the cross-linguistic acoustic space for stop contrasts and provides a low-dimensional parameterization of their complex acoustic space, which can be used in other analyses.

5.5 Future Work

The analysis of language-specific effects on the MDS space did not return interpretable results. This is likely due to the nature of dataset, which was constructed by imputation from measured values. As discussed in Section 5.2, the studies from which the acoustic data were compiled varied to a great degree in methodology. Future work will be conducted on a more controlled dataset, with primary acoustic measurements as in Keating et al. (2023) rather than imputing them based on compiled data. The analysis of such a dataset is more likely to reveal consistent effects of language and inventory type on the acoustic space of stop contrasts.

CHAPTER 6

Stop inventories in acoustic space

Lack of imagination is not itself an argument.

– DAVID GRAEBER & DAVID WENGROW,
The Dawn of Everything

This chapter presents a way of evaluating two different aspects of inventory attestedness – whether a given inventory is attested at all, and if it is, how frequently it is attested among the world's languages. I show that the factors that determine whether an inventory is possible differ from those that determine whether it is preferred across languages. I evaluate stop inventories in terms of their acoustic properties and show that whether an inventory is attested is governed by its dispersion, articulatory effort, size and acoustic symmetry. However, of these, only the dispersion-theoretic factors – dispersion, articulatory effort and size – govern whether an inventory is common or rare. This parallels the findings from the MaxEnt model in Chapter 3, which found that featural economy is not required as a separate constraint, although its effects emerge from the three dispersion-theoretic constraints. Thus, the results presented in this chapter establish similar patterns for stop inventory typology in terms of their acoustic properties, as Chapter 3 did in terms of their featural properties.

6.1 Background

In contrast to the model presented in Chapter 3, the model presented in this chapter operates in the acoustic, rather than featural, space. It thus considers the MDS space parameterized in Chapter 5 as the acoustic space of stop contrasts, from which languages draw their inventories. It evaluates the effect of four factors on how languages devise their stop inventories. The methodology is described in detail in Section 6.2.

The first factor included here is inventory size, which is the same as that in the MaxEnt model in Chapter 3, except, instead of being implemented as a constraint, it is simply included as a predictor. Size, or the number of contrasts in an inventory, is a dispersion-theoretic predictor that represents the functional tendency towards larger inventories so as to allow for a sufficiently expressive lexicon (Lindblom, 1986, 1990). Therefore, we expect this predictor to have a positive effect on frequency of attestation.

The second factor is maximizing dispersion. In contrast to the MaxEnt model, here dispersion is measured as the distance in the 4D acoustic space. Like the MaxEnt model, two measures of dispersion, total and minimum, are calculated. Since we expect a preference for more dispersed inventories, dispersion is expected to have a positive effect on inventory frequency. Recall, also, that the MaxEnt model found that minimum, rather than total, dispersion captured the distribution of inventories better in featural space. We may expect this to be the case in acoustic space as well.

The third factor is minimizing articulatory effort, which in the MaxEnt model was evaluated by a set of markedness constraints. Here, effort associated with an inventory is calculated as the sum of the effort involved in producing each stop in it. More details are provided in Section 6.2. Since we expect languages to prefer simpler stops, we expect a negative effect of articulatory effort on inventory frequency.

The final factor is symmetry in the acoustic space, which is analogous to feature economy in the acoustic space. As noted in Chapter 1, statistical studies of inventory databases have found support for the effect of structural constraints on inventory composition, in the form of feature economy (Dunbar and Dupoux, 2016; Mukherjee et al., 2007; Nikolaev and Grossman, 2020) and feature symmetry (Dunbar and Dupoux, 2016) in both consonant and vowel inventories. In acoustic terms, a large-scale acoustic investigation of vowels found evidence for adaptive dispersion, with a minor effect of symmetry (Becker-Kristal, 2010).

Note, however, that the acoustic space is not obviously symmetrical, like the featural space.

Features lend themselves well to symmetry as they are typically binary[1]. Acoustic space, by contrast, is not as obviously symmetrical. The acoustic space of stops, in particular, as formulated in Chapter 5 and shown in Figure 6.1, is quite complex and it is not immediately apparent what a symmetric inventory may look like in this space.

Take the inventory /p t k/, for example, which is symmetrical in featural terms. The 4D co-ordinates of the category centroids of this inventory are shown in Table 6.1. Even though

Stop	**MDS1**	**MDS2**	**MDS3**	**MDS4**
p	-1.17	-0.62	-2.60	2.40
t	-3.96	0.33	0.54	-1.58
k	3.01	3.25	-1.11	-0.92

Table 6.1: *MDS co-ordinates of /p/, /t/ and /k/ in the 4D space.*

the three categories are approximately evenly spaced along dimensions MDS1 (the distance between p-t ≈ the distance between p-k) and MDS4 (the distance between p-k ≈ the distance between t-k), their co-ordinates along the two other dimensions do not display such symmetry. Moreover, the symmetrical pairs of the categories along MDS1 and MDS4 are contradictory – the symmetrical partner of /t/ (at -3.96) along MDS1 is /k/ (at 3.01); but along MDS4, the symmetrical partner of /t/ (at -1.58) is /p/ (at 2.40).

Compare this to the case of vowels, which exhibit symmetry in the acoustic space (Becker-Kristal, 2010). Consider the vowel acoustic space shown in Figure 6.2. This figure shows the centroids 22 of the 28 IPA vowels in the $F_1 - F_2'$ space from Becker-Kristal (2010). The six remaining vowels[2], which occur in fewer than 5% of the languages in PHOIBLE, are not shown to make the figure more legible. Note that the figure shows log-transformed values of F_1 and F_2', since Becker-Kristal (2010) found evidence for acoustic symmetry when the formants were log-transformed.

Consider the pair of vowels /i/-/u/. This pair, which differs in backness and shares height, is "symmetrical" in the sense that their positions along F_2', are more or less equidistant from a

[1]Even in the case of privative features, there are exactly two possibilities for a feature value.

[2]These are /ɜ/, /ɘ/, /ɵ/, /ʏ/, /ɞ/ and /ɶ/.

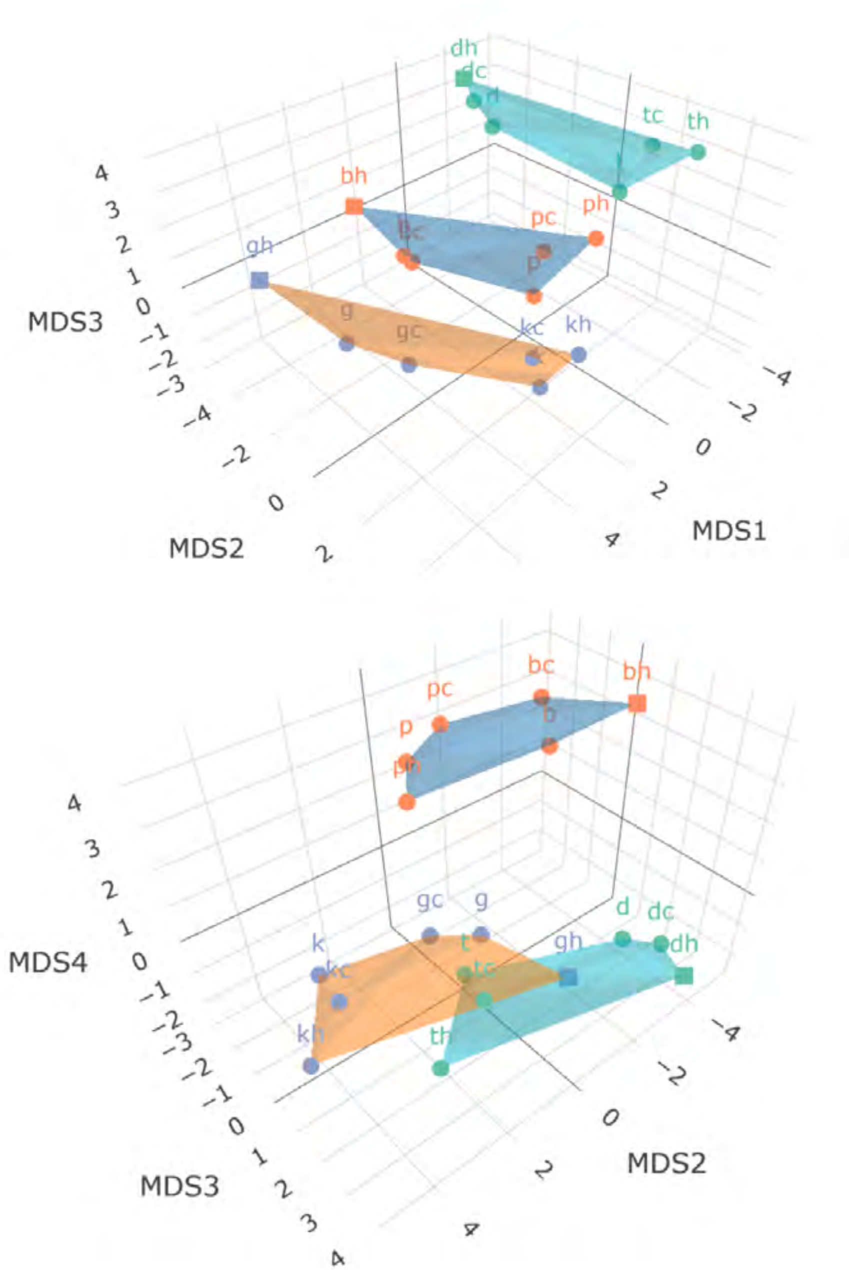


Figure 6.1: *The phonetic space of stop contrasts.*

central high vowel like /ɨ/ or /ʉ/. This is analogous to the positions of /t/ and /k/ being approximately symmetrical about /p/, along the MDS1 dimension. Notice, however, the difference between the stop and vowel space is apparent along the dimensions that are not relevant for a pair of contrasts. In the case of vowels, the pair /i/-/u/ does not differ in height and the

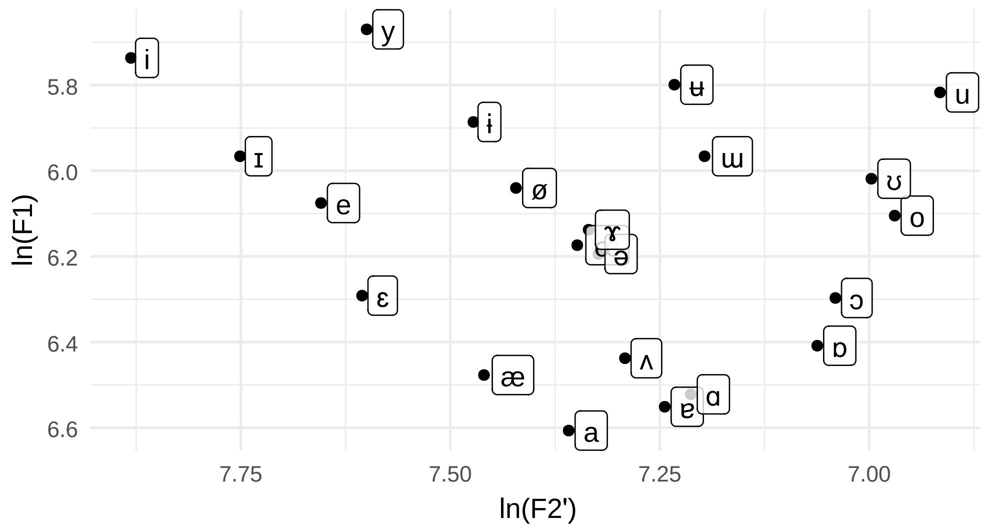


Figure 6.2: *The phonetic space of vowels from Becker-Kristal (2010).*

positions of these vowels along F_1 are comparable. Indeed, this is what makes them symmetrical. However, this is not the case for stops, as is apparent in Table 6.1. The stops /t/ and /k/ do not differ in terms of voicing (represented by MDS2) or laryngeal quality (represented by MDS3), but they still have very different positions along these dimensions. This is because the dimensions of the stop space are more complex, compared to the vowel space. Even though MDS2 mainly represents voicing contrasts, it does not do so exclusively. In fact, the biggest contributor to this dimension is VOT, which although a correlate of voicing, is also an acoustic correlate of stop place of articulation (e.g., Cho and Ladefoged, 1999). Thus, the dimensions of the stop acoustic space are not as readily separable as the vowel space. Given this complex structure of the stop acoustic space, we do not necessarily expect featurally symmetrical stop inventories to also be acoustically symmetrical.

In addition to characterizing stops in terms of their acoustic properties, the approach taken in this chapter differs from previous work on inventories in a key way. Whereas in previous work the question of why some inventories occur and others don't is addressed implicitly, the comparison is made explicit here. Consider the approach taken by Liljencrants and Lindblom (1972), who modeled vowel inventories by simulating vowel systems in a transformed $F_1 - F_2$ space such that the optimal vowel inventory of a particular size was one that maximized the perceptual distance between each pair of vowels. They compared, for each inventory size, the systems identified as optimal by the simulation with the most commonly attested vowel systems of that size. For example, they showed that the three-vowel system with the largest perceptual separation in $F_1 - F_2$, /a-i-u/, is the most commonly attested three-vowel inventory among the world's languages. Thus, Liljencrants and Lindblom (1972) argued that maximizing perceptual contrast accounts for the structure of the most commonly attested vowel inventories.

However, optimal inventory composition is variable – while some systems are clear winners, others only marginally outperform alternatives. Moreover, many logically possible combinations of segments are never attested within a single inventory. In this chapter, I explicitly compare these two aspects, attestedness and optimality, in both stop and vowel inventories. What determines whether a particular combination of stops or vowels occurs at all? And among the combinations that do occur, what makes some more likely than others?

6.2 Method

6.2.1 Data

As in previous chapters, all the data come from PHOIBLE (Moran and McCloy, 2019), a comprehensive, openly accessible cross-linguistic repository of 3020 phonological inventories, based on data from descriptions and other databases. The general methodology employed here is the following. A set of n segments can yield a possible set of 2^n inventories, assuming each segment is either present or absent in any given inventory. The task is to determine: (1) which properties separate the subset of inventories, of the full set of 2^n, that are attested in the languages of the world; and (2) which properties separate the most commonly attested inventories, such as /p t k/, from less common ones, such as /b t k/.

It was assumed that inventories can have at most 18 stops (3 places and 6 laryngeal settings). These were assumed to have the organization in Figure 6.1, formulated in Chapter 5. Since there are 18 possible stops, we can generate 2^{18} possible stop inventories. This far outnumbers the total number of languages in the world. If the contents of inventories were random, every language, in principle, could have a unique inventory. This, however, is not the case and we find that the same inventories appear over and over again in languages. The task here is to uncover the underlying organizational principles that lead to the uniformity in the contents of attested inventories. In other words, the task is to uncover what separates the blue curve in Figure 6.3, showing the actual distribution of attested stop inventories, from the orange curve, showing the possible distribution of stop inventories.

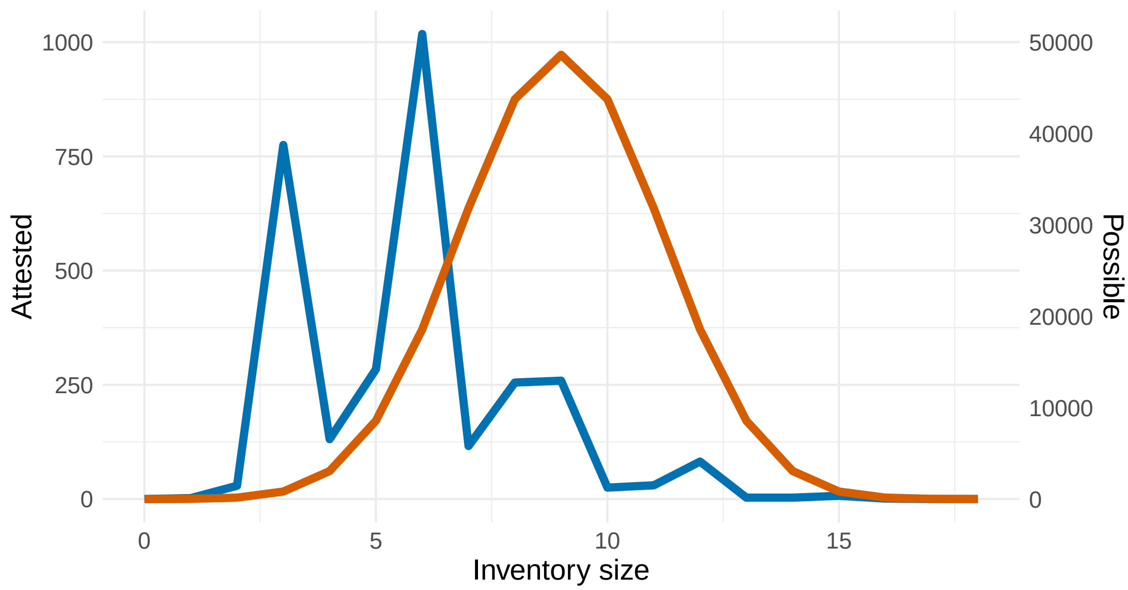


Figure 6.3: *Attested versus possible stop inventories.*

Thus, the idea is that every language picks a subset from this set of 18 stops and we want to determine what separates attested stop systems from unattested ones, and what explains the higher frequency of the most common stop systems compared to those that are attested but rare. Each of the 2^{18} inventories was specified as either having or not having each of the 18 stops with 1s and 0s (representing presence and absence respectively). An illustrative table is shown in 6.2 below.

The first part of the table shows attested inventories. It shows that there are 801 PHOIBLE stop inventories that have /p t k b d g/, 752 that have /p t k/ and so on. The second part of the table shows possible but unattested inventories. As shown here, inventories with no stops, all 18 stops, /p dh gc/ and /p pc b bc t tc d dc k kc g gc/ are never attested. The vast majority of the dataset consists of inventories of the second type, which are never attested,

p	ph	pc	b	bh	bc	t	th	tc	d	dh	dc	k	kh	kc	g	gh	gc	freq
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	801
1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	752
1	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	0	138
1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	92
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Table 6.2: *Example table showing the stop data.*

with frequencies of zero.

Each inventory was defined as a vector of 1s and 0s representing the presence and absence of each of the 18 stops. Thus, every entry in Table 6.2 was converted into a vector, yielding a total of 2^{18} inventories as shown in (6.1). Dispersion, articulatory effort, symmetry and size were calculated for each vector.

(6.1) a. $L_1 = [1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0]$

b. $L_2 = [1\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0]$

c. $L_3 = [1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 0]$

d. $L_4 = [1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0]$

e. ...

f. ...

g. ...

h. $L_n = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]$

The first predictor was inventory size, which represents the number of stops in a given inventory. This is equivalent to the constraint MAXIMIZECONTRASTS in Chapter 3. A few example inventories with their associated frequencies and sizes are shown in Table 6.3. L_1 represents the inventory /p t k b d g/ and thus has a size of 6, L_2 represents /p t k/ and has a size of 3, and so on.

Inventory vector	Frequency	Size
$L_1 = [1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0]$	801	6
$L_2 = [1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0]$	752	3
...	...	...
$L_n = [1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1]$	0	18

Table 6.3: *Example table showing stop inventories, their frequencies of occurrence and sizes.*

Two measures of dispersion were calculated. The first, which I will refer to as "total dispersion," was calculated as the sum of pairwise Euclidean distances between the stops of the vector in the four-dimensional space in Figure 6.1, following Liljencrants and Lindblom (1972) and subsequent work. Thus, the total dispersion of each stop inventory is given by (6.2).

(6.2)

$$r = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 + (w_i - w_j)^2}$$

As in Chapter 3, instead of measuring the total pairwise distances, "minimum dispersion" was based on the minimum distance between any pair of contrasts in an inventory (Becker-Kristal, 2010; Flemming, 2005). The minimum dispersion of a given inventory was calculated as

(6.3)

$$r_{\text{min}} = \min_{i,j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 + (w_i - w_j)^2}$$

Table 6.4 shows a handful of illustrative inventories and their associated frequencies and distances. In this table, L_1 represents /p t k b d g/, L_2 represents /p t k/ and L_n represents the inventory with all 18 stops.

Next, the articulatory effort associated with each inventory was calculated as the total articulatory cost in accordance with Table 6.5. The cost in this table represents the penalty for the use of effortful articulations in an inventory. This includes the use of any laryngeal articulation – glottal vibration (i.e. voicing), glottal spreading (i.e. aspiration) and glottal constriction (i.e. constriction). I assume that the concurrent articulation of two laryngeal maneuvers

Inventory vector	**Frequency**	r	$r_{\min}$
$L_1 = [1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0]$	801	108.59	4.59
$L_2 = [1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0]$	752	20.39	5.86
...	...	...	...
$L_n = [1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1]$	0	1169.31	1.13

Table 6.4: *Example table showing stop inventories, their frequencies of occurrence and dispersion.*

requires an articulatory effort greater than the effort involved in the articulation of each laryngeal maneuver separately. For instance, the effort required to produce a voiced-aspirated stop is greater than the sum of the effort involved in producing a voiced stop and a voiceless-aspirated stop. Therefore, voiced-aspirated and voiced-constricted stops are associated with specific articulation costs beyond the costs associated with voicing, aspiration and constriction.

Articulatory maneuver	**Cost**
Voicing	1
Aspiration	1
Constriction	1
Voiced-aspiration	1
Voiced-constriction	1
Labial closure	1
Denti-alveolar closure	1
Velar closure	1
Voiced velar	1
Voiceless-constricted labial	1
Voiced-constricted velar	1

Table 6.5: *Cost associated with each articulatory maneuver.*

It was assumed that each place is associated with an articulatory cost, but this does not differ by location of the closure. Finally, combinations of place and laryngeal features that are known to be effortful were assumed to incur an additional cost. There is a dispreference for voicing at posterior places due to its aerodynamic disadvantage, since it is difficult to maintain voicing when the space between the glottis and the closure is small (Keating, 1984a; Westbury and Keating, 1986). Bilabial ejectives are also articulatorily marked, for the exact opposite reason. The closure at the bilabial place requires more intra-oral pressure than at other places since the lips are both active articulators and more volume must be filled with air. Note that even though markedness arising due to perceptual properties, such as the diffuse spectrum and lower intensity of bilabial stops making them more difficult to be perceived, was not included in the cost in Table 6.5, their inclusion would not significantly change the markedness of the various stops relative to each other.

Table 6.6 shows the effort associated with the same example inventories. L_1, which represents /p t k b d g/, has an articulatory cost of 11 since it employs voicing in three of its stops, yielding three units of effort; /p/ and /g/ which are especially effortful, yielding 2 more units; and the use of each of the three places twice yields 6 additional units of effort. This leads to a total of 11 units of articulatory effort for L_1. The articulatory effort associated with each inventory in the dataset was calculated in this manner.

Inventory vector	**Frequency**	**Effort**
$L_1 = [1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0\,1\,0\,0]$	801	11
$L_2 = [1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0]$	752	4
...	...	...
$L_n = [1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1]$	0	51

Table 6.6: *Example table showing stop inventories, their frequencies of occurrence and articulatory effort.*

Finally, the following method was used to quantify how symmetrical an inventory was. The symmetry penalty of each inventory was calculated as the degree of asymmetry exhibited by its stops in the 4D space. For an inventory to be perfectly symmetric, each stop p_i must have a partner at $-p_i$. Therefore, the symmetry penalty of an inventory is given by:

(6.4)
$$P = \frac{1}{n}\sum_{i=1}^{n}\min_{j\neq i}\sqrt{(x_j + x_i)^2 + (y_j + y_i)^2 + (z_j + z_i)^2 + (w_j + w_i)^2}$$

For each point p_i, the formula in (6.4) finds the minimum distance from its perfect pair ($-p_i$) to any other point (p_j) in the inventory and returns the average of these distances. In other words, it returns the average distance that represents how far each point in an inventory is from its perfectly symmetrical partner in the acoustic space. This yields a score which has a value of zero for a perfectly symmetric inventory. Larger values indicate worse overall symmetry, while values closer zero indicate good symmetry. The symmetry of an inventory was then taken to be the negative value of the symmetry penalty given by 6.4.

Table 6.7 shows example inventories with their associated frequencies and symmetry. Recall that acoustic symmetry is separate from feature economy. Even though the inventories L_1 and L_2 are perfectly economical, they are not symmetrical in the acoustic space. However, the symmetry of all three inventories is still comparable.

Inventory vector	**Frequency**	**Symmetry**
$L_1 = [1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0]$	801	-4.19
$L_2 = [1\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0]$	752	-4.71
...	...	...
$L_n = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]$	0	-4.00

Table 6.7: *Example table showing stop inventories, their frequencies of occurrence and symmetry penalties.*

The full dataset was assembled in this manner. To determine how the four factors – dispersion (total or minimum), articulatory effort, symmetry and size – influenced the content of stop inventories, the attested frequency was modeled as a function of these factors, as described in Section 6.2.2. Notice that inventory size is correlated with the other factors, such that larger sized inventories have greater values of total dispersion and effort. Therefore, the values of these variables were normalized within a given size to the $[0,1]$ range using

(6.5)

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

This reduces the influence of inventory size and ensures that these predictors capture structural properties of inventories rather than differences arising solely from the number of stops. This was crucial as inventory size was also included as a separate predictor in the model. Note that the values of minimum distance and symmetry do not similarly increase with inventory size. Therefore, these were not normalized relative to size.

Following size normalization, all predictors were scaled to the $[0,1]$ range so that their coefficients were comparable. As the symmetry measure is defined on a negative scale, this transformation reverses its interpretation: a value of 0 corresponds to the most asymmetrical inventory in the dataset, whereas a value of 1 corresponds to the most symmetrical inventory. This must be kept in mind when interpreting results from the statistical analysis.

6.2.2 Statistical Analysis

To model the relationship between the dependent variable, inventory frequency, and the predictors, dispersion, articulatory effort, symmetry and size, a hurdle model (Cragg, 1971; Howell et al., 2017; Mullahy, 1986) was implemented using the R package `pscl`. Two separate models were run corresponding to the two ways of measuring dispersion, total and minimum. The performance of these models was compared using AIC.

Hurdle models are a class of models for count data with excess zeros and over-dispersion, and are typically used to model rare events. The events being modeled here are "rare" in that when creating an inventory by sampling from the full set of stops, the probability of creating an inventory that occurs in a real language is extremely low, compared to the inventories that are combinatorially possible from segments attested in languages. The creation of an attested inventory is, thus, a rare event from the point of view of modeling. That is not to say that languages arrive at their inventories by randomly sampling from a predetermined set of segments. Indeed, the model presented here, like Liljencrants and Lindblom (1972) and subsequent work in that tradition, is not meant to be a model of *how* languages arrive at their inventories, but of *why*.

Hurdle models are well suited to the structure of the dataset under consideration here for two reasons. First, the response variable – the frequency with which each inventory type is attested in the data, consists of count data, which violates the assumptions of ordinary least squares regression. Second, the dataset consists of two types of inventories, attested and unattested, with the latter type being far more common. The hurdle model addresses this imbalance by fitting two components separately – a binomial logistic regression modeling the probability that an inventory is attested at all, and a zero-truncated negative binomial regression modeling frequency conditional on attestation. For the second part of the model, the negative binomial distribution was chosen over a different distribution that can model count data (e.g., Poisson), given the over-dispersion in the inventory frequency data.

Conceptually, the hurdle model first addresses what makes an inventory possible at all. The idea is that a set of segments can become an attested inventory that is associated with positive counts once some hurdle is cleared. If the hurdle is not cleared, that set of segments is associated with a count of zero, meaning the model predicts that this set cannot occur as inventory of a real language. The second part of the model captures what happens after the hurdle is cleared – among the attested inventories, what explains the high frequency of occurrence of some inventories and the rarity of others?

6.3 Results

Recall that the hurdle model separates the analysis into two parts. The first, called the "zero model," addresses what predicts whether an inventory is attested at all. The second, called the "count model," addresses what predicts how common the attested inventories are. In other words, the first part models the predictors that explain why certain inventories are attested in the world's languages and certain others are not, without considering how common the attested inventories are. The second part models why some attested inventories are more common than others.

Model comparison between the model with total dispersion ($AIC = 2350.906$) and minimum dispersion ($AIC = 2510.032$) showed that minimum dispersion captured the data better, in line with results from the MaxEnt model presented in Chapter 3. This also mirrors the tendency of dispersion in vowel inventories to be based on the bottleneck (Becker-Kristal, 2010). Table 6.8 shows the results from the hurdle model for stops, based on minimum dispersion. Results from the model with total dispersion are included in the Appendix.

In the zero component, all the predictors significantly influenced the probability of a non-zero outcome. Minimum distance had a positive effect, suggesting stop inventories whose most confusable pair of contrasts are well dispersed are more likely to be attested. Articulatory effort had a very strong negative effect, indicating that stop inventories that are articulatorily effortful are less likely to be attested. The effects of articulatory effort and dispersion on

Zero hurdle model – attestedness					
Predictor	Estimate	Std. Error	z value	Pr(> |z|)	Significance
(Intercept)	1.1533	1.0876	1.060	0.289	n.s.
dispersion	9.1556	0.7062	12.965	$< 2 \times 10^{-16}$	***
effort	-19.7617	0.8465	-23.345	$< 2 \times 10^{-16}$	***
symmetry	-6.0214	1.3244	-4.546	5.46×10^{-6}	***
size	3.2713	0.6775	4.828	1.38×10^{-6}	***
Count model – frequency					
Predictor	Estimate	Std. Error	z value	Pr(> |z|)	Significance
(Intercept)	-16.151	49.997	-0.323	0.747	n.s.
dispersion	11.898	1.799	6.614	3.75×10^{-11}	***
effort	-7.774	2.249	-3.456	0.0006	***
symmetry	4.130	2.665	1.550	0.121	n.s.
size	7.119	2.095	3.398	0.0007	***

Table 6.8: *Hurdle model summary for stops with minimum dispersion.*

attestedness are in line with Lindblom's (1986) theory of *Adaptive Dispersion.* Symmetry had a negative effect, suggesting that more symmetrical inventories are *less* likely to be attested. Once again, recall that symmetry here refers to geometric symmetry in the acoustic space, about which we did not have any a priori expectations. However, to the extent that geometric symmetry reflects feature economy, this result is the opposite of what would be expected per Ohala (1980) and Clements (2003), even though the count component found the expected direction of effect, which was not significant. Finally, size had a small positive effect on attestedness, meaning real stop inventories have a tendency to be larger rather than smaller, compared to unattested ones.

In contrast to these results, in the count component, only minimum dispersion, articulatory effort and size had statistically significant effects. This is in line with the MaxEnt results from Chapter 3, which found that the phonological constraints corresponding to these three variables adequately captured the frequencies of attested and unattested inventories. Moreover, this model found that the effects of economy and symmetry emerged from the interaction of the three phonetically defined constraints. Similarly, in the count component of the hurdle model, symmetry had a positive effect, meaning more common inventories exhibit more symmetry, but it was not significant. This suggests that even in the acoustic space, the structural parameter of symmetry is not necessary to explain the frequency distribution of attested inventories. By contrast, the parameters reflecting functional pressures – maximizing dispersion, minimizing effort and maximizing the number of contrasts – sufficiently explain why some stop inventories are more common than others.

Since the hurdle model treats attested and unattested inventories differently, we can probe how robust its results are to analyst error or typologically odd inventories that happen to come into existence. Is it possible that an inventory that really does not exist was erroneously described or recorded? Are the one-off inventories in PHOIBLE exerting too much of an influence on the results of the hurdle model? To test this, I randomly sampled 25% of the stop inventories that only occur once in PHOIBLE and removed them from the model. There are 84 such unique inventories, only 63 of which were retained in the model. The remaining 21 were assigned frequencies of zero and added back to the dataset[3].

The results from this model are shown in Table 6.9. The coefficients in this table are very

Zero hurdle model – attestedness					
Predictor	Estimate	Std. Error	z value	$Pr(>\|z\|)$	Significance
(Intercept)	0.7187	1.1779	0.610	0.542	n.s.
dispersion	9.7002	0.7451	13.019	$< 2 \times 10^{-16}$	***
effort	-19.7903	0.9036	-21.901	$< 2 \times 10^{-16}$	***
symmetry	-5.9696	1.4260	-4.186	2.84×10^{-5}	***
size	3.6935	0.7129	5.181	2.21×10^{-7}	***
Count model – frequency					
Predictor	Estimate	Std. Error	z value	$Pr(>\|z\|)$	Significance
(Intercept)	-15.364	41.951	-0.366	0.714	n.s.
dispersion	10.798	1.952	5.532	3.17×10^{-8}	***
effort	-7.513	2.369	-3.172	0.002	***
symmetry	4.667	2.856	1.634	0.102	n.s.
size	6.163	2.178	2.830	0.005	**

Table 6.9: *Hurdle model summary for stops with 25% of one-off inventories assigned zero frequencies.*

similar to those in the original model, shown in Table 6.8. As in the original model, all four factors significantly influence the possibility of a non-zero outcome. Larger, more dispersed, asymmetrical and articulatorily simpler inventories are more likely to occur in languages. The

[3]A model that left these out of the dataset entirely also returned nearly identical results.

count component also found results comparable to those from the original model. In this component, the effect of symmetry was not significant, meaning that only the dispersion-theoretic predictors suffice to explain the greater preference for some inventories across the world's languages.

Since the results in Table 6.9 are a very close match to those from the model with all the data included (shown in Table 6.8), the results from the hurdle model are robust against one-off oddities.

6.4 Conclusion

This chapter has evaluated how four factors that have been proposed to explain the contents of inventories, dispersion, articulatory effort, symmetry and inventory size, influence stop inventories. The analysis presented here addressed this in two parts – what makes an inventory possible and what makes it preferred?

First, it was shown that minimum, rather than total, dispersion plays a central role in both aspects – it influences whether a stop inventory is attested at all, and positively affects the count frequencies of the attested inventories. Stop inventories that maximize the distance between the pair of stops closest in acoustic space are more likely to be attested and more common.

I also showed that the hurdle for stop inventories is characterized by all four variables – high minimum dispersion, low effort, greater inventory size and *lower* symmetry in acoustic space. Larger, more dispersed stop inventories in terms of minimum dispersion that consist of stops that are easy to produce and are acoustically asymmetric are more likely to be attested, rather than unattested. The second part of the model showed that the higher frequencies of the most commonly attested inventories were driven by maximizing size and minimum distance, and minimizing articulatory effort.

These results mirror the results from the MaxEnt model in Chapter 3. The count component of the hurdle model is more directly comparable with the MaxEnt model, as both of these model inventory frequency, albeit in slightly different ways. Recall, also, that the MaxEnt model operates in the featural space and the hurdle model in the acoustic space. Despite these differences, both models found that the dispersion-theoretic functional factors – maximizing dispersion and inventory size, and minimizing articulatory cost – adequately capture inventory frequency. This suggests that structural constraints like featural economy and symmetry are not required as predictors and their effects may emerge from the aforementioned phonetically defined predictors.

The method introduced in this chapter also shows that the factors that affect inventory content differ in terms of attestedness versus frequency of occurrence. The factors that determine whether an inventory is possible may differ from those that determine whether it is preferred.

CHAPTER 7

Searching the acoustic space

All models are wrong, but some are useful.

– George E. P. Box

In this chapter, I present a model of stop inventories in terms of their acoustic properties. The problem of mapping the continuous acoustic space on to a discrete number of categories to form an inventory is conceptualized as a search problem. The idea is that individual languages search the same cross-linguistic acoustic space to devise their inventories. The model presented in this chapter is based on genetic algorithms, a class of evolutionary algorithms inspired by the principles of biological evolution and used in search and optimization problems. This model is not meant to simulate processes of acquisition or sound change, but to investigate the factors that influence the contents of inventories in terms of their phonetic properties.

7.1 Background

7.1.1 Inventory Structure

The hurdle models in Chapter 6 found that the contents of stop inventories are influenced by articulatory cost and featural economy, along with a modest contribution of dispersion. The MaxEnt model in Chapter 3 found that the interaction of constraints based on minimizing articulatory markedness, maximizing dispersion and maximizing the number of stops in an inventory predict economical inventories to be optimal. Thus, according to the MaxEnt model, constraints that penalize the uneconomical use of features in an inventory are not required, since they emerge from the interaction of the other three dispersion-theoretic constraints.

In this chapter, I characterize the influence of these factors on the contents of stop inventories based on the acoustic space of stops developed in Chapter 5. Since the space under consideration is defined in terms of continuous acoustic variables, not features as in in Chapter 3, I use geometric symmetry (Dunbar and Dupoux, 2016) as a measure of how "economically" a given inventory utilizes the space. Additionally, I evaluate the hierarchical nature of inventories – the tendency for more complex articulations to be employed only after simpler articulations have been employed maximally (Lindblom and Maddieson, 1988).

Lindblom and Maddieson (1988) proposed that as the size of an inventory increases, a basic phonetic space must first be expanded and saturated. Beyond this saturation point, the space itself must become more complex by incorporating additional dimensions of contrast. They evaluated the contents of consonant inventories and noted that obstruents and sonorants can be classified into three sets. They defined Set I as the class of consonants that use "basic articulations." This set consists of consonants that are phonetically simple. Table 7.1 shows the obstruents of Set I.

p	t	k	ʔ
b	d	g	
f	s		h
	tʃ		

Table 7.1: *Set I obstruents proposed by Lindblom and Maddieson (1988).*

Set II consonants are described as having "elaborated articulations" such as breathy voice and creaky voice, places beyond those in Table 7.1, as well as ejectives, implosives and clicks. Finally, consonants in Set III are described as involving "complex articulations" that combine at least two Set II dimensions. For example, the voiced-aspirated retroflex stop /ɖʱ/ is a Set III consonant as it combines the retroflex place with breathy voicing, both of which are Set II articulations.

Thus, Lindblom and Maddieson (1988) proposed that as the size of an inventory increases, the phonetic space available for creating contrasts expands from Set I to Set II and then to Set III. They proposed that such a mechanism drives the tendency of consonant inventories to achieve maximal dispersion at minimum articulatory cost.

Applying this framework to stop contrasts, I classify stops into sets, going from basic articulations to more elaborated and complex ones. This "hierarchy" of the phonetic space of stops is developed in Section 7.2.2. In addition to predicting that inventories that employ marked articulations should also employ unmarked ones, the "saturation" aspect of Lindblom and Maddieson's proposal predicts that inventories that employ higher dimensions should not exhibit gaps in the lower dimensions. Therefore, their proposal can potentially explain the tendency of larger inventories to be more economical, compared to smaller inventories.

Thus, we can evaluate how stop inventory content is influenced by five factors – (i) articulatory cost, (ii) dispersion, (iii) symmetry, (iv) maximizing the number of stops, and (v) the hierarchical nature of articulations in Sets I, II and III. Since we are operating in the continuous acoustic space, the task can be formulated as searching this space for stop categories, subject to the pressures represented by (i)-(v). Since the search space is high-dimensional and continuous, and the pressures in (i)-(v) are potentially conflicting, the problem can be thought of as an optimization problem. How might we search the acoustic space of stops to optimize trade-offs between (i)-(v) to create an inventory? To solve this optimization problem, I use genetic algorithms, a class of evolutionary algorithms, which are used to generate solutions to optimization and search problems.

7.1.2 Simple Genetic Algorithms

Genetic algorithms (GAs) (Holland, 1975; Goldberg, 1989) are population-based stochastic algorithms modeled on mechanisms of biological evolution. They are typically used to find approximate solutions to optimization and search problems. In our case, we are interested in understanding how languages search the space of possible stops to devise their inventories. The search algorithm is presented with this problem – "Given a continuous phonetic space, choose an inventory." This "choice" is subject to several different, often conflicting, demands, as we have seen in preceding chapters. The algorithm must find a way to balance these demands and generate well-formed inventories.

GAs work by simulating natural selection over a population of solutions identified by their *genomes*[1]. The algorithm starts with a random population, $P^{(t)} = \{\mathrm{x}_1^{(t)}, \ldots, \mathrm{x}_n^{(t)}\}$ of n candidate solutions, which evolves over discrete generations $t = 0, 1, 2, \ldots$, each representing a potential solution. Each individual x_i is assigned a scalar fitness value $F(\mathrm{x}_i)$, based on some objective. For example, if the goal is to choose an inventory that minimizes articulatory cost, the population of solutions must be evaluated against this objective. Based on the articulatory complexity of every phoneme in an inventory, the corresponding genome is assigned a fitness value that represents its performance on the objective of minimizing the total articulatory complexity. The function that evaluates the performance of potential solutions on the objective is called the *objective function* or the *fitness function*. The fitness of every genome is evaluated against the objective function and the *fittest* solutions are chosen as *parents* from that generation.

The next generation of solutions is generated from the parents identified in the previous generation. The *offspring* of these parents are generated via two processes – *cross-over* and *mutation*. Figure 7.1 illustrates these processes of producing offspring. Say we start with the

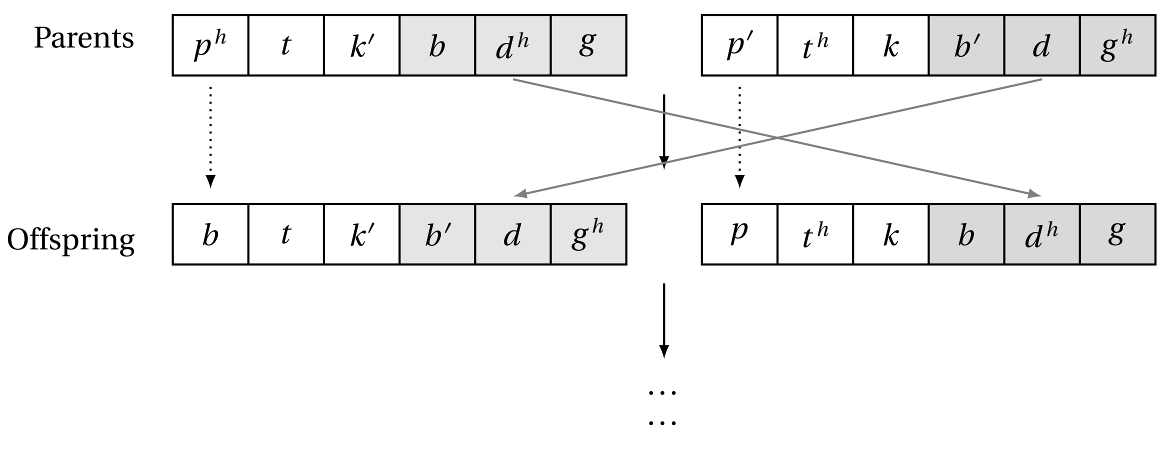


Figure 7.1: *Illustration of cross-over and mutation in a simple genetic algorithm.*

parents /p^h t k’ b d^h g/ and /p’ t^h k b’ d g^h/. Cross-over combines parts of the parents to produce distinct offspring that retain some properties of the parents. In Figure 7.1, the

[1] Note that in literature on genetic algorithms, the terms *genome* and *chromosome* are often used interchangeably. Throughout this chapter I will refer to the identifier of a particular solution as its genome.

grey parts of the genome in the parents' generation are crossed-over in the next generation to produce the offspring /$\mathrm{p^h}$ t k' b' d $\mathrm{g^h}$/ and /p' $\mathrm{t^h}$ k b $\mathrm{d^h}$ g/. Additionally, a mutation occurs at random that changes a part of each parent. In Figure 7.1, the /$\mathrm{p^h}$/ and /p'/ of the parents mutate into /b/ and /p/ respectively. Thus, the offspring in the figure, following cross-over and mutation, end up as /b t k' b' d $\mathrm{g^h}$/ and /p $\mathrm{t^h}$ k b $\mathrm{d^h}$ g/.

In addition to cross-over and mutation, some genetic algorithms preserve top-performing individuals from the previous generation. I will refer to this strategy, which ensures that good solutions are not lost due to cross-over or mutation, as *conservation*, even though the term *elitism* is more common. Thus, generations of potential solutions are evolved by combining parts of good solutions, preserving the especially good solutions and adding occasional changes via mutation into the mix.

Simple genetic algorithms are typically used in problems which have a single objective, even though they may also be used to solve problems with multiple objectives by combining the individual objectives into a scalar objective function (e.g., Ke et al., 2003). However, in cases that have multiple objectives, multi-objective genetic algoritms (MOGAs) (Fonseca et al., 1993) are more suitable. These are discussed in more detail below.

7.1.3 Multi-objective Genetic Algorithms

Many optimization problems involve the simultaneous optimization of multiple, often conflicting, objective functions. Selecting inventories, and even phonologies and grammars more generally, is such a problem. As we have seen in the preceding chapters, inventory content is minimally subject to maximizing perceptual contrast while minimizing articulatory effort. An inventory must find a way to balance these objectives, since minimizing effort often means reducing perceptual distance. So the goal becomes to find a set of optimal trade-offs, as no single solution simultaneously achieves both objectives.

Since no one solution is better than another, as performance is measured on different, conflicting objectives, MOGAs evaluate solutions in terms of *dominance* rather than reducing them to a single fitness value. This way of comparing solutions in a multi-objective setting is implemented with *Pareto optimality*[2]. A solution is said to *dominate* another if it is at least as good on every objective and strictly better on at least one. A solution is considered *Pareto optimal* if no other solution dominates it. The set of all such solutions is called the *Pareto set*, and their corresponding values in the objective space are called the *Pareto front*. These represent the set of best achievable trade-offs in the multi-objective problem. Instead of converging to one optimal solution, the goal in a MOGA is to maintain a population of solutions that approximates the Pareto front.

The Pareto front is illustrated in Figure 7.2. In this figure, each point represents a possible solution characterized by its performance on two competing objectives – f_1 and f_2 – which represent, say, minimizing articulatory cost and minimizing perceptual confusion[3] respectively. Thus, points that are closer to the origin represent better solutions. The filled circles on the staircase-shaped curve form the Pareto-optimal front. That is, these solutions are non-dominated as no other solution can improve one objective without worsening the other. The open circles represent dominated solutions, which are worse than at least one Pareto-optimal point on both objectives simultaneously.

There are various ways of determining the Pareto optimal set, but a common approach is to assign solutions to different ranks based on their dominance relationships (Goldberg, 1989). To assign Pareto ranks, the non-dominated solutions in the population are first identified. These are assigned the rank of 1. Rank 1 solutions are subsequently removed from the population and the same process is repeated to identify the second rank, and so on. In this way, the population is ranked such that lower-ranked solutions are better. These ranks are then used to inform selection of solutions to serve as parents in their generation.

Because many solutions may share the same rank, MOGAs risk the population becoming too concentrated in a small region of the solution space and thus may fail to adequately explore the entire range of possible solutions. One way of avoiding this is by implementing

[2]The notion of pareto optimality may be more familiar to phonologists as *harmonic bounding*.

[3] f_2 represents the objective of minimizing confusion rather than maximizing contrast as it is easier to mathematically devise an algorithm where both objectives are to be minimized or maximized, compared to a case where one objective is to be minimized the other is to be maximized.

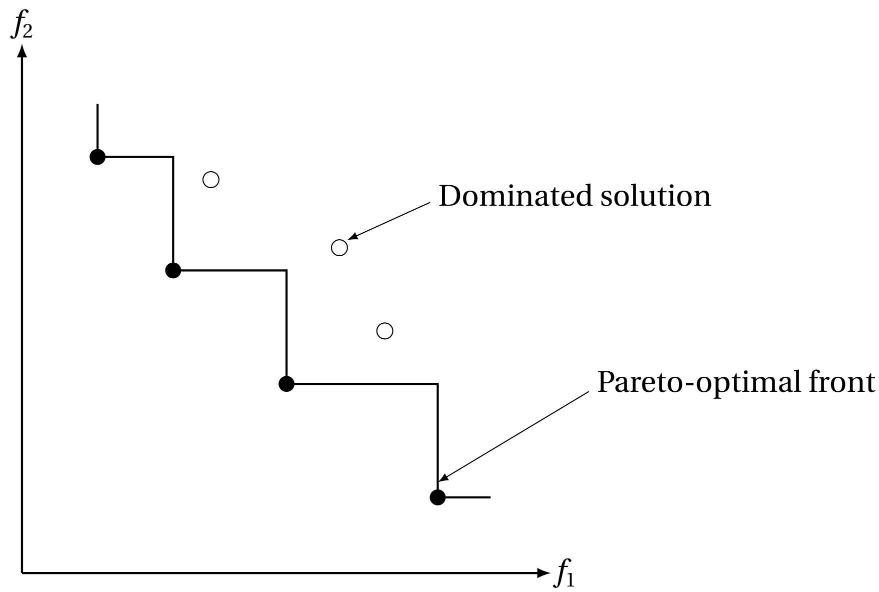


Figure 7.2: *Pareto ranking*

diversity so that solutions in less crowded regions are preferred, since they help maintain a diverse and well-distributed Pareto front.

7.1.4 Summary

In sum, a MOGA proceeds iteratively as follows:

1. An initial population of candidate solutions is generated.
2. Each solution is evaluated on all objectives.
3. Solutions are ranked based on dominance relationships.
4. Diversity is used to inform ranks.
5. A subset of solutions is selected as parents.
6. New solutions are generated through crossover and mutation and by conservation of some parents.
7. A new population is formed and the process repeats until a threshold is reached.

7.1.5 Previous Work

Genetic algorithms enjoyed a brief period of popularity in linguistics in the 90s and early 2000s, when they were used to model vowel and tone inventories (Ke et al., 2003), phonotactic learning (Belz, 1998), constraint ranking in OT (Pulleyblank and Turkel, 1998, 2000), the typology of syllable structure (Redford et al., 2001), and syntactic learning (Clark, 1992) and change (Clark and Roberts, 1993).

The work most relevant to the present case is Ke et al.'s (2003) model of vowel and tone inventories, which showed limited but promising outcomes. Ke et al. (2003) adopted a weighted scalar function in simple GA to combine two criteria influencing vowel inventory organization – perceptual contrast maximization (Liljencrants and Lindblom, 1972) and focalization (Schwartz et al., 1997). They found that the results from their GA matched the most frequently observed three- and four-vowel inventories in UPSID_{451}. The performance of their model on larger inventories, however, did not match the observed typology. It is possible that a multi-objective, rather than simple, GA could have yielded better results.

Ke et al. (2003) did use a multi-objective GA to model tone inventories where the two criteria baked into the fitness function were maximizing Euclidean perceptual distance (as a measure of dispersion) and minimizing articulatory markedness, calculated as a measure of effort associated with tonal complexity. This MOGA predicted tone inventories to be more peripheral than the observed inventories. Since this result suggests that perceptual contrast was weighted as being more important by the model, it is possible that the markedness measure used by Ke et al. (2003) duplicated the effect of perceptual contrast maximization. Moreover, determining the aticulatory cost associated with the production of different tones is non-trivial, and it is possible that a different method of measuring it would have returned more typologically consistent results. Additionally, Ke et al. (2003) did not use conservation as a means to generate children for the new generation, which means they might have possibly abandoned optimal solutions earlier in the evolution.

A better performing GA was implemented by Pulleyblank and Turkel (2000) in simulating the learning of the tongue-root-harmony system of Yoruba. However, they used only mutations and did not use conservation or cross-over in their algorithm, significantly downgrading its power. Their simple algorithm was still able to find the correct ranking of ten relevant phonological constraints.

The body of previous work on the application of GAs to linguistic problems suggests that carefully defined fitness functions and the use of MOGAs show promise for modeling inventories and other linguistic phenomena.

7.2 Method

The task here is to select an inventory by sampling from the continuous acoustic space of stops developed in Chapter 5. Therefore, the optimization problem was formulated as a MOGA that selects a set of 18 points in the 4-dimensional space defined by the Multidimensional Scaling (MDS) coordinates from Chapter 5, shown in Figure 7.3. Each of the 18 points corresponds to a possible stop out of the set [p, t, k, b, d, g, ph th kh, pc, tc, kc, bh, dh, gh, bc, dc, gc]. Each solution (or genome) consists of 18 points, each with (x, y, z, w) coordinates. Since each solution consists of 18 points in the acoustic space, the inventory it represents may have anywhere between one and 18 stops, since different points, as identified by their co-ordinates, may share a category label. Thus, a genome may consist of 18 points with unique 4D co-ordinates that all belong to the same category. Such a genome would represent a stop inventory consisting of a single stop category, despite this category having 18 distinct co-ordinates in the genome. A genome with 18 points that each correspond to a unique category label represents a stop inventory with all 18 stops. The algorithm evaluated the genomes by simultaneously optimizing the five competing objectives, discussed in 7.2.2, using Pareto-based selection. At the end of the optimization, each point was assigned the label of the stop category whose centroid it was closest to. The MOGA was implemented with a custom Python script.

7.2.1 Defining the search space

The bounds of the search space were derived from the co-ordinates of the 18 stop category centroids from the MDS solution, with a buffer of 1 additional unit in each direction. The MOGA started with an initial population consisting of 100 randomly generated genomes. Each genome was created by sampling 18 independent points, randomly drawn from the 4D space given by:

$$X \in [X_{\text{min}} - 1.0, X_{\text{max}} + 1.0]$$
$$Y \in [Y_{\text{min}} - 1.0, Y_{\text{max}} + 1.0]$$
$$Z \in [Z_{\text{min}} - 1.0, Z_{\text{max}} + 1.0]$$
$$W \in [W_{\text{min}} - 1.0, W_{\text{max}} + 1.0]$$

Recall that this MDS space was constructed by including data from a wide range of languages, which did not reflect the typological distribution of various stop inventory types. For example, there are 138 languages in PHOIBLE that contrast plain (T), voiceless-aspirated (Th) and voiced (D) stops, but only 2 such languages were included in the MDS solution. Conversely, there are only four languages in PHOIBLE that contrast T, Th, D, Dh (voiced-aspirated) and Dc (voiced-constricted) stops, and 2 of these were included in the MDS solution. Thus, there are just as many languages of the T-Th-D type, which is a far more common type of inventory, as there are of the type T-Th-D-Dh-Dc, which is much rarer.

To incorporate the relative importance of the different dimensions from the MDS solution as represented by typological frequency, a weighted Euclidean distance metric was used. A Poisson regression analysis was conducted to determine the relative contribution of each acoustic dimension from the 4D MDS solution to inventory frequency. The frequency of occurrence of each type of inventory, as attested in the PHOIBLE data, was modeled as the dependent variable, given by the logarithm of the expected count (e.g., 138 inventories of the T-Th-D type, 4 of the T-Th-D-Dh-Dc type etc.). This dependent variable was modeled as a linear function of the Euclidean distances along each dimension of the 4D acoustic space.

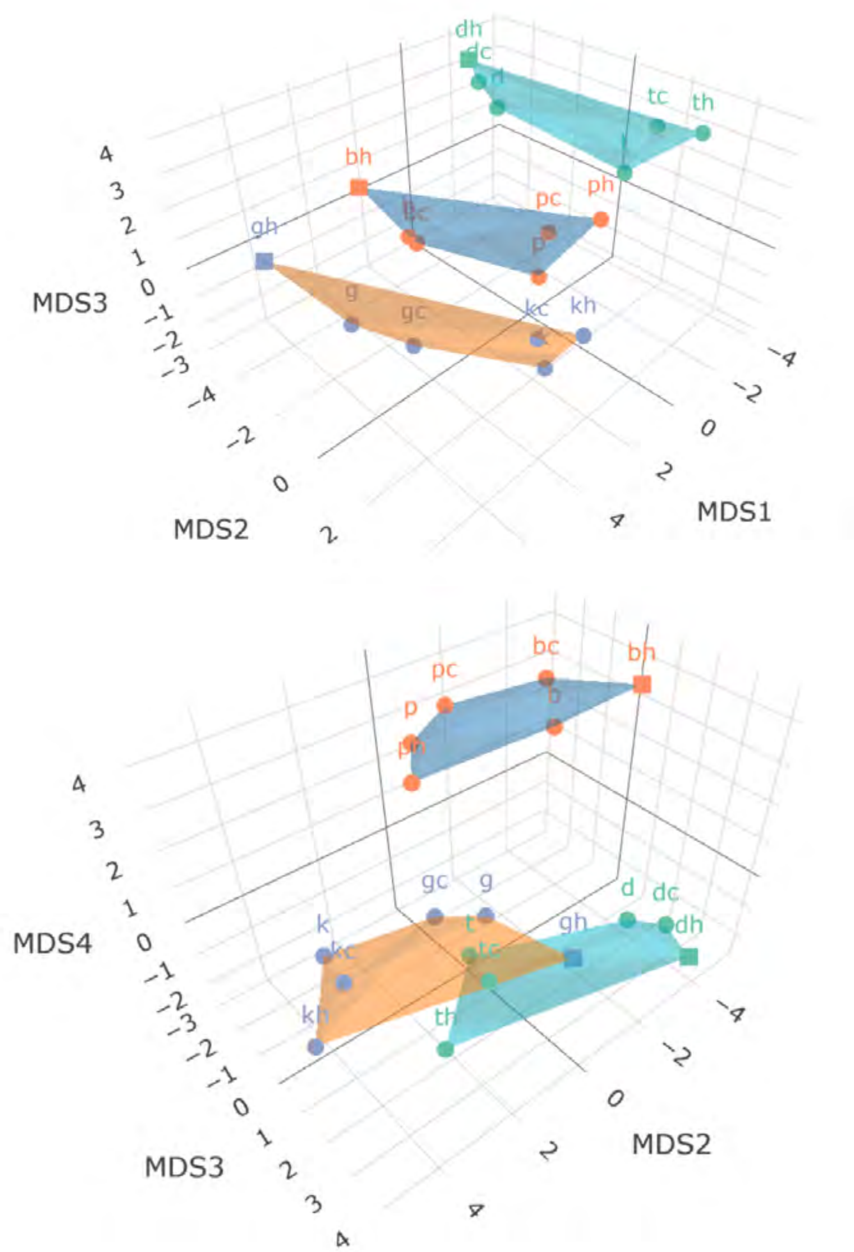


Figure 7.3: *4D MDS search space with reference points defining the 18 stop centroids.*

To assess the relative contribution of each dimension, the absolute z-scores from the coefficient estimates were calculated and normalized to proportions. This allowed direct comparison of relative importance regardless of scale differences, with higher proportions indicating stronger associations with inventory frequency.

The results showed that distance along the four MDS dimensions had varying effects on inventory frequency. The relative contribution of distance along the x, y, z, w dimensions was found to be [0.11, 0.29, 0.32, 0.28]. This weighted distance was used in the calculation of dispersion. For two points $\mathrm{p}_i = (x_i, y_i, z_i, w_i)$ and $\mathrm{p}_j = (x_j, y_j, z_j, w_j)$, the weighted Euclidean distance was computed as:

$$d_w(\mathrm{p}_i, \mathrm{p}_j) = \sqrt{w_x(x_i - x_j)^2 + w_y(y_i - y_j)^2 + w_z(z_i - z_j)^2 + w_w(w_i - w_j)^2} \tag{7.1}$$

7.2.2 Objective Functions

The following objectives (7.1-7.5) were defined. Note that all five objectives were formulated as minimization problems. The first objective (maximizing dispersion) was transformed by returning a negative value, in effect minimizing the negative Euclidean distance.

(7.1) Objective 1 (f_1): Maximize total weighted distance

The total weighted distance was defined as the sum of all pairwise weighted Euclidean distances between the 18 points of the genome. The objective of maximizing this distance was implemented as

$$f_1(\mathrm{G}) = -\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_w(\mathrm{p}_i, \mathrm{p}_j) \tag{7.2}$$

where G represents the genome (the set of N points; $N = 18$) and p_i and p_j represent the i^{th} and j^{th} points respectively. The negative sign ensures that minimizing f_1 maximizes the actual total distance, thus encouraging points to be dispersed in the 4D space.

(7.2) Objective 2 (f_2): Minimize total articulatory cost

Each point of the genome was assigned the articulatory cost of its nearest category centroid in the 4D acoustic space. Table 7.2 shows the articulatory cost associated with each of the 18 stop categories. Articulatory cost is assumed to have four components – cost associated with glottal vibration or voicing, glottal state, aerodynamics and laryngeal movement. In this table “u” (which stands for “unmarked”) denotes an associated articulatory cost of zero. Each

"+" denotes one unit of articulatory cost. Voicing is assumed to require articulatory effort and thus incurs a cost of 1 unit. Similarly, glottal constriction and spread each incur a cost of 1 unit. Stops that are known to provide aerodynamic challenges, such as voicing at posterior places of articulation or ejectives at fronter places, incur additional articulatory costs. Finally, stops that involve raising or lowering the larynx are associated with additional costs. The resulting total cost associated with each of the 18 stops is shown in the last column of Table 7.2.

Category	**Voicing**	**State**	**Aerodynamics**	**Larynx movement**	**Articulatory Cost**
p	u	u	u	u	0
t	u	u	u	u	0
k	u	u	u	u	0
b	+	u	u	u	1
d	+	u	+	u	2
g	+	u	++	u	3
ph	u	+	u	u	1
th	u	+	u	u	1
kh	u	+	u	u	1
pc	u	+	++	+	4
tc	u	+	+	+	3
kc	u	+	u	+	2
bh	+	+	u	u	2
dh	+	+	+	u	3
gh	+	+	++	u	4
bc	+	+	u	+	3
dc	+	+	+	+	4
gc	+	+	++	+	5

Table 7.2: *Category markedness values used in the MOGA.*

Let $\mathscr{C}(\mathrm{p})$ denote the cost of the nearest stop category. The total articulatory cost of a genome G is given by the sum of the articulatory cost incurred by each point p:

$$f_2(\mathrm{G}) = \sum_{i=1}^{N} \mathscr{C}(\mathrm{p}_i) \tag{7.3}$$

Objective 2 thus favors points in the space that lie near stop category centroids with low aric-ulatory cost.

(7.3) Objective 3 (f_3): Maximize Symmetry

For an inventory to be perfectly symmetric, each point p_i in the genome must have a partner at $-\mathrm{p}_i$. The objective of maximizing symmetry was implemented as minimizing the symmetry error – the average weighted distance from each point to its nearest reflected counterpart – given by:

$$f_3(\mathrm{G}) = -\frac{1}{N}\sum_{i=1}^{N}\min_{j\neq i} d_w(\mathrm{p}_j, -\mathrm{p}_i) \tag{7.4}$$

The negative sign converts the objective of maximizing symmetry to minimizing this error. A value of $f_3 = 0$ indicates perfect symmetry.

(7.4) Objective 4 (f_4): Maximize distinct categories

Since the algorithm operates in continuous acoustic space, it can generate distinct genomes that all share the same category label, since every point receives a category label based on the stop category centroid it is closest to. Thus, two genomes can have points with distinct (x, y, z, w) co-ordinates but can be identical in terms of category labels. Objective 4 ensures that the 18 points that make up a genome are associated with as many distinct stop categories as possible. This objective is the equivalent of the MAXIMIZECONTRASTS constraint in Chapter 3.

This objective was implemented as minimizing repetition of the same category in an inventory. Let $\mathscr{K}(\mathrm{p})$ denote the category label of the nearest stop centroid. For a total of N points, the repetition penalty is given by:

$$f_4(\mathrm{G}) = N - \left|\mathscr{K}(\mathrm{p}_i) : \mathrm{p}_i \in \mathrm{G}\right| \tag{7.5}$$

where $\mathscr{K}(\mathrm{p}_i)$ is the set of distinct category labels assigned to the points in G. This penalty effectively counts the number of repeated category labels. An inventory that employs all 18 stop categories yields a penalty of 0 as it has no repeated labels.

(7.5) Objective 5 (f_5): Minimize hierarchy penalty

The objective of maintaining the hierarchy of Lindblom and Maddieson (1988) was implemented as minimizing the hierarchy penalty. I used a different classification scheme here.

The contents of Lindblom and Maddieson's Set I are "intuitively-based" (Lindblom and Maddieson, 1988:68). Instead, the scheme used here directly refers to the articulatory cost shown in Table 7.2. I will refer to the sets based on this scheme as "levels" to distinguish them from Lindblom and Maddieson's sets. Thus, Level 0 consists of stops employing the most basic articulations – /p t k/. The addition of glottal vibration is an "elaborated" articulation that expands the contents of Level 0 to include /b d g/. Thus, the contents of Lindblom and Maddieson's Set I are now split into two separate levels – Level 0 and Level 1. The next set of articulations, Level 2, which is in line with Lindblom and Maddieson's Set II, includes aspiration, glottalization, ejectives and implosives (Lindblom and Maddieson, 1988:67). Thus, /pʰ tʰ kʰ p' t' k' ɓ ɗ ɠ/ are added to the set /p t k b d g/ in Level 2. Finally, /bʰ dʰ gʰ/, which combine the elaborated articulations of voicing and aspiration, make up the highest level, Level 3. These levels are summarized in the table below.

Laryngeal category	**Level**
T	0
D	1
Th, Tc, Dc	2
Dh	3

Table 7.3: *Levels used in the evaluation of hierarchy penalty.*

Thus, different categories were assigned hierarchical levels based on their laryngeal type in evaluating f_5:

$$L(\text{category}) = \begin{cases} 0 & \text{if a stop belongs to a T category} \\ 1 & \text{if a stop belongs to a D category} \\ 2 & \text{if a stop belongs to a Th or Tc or Dc category} \\ 3 & \text{if a stop belongs to a Dh category} \end{cases}$$

Let U_ℓ be the set of unique categories used at level ℓ, and let T_ℓ be the total number of categories available at that level ($T_0 = 3,\ T_1 = 3,\ T_2 = 9,\ T_3 = 3$). The hierarchy penalty was calculated separately for Levels 1 and 2 on the one hand, and Level 3 categories on the other, since they have different dependency requirements. If any category from Level 1 or Level 2 is used, all lower-level categories must be fully used. The penalty is thus given by:

$$P_{\text{level}\leq 2} = \sum_{\ell=1}^{2} \mathbb{I}(\ell \in \mathscr{L}_{\text{used}}) \sum_{\ell'=0}^{\ell-1} \max(0, T_{\ell'} - |U_{\ell'}|) \times (\ell - \ell') \tag{7.6}$$

Effectively, if a Level 1 category (D) is used without fully using all Level 0 categories (T), a penalty proportional to the number of missing Level 0 categories is applied, multiplied by the level difference $(1-0)=1$. Similarly, if a Level 2 category (Th, Tc, or Dc) is used without using all Level 0 and Level 1 categories, penalties are applied for missing categories at both lower levels, weighted by their respective level differences ($(2-0)=2$ for Level 0, $(2-1)=1$ for Level 1).

Along similar lines, if any Level 3 category (Dh) is used, all Level 0, Level 1 and Th categories from Level 2 must be fully used. Crucially, Dh stops do not require Tc or Dc categories to be used for the hierarchy to be satisfied. Let U_{Th} be the set of Th categories used, and $T_{\text{Th}}=3$ be the total number of Th categories. The penalty is given by:

$$\mathbb{I}(3 \in \mathscr{L}_{\text{used}})[\max(0, T_0 - |U_0|) \times (3-0) + \max(0, T_1 - |U_1|) \times (3-1) + \max(0, T_{\text{Th}} - |U_{\text{Th}}|) \times (3-2)] \tag{7.7}$$

where $\mathscr{L}_{\text{used}}$ is the set of levels (0–3) with at least one used category, and $\mathbb{I}(\cdot)$ is the indicator function. The total hierarchy penalty is the sum of the two penalties:

$$f_5(\text{G}) = P_{\text{level2}} + P_{\text{level3}} \tag{7.8}$$

Thus, the hierarchy penalty forces the full use of lower levels if at least one category of level ℓ is used. This penalty increases when higher-level categories are used while lower-level categories remain incomplete, with larger penalties for larger gaps in the hierarchy (as indicated by the terms $(2-0)$, $(2-1)$ etc. in (7.6) and 7.7)). Thus, the use of more complex spaces before the lower spaces are fully saturated is penalized, in line with Lindblom and Maddieson (1988).

7.2.3 Pareto Dominance

Over the course of evolution, solutions were chosen based on Pareto rank (lower is better) by comparing pairs of randomly chosen solutions, as described in Section 7.1.3. Additionally, crowding distance (higher is better) was calculated within a front to maintain diversity. This is a measure of how isolated a solution G_i is from its neighbors in the Pareto front, given by:

$$\mathrm{crowd}(\mathrm{G}_i) = \sum_{m=1}^{5} \frac{f_m(\mathrm{G}_{i+1}) - f_m(\mathrm{G}_{i-1})}{f_m^{\max} - f_m^{\min}} \tag{7.9}$$

where m indexes the objectives, $f_m(\mathrm{G}_{i+1})$ and $f_m(\mathrm{G}_{i-1})$ are the objective values of the neighboring solutions when sorted by objective m and $f_m^{\max}$ and $f_m^{\min}$ are the maximum and minimum values of objective m in the current Pareto front. For each objective, the crowding distance calculates the normalized gap between a solution and its neighbors. Note that solutions at the extremes (first or last in sorted order) received infinite crowding distance to preserve boundary points. The final crowding distance was the sum of these normalized gaps across all the objectives. Solutions with larger crowding distance are preferred because they maintain diversity in the population.

In this MOGA, each of the five objectives (7.1–7.5) was treated as a separate dimension of fitness. However, before performing Pareto dominance comparisons, each objective value was multiplied by a weight, derived from the MaxEnt model in Chapter 3. The weights for dispersion (f_1), articulatory cost or markedness (f_2) and maximizing the number of distinct stops (f_4) were implemented as the sum of the individual MINDIST, markedness and MAXIMIZECONTRASTS constraints respectively. The remaining objectives, symmetry (f_3) and respecting Lindblom and Maddieson's hierarchy (f_5) were assigned weights that were the average of the other three weights. For a genome G with raw objective values $\mathrm{f}(\mathrm{G}) = [f_1, f_2, \ldots, f_5]$, the weighted objective vector was computed as:

$$\mathrm{f}_{\mathrm{weighted}}(\mathrm{G}) = [w_1 f_1, w_2 f_2, \ldots, w_5 f_5] \tag{7.10}$$

where $[w_1, w_2, \ldots, w_5]$ denote weights from the MaxEnt model in Chapter 3. Pairs of solutions were then compared using *weighted* Pareto dominance. The MaxEnt weights served to scale each objective to a proportional magnitude. Higher weights increased the relative importance of an objective in the calculation of crowding distance, biasing diversity preservation towards the higher weighted objectives.

Note that despite using weights to scale the objectives in this manner, the algorithm used here is still a MOGA and not a simple GA, as the weights did not collapse the problem into a single scalar objective as in a simple GA. Instead, the weights scale the range of each objective such that improvements in higher weighted objectives (such as markedness, as shown in Chapter 3), are more likely to influence dominance, compared to improvements in lower weighted ones (such as dispersion). The algorithm still produces a full Pareto front of trade-off solutions, as in any MOGA, but the weights bias the search toward regions where performance on the higher weighted objectives is better.

Thus, the weights used in the Pareto front selection guide the search without collapsing the Pareto front into a single solution. The output is a set of inventories that are optimal under different trade-offs, not a single inventory that optimizes a fixed scalar objective function. The weights used here do not influence the calculation of the objectives themselves, but rather make some trade-offs more attractive than others. Conceptually, this means that a phonological filter is applied at the end of each generation to choose solutions for the next generation in a way that stretches the objective space along the dimensions that are more important (according to findings from the MaxEnt model) and contracts it along the dimensions that are less important.

7.2.4 Genetic Operators & Termination

Cross-over was implemented to produce offspring as linear combinations of the parents. Gaussian mutation was implemented by adding noise to the co-ordinate of each point in the genome with probability 0.3. If either cross-over or mutation produced co-ordinates beyond the bounds of the 4D space, these out-of-bounds co-ordinates were reduced to the outer bound of the corresponding dimension.

Finally, conservation was implemented by combining the parents and offspring into a single population and selecting the best 100 individuals by prioritizing lower ranks and higher crowding distances within the last front. These individuals made up the next generation of solutions.

The algorithm ran for 200 generations. The final population was analyzed to extract the Pareto front using true Pareto dominance on the raw objective values. The solutions were reported in terms of stop category labels, based on the category centroid in the 4D space the solution was closest to. Extreme solutions on each objective were identified: maximum total distance, minimum articulatory cost, best symmetry, minimum repetition, and minimum hierarchy penalty. Additionally, the most balanced solution was found by normalizing all objectives to $[0,1]$ and minimizing the Euclidean distance to the ideal objective values $(1,0,0,0,0)$. The ideal objective represents an inventory that is maximally dispersed, minimally articulatorily complex, maximally symmetric and diverse and incurs minimum penalty on Lindblom and Maddieson's hierarchy.

7.2.5 Summary

In summary, the MOGA works by searching the 4D MDS space for stop categories. It starts with a randomly generated population of 100 genomes, each consisting of 18 randomly generated points. The algorithm selects points in the continuous space but assigns category labels based on the *x, y, z, w* coordinates of the centroids of each category. In case of a tie, it chooses the less marked stop, although cases of ties are very rare. The population evolves over 200 generations. The 18 points in every genome at the end of the evolution are given category labels based on proximity to the closest category centroid. This is the set of optimal stop inventories identified by the algorithm.

7.3 Results

The MOGA identified key trade-offs between the objectives. The strongest conflict was between the objectives f_1 and f_2 – maximizing distance while minimizing articulatory costs, in line with Lindblom and Maddieson (1988) and echoing results from previous chapters. Figures 7.4-7.6 show how the Pareto-optimal solutions trade-off with each other. Each point in these plots represents one non-dominated set of 18 selected in the 4D space at the end of 200 generations of evolution. In other words, each point in these plots represents a possible stop inventory.

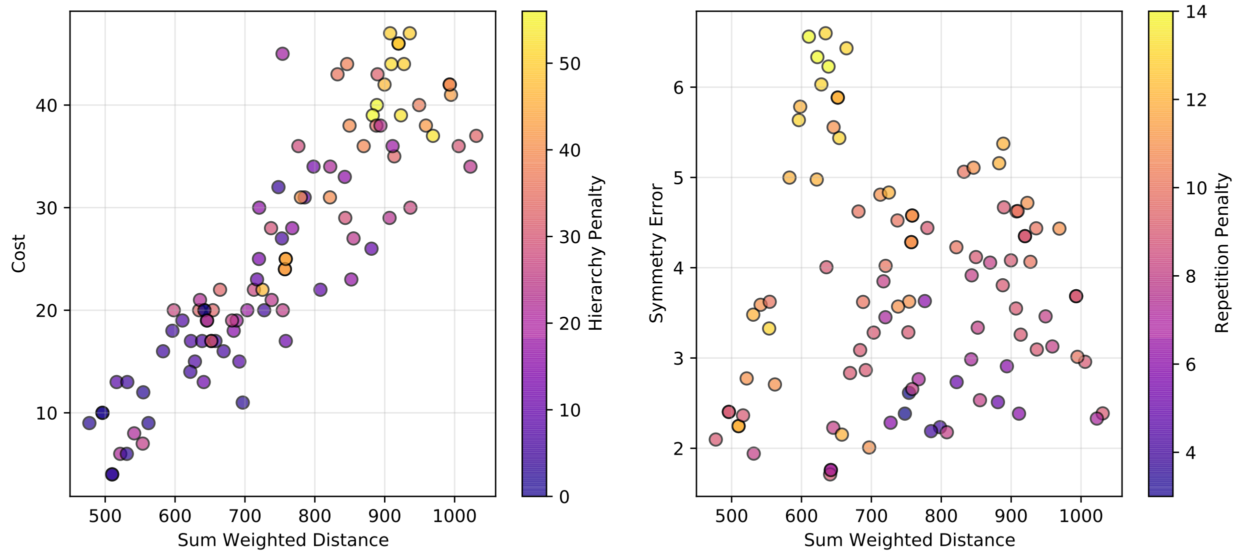


Figure 7.4: *Trade-offs between dispersion, articulatory cost and symmetry.*

The left panel in Figure 7.4 shows the relationship between dispersion (x-axis), articulatory cost (y-axis) and hierarchy penalty (color gradient). The perfect inventory would appear in the lower-right corner of this plot, representing high distance and low complexity, while the least optimal inventories would appear in the top-left corner. Both of these parts of the plot are sparsely populated as the points cluster along the diagonal. The more dispersed inventories in this plot generally have high hierarchy penalties as the color gradient goes from mostly blue (representing lower hierarchy penalties) to mostly yellow (representing higher hierarchy penalties).

The right panel of this figure shows dispersion against symmetry error on the y-axis and repetition penalty on the color gradient. Good symmetry, as evidenced by lower symmetry error, is associated with a wide range of dispersion, since lower values on the y-axis are spread out over a wide range on the x-axis. Repetition penalty is generally greater at lower distances as the colors go from more yellow at low values on the x-axis to more blue at higher values. This trend is largely driven by size, since smaller inventories have lower dispersion and more repetition. There is a cluster of inventories with relatively low dispersion (between approximately 600 and 700 units on the x-axis) with high symmetry error (5 and over) and high repetition (mostly yellow). This cluster represents small inventories that are asymmetrical and less dispersed in the space, such as /k $\mathrm{p^h}$ $\mathrm{k^h}$ k' $\mathrm{b^h}$/ (dispersion= 664.4, Symmetry= 6.56, Repetition= 13) and /k $\mathrm{p^h}$ $\mathrm{k^h}$ k'/ (dispersion= 610.4, Symmetry= 6.43, Repetition= 14). Notice that these inventories also have high cost and hierarchy violations. This cluster thus represents the inventories that are farthest from the ideal point $(1,0,0,0,0)$ (also confirmed by their Euclidean distance to the ideal point). They still make their way to the end of the evolution as they are still better than inventories like /k $\mathrm{p^h}$ $\mathrm{k^h}$ p' $\mathrm{g^h}$/ and /g $\mathrm{b^h}$ $\mathrm{g^h}$ p'/, for example.

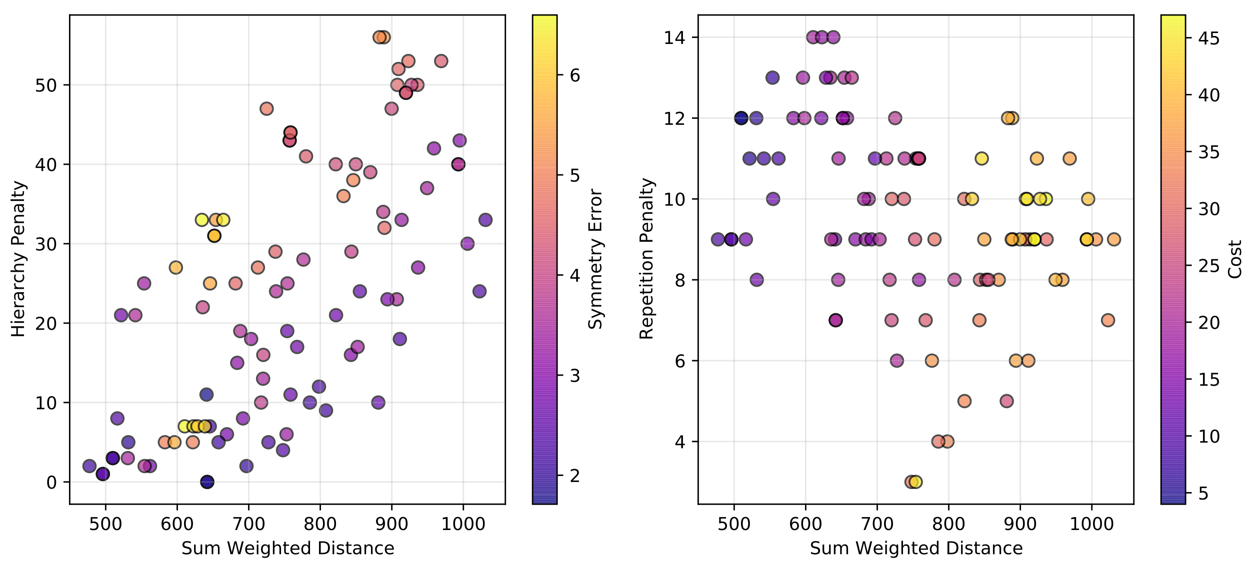


Figure 7.5: *Trade-offs between dispersion and hierarchy (left); dispersion and repetition (right).*

The left panel of Figure 7.5 shows dispersion on the x-axis, hierarchy penalty on the y-axis, and symmetry error on the color gradient. The hierarchy penalty for more dispersed inventories is generally worse. The low-dispersion, low symmetry cluster from the right panel of Figure 7.4 is now broken up into two clusters – one with low hierarchy penalty and the other with high hierarchy penalty. Clusters with lower penalties are of the type /k p^h k^h k’/, which employ Level 2 stops, and those with greater penalties are of the type /k p^h k^h k’ b^h/, which employ both Level 2 and Level 3 stops. The use of higher-level stops despite gaps in Levels 0, 1 and 2 penalize inventories of the second type more than the first, for which only Level 0 and 1 gaps are relevant, in terms of their hierarchy penalties.

The right panel shows the relationship between dispersion, repetition and cost. Less dispersed inventories incur lower articulatory costs but greater repetition penalties, whereas more dispersed inventories employ a greater number of stops at a higher total articulatory cost.

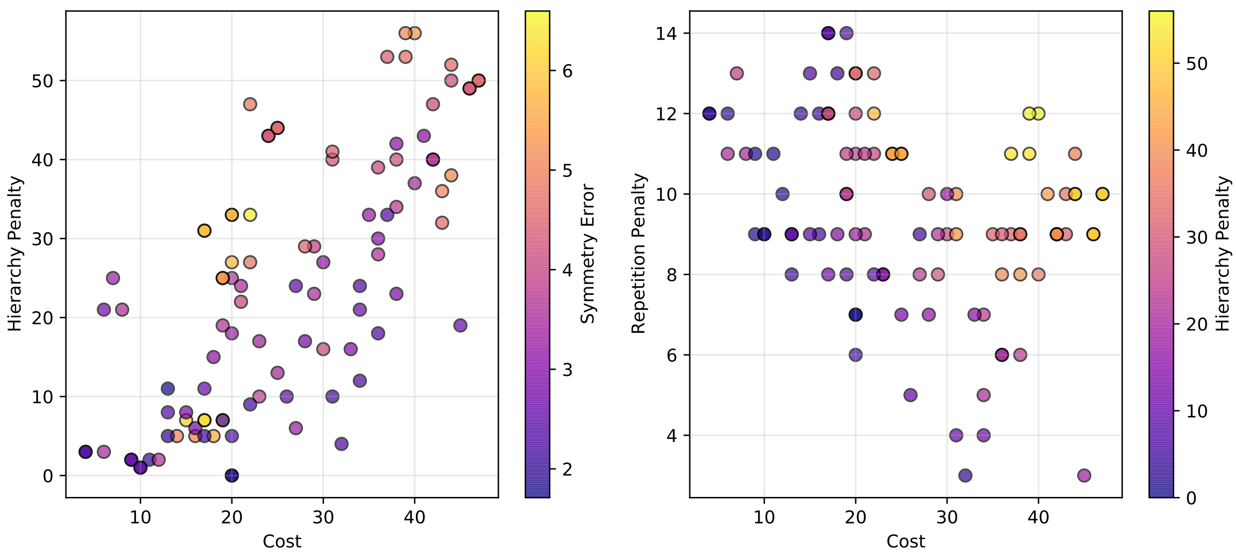


Figure 7.6: *Trade-offs between cost, hierarchy and symmetry (left); cost, repetition and hierarchy (right).*

Finally, Figure 7.6 shows the relationship between the objectives excluding dispersion. The left panel shows articulatory cost on the x-axis, hierarchy penalty on the y-axis and symmetry error on the color gradient. This plot shows that low-cost, symmetrical inventories that adhere to the hierarchy, like /p t k p^h t^h k^h/, are possible, as indicated by the cluster of blue dots in the lower left. The right panel shows that these objectives must trade-off against diversity. This is evident as the lower left zone in this plot is empty, which shows repetition penalty on the y-axis, as low-cost inventories must repeat the same stop categories.

Table 7.4 shows the extreme solutions identified by the algorithm. The first column shows the objective, the second shows the inventory identified as being optimal on that objective, and the third shows the attested number of inventories of that composition in PHOIBLE.

Objective	**Inventory**	**Attested N**
Best cost	p t k, ph th kh	73
Best dispersion	_ _ k, _ _ g, ph th kh, _ _ kc, bh _ _, bc dc gc	0
Best diversity	p t k, b d g, ph th kh, pc _ kc, bh _ gh, bc dc gc	0
Best symmetry	p t k, b d _, ph _ kh, pc tc kc, bh _ _, _ _ gc	0
Best hierarchy	p t k, b d g, ph th kh, _ tc kc	1
Best overall	p t k, _ d g, ph _ kh, pc tc kc, bh _ _, bc dc gc	0

Table 7.4: *Best solution for each objective.*

As seen in this table, the MOGA identified /p t k pʰ tʰ kʰ/ as the optimal solution based on prioritizing minimization of articulatory cost. This inventory type, which occurs in languages like English and Mandarin, is well-attested, as 73 languages in PHOIBLE have /p t k pʰ tʰ kʰ/ as their stop inventory. However, the inventories identified as optimal on the other objectives are almost never attested.

Table 7.5 shows results of a MOGA that implemented the dispersion objective as maximizing the minimum distance between the closest pair of points. The results from this MOGA do not match any attested inventories, but it is worth noting that two of the objectives – maximizing diversity and adherence to the hierarchy – produce identical inventories.

Solution	**Inventory**	**Attested N**
Best complexity	p t k, ph _ _, _ _ kc	0
Best dispersion	p _ _, b d g, ph th kh, _ _ kc, bh dh gh, bc dc _	0
Best diversity	p t k, b d g, ph th kh, pc tc kc, bh dh gh, _ _ gc	0
Best symmetry	p t k, b d g, ph th kh, pc tc kc, _ _ gh	0
Best hierarchy	p t k, b d g, ph th kh, pc tc kc, bh dh gh, _ _ gc	0
Best overall	p t k, b d g, ph th kh, pc tc kc, bh dh gh, _ _ gc	0

Table 7.5: *Best solution for each objective with minimum distance.*

To test whether a different set of objectives would produce more typologically consistent results, I iteratively dropped objectives f_1 through f_5 from the model. Table 7.6 shows the results when f_1, the objective of maximizing total distance was dropped.

Solution	Inventory	Attested N
Best cost	p t k, ph th kh	73
Best symmetry	p t k, b d g, ph th kh, pc tc kc, bh dh gh, bc _ _	1
Best diversity	p t k, b d g, ph th kh, pc tc kc, bh dh gh, bc _ _	1
Best hierarchy	p t k, b d g, ph th kh, pc tc kc, bh dh gh, bc _ _	1
Best overall	p t k, b d g, ph th kh, pc tc kc	8

Table 7.6: *Best solution for each objective without the dispersion objective.*

This table shows that dropping dispersion as an objective improves the predictions as the inventories resemble attested ones. The best inventory from the point of view of reducing articulatory cost remains the same – the English-type inventory /p t k p^h t^h k^h/. The inventory closest to the ideal point – /p t k b d g p^h t^h k^h p' t' k'/ – is also attested in 8 PHOIBLE inventories. Finally, the symmetry, diversity and hierarchy objectives, all converge on the same inventory, /p t k b d g p^h t^h k^h p' t' k' b^h d^h g^h b/. This matches the largest inventory attested in PHOIBLE, Santali, which has the same 16 stops.

Dropping the other objectives returned less favorable results. Table 7.7 shows the result without the objective of minimizing articulatory cost. Without this objective, dispersion introduces several gaps at unmarked stops to increase the total distance at the cost of the other objectives. The best inventory for maximizing symmetry is the same inventory that maximizes diversity. This inventory contains 17 of the 18 stops. However, without f_1, the stop that gets left out is /t/, possibly the least marked of all the stops and the second most common stop in the world's languages after /k/. Hierarchy fixes this issue by moving the gap higher up in the levels, since gaps at lower levels are penalized more. Interestingly, it predicts a gap at /p'/ despite not being given any information about the higher markedness of the bilabial ejective. However, none of the inventories predicted by this MOGA are attested. Without articulatory complexity to keep the overuse of stops in check, all the inventories chosen by the MOGA simply have too many stops.

Table 7.8 shows the optimal inventories chosen by a MOGA without the symmetry objective. This result is slightly better than the result in Table 7.7 since the least articulatorily marked inventory it chooses is attested and sensible. However, none of the other inventories are attested. Recall that here "symmetry" refers to geometrical symmetry in the 4D search

Solution	Inventory	Attested N
Best dispersion	_ t k, _ _ g, ph th kh, _ tc kc, bh _ gh, bc dc _	0
Best symmetry	p _ k, b d g, ph th kh, pc tc kc, bh dh gh, bc dc gc	0
Best diversity	p _ k, b d g, ph th kh, pc tc kc, bh dh gh, bc dc gc	0
Best hierarchy	p t k, b d g, ph th kh, _ tc kc, bh dh gh	0
Best overall	p t k, b d g, ph th kh, _ tc kc, _ dh _, bc dc gc	0

Table 7.7: *Best solution for each objective without the articulatory cost objective.*

space and not featural symmetry, even though the dimensions of the MDS space closely match some features. In this MOGA, the same inventory performs best on two different objectives, maximizing diversity and following the hierarchy, although this inventory is not attested.

Solution	Inventory	Attested N
Best cost	p t k, b d _, ph th kh	44
Best dispersion	p t k, b d g, _ th kh, _ _ kc, bh _ gh, bc dc _	0
Best diversity	p t k, b d g, ph th kh, pc tc kc, bh _ gh, bc dc _	0
Best hierarchy	p t k, b d g, ph th kh, pc tc kc, bh _ gh, bc dc _	0
Best overall	p t k, b d g, ph th kh, _ tc kc, bh _ gh	0

Table 7.8: *Best solution for each objective without the symmetry objective.*

Next consider results from the MOGA without the diversity objective, which promotes the use of distinct categories. Recall that since the algorithm operates in continuous space, it selects points represented by (x, y, z, w) co-ordinates in this space, which are then assigned category labels. Without this objective, the MOGA selects smaller stop inventories, as seen in Table 7.9. The least marked inventory as in the cases with all five objectives and without dispersion, /p t k p^h t^h k^h/, but the best inventories on the other objectives are notably smaller. None of these are attested, however, even though some of the gaps are sensible. Note also that in this case, there are gaps at all levels, suggesting that removing diversity from the objectives set degrades adherence to the hierarchy. Also note that the inventory closest to the ideal is the one that is optimal from the point of view of hierarchy.

Table 7.10 shows the results from the MOGA without the objective of following the hierarchy. The least-cost inventory remains /p t k p^h t^h k^h/ and none of the others are attested. Moreover, the removal of the hierarchy objective not only allows gaps to proliferate, but the gaps themselves do not seem sensible, as they occur at less marked rather than marked artic-

Solution	Inventory	Attested N
Best cost	p t k, ph th kh	73
Best dispersion	p _ k, _ _ g, ph _ kh, _ _ kc, bh _ gh, bc _ _	0
Best symmetry	p t k, b d _, _ _ kh, bc _ _	0
Best hierarchy	p t k, b d g, ph _ kh, bh _ _	0
Best overall	p t k, b d g, ph _ kh, bh _ _	0

Table 7.9: *Best solution for each objective without the diversity objective.*

ulations, i.e. at /b/ rather than /g/.

Solution	Inventory	Attested N
Best cost	p t k, ph th kh	73
Best dispersion	_ t k, _ d g, ph th kh, _ _ kc, bh _ gh, bc dc gc	0
Best symmetry	p t k, b d g, ph th _, pc tc kc, bh _ _, _ dc _	0
Best diversity	p t k, _ d g, ph th kh, pc _ kc, bh _ gh, bc dc gc	0
Best overall	p t k, _ d g, ph th kh, pc tc _, bh _ gh, bc dc _	0

Table 7.10: *Best solution for each objective without the hierarchy objective.*

Overall, as seen in Table 7.6, the MOGA with four objectives – minimizing articulatory cost, maximizing the symmetry and diversity of stops, and adhering to the hierarchy – produces results that resemble real inventories. The lack of a role for dispersion is somewhat unsurprising since it was found to play a limited role in determining the makeup of stop inventories in previous chapters as well. Notice, however, that the diversity objective seems to be doing some of the work done by dispersion. As noted before, since the inventory genomes consist of points in the continuous space, a single genome can have 18 distinct points that all correspond to the same category. Thus, the diversity objective is needed as omitting it leads to ensure that the points in the genome belong to different stop categories. However, in doing so, the diversity objectives pushes promotes genomes with points that are farther away from each other. In other words, it promotes more dispersed inventories. The addition of a distinct dispersion objective then duplicates the effect of well-separated points surviving the over the course of evolution. Additionally, hierarchy also promotes dispersion while still constraining it, since it allows more peripheral categories to occur so long as the categories in lower levels also occur. Therefore, dispersion may still be an objective in the search for stop inventories.

One outstanding issue is justifying the use of weights from the MaxEnt model in the de-termination of optimal solutions. Table 7.11 shows the optimal inventories on each objective selected by the MOGA without the dispersion objective, now without the weighting scheme. This result offers a much poorer match to the attested inventory types, compared to the result in Table 7.6, which predicted reasonable inventories. This suggests that filtering the inventories through the MaxEnt weights at the end of each generation produces more sensible results.

Solution	**Inventory**	**Attested N**
Best cost	p t k, _ th kh	0
Best symmetry	p t k, b d g, ph th kh, pc _ kc, bh dh gh	0
Best diversity	p t k, b d g, ph th kh, pc tc kc, bh dh gh, _ _ gc	0
Best hierarchy	p t k, b d g, ph th kh, pc tc kc, bh dh gh, _ _ gc	0
Best overall	p t k, b d g, ph th kh, pc tc kc, bh _ _	0

Table 7.11: *Best solution for each objective without the dispersion objective and weighting scheme.*

7.4 Discussion

This chapter has presented results from a multi-objective genetic algorithm that was tasked with finding stops in the 4-dimensional MDS space with the objectives of minimizing articulatory cost, maximizing symmetry and inventory size, and following the kind of hierarchy of complexity proposed by Lindblom and Maddieson (1988). Although it was found that including it as an objective returned unattested inventories, dispersion may still be thought of as an objective as the diversity objective is implemented in a way that promotes dispersion.

These results are in line with Lindblom and Maddieson's proposal that consonant inventories are not governed by maximal dispersion but by processes of subspace saturation and expansion. For small stop inventories, sufficient contrast is achieved by employing an unmarked inventory that uses Set I segments, such as /p t k/. Beyond this size, more elaborated articulations must be employed. The use of these articulations can crowd the perceptual space until a saturation point beyond which the space must incorporate additional dimensions to sustain the contrast in the system.

The MOGA needs four explicit components to be successful. The first is articulatory simplicity to ensure the smallest systems employ /p/, /t/ and /k/. It also needs Lindblom and Maddieson's hierarchical sets (implemented as more elaborated levels here) to ensure saturation of simpler spaces before more complex articulations are employed. Beyond these, it also needs an objective maximizing the number of contrasts (equivalent to the phonological constraint MAXIMIZECONTRASTS from Chapter 3) to ensure that the algorithm does not get stuck at /p t k/. Note again that this objective implicitly includes dispersion. The symmetry objective ensures that mismatched sets of higher-level categories do not occur in the same inventory. For example, the set /p' t' k'/ has the same total articulatory complexity as /p' k' ɓ/ since /t'/ and /ɓ/ have the same articulatory cost (see Table 7.2). The set /p' k' ɓ/ does not violate the hierarchy objective since the ejectives and implosives occur in higher levels, which are not subject to the hierarchy. An inventory that adds these sets to say /p t k b d g/ is also of the same size. In such a case, it is symmetry that makes the inventory /p t k b d g p' t' k'/ more attractive over /p t k b d g p' k' ɓ/. In other words, since hierarchy does not apply to the higher levels, symmetry promotes solutions without gaps in the higher levels, thus promoting more economical inventories.

Another characteristic of the most viable MOGA is that three different objectives, maximizing the diversity and symmetry and adherence to the hierarchy, seem to act in concert to converge on the same inventory. This stop inventory, the largest one attested in any language in PHOIBLE, is found in Santali. It is possible that this highly marked inventory with sub-optimal featural inventory exists as it strikes a balance between objectives f_3, f_4 and f_5.

One seemingly surprising result of the MOGA is that the inventory /p t k p^h t^h k^h/ was found to be the optimal inventory on the objective of minimizing articulatory cost. The most commonly occurring inventory of size 6, however, is /p t k b d g/. The reason for this is the imbalance in cost across the places of articulation in the voiced stops, which leads to a higher total cost for /p t k b d g/ compared to /p t k p^h t^h k^h/. However, it is possible that the former is more commonly attested in languages not because it is more optimal as a solution to the acoustic search problem, but because the stops /b d g/ tend to occur in phonological systems that employ /p t k/, as variants of intervocalic stops, for instance. Therefore, /b d g/ can find their way into the inventories of languages that already have /p t k/. Such a path is less readily available to stops like /p^h t^h k^h/, whose frequency of co-occurrence with /p t k/ is probably made possible by their relatively lower articulatory cost in an inventory that is trying to balance this objective with maximizing the diversity and symmetry, and adherence to Lindblom and Maddieson's hierarchy.

7.5 Conclusion

This chapter has presented a model of stop inventories that searches the acoustic space of stops to devise inventories that simultaneously maximize dispersion, symmetry, diversity and adherence to a hierarchy of articulatory complexity while minimizing the total articulatory cost. It has shown that although dispersion in the acoustic space does not emerge as an explicit objective, its effects are implicitly captured by the implementation of the diversity objective. Moreover, the hierarchy objective, modeled on Lindblom and Maddieson's proposal, sufficiently captures the goal of increasing perceptual distance and maximizing economy, by annexing more of the acoustic space when simpler options have been exhausted. Finally, the objectives of maximizing the diversity and symmetry and adherence to the hierarchy may work in concert to prefer the same stop inventories, which may explain why some highly marked stop inventories are attested in the world's languages.

CHAPTER 8

Conclusion

Tiger got to sleep, bird got to land;

Man got to tell himself he understand.

– The Books of Bokonon

KURT VONNEGUT, Cat's Cradle

This dissertation started out by asking why languages have the sounds they do. The basic approach taken in answering this question was comparing the set of attested inventories with the set of possible but unattested inventories. Consider, one last time, the distribution of attested stop inventory sizes (shown in blue) and possible sizes (shown in orange) in Figure 8.1. I have presented three different methods for understanding what separates the blue curve from the orange curve.

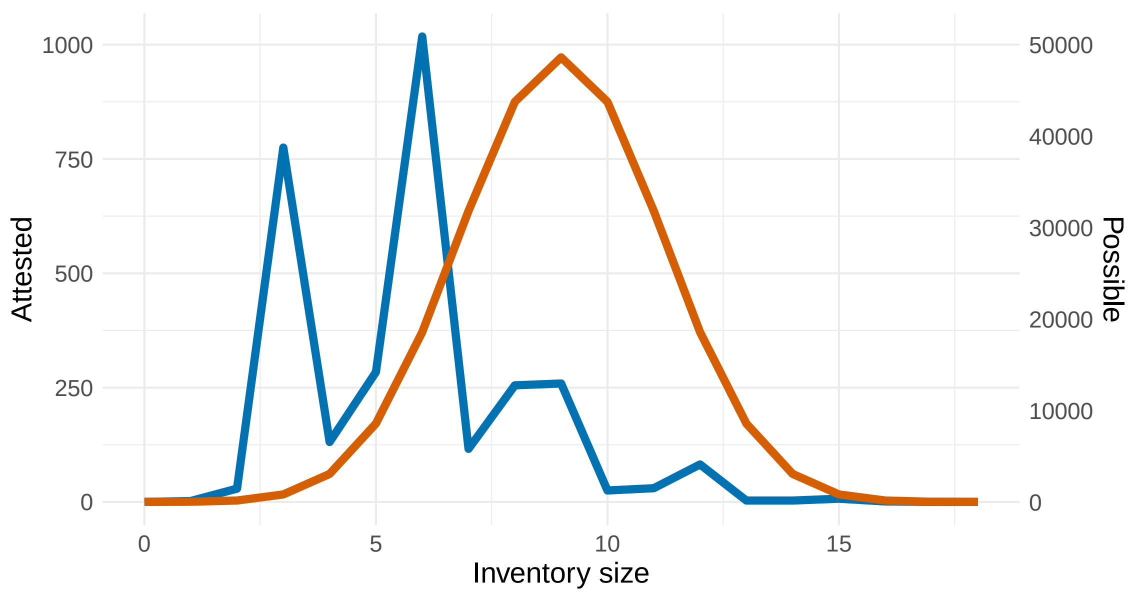


Figure 8.1: *Attested versus possible stop inventories.*

Overall, it was found that attested stop inventories are shaped by striking a balance between dispersion, articulatory markedness and maximizing the size of the inventory. The effect of dispersion on stop inventories was found to be limited to maximizing the difference between the closest pair of stops, in line with results on vowel inventories (e.g., Becker-Kristal, 2010). Moreover, the effects of feature economy and symmetry can be captured by the interaction of phonetically-based constraints that maximize the minimum dispersion and inventory size and minimize articulatory markedness. Thus, these structural properties are not needed as independent factors.

Results from the three models are summarized in Table 8.1. As shown in this table, the models differed in their architecture as they investigated different aspects of stop inventory composition. The MaxEnt model dealt with traditional phonological objects in the form of stop categories and associated features, while the hurdle model and MOGA operated in the continuous acoustic space. The hurdle model, like the MaxEnt model, modeled the inventories as categorical and fit their attested frequencies, while the MOGA devised inventories based on their co-ordinates in the continuous acoustic space, and was blind to the distribution of different stop inventories in the languages of the world.

	MaxEnt	**Hurdle**	**MOGA**
Architecture			
Space	Featural	Acoustic	Acoustic
Inventories	Categories	Categories	4D co-ordinates
Results			
Dispersion	✓	✓	✓
Articulatory effort	✓	✓	✓
Inventory size	✓	✓	✓
Feature economy	×	–	–
Acoustic symmetry	–	×	✓

Table 8.1: *Summary of results.*

Table 8.1 shows the different factors that were found to influence inventory contents with a '✓' and those that were not found to affect inventory contents with a '×.' Empty cells indicate that the corresponding factors were not included as predictors. As seen in this table, all three models found evidence for the role of dispersion, articulatory effort and inventory size in determining inventory content, although the role of dispersion could not be disentangled from the role of size in the MOGA. Results from the MaxEnt and hurdle models also found that minimum rather than total dispersion best characterizes the effect of maximizing contrast in stop inventories. A major finding across these two models is that the effects of structural properties like economy and symmetry are derivable from the effects of the dispersion-theoretic predictors. The MOGA, which tackles a much more difficult problem, needs all the objectives to be successful, in addition to an explicit mechanism to constrain the search in a way that first maximizes the use of articulatorily simple stops and then adds a phonetic dimension to the basic space. Additionally, an explicit symmetry objective is required to guide the search towards dimensions that are symmetric to the ones already in use.

Some of the key findings are discussed in further detail below.

8.1 The cross-linguistic acoustic space of stop contrasts

In Chapter 5 I characterized the cross-linguistic acoustic space of stop contrasts by transforming data collected from published acoustic studies using multi-dimensional scaling (MDS). The space of stop place and laryngeal contrasts was found to be largely reducible to four dimensions which were phonetically interpretable. The first dimension was primarily defined by measures of spectral shape, with particularly strong positive loadings for skewness and kurtosis and negative loadings for COG and standard deviation. Thus, it was correlated with stop place of articulation, representing a continuum from diffuse to compact spectral shape. This dimension corresponds to the classical dimension of diffuseness (Jakobson et al., 1952), separating the more diffuse labials and denti-alveolars from the more compact velars. The second dimension was found to represent a voicing dimension, which resembled the traditional VOT continuum (Lisker and Abramson, 1964), going from voiced to voiceless stops. The third dimension was correlated with voice quality and voicing such that aspirated stops were separated from modal and constricted stops, although there was some place specificity. Finally, the fourth dimension, which mostly captured some residual variation, was still interpretable as distinguishing place contrasts, separating burst properties such that shorter, more diffuse bursts were separated from longer, more intense bursts.

Putting these together, we can think of a somewhat simplified stop phonetic space as consisting of two dimensions that separate place – the first (corresponding to the first dimension in the MDS solution) distinguishes coronals, labials and velars along a continuum, and the second (corresponding to the fourth dimension in the MDS solution) separates the labials from the coronals and velars. Putting these together, we can formulate the space of stop place contrasts as a two-dimensional space whose x-axis represents spectral shape and whose y-axis represents burst properties.

Similarly, we can think of stop laryngeal contrasts as a two-dimensional space. The first dimension (corresponding to the third dimension in the MDS solution) distinguishes voiced stops from voiceless ones, and the second dimension (corresponding to the fourth dimension in the MDS solution) distinguishes laryngeal category. These two spaces can be combined to create the full space of stop place and laryngeal contrasts.

The previous unavailability of such a space, which can function as the equivalent of the $F_1 - F_2'$ vowel space for stops, has been a major bottleneck in the study of stop inventories. The parameterization of such a space for stops can enable future studies on stops.

To build a more realistic acoustic space of stops, future work will be conducted on a more controlled dataset with primary acoustic measures rather than imputing them based on data compiled from other studies. One limitation of the current study is that place contrasts were restricted to the near-universal labial-coronal-velar series. While future work can add other places of articulation, the coronal place itself needs to be further elaborated on, since the current study collapses dental and alveolar stops into a single category. However, the acoustic correlates that explain most of the variance in the data, namely the spectral moments, differ between dentals and alveolars (Sundara, 2005). Therefore, future work should model the various coronals as separate place categories.

The use of a controlled dataset will also reduce the noise in the present data, since the studies from which the acoustic data were compiled varied in their methodology. This will also increase the signal, as the imputation procedure used in Chapter 5 may have led to the loss of language-specific information. The analysis of such a newly compiled and better controlled dataset is more likely to reveal consistent effects of language and inventory type on the acoustic space of stop contrasts.

Such a dataset will also be capable of generating probability distributions in the acoustic space associated with different stop categories, as opposed to the centroids used here. This can enable the modeling of dispersion as measured by degree of overlap rather than distance between centroids[1].

8.2 Functional and structural properties

To understand which factors influence stop inventory contents, the effects of two types of predictors on inventory frequency were evaluated. The first set of predictors were dispersion-theoretic factors representing the functional pressures on inventories. These included maximizing dispersion, minimizing articulatory effort and maximizing inventory size. The second set of factors included the structural properties of maximizing economy and symmetry. Economy was evaluated in featural terms, while symmetry was evaluated in both featural and acoustic terms. Two different models investigated the effect of these functional and structural properties on stop inventory typology.

8.2.1 Featural space

A model couched in MaxEnt grammars modeled the frequency of all possible stop inventories, based on their featural properties. It found that structural constraints like feature economy and symmetry, which promote the maximal use of distinctive features, though positively correlated with the frequency of stop inventories, are not needed in a model that explicitly models dispersion-theoretic constraints. Moreover, dispersion based on minimum rather than global distance (also known as the *bottleneck*) provided a better fit to the data.

[1]See McGahay (2025) for an example of modeling dispersion in vowel inventories as category overlap rather than distance between centroids.

This model predicted the most economical inventory within each size to be optimal. In sizes that do not lend themselves well to being economical (i.e. when the number of stops is smaller than the maximum number of stops that can be derived from the features), the gaps are determined by markedness. Inventory frequency predicted by this model was positively correlated with both economy and symmetry, meaning the interaction of phonetically defined constraints predicts more economical and symmetric inventories to be more frequent in the world's languages, without the need for specific constraints that enforce these structural properties.

Weights from the MaxEnt model were also useful in hypothesis testing of different ancestral states. Chapter 4 demonstrated how the typological generalizations from the MaxEnt model can be incorporated into models comparing different ancestral inventories in terms of how well they predicted the distribution of later inventories. This approach, whereby weights from a general model of a number of languages is first built to arrive at a set of weights that represent typological tendencies across languages, can be used in future models of sound change.

In future work, the model sketched out in Chapter 4 can be made more realistic by incorporating priors that represent perceptual confusability between pairs of contrasts that are the inputs and outputs of sound change. Moreover, simulating a series of sound changes can benefit from iterative models that feed into each other. More generally, more sophisticated models of sound change can be built, which can take general typological tendencies of inventories into account by specifying these are priors, as was done in the hypothesis-testing model.

8.2.2 Acoustic space

In contrast to the MaxEnt Model, which was defined in featural terms, hurdle models were employed to understand the contents of stop inventories in terms of their acoustic properties. The acoustic space constructed in Chapter 5 was used to evaluate the factors influencing the contents of stop inventories. The effects of dispersion in the 4D space, articulatory effort, inventory size and acoustic symmetry were investigated in predicting two separate outcomes

– whether an inventory is attested, and if it is, how frequently it is attested.

Results from the zero component of the hurdle model showed that all four factors influence whether or not an inventory is attested versus unattested. In a somewhat surprising result, the acoustic symmetry of an inventory had a *negative* effect on attestedness, meaning that the more symmetric an inventory is in acoustic space, the *less* likely it is to be attested. Beyond this, larger, more dispersed and less articulatorily effortful inventories are more likely to be attested, in line with results from the MaxEnt model.

The count component of the hurdle model found that frequent inventories are differentiated from rare inventories by the same three dispersion-theoretic factors. In this component, symmetry was not found to significantly influence the frequency of attested inventories, although the direction of its effect suggested that more common inventories may be more acoustically symmetrical. The count component is more directly comparable to the MaxEnt model, as they both model frequencies of occurrence, rather than a binary outcome like the zero component of the hurdle model. Thus, results from these two models – the MaxEnt model and the count component of the hurdle model – found the same pattern, whereby structural constraints or predictors are not necessary in modeling stop inventory frequency.

Also in line with results from the MaxEnt model, the hurdle model found that the bottleneck approach to maximizing dispersion better characterized the frequency of attestation of stop inventories.

The results from the hurdle model were also found to be robust to one-off inventories, meaning that erroneously described inventories likely did not have a noticeable effect on the results from this model.

Despite their apparent success in modeling inventory attestedness and frequency, the deployment of hurdle models should be approached with caution when modeling linguistic data. They are a promising way of modeling in cases where zero and non-zero values are meaningfully different. However, since the zero component of the hurdle model treats all zero outcomes as arising from a process that may be separate from the process that produces non-zero outcomes, a hurdle model will treat something like an accidental gap in the same way it treats a principled, theoretically expected gap. Nevertheless, approaches that treat linguistic phenomena as strictly categorical (e.g., Du and Durvasula, 2022) may find the use of hurdle models fruitful.

8.3 Acoustic search

To model how inventory formation works in the continuous acoustic space without a priori knowledge of the distribution of attested inventories, a search algorithm inspired by mechanisms of biological evolution was implemented. This multi-objective genetic algorithm was able to search the acoustic space developed in Chapter 5 to construct inventories that could simultaneously balance various conflicting objectives that were found to influence the contents of stop inventories.

This algorithm differed from the hurdle and MaxEnt models in two key ways. First, the algorithm selected points in the 4D MDS space, not categories. Second, it does not fit any inventory frequency data. It simply searched the space for points which would balance five objectives – (i) maximize dispersion (ii) minimize articulatory cost (iii) maximize symmetry (iv) maximize the number of distinct categories (v) follow a hierarchy based on Lindblom and Maddieson (1988), enforcing the comprehensive use of simpler categories before including more articulatorily complex categories. The only information about which inventories are frequent the algorithm had access to was in the form of weighted distances in the calculation of the dispersion objective.

Despite this impoverished knowledge of the contents of real stop inventories, the search algorithm devised inventories that were typologically reasonable. The set of constraints that produced the most reasonable inventories did not include dispersion as an objective. Moreover, some of the other objectives converged on the same inventory, whose contents matched that of the largest stop inventory attested in PHOIBLE.

Although the exclusion of dispersion was found to produce more unattested inventories, dispersion is still likely required as an objective in the genetic search. The objective of maximizing inventory size implicitly includes dispersion as it pushes points in the 4D space away from each other over the course of evolution. Including dispersion as a separate objective, then, duplicates its effects and produces predicted inventories that are far too dispersed than attested stop inventories are.

The results from the genetic search can be interpreted as evidence for Lindblom and Maddieson's proposal that the effects of economy can be derived from a mechanism whereby a growing inventory first saturates a basic phonetic space and then expands it. That symmetry and maximizing diversity are necessary objectives in addition to Lindblom and Maddieson's hierarchy hints at the possible mechanism underlying the expansion of the phonetic space once the basic space is saturated. As diversity increases, the basic space is saturated but its expansion may be influenced by symmetry. That is, once the space is saturated, phonetic symmetry in the acoustic space guides its expansion into areas of the space that are symmetrical to those that are already employed. Symmetry also allows saturation of the most complex spaces as the hierarchy objective does not enforce saturation beyond the saturation point. That is, an inventory like /p t k $\mathrm{p^h}$/ satisfies symmetry as the basic space /p t k/ is fully saturated, but this objective then forces saturation in the higher dimension, yielding /p t k $\mathrm{p^h}$ $\mathrm{t^h}$ $\mathrm{k^h}$/.

Finally, it was found that scaling the objective space with the weights from the MaxEnt model in Chapter 3 yielded more realistic inventories. Thus, filtering the inventories found by the search algorithm iteratively through the phonological constraints yields better results. This suggests that scaling the objective space to bias the search towards regions that receive higher weights in the MaxEnt model generates more realistic inventories.

Note that the effect of adherence to Lindblom and Maddieson's hierarchy was included as a specific objective only in the MOGA. The MaxEnt and hurdle models did not specify it as a constraint or predictor. Future work can evaluate its effect on inventory likelihood explicitly using these other models.

Despite the fairly complex setup compared to previous work using GAs, the performance of the MOGA was poorer than expected. Even though the best MOGA found attested inventories, they did not match the most commonly attested inventories. A better performing search would have found inventories that more closely resemble the most commonly attested inventories in the world's languages, such as /p t k b d g/ over the MOGA's preferred /p t k ph th kh/.

Part of the issue is with the model setup. The inventory /p t k ph th kh/ is associated with a lower articulatory cost than /p t k b d g/, as the voiced stops incur greater costs as the place of articulation moves from front to more posterior closures, while the aspirated stops all have the same articulatory cost as the simplest voiced stop (/b/). However, it is well known that maintaining voicing is less effortful than initiating it (Westbury and Keating, 1986). The way the articulatory effort associated with a given inventory is calculated in the MOGA does not include such effects of position on articulatory effort. Enriching the model with variation in articulatory cost in different environments (e.g., Hayes, 1999) may improve its performance.

Another potential source of poor performance is the mechanism mapping of the individual points in the genome to the respective stop category labels. Recall that the genome consists of 18 different points that can be drawn from the MDS space. The points, identified by their (x, y, z, w) co-ordinates, receive category labels based on the category centroids they are closest to in MDS space. Without the diversity objective, a genome can consist of 18 points with unique 4D co-ordinates that all belong to the same category. Such a genome would represent a stop inventory consisting of a single stop. The diversity objective pushes individual points in a genome away from each other so that they are more likely to be associated with distinct category labels.

This approach is potentially problematic. First, it duplicates the effect of dispersion, as discussed earlier. The second issue arises specifically from the process used to assign category labels to points. Labeling is based on calculating the distance of a given point to each of the 18 category centroids and assigning the label of the closest centroid. This approach is unable to account for the variability in acoustic space exhibited by various categories. For example, it is possible that, given the articulatory variation in the production of, say, implosives (Nihalani, 1986), their distribution in the acoustic space displays greater variability than that of, say, plain voiced stops. Therefore, it is possible for a point representing an implosive to lie farther from its own category centroid while still falling within that category's distribution in the space, and at the same time to lie closer to a different category centroid while remaining outside that category's distribution.

Finally, this method of assigning category labels to unique points introduces a challenge in evaluating the results. At the end of the evolution, the MOGA produces a set of pareto-optimal solutions. Each of these is a unique set of 18 points in the MDS space, not 18 unique stop categories. This means that distinct solutions may represent the same inventory. Practically, this means that despite a large number of solutions entering into the pareto-optimal set, the categorical inventories they represent form a much smaller set. Thus, some sizes are not represented in the pareto-optimal set, making it difficult to identify the most optimal inventories at each size, as was done in the MaxEnt model and in previous work on vowels (Liljencrants and Lindblom, 1972).

Future work on inventory structure using genetic algorithms may be successful if the shortcomings discussed so far can be addressed. The acoustic space itself must be enriched with information about category distributions, rather than simply centroid locations. Second, the calculation of articulatory effort must take into account the effect of position. And finally, the diversity objective must be implemented in a way that does not implicitly include dispersion. Beyond these immediate fixes, a search algorithm that operates in continuous space that does not represent a reduced acoustic space, but rather directly refers to acoustic values, may be capable of constructing a universal prior distribution from which inventories are drawn, a proper goal of typological research (Cotterell and Eisner, 2017). This is particularly difficult to do for stops given their complex acoustic properties, but the search algorithm presented in this dissertation may offer a starting point on how to tackle such a problem.

8.4 Typological generalizations

The statistical modeling in Chapter 2 and the MaxEnt model in Chapter 3 also provided evidence towards previously proposed typological generalizations. Chapter 2 found evidence for the influence of phonetic factors, as evidenced by gaps at /p/, /p'/ and /g/. The predicted gaps match Maddieson's (1984) findings on UPSID. Additionally, I also found that in all three cases, the plain voiceless (T), plain voiced (D) and voiceless constricted (Tc) series are just as likely to occur without any gaps. This chapter also found evidence for the feature economy principle, as indicated by co-occurrence and implicational relationships between different laryngeal categories. Broadly, the trends found in this chapter were a close match to Maddieson's findings, suggesting that the genetically and areally imbalanced PHOIBLE sample is not skewed towards language- or area-specific patterns in stops.

The MaxEnt model found additional support for the implicational statement that the presence of voiced-aspirated (Dh) stops implies the presence of voiced (D) and voiceless-aspirated (Th) stops. Such an implicational relationship was not found to hold for the glottalic stops – the presence of voiced-constricted (Dc) stops did not imply the presence of voiced (D) and voiceless-constricted (Tc) stops. Finally, the MaxEnt model also found some evidence for the claim that implosives pattern unlike other stops (Clements and Osu, 2002), since when implosives were excluded from the model, the implicational relationship was found to hold for pulmonic and glottalic stops alike.

8.5 Finale

We have seen that the basic factors that shape inventory content are minimizing articulatory effort, maximizing minimum distance and maximizing the size of the inventory. We also found evidence for the mechanism of inventory size expansion proposed by Lindblom and Maddieson (1988). Inventories may acquire more stops by first saturating basic spaces and then expanding into symmetrical regions.

The three models presented in this dissertation are explicitly acknowledged as models that generate the typology of stop inventories, and are not claimed to be models of learning or change. I do not claim that the above factors are directly implicated in the grammars, whether phonological or phonetic, of speakers, listeners and learners. Language users do not generate all possible inventories and then choose the most optimal one, nor do they blindly search the universal, cross-linguistic acoustic space for stop categories. Instead, I propose that these factors must constrain change in some way. While this dissertation has treated inventories as static objects, the factors uncovered here can be interpreted as shaping inventories through sound change. That is, there must be a tendency towards larger, more dispersed and less articulatorily effortful stop systems. The structural properties, by contrast, must emerge via the interaction of these factors, without themselves constraining change in any way.

A.1 Models for evaluating place co-occurrence

1. Voiceless, unaspirated stops (T)

   (a) Model t1 (all two-way interactions): `count ~ p + t + k + p*t + t*k + p*k`

   (b) Model t2 (no labial x velar): `count ~ p + t + k + p*t + t*k`

   (c) Model t3 (no labial x denti-alveolar): `count ~ p + t + k + p*k + t*k`

   (d) Model t4 (no denti-alveolar x velar): `count ~ p + t + k + p*t + p*k`

   (e) Model t5 (main effects only): `count ~ p + t + k`

   The specified priors incorporate a gap at [p].

2. Voiced, unaspirated stops (D)

   (a) Model d1 (all two-way interactions): `count ~ b + d + ɡ + b*d + d*ɡ + b*ɡ`

   (b) Model d2 (no labial x velar): `count ~ b + d + ɡ + b*d + d*ɡ`

   (c) Model d3 (no labial x denti-alveolar): `count ~ b + d + ɡ + d*ɡ + b*ɡ`

   (d) Model d4 (no denti-alveolar x velar): `count ~ b + d + ɡ + b*d + b*ɡ`

   (e) Model d5 (main effects only): `count ~ b + d + ɡ`

   The specified priors incorporate a gap at [ɡ].

3. Voiceless, aspirated stops (Th)

   (a) Model th1 (all two-way interactions): count $\sim$ $p^h + t^h + k^h + p^h{*}t^h + t^h{*}k^h + p^h{*}k^h$

   (b) Model th2 (no labial x velar): count $\sim$ $p^h + t^h + k^h + p^h{*}t^h + t^h{*}k^h$

   (c) Model th3 (no labial x denti-alveolar): count $\sim$ $p^h + t^h + k^h + t^h{*}k^h + p^h{*}k^h$

   (d) Model th4 (no denti-alveolar x velar): count $\sim$ $p^h + t^h + k^h + p^h{*}t^h + p^h{*}k^h$

(e) Model th5 (main effects only): `count` $\sim$ $p^h + t^h + k^h$

The specified priors do not incorporate any gaps.

4. Voiceless, constricted stops (Tc)

   (a) Model tc1 (all two-way interactions): `count ~ p' + t' + k' + p'*t' + t'*k' + p'*k'`

   (b) Model tc2 (no labial x velar): `count ~ p' + t' + k' + p'*t' + t'*k'`

   (c) Model tc3 (no labial x denti-alveolar): `count ~ p' + t' + k' + p'*k' + t'*k'`

   (d) Model tc4 (no denti-alveolar x velar): `count ~ p' + t' + k' + p'*t' + p'*k'`

   (e) Model tc5 (main effects only): `count ~ p' + t' + k'`

   The specified priors incorporate a gap at [p'].

5. Voiced, aspirated stops (Dh)

   (a) Model dh1 (all two-way interactions): `count` $\sim$ $b^h + d^h + g^h + b^h{*}d^h + d^h{*}g^h + b^h{*}g^h$

   (b) Model dh2 (no labial x velar): `count` $\sim$ $b^h + d^h + g^h + b^h{*}d^h + d^h{*}g^h$

   (c) Model dh3 (no labial x denti-alveolar): `count` $\sim$ $b^h + d^h + g^h + d^h{*}g^h + b^h{*}g^h$

   (d) Model dh4 (no denti-alveolar x velar): `count` $\sim$ $b^h + d^h + g^h + b^h{*}d^h + b^h{*}g^h$

   (e) Model dh5 (main effects only): `count` $\sim$ $b^h + d^h + g^h$

   The specified priors incorporate a gap at [g^h].

6. Voiced, constricted stops (Dc)

   (a) Model dc1 (all two-way interactions): `count` ∼ ɓ + ɗ + ɠ + ɓ*ɗ + ɗ*ɠ + ɓ*ɠ

   (b) Model dc2 (no labial x velar): `count` ∼ ɓ + ɗ + ɠ + ɓ*ɗ + ɗ*ɠ

   (c) Model dc3 (no labial x denti-alveolar): `count` ∼ ɓ + ɗ + ɠ + ɗ*ɠ + ɓ*ɠ

   (d) Model dc4 (no denti-alveolar x velar): `count` ∼ ɓ + ɗ + ɠ + ɓ*ɗ + ɓ*ɠ

(e) Model dc5 (main effects only): `count` ∼ ɓ + ɗ + ɠ

The specified priors incorporate a gap at [ɠ].

A.2 Best performing models for each laryngeal category

I Plain voiceless stops

	Estimate	**Error**	**95% CI**
Intercept	1.03	0.17	[0.71, 1.35]
p	-0.94	0.20	[-1.33, -0.56]
t	-0.83	0.20	[-1.20, -0.44]
k	-0.96	0.21	[-1.37, -0.55]
pt	**1.01**	**0.20**	**[0.63, 1.40]**
tk	**1.96**	**0.27**	**[1.42, 2.49]**
pk	**0.88**	**0.19**	**[0.50, 1.25]**

Table A.1: *T1 result summary (all reference levels are 0, i.e., absence)*

For the plain voiceless stops, results from the best performing model (T1), which has all two-way interactions, are shown in Table A.1. Since this is a negative binomial model, the coefficients are log-transformed expected counts. All effects are credible since none of the 95% Credible Intervals (CI) contain zero. The crucial coefficients, those of the interaction terms, are shown in bold. The main effects cannot be interpreted by themselves due to the presence of credible interactions. The coefficients of all three interaction terms are positive, meaning the presence of each interaction is associated with a greater count of languages. Greater values of the estimate can be interpreted as greater likelihood of the corresponding predictor occurring in a language since it has a greater effect on the dependent variable. According to Table A.1, the interaction `tk` has the greatest effect on the number of languages, followed by `pt`, and then `pk`. This means that languages with /t/ and /k/ outnumber those with /p/ and /t/ which, in turn, outnumber those with /p/ and /k/. This means that a gap at /p/ is more likely than a gap at /t/ or /k/ since the interaction terms that have /p/ have lower estimates than the interaction term without /p/. Moreover, the interaction of /t/ with /p/ has a greater effect on `count` than the interaction of /k/ with /p/, meaning /t/ is more likely than /k/ to occur with /p/. Therefore, based on this result, an implicational hierarchy can be set up such that presence of /p/ implies the presence of /k/, which similarly implies presence of /t/. This is in line with Maddieson's findings on UPSID.

II Plain voiced stops

	Estimate	Error	95% CI
Intercept	2.26	0.12	[2.03, 2.50]
b	-1.14	0.18	[-1.49, -0.78]
d	-1.28	0.19	[-1.65, -0.90]
g	-1.22	0.18	[-1.56, -0.86]
bd	**1.53**	**0.26**	**[1.03, 2.02]**
dg	**0.84**	**0.19**	**[0.48, 1.20]**
bg	**0.91**	**0.19**	**[0.55, 1.28]**

Table A.2: *D1 result summary (all reference levels are 0, i.e., absence)*

Table A.2 shows results from the best performing model (D1), which has all two-way interactions. All effects are credible since none of the 95% Credible Intervals (CI) contain zero. The coefficients of all three interaction terms are positive, meaning the presence of each interaction is associated with a greater count of languages. Moreover, the estimates of the interactions are in the order `bd` > `bg` > `dg`. This means that /b/ and /d/ are more likely to co-occur with each other than with /g/, and /b/ is more likely to co-occur with /g/ than /d/ is with /g/. Therefore, consistent with Maddieson (1984), the implicational hierarchy for plain voiced stops is: the presence of /g/ implies presence of /d/, which implies presence of /b/.

III Voiceless-aspirated stops

Table A.3 shows results from the best performing model (Th1), which has all two-way interactions. All effects are credible since none of the 95% Credible Intervals (CI) contain zero. The coefficients of all three interaction terms are positive and follow the order $\mathtt{p^hk^h} > \mathtt{p^ht^h} \approx \mathtt{t^hk^h}$. Therefore, the implicational hierarchy for voiceless aspirated stops is: the presence of /t^h/ implies presence of both /p^h/ and /k^h/.

	Estimate	**Error**	**95% CI**
Intercept	2.23	0.08	[2.07, 2.39]
p^h	-1.30	0.29	[-1.85, -0.73]
t^h	-1.43	0.25	[-1.91, -0.95]
k^h	-1.33	0.26	[-1.84, -0.80]
$p^h t^h$	**1.04**	**0.22**	**[0.59, 1.46]**
$t^h k^h$	**1.02**	**0.22**	**[0.59, 1.45]**
$p^h k^h$	**1.15**	**0.22**	**[0.74, 1.58]**

Table A.3: *Th1 result summary (all reference levels are 0, i.e., absence)*

IV Voiceless-constricted stops

	Estimate	Error	95% CI
Intercept	2.11	0.07	[1.97, 2.26]
p’	-1.37	0.29	[-2.05, -0.67]
t’	-1.26	0.25	[-1.85, -0.67]
k’	-1.31	0.26	[-1.84, -0.75]
p’t’	**0.83**	**0.36**	**[0.14, 1.53]**
t’k’	**1.06**	**0.23**	**[0.60, 1.51]**
p’k’	**0.89**	**0.35**	**[0.21, 1.58]**

Table A.4: *Tc1 result summary (all reference levels are 0, i.e., absence)*

Table A.4 shows results from the best performing model (Tc1), which has all two-way interactions. All effects are credible since none of the 95% Credible Intervals (CI) contain zero. The coefficients of the interaction terms are positive, meaning the presence of each interaction is associated with a greater count of languages, and follow the order t’k’ > p’k’ > p’t’. This suggests that /t’/ and /k’/ are more likely to co-occur, and /k’/ is more likely to co-occur with /p’/ than /t’/ is. Therefore, the presence of /p’/ implies the presence of /t’/, which implies the presence of /k’/, consistent with Maddieson (1984).

V Voiced-aspirated stops

Table A.5 shows results from the best performing model (Dh1), which has all two-way interactions. All effects are credible since none of the 95% Credible Intervals (CI) contain zero. The coefficients of all three interaction terms are positive, in the order $b^h d^h > d^h g^h = b^h\ g^h$. This suggests that /$b^{ɦ}$/ and /$d^{ɦ}$/ are more likely to co-occur with each other than with /$g^{ɦ}$/,

	Estimate	Error	95% CI
Intercept	1.94	0.07	[1.81, 2.07]
bɦ	-1.89	0.23	[-2.37, -1.44]
dɦ	-1.88	0.23	[-2.33, -1.44]
gɦ	-1.84	0.23	[-2.30, -1.37]
bɦdɦ	**2.43**	**0.36**	**[1.71, 3.14]**
dɦgɦ	**1.16**	**0.23**	**[0.70, 1.62]**
bɦgɦ	**1.16**	**0.24**	**[0.69, 1.63]**

Table A.5: *Dh1 result summary (all reference levels are 0, i.e., absence)*

supporting the implicational hierarchy: the presence of /gɦ/ implies the presence of /bɦ/ and /dɦ/.

VI Voiced-constricted stops

Results from the voiced-constricted model are included in Chapter 2 as they pattern unlike any of the other categories.

A.3 Additional MaxEnt models

I MaxEnt model with implosives separated

A different way of thinking about implosives is to think of them as variants of voiced stops, separate from the voiced-constricted stops that make up the Dc category.

Since this model models the same dataset as the optimal model in Chapter 3, we can compare their performance directly. In the table below, Model 1 refers to the model from Chapter 3 and Model 1′ refers to the model where implosives are treated as a variant of voiced stops.

II MaxEnt model with prenasalized stops

I also ran a model where prenasalized stops were more strongly voiced variants of voiced stops and implosives were grouped with other Dc stops.

This yields a different dataset whose performance cannot be compared directly to the

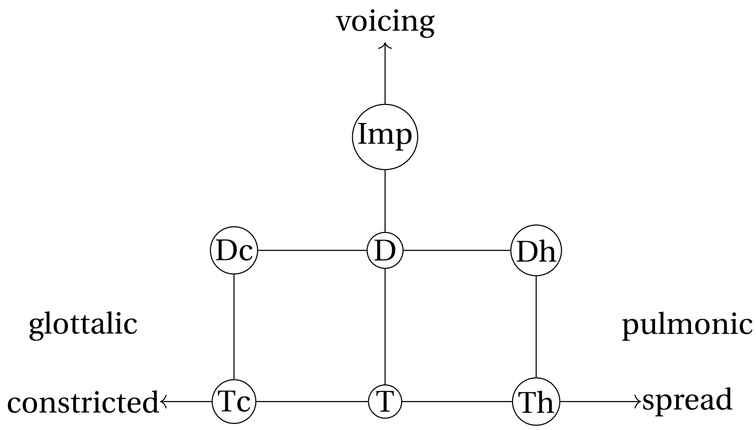


Figure A.1: *Model with implosives separated from voiced-constricted stops.*

Constraint	**Weight**
MAXIMIZECONTRASTS	1.73
MAXIMIZEPLACECONTRASTS	4.52
MINDIST=PLACE	0.46
MINDIST=VOICE	2.59
MINDIST=WIDTH	1.75
*VOICE	4.85
*IMPLOSIVE	3.43
*ASPIRATION	6.09
*CONSTRICTION	6.27
*VOICED-ASPIRATION	4.90
*VOICED-CONSTRICTION	1.24
*G	0.25
*PC	0.67
*GC	3.19
*LABIAL	0.7
*VELAR	0.12
MAXIMIZEWIDTHCONTRASTS	0
MAXIMIZEVOICECONTRASTS	0
*CORONAL	0
*G	0

Table A.6: *Constraint weights in the model with implosives separated from voiced-constricted stops.*

other models. Weights from this model are provided below.

Model	AIC	Δ(AIC)	Log-likelihood
Model 1	21579.54	0	-10770.77
Model 1′	21767.72	188.18	-10860.86

Table A.7: *Model comparison testing implicational relationships. Models are ordered from best to worst.*

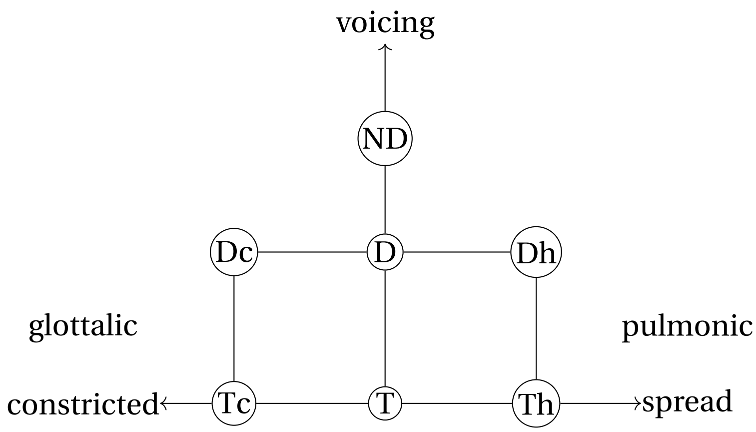


Figure A.2: *Model with prenasalized stops.*

Constraint	Weight
MAXIMIZECONTRASTS	1.06
MAXIMIZEPLACECONTRASTS	6.37
MINDIST=VOICE	1.90
*VOICE	3.43
*ASPIRATION	3.88
*CONSTRICTION	4.43
*VOICED-ASPIRATION	2.20
*VOICED-CONSTRICTION	2.82
*GC	2.18
*LABIAL	0.76
*VELAR	0.12
MAXIMIZEWIDTHCONTRASTS	0
MAXIMIZEVOICECONTRASTS	0
MINDIST=PLACE	0
MINDIST=WIDTH	0
*CORONAL	0
*PC	0
*G	0

Table A.8: *Constraint weights in the model with pre-nasalized stops.*

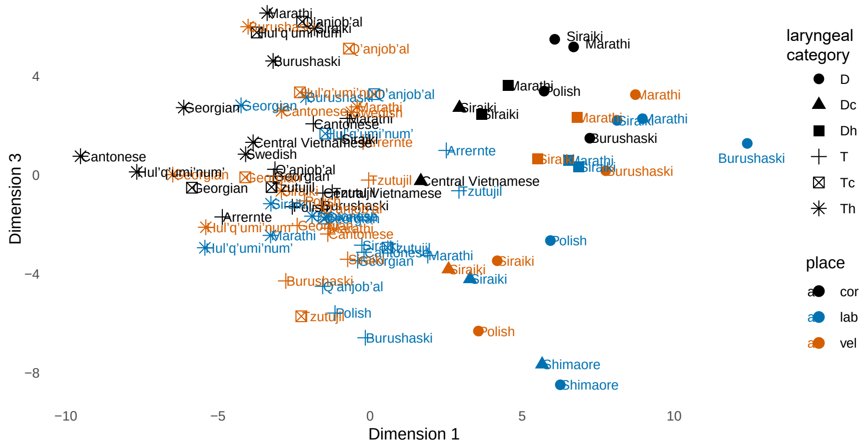
A.4 Languages and acoustic measures used in the MDS solution

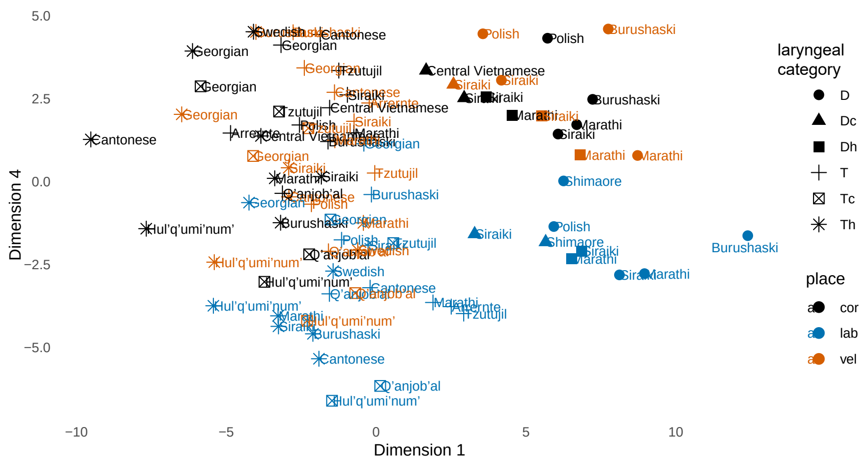
Language	Source	Acoustic Measures
Afrikaans	Coetzee et al. (2018)	VOT, Cf0
Amharic	Negesse (2021); Seid et al. (2009)	VOT, CD
Anyi	Koffi (2023)	VOT
Ao-Naga	Coupe (2003)	VOT
Armenian	Seyfarth and Garellek (2018)	VOT, CD
Arrernte	Tabain et al. (2016)	burst duration, COG, SD
Bengali	Narkar (2025); Yu et al. (2014)	VOT, Cf0, CPP, H1-H2
Burushaski	Hussain (2021)	VOT, COG, SD, skewness, F2
Cambodian	Kirby (2018)	VOT, Cf0
Canadian French	Sundara (2005)	VOT, burst intensity, COG, SD, skewness, kurtosis
Cantonese	Kwok and Stokes (1997); Ng and Wong (2009)	VOT, CD
Catalan	Pricop and Chodroff (2024); Recasens and Mira (2012)	VOT, Cf0
Central Vietnamese	Kirby (2018)	VOT, Cf0
Dahalo	Maddieson et al. (1993)	VOT
Dutch	Kuijpers (1996); Lisker and Abramson (1965); Pinget and Quené (2023); Van Alphen and Smits (2004)	VOT, Cf0, ΔCf0, CD burst duration
English	Chodroff and Wilson (2014); Lisker and Abramson (1964); Tabain et al. (2016)	VOT, Cf0, ΔCf0, CD burst intensity, burst duration, H1-H2, COG, SD, skewness, kurtosis, F2
Estonian	Ermus (2023)	VOT, burst intensity, COG, SD, skewness,
Finnish	Suomi et al. (2009)	VOT
Georgian	Vicenik (2010)	VOT, ΔCf0, CD, burst intensity, H1-H2, COG, skewness, kurtosis
German	Jessen (1999); Stoehr et al. (2017)	VOT, CD
Gipuzkoan Basque	Souganidis et al. (2024)	VOT
Gujarati	Rami et al. (1999)	VOT
Gurindji	Ennever et al. (2017)	VOT, CD
Hanoi Vietnamese	Tran et al. (2019)	VOT, CD, burst intensity
Hindi	Patil and Rao (2016)	VOT
Hul'q'umi'num'	Percival (2024)	VOT, Cf0, ΔCf0, CD burst intensity, H1-H2
Hungarian	Gósy and Ringen (2009); Gráczi and Kohári (2014)	VOT, CD
Igbo	Nkamigbo (2011)	VOT
Indonesian	Li et al. (2019)	VOT, CD
Italian	Bortolini et al. (1995) Cerrato and Falcone (1998); Kirby and Ladd (2016)	VOT, Cf0, CD, burst duration
Japanese	Homma (1981); Hussain and Shinohara (2019); Zhang (2020)	VOT, Cf0, CD

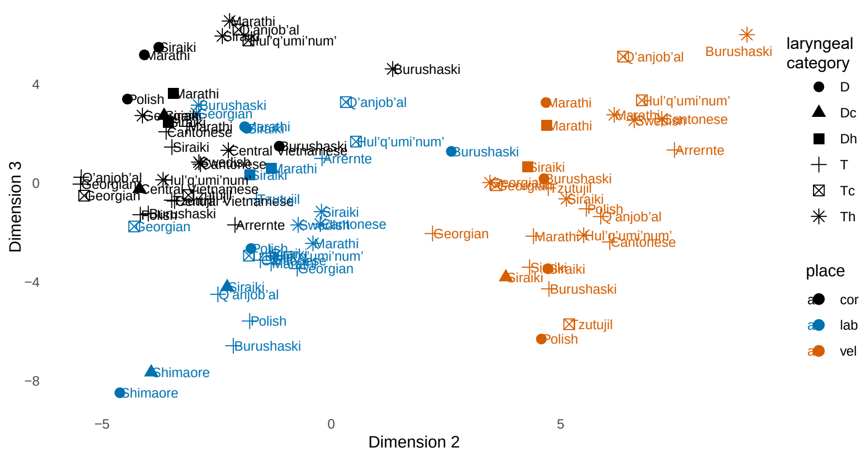
Kalasha	Hussain and Mielke (2020)	VOT, ΔCf0, COG, SD skewness, kurtosis
Kannada	Kochetov and Sreedevi (2016)	CD
Korean	Cho et al. (2002); Han and Weitzman (1970); Kang (2014); Kim et al. (2002, 2003)	VOT, Cf0, CD, H1-H2
Makasar	Tabain et al. (2016)	burst duration, COG, SD
Malagasy	Howe (2017)	VOT, Cf0, ΔCf0, CD, burst intensity, H1-H2
Mandarin Chinese	Deterding and Nolan (2007); Gu (2023)	VOT, Cf0
Maori	Maclagan and King (2007)	VOT
Marathi	Berkson (2012); Dmitrieva and Dutta (2020); Lisker and Abramson (1964)	VOT, Cf0, CPP, H1-H2
Mixtec (Yoloxóchitl)	DiCanio et al. (2020)	VOT, CD
Navajo	Cho and Ladefoged (1999)	VOT
Oromo	Negesse (2021)	VOT, Cf0, burst intensity
Peninsular Spanish	Casillas et al. (2015); Souganidis et al. (2024)	VOT, burst intensity, COG, SD, skewness, kurtosis
Persian	Bijankhan and Nourbakhsh (2009)	VOT, CD
Pitjantjatjara	Tabain and Butcher (2015); Tabain et al. (2016)	burst duration, COG, SD kurtosis, skewness, F2
Polish	Keating et al. (1981)	VOT
Q'anjob'al	Percival (2024)	VOT, Cf0, ΔCf0, CD burst intensity, H1-H2
Russian	Ringen and Kulikov (2012)	VOT
Somali	Abdiraman and Koffi (2012)	VOT
Shimaore	Mori (2023)	VOT, burst intensity, CPP
Sindhi	Hussain (2018)	VOT
Siraiki	Hussain (2018)	VOT
Standard Modern Greek	Nicolaidis et al. (2019); Themistocleous (2016)	VOT, CD, burst intensity, COG, SD, skewness, kurtosis
Swahili	Alsamaani (2023)	VOT
Swedish	Beckman et al. (2011); Lundeborg et al. (2012)	VOT
Swiss German	Ladd and Schmid (2018); Zebe-Sheng et al. (2025)	VOT, Cf0, ΔCf0, CD
Tamil	Lisker and Abramson (1964); Narayan (2023)	VOT, burst intensity
Telugu	Reddy (1985); Reddy et al. (2013)	VOT, CD
Tigrinya	Woldu (1985)	VOT, burst duration
Turkish	Öğüt et al. (2006)	VOT
Tzutujil	Bennett (2010)	VOT
Uyghur	Wang and Jia (2022)	VOT
Warlpiri	Tabain et al. (2016)	VOT, CD, burst duration, COG, SD
Wubuy	Bundgaard-Nielsen et al. (2016)	VOT, CD, COG
Yoruba	Grawunder et al. (2011)	VOT
Zuberoan Basque	Mounole (2004)	VOT

Table A.9: *Languages used in the MDS analysis.*

A.5 Additional MDS plots







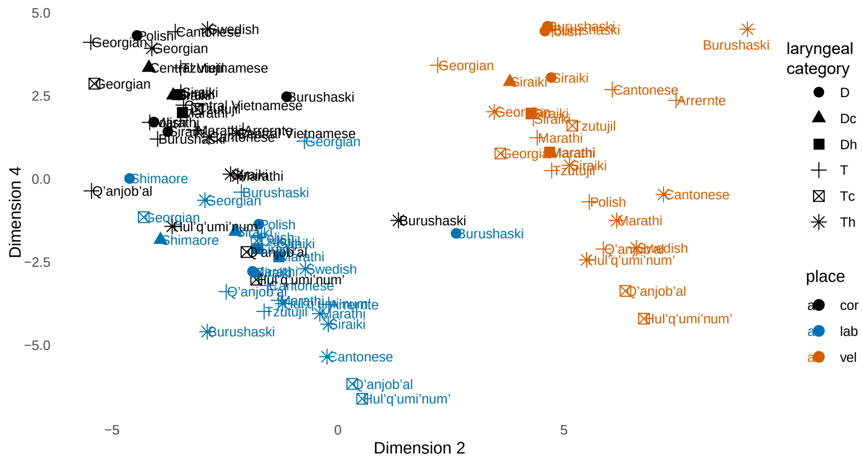


Figure A.3: *Additional 2D plots of dimensions from the MDS solution on representative languages.*

A.6 Hurdle model with total distance

Zero hurdle model – attestedness					
Predictor	Estimate	Std. Error	z value	Pr(> |z|)	Significance
(Intercept)	2.8817	1.0519	2.740	0.006153	**
total dispersion	2.6548	0.8058	3.295	0.000985	***
effort	-19.6325	0.9375	-20.942	$< 2 \times 10^{-16}$	***
symmetry	-6.3175	1.4914	-4.236	2.28×10^{-5}	***
size	-1.6438	0.6644	-2.474	0.013363	*
Count model – frequency					
Predictor	Estimate	Std. Error	z value	Pr(> |z|)	Significance
(Intercept)	-7.984	44.877	-0.178	0.8588	n.s.
total dispersion	7.454	1.425	5.231	1.69×10^{-7}	***
effort	-17.374	2.736	-6.350	2.15×10^{-10}	***
symmetry	-4.070	3.150	-1.292	0.1964	n.s.
size	3.149	1.700	1.853	0.0639	.

Table A.10: *Hurdle model summary for stops with total dispersion.*

Bibliography

Abdiraman, M. and Koffi, E. (2012). An exploratory VOT analysis of Somali stop consonants. *Linguistic Portfolios*, 1(1):2.

Alsamaani, M. (2023). The voice onset time of Arabic loanwords in Swahili. *Migration Letters*, 20(8):1225–1238.

Bally, C. and Gautier, L. e. (1922). *Recueil des publications scientifiques de Ferdinand de Saussure*. Geneva: Société Anonyme des Éditions Sonor.

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus*. University of California, Los Angeles.

Beckman, J., Helgason, P., McMurray, B., and Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of phonetics*, 39(1):39–49.

Belz, A. (1998). Discovering phonotactic finite-state automata by genetic search. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 2*, pages 1472–1474.

Bennett, R. (2010). Contrast and laryngeal states in Tzutujil. *The UCSC Linguistics Research Center*, pages 93–120.

Berkson, K. H. (2012). Capturing breathy voice: Durational measures of oral stops in Marathi.

Bijankhan, M. and Nourbakhsh, M. (2009). Voice onset time in Persian initial and intervocalic stop production. *Journal of the International Phonetic Association*, 39(3):335–364.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Blevins, J. (2018). Evolutionary phonology and the life cycle of voiceless sonorants. *Typological hierarchies in synchrony and diachrony*, pages 29–60.

Boersma, P. and Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, 25(2):217–270.

Bortolini, U., Zmarich, C., Fior, R., and Bonifacio, S. (1995). Word-initial voicing in the productions of stops in normal and preterm Italian infants. *International Journal of Pediatric Otorhinolaryngology*, 31(2-3):191–206.

Brandt, E. and Simpson, A. P. (2021). The production of ejectives in German and Georgian. *Journal of Phonetics*, 89:101111.

Brown, J. (2006). An evolutionary perspective on stop inventories. *Theoretical Linguistics*, 32:175–84.

Brugmann, K. (1897). *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. Strassburg: K. J. Triibner.

Bundgaard-Nielsen, R. L., Kroos, C., Baker, B. J., Best, C. T., and Harvey, M. (2016). Consonantal timing and release burst acoustics distinguish multiple coronal stop place distinctions in Wubuy (Australia). *The Journal of the Acoustical Society of America*, 140(4):2794–2809.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, 80:1–28.

Bybee, J. and Easterday, S. (2022). Primal consonants and the evolution of consonant inventories. *Language Dynamics and Change*, 1:1–33.

Casillas, J. V., Díaz, Y., and Simonet, M. (2015). Acoustics of Spanish and English coronal stops. In *ICPhS*.

Cerrato, L. and Falcone, M. (1998). Acoustic and perceptual characteristic of Italian stop consonants. In *ICSLP*.

Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, 39(4):612–625.

Cho, T., Jun, S.-A., and Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of phonetics*, 30(2):193–228.

Cho, T. and Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of phonetics*, 27(2):207–229.

Chodroff, E. and Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America*, 136(5):27622772.

Chodroff, E. and Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant vot in american english. *Journal of Phonetics*, 61:30–47.

Chodroff, E. and Wilson, C. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in american english. *Language*, 98(2):250–289.

Clark, R. (1992). The selection of syntactic knowledge. *Language acquisition*, 2(2):83–149.

Clark, R. and Roberts, I. (1993). A computational model of language learnability and language change. *Linguistic inquiry*, 24(2):299–345.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, 20(3):287–333.

Clements, G. N. and Osu, S. (2002). Explosives, implosives and nonexplosives: The linguistic function of air pressure differences in stops. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology 7*, pages 299–350. Mouton de Gruyter, Berlin and New York.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., and Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, 66:185–216.

Comrie, B. (2014). Typology and reconstruction. In *Historical Linguistics*, pages 74–97. Routledge.

Cotterell, R. and Eisner, J. (2017). Probabilistic typology: Deep generative models of vowel inventories. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1182–1192.

Coupe, A. R. (2003). *A phonetic and phonological description of Ao: a Tibeto-Burman language of Nagaland, north-east India*. Pacific linguistics, Research School of Pacific and Asian Studies, The .

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: journal of the Econometric Society*, pages 829–844.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of phonetics*, 28(4):441–465.

de Groot, A. W. (1931). Phonologie und phonetik als Funktionswissenschaften. *Travaux du Cercle Linguistique de Prague*, 4:116–147.

Deterding, D. and Nolan, F. (2007). Aspiration and voicing of Chinese and English plosives. In *Proceedings of the 16th international congress of phonetic sciences*, pages 385–388. Universität des Saarlandes Saarbrücken Germany.

DiCanio, C. T., Zhang, C., Whalen, D. H., and García, R. C. (2020). Phonetic structure in Yoloxóchitl Mixtec consonants. *Journal of the International Phonetic Association*, 50(3):333–365.

Dmitrieva, O. (2012). *Geminate typology and the perception of consonant duration*. PhD thesis, Stanford University.

Dmitrieva, O. and Dutta, I. (2020). Acoustic correlates of the four-way laryngeal contrast in Marathi. *Phonetica*, 77:209 – 237.

Dmitrieva, O., Llanos, F., Shultz, A. A., and Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49:77–95.

Du, N. and Durvasula, K. (2022). Phonetically incomplete neutralisation can be phonologically complete: evidence from huaian mandarin. *Phonology*, 39(4):559–595.

Dunbar, E. and Dupoux, E. (2016). Geometric constraints on human speech sound inventories. *Frontiers in psychology*, 7:1061.

Easterday, S. and Bybee, J. (2023). Diachronic phonological typology: understanding inventory structure through sound change dynamics. *Linguistic Typology*, 27(2):405–427.

Elbourne, P. (1998). Proto-Indo-European voiceless aspirates. *Historische Sprachforschung/Historical Linguistics*, 111(1. H):1–30.

Ennever, T., Meakins, F., and Round, E. R. (2017). A replicable acoustic measure of lenition and the nature of variability in Gurindji stops. *Laboratory Phonology*, 8(1).

Ermus, L. (2023). *Acoustic characteristics of Estonian plosive bursts*, page 94105. Sciendo.

Faytak, M. and Steffman, J. (2024). Voiced aspirates with mixed voicing in Yemba, a Grassfields Bantu language of Cameroon. *Journal of the International Phonetic Association*, 54(1):189–226.

Flemming, E. (2005). A phonetically-based model of phonological vowel reduction. *Manuscript. MIT*.

Flemming, E. (2021). Comparing maxent and noisy harmonic grammar. *Glossa*, 6(1):1–42.

Flemming, E. S. (2013). *Auditory representations in phonology*. Routledge.

Fonseca, C. M., Fleming, P. J., et al. (1993). Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. In *Icga*, volume 93, pages 416–423.

Francis, A. L., Ciocca, V., Wong, V. K. M., and Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America*, 120(5):2884–2895.

Gamkrelidze, T. V. and Ivanov, V. V. (1995). *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-Language and Proto-Culture*. De Gruyter Mouton.

Garellek, M. (2019). The phonetics of voice. In *The Routledge handbook of phonetics*, pages 75–106. Routledge.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading.

Goldwater, S. and Johnson, M. (2003). Learning ot constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, pages 111–120.

Gordon, M. K. (2016). *Phonological typology*, volume 1. Oxford University Press.

Gósy, M. and Ringen, C. O. (2009). Everything you always wanted to know about VOT in Hungarian. In *IXth International Conference on the Structure of Hungarian*.

Gráczi, T. E. and Kohári, A. (2014). Multiple bursts in Hungarian voiceless plosives and vot measurements. In *Proceedings of the 10th international seminar on speech production (ISSP)*, pages 158–161.

Grawunder, S., Winter, B., and Atoyebi, J. (2011). Voicing of labiovelar stops in Yoruba. In *ICPhS Proc.*, pages 767–770.

Greenberg, J. H. (1978). Diachrony, synchrony and language universals. *Universals of human language*, 1:61–92.

Groenen, P. J. and Borg, I. (2013). *The past, present, and future of multidimensional scaling*. Econometric Institute.

Gu, X. (2023). An acoustic study of the stop consonants in lettered-words produced by Chinese Mandarin speakers. *Theory & Practice in Language Studies (TPLS)*, 13(11).

Hall, D. C. (2007). *The role and representation of contrast in phonological theory*. PhD thesis, University of Toronto.

Han, M. S. and Weitzman, R. S. (1970). Acoustic features of Korean/p, t, k/,/p, t, k/and/ph, th, kh/. *Phonetica*, 22(2):112–128.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, 39(3):379–440.

Hayes, B. P. (1999). Phonetically driven phonology. *Functionalism and formalism in linguistics*, 1:243–285.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4):769–778.

Hockett, C. F. (1955). *A manual of phonology*. Waverly Press, Baltimore.

Holland, J. H. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, 9(3):273–281.

Hopper, P. J. (1973). Glottalized and murmured occlusives in Indo-European. *Glossa*, 7(2):141–66.

House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1):105–113.

Howe, P. J. (2017). *Tonogenesis in Central dialects of Malagasy: Acoustic and perceptual evidence with implications for synchronic mechanisms of sound change*. PhD thesis, Rice University.

Howell, P., Tang, K., Tuomainen, O., Chan, S. K., Beltran, K., Mirawdeli, A., and Harris, J. (2017). Identification of fluency and word-finding difficulty in samples of children with diverse language backgrounds. *International journal of language & communication disorders*, 52(5):595–611.

Hussain, Q. (2018). A typological study of voice onset time (VOT) in Indo-Iranian languages. *Journal of Phonetics*, 71:284–305.

Hussain, Q. (2021). Phonetic correlates of laryngeal and place contrasts of Burushaski. *Speech Communication*, 126:7189.

Hussain, Q. and Mielke, J. (2020). An acoustic and articulatory study of laryngeal and place contrasts of Kalasha (Indo-Aryan, Dardic). *The Journal of the Acoustical Society of America*, 147(4):28732890.

Hussain, Q. and Shinohara, S. (2019). Partial devoicing of voiced geminate stops in Tokyo Japanese. *The Journal of the Acoustical Society of America*, 145(1):149163.

Jaeger, J. J. (1980). Testing the psychological reality of phonemes. *Language and Speech*, 23(3):233–253.

Jakobson, R., Fant, G., and Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates.* MIT Press.

Jakobson, R. O. (1957). *Typological studies and their contribution to historical comparative linguistics.* University Press.

Jessen, M. (1999). Redundant aspiration in german is primarily controlled by closure duration. In *Proc. 14th ICPhS*, pages 993–996.

Jongman, A., Blumstein, S. E., and Lahiri, A. (1985). Acoustic properties for dental and alveolar stop consonants: A cross-language study. *Journal of Phonetics*, 13(2):235–251.

Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, 45:76–90.

Kang, Y. and Han, S. (2013). Tonogenesis in early contemporary Seoul Korean: A longitudinal case study. *Lingua*, 134:62–74.

Ke, J., Ogura, M., and Wang, W. S.-Y. (2003). Optimization models of sound systems using genetic algorithms. *Computational Linguistics*, 29(1):1–18.

Keating, P., Kuang, J., Garellek, M., Esposito, C. M., and Khan, S. (2023). A cross-language acoustic space for vocalic phonation distinctions. *Language*, 99(2):351–389.

Keating, P., Wymark, D., and Sharif, R. (2021). Proposal for superscript diacritics for prenasalization, preglottalization and preaspiration. *Journal of the International Phonetic Association*, 51(1):75–90.

Keating, P. A. (1984a). Phonetic and phonological representation of stop consonant voicing. *Language*, pages 286–319.

Keating, P. A. (1984b). Physiological effects on stop consonant voicing. *UCLA working papers in phonetics*, 59:29–34.

Keating, P. A. (2003). Phonetic and other influences on voicing contrasts. In *Proceedings of the 15th international congress of phonetic sciences*, pages 20–23.

Keating, P. A., Mikoś, M. J., and Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *The Journal of the Acoustical Society of America*, 70(5):1261–1271.

Kessinger, R. H. and Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of phonetics*, 25(2):143–168.

Kim, M.-R., Beddor, P. S., and Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30(1):77–100.

Kim, M.-R. C., Lee, J., and Lotto, A. J. (2003). An investigation of the manner contrast in intervocalic Korean stops. *The Korean Language in America*, 8:75–84.

Kingston, J. (1985). The ineffectiveness of larynx movement. *The Journal of the Acoustical Society of America*, 77(S1):S86–S87.

Kingston, J. (1993). Size, structure, and markedness in phonological inventories. *University of Massachusetts Occasional Papers in Linguistics*, 16(1):4.

Kingston, J. (2005). The phonetics of Athabaskan tonogenesis. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 269:137.

Kiparsky, P. (2018). Formal and empirical issues in phonological typology. *Phonological typology*, 23:9783110451931–003.

Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics*, 71:326–354.

Kirby, J. P. and Ladd, D. R. (2016). Effects of obstruent voicing on vowel f0: Evidence from true voicing languages. *The Journal of the Acoustical Society of America*, 140(4):2400–2411.

Kochetov, A. and Sreedevi, N. (2016). Articulation and acoustics of Kannada affricates: A case of geminate /tʃ/. *Clinical linguistics & phonetics*, 30(3-5):202–226.

Koffi, E. (2023). The acoustic phonetic properties of singly and doubly articulated stops in Anyi: A data-driven analysis. *Linguistic Portfolios*, 12(1):10.

Kokkelmans, J. H. (2021). *The phonetics and phonology of sibilants: A synchronic and diachronic OT typology of sibilant inventories.* PhD thesis, Università degli Studi di Verona.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional scaling.* Number 11. Sage.

Kuijpers, C. T. (1996). Perception of the voicing contrast by dutch children and adults. *Journal of Phonetics*, 24(3):367–382.

Kuo, J. (2023). *Phonological markedness effects in reanalysis.* PhD thesis, University of California, Los Angeles.

Kwok, I. C.-L. and Stokes, S. F. (1997). An electropalatographic description of Cantonese speech. *Asia Pacific Journal of Speech, Language and Hearing*, 2(2):125–138.

Ladd, D. R. and Schmid, S. (2018). Obstruent voicing effects on f0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics*, 71:229–248.

Ladefoged, P. and Loeb, G. (2002). Preliminary studies on respiratory activity in speech. *UCLA Working Papers in Phonetics*, pages 50–60.

Ladefoged, P. and Maddieson, I. (1996). *The sounds of the worlds languages*. Oxford: Blackwell Publishers.

Li, H., Baryadi, I. P., and Wijana, I. D. P. (2019). Sound pattern of Indonesian plosives. *Linguistik Indonesia*, 37(1):1–12.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, pages 839–862.

Lindau, M. (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics*, 12(2):147–155.

Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental phonology*, pages 13–44.

Lindblom, B. (1990). Phonetic content in phonology. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm*, 11:101–118.

Lindblom, B. and Maddieson, I. (1988). Phonetic universals in consonant systems. *Language, speech and mind*, 6278.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and speech*, 29(1):3–11.

Lisker, L. and Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20:384–422.

Lisker, L. and Abramson, A. S. (1965). Stop categorization and voice onset time. In *ICPhS*, pages 389–391.

Lundeborg, I., Larsson, M., Wiman, S., and Mcallister, A. M. (2012). Voice onset time in Swedish children and adults. *Logopedics Phoniatrics Vocology*, 37(3):117–122.

Mackie, S. and Mielke, J. (2011). Feature economy in natural, random, and synthetic inventories. *Where do phonological features come from*, pages 43–63.

Maclagan, M. and King, J. (2007). Aspiration of plosives in Māori: Change over time. *Australian Journal of Linguistics*, 27(1):81–96.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge University Press, Cambridge.

Maddieson, I. (1995). Gestural economy. In *ICPhS Proc.*, pages 574–77.

Maddieson, I. (2019). The sounds of the Bantu languages. In Nurse, D. and Philippson, G., editors, *The Bantu languages*, pages 15–41. Routledge.

Maddieson, I. and Precoda, K. (1989). Updating UPSID. *The Journal of the Acoustical Society of America*, 86(S1):S19–S19.

Maddieson, I., Spajić, S., Sands, B., and Ladefoged, P. (1993). Phonetic structures of Dahalo. *UCLA Working Papers in Phonetics*, 84:25–66.

Martin, A. (2011). Grammars leak: Modeling how phonotactic generalizations interact within the grammar. *Language*, 87(4):751–770.

Martinet, A. (1939). *Rôle de la corrélation dans la phonologie diachronique*. Verlag nicht ermittelbar.

Mayer, C., Tan, A., and Zuraw, K. R. (2024). Introducing maxent.ot: an R package for maximum entropy constraint grammars. *Phonological Data and Analysis*, 6(4):1–44.

McGahay, J. (2025). Modeling vowel system typology using iterated confusion minimization. In *Proc. Interspeech*.

Melchert, H. C. (1994). *Anatolian historical phonology*, volume 3. Rodopi.

Moran, S. and McCloy, D., editors (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History, Jena.

Mori, M. (2023). The acoustic characteristics of implosive and plosive bilabials in Shimaore. *Journal of the International Phonetic Association*, 53(3):950–976.

Mounole, C. (2004). Zubererazko herskarien azterketa akustikoa. *Anuario del Seminario de Filología Vasca" Julio de Urquijo"*, 38(1):207–248.

Mukherjee, A., Choudhury, M., Basu, A., and Ganguly, N. (2007). Emergence of community structures in vowel inventories: an analysis based on complex networks. In *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, pages 101–108.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3):341–365.

Narayan, C. R. (2023). Speaking rate, oro-laryngeal timing, and place of articulation effects on burst amplitude: Evidence from English and Tamil. *Language and Speech*, 66(4):851–869.

Narkar, J. (2025). Reconceptualizing VOT: Further contributions to marking 50 years of research on voice onset time. *Journal of Phonetics*, 108:101387.

Narkar, J. and Meszaros, M. (under review). A perceptual explanation for Hindi loanword adaptation. Language: Research Reports.

Negesse, F. (2021). An an acoustic analysis of Oromo and Amharic ejective stops. *Canadian Acoustics*, 49(4).

Ng, M. L. and Wong, J. (2009). Voice onset time characteristics of esophageal, tracheoesophageal, and laryngeal speech of Cantonese. *Journal of Speech, Language, and Hearing Research*, 52(3):780–789.

Nicolaidis, K., Sfakianaki, A., Vlahavas, G., and Kafentzis, G. (2019). An acoustic study of Greek voiceless stops. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, pages 1184–1188.

Nihalani, P. (1986). Phonetic implementation of implosives. *Language and Speech*, 29(3):253–262.

Nikolaev, D. and Grossman, E. (2020). Consonant co-occurrence classes and the feature-economy principle. *Phonology*, 37(3):419–451.

Nkamigbo, L. (2011). Experimental analysis of voicing contrasts in Igbo. *UJAH: Unizik Journal of Arts and Humanities*, 12(2):189–203.

Öğüt, F., Kiliç, M. A., Engin, E. Z., and Midilli, R. (2006). Voice onset times for Turkish stop consonants. *Speech Communication*, 48(9):1094–1099.

Ohala, J. J. (1980). Moderators introduction to symposium on phonetic universals in phonological systems and their explanation. In *Proceedings of the 9th International Congress of Phonetic Sciences*, volume 3, pages 181–185.

Padgett, J. (2003). Contrast and post-velar fronting in Russian. *Natural Language & Linguistic Theory*, 21(1):39–87.

Patil, V. V. and Rao, P. (2016). Detection of phonemic aspiration for spoken Hindi pronunciation evaluation. *Journal of Phonetics*, 54:202–221.

Percival, M. (2024). *Production and perception of ejective stops across languages.* PhD thesis, University of Toronto (Canada).

Pinget, A.-F. and Quené, H. (2023). Effects of obstruent voicing on vowel fundamental frequency in Dutch. *The Journal of the Acoustical Society of America*, 154(4):2124–2136.

Pricop, B. and Chodroff, E. (2024). Consonant f0 effects. a case study on Catalan. In *Proc. SpeechProsody 2024*, pages 329–333.

Prince, A. and Smolensky, P. (1994). Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in phonology: A reader*, pages 1–71.

Pulleyblank, D. and Turkel, W. J. (1998). Language acquisition in optimality theory. *Is the Best Good Enough?: Optimality and Competition in Syntax*, page 399.

Pulleyblank, D. and Turkel, W. J. (2000). Learning phonology: Genetic algorithms and Yoruba tongue-root harmony. *Optimality Theory: Phonology, syntax, and acquisition*, pages 554–591.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rami, M. K., Kalinowski, J., Stuart, A., and Rastatter, M. P. (1999). Voice onset times and burst frequencies of four velar stop consonants in Gujarati. *The Journal of the Acoustical Society of America*, 106(6):3736–3738.

Recasens, D. and Mira, M. (2012). Voicing assimilation in Catalan two-consonant clusters. *Journal of Phonetics*, 40(5):639–654.

Reddy, B., Kumar, N., and Sreedevi, N. (2013). Voice onset time across gender and different vowel contexts in Telugu. *Lang. India*, 14(12):252–63.

Reddy, N. (1985). Phonetic conditioning of duration of vowels and consonants in Telugu. *Osmania Papers in Linguistics*, 11:54–83.

Redford, M. A., Chen, C. C., and Miikkulainen, R. (2001). Constrained emergence of universals and variation in syllable systems. *Language and Speech*, 44(1):27–56.

Rice, K. and Avery, P. (1993). Segmental complexity and the structure of inventories. *Toronto Working Papers in Linguistics*, 12(2).

Ridouane, R. (2010). Geminates at the junction of phonetics and phonology. *Laboratory phonology*, 10:61–90.

Ringen, C. and Kulikov, V. (2012). Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic linguistics*, 20(2):269–286.

Sande, H. and Oakley, M. (2023). A typological survey of the phonological behavior of implosives: Implications for feature theories. *Phonological Data and Analysis*, 5(7):1–46.

Sapir, E. (1933). *La réalité psychologique des phonèmes*. Presses universitaires de France.

Schwartz, J.-L., Boë, L.-J., Badin, P., and Sawallis, T. R. (2012). Grounding stop place systems in the perceptuo-motor substance of speech: On the universality of the labial–coronal–velar stop series. *Journal of Phonetics*, 40(1):20–36.

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of phonetics*, 25(3):255–286.

Seid, H., Rajendran, S., and Yegnanarayana, B. (2009). Acoustic characteristics of ejectives in Amharic. In *Proc. Interspeech*, pages 2287–2290.

Seyfarth, S. and Garellek, M. (2018). Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics*, 71:425–450.

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, pages 287–308.

Simpson, A. P. (1999). Fundamental problems in comparative phonetics and phonology: does UPSID help to solve them. In *Proceedings of the 14th international congress of phonetic sciences*, volume 1, pages 349–352. University of California Berkeley.

Souganidis, C., Molinaro, N., and Stoehr, A. (2024). Bilinguals produce language-specific voice onset time in two true-voicing languages: The case of Basque-Spanish early bilinguals. *Linguistic Approaches to Bilingualism*, 14(3):370–399.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. *Human communication: A unified view*, pages 51–66.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, 17(1):3–45.

Stoehr, A., Benders, T., Van Hell, J. G., and Fikkert, P. (2017). Second language attainment and first language attrition: The case of VOT in immersed Dutch–German late bilinguals. *Second Language Research*, 33(4):483–518.

Stuart-Smith, J. (2004). *Phonetics and philology: sound change in Italic*. OUP Oxford.

Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *The Journal of the Acoustical Society of America*, 118(2):10261037.

Suomi, K., Toivanen, J., and Ylitalo, R. (2009). *Finnish sound structure: Phonetics, phonology, phonotactics and prosody*. University of Oulu.

Tabain, M., Breen, G., Butcher, A., Jukes, A., and Beare, R. (2016). Stress effects on stop bursts in five languages. *Laboratory Phonology*, 7(1):16.

Tabain, M. and Butcher, A. (2015). Stop bursts in Pitjantjatjara. *Journal of the International Phonetic Association*, 45(2):149176.

Themistocleous, C. (2016). The bursts of stops can convey dialectal information. *The Journal of the Acoustical Society of America*, 140(4):EL334EL339.

Ting, C., Sonderegger, M., Clayards, M., and McAuliffe, M. (2025). The crosslinguistic distribution of vowel and consonant intrinsic f0 effects. *Language*, 101(1):1–36.

Tran, T. T. H., Vallée, N., and Granjon, L. (2019). Effects of word position on the acoustic realization of Vietnamese final consonants. *Phonetica*, 76(1):1–30.

Van Alphen, P. M. and Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of phonetics*, 32(4):455–491.

Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 45:1–67.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., and Gelman, A. (2024). loo: Efficient leave-one-out cross-validation and waic for bayesian models. R package version 2.8.0.

Vicenik, C. (2010). An acoustic study of Georgian stop consonants. *Journal of the International Phonetic Association*, 40(1):5992.

Volaitis, L. E. and Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America*, 92(2):723–735.

Wang, S.-F. (2020). *Feature systems, inventory structures, and phonological classes.* PhD thesis, New York University.

Wang, Z. and Jia, Y. (2022). Acoustic analysis of the acquisition of plosives and affricates by preschool teachers. In *2022 International Conference on Asian Language Processing (IALP)*, pages 384–388. IEEE.

Wedel, A. B. (2006). Exemplar models, evolution and language change. *Linguistic Review*, 23(3).

Westbury, J. R. and Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of linguistics*, 22(1):145–166.

Whiteside, S. P. and Irving, C. J. (1997). Speakers' sex differences in voice onset time: Some preliminary findings. *Perceptual and motor skills*, 85(2):459–463E.

Winter, B. and Bürkner, P.-C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and linguistics compass*, 15(11):e12439.

Woldu, K. F. (1985). The perception and production of Tigrinya stops. *RUUL. Reports from Uppsala University Department of Linguistics*, (13).

Yates, A. D. (2017). *Lexical Accent in Cupeño, Hittite, and Indo-European.* PhD thesis, University of California, Los Angeles.

Yu, A. C. (2023). The actuation problem. *Annual Review of Linguistics*, 9(1):215–231.

Yu, K. M., Khan, S. U. D., and Sundara, M. (2014). Intonational phonology in Bengali and English infant-directed speech. In *Proc. Speech Prosody 2014*, pages 1130–1134.

Zebe-Sheng, F., Watter, C., Schmid, S., and Ladd, D. R. (2025). Transitory and sustained cf0 effects: Evidence from Swiss German. *Journal of Phonetics*, 113:101453.

Zhang, M. (2020). Effect of consonants on onset f0: Evidence from Kansai Japanese. *The Journal of the Acoustical Society of America*, 148(4):2472–2472.

Zuraw, K. (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory*, 28(2):417–472.

Zuraw, K. (2013). Map constraints (ms.). *University of California, Los Angeles*.