2018

A MARKEDLY DIFFERENT APPROACH: INVESTIGATING PIE STOPS USING MODERN EMPIRICAL METHODS

Phillip Barnett
University of Kentucky, phillip.barnett@uky.edu
Digital Object Identifier: https://doi.org/10.13023/ETD.2018.203

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Phillip Barnett, Student

Dr. Andrew Miles Byrd, Major Professor

Dr. Rusty Barrett, Director of Graduate Studies

A MARKEDLY DIFFERENT APPROACH: INVESTIGATING PIE STOPS USING MODERN EMPIRICAL METHODS

THESIS

A thesis submitted in partial
fulfillment of the requirements for
the degree of Master of Arts in the
College of Arts and Sciences at the
University of Kentucky

By
Phillip Barnett
Lexington, Kentucky

Co-Directors: Dr. Andrew Miles Byrd, Professor of Linguistics
and Dr. Kevin B. McGowan, Professor of Linguistics
Lexington, Kentucky 2018

ABSTRACT OF THESIS

A Markedly Different Approach:
Investigating PIE Stops Using Modern Empirical Methods

In this thesis, I investigate a decades-old problem found in the stop system of Proto-Indo-European (PIE). More specifically, I will be investigating the paucity of */b/ in the forms reconstructed for the ancient, hypothetical language. As cross-linguistic evidence and phonological theory alone have fallen short of providing a satisfactory answer, herein will I employ modern empirical methods of linguistic investigation, namely laboratory phonology experiments and computational database analysis. Following Byrd 2015, I advocate for an examination of synchronic phenomena and behavior as a method for investigating diachronic change.

In Chapter 1, I present an overview of the various proposed phonological systems of PIE and some of the explanations previously given for the enigmatic rarity of PIE */b/. Chapter 2 presents a detailed account of three lab phonology experiments I conducted in order to investigate perceptual confusability as a motivator of asymmetric merger within a system of stop consonants. Chapter 3 presents the preliminary form and findings of a computational database of reconstructed forms in PIE that I created and have named the Database of Etymological Reconstructions Beginning in Proto-Indo-European (DERBiPIE). The final chapter, Chapter 4, offers a summary of the work presented herein and conclusions that may be drawn, offering suggestions for continued work on the topic and others like it.

KEYWORDS: Proto-Indo-European, PIE, phonology, perceptual confusability, sound merger, computational lingusitics

Phillip Barnett

May 10, 2018

A Markedly Different Approach:
Investigating PIE Stops Using Modern Empirical Methods

By
Phillip Barnett

Co-Director of Thesis: Dr. Andrew Miles Byrd

Co-Director of Thesis: Dr. Kevin B. McGowan

Director of Graduate Studies: Dr. Rusty Barrett

Date: May 10, 2018

Dedicated to Danna, who has always supported me in following my dreams, even when I haven't.

ACKNOWLEDGMENTS

A little more than two years ago, I was an undergrad, finishing a double degree in Theatre and Communication at the University of Kentucky. I'd realized too late that I was passionate about linguistics to major in it, but that didn't stop me from taking Andrew Byrd's upper-level Historical Linguistics course. The course was a trial-by-fire and forced me to teach myself, with Andrew's guidance, the foundations of linguistic theory and typology. Seeing my enthusiasm and acumen, Andrew suggested I apply for the University's Master of Arts in Linguistic Theory and Typology, which I didn't even realize was an option. Naturally, I applied and, to my surprise, was accepted. These past two years in the program have been filled with challenges and opportunities I could have never imagined and that would not have been possible without the incredible faculty, students, and staff here.

The largest thanks go to my committee, Andrew Miles Byrd, Kevin B. McGowan, Mark Richard Lauersdorf, and Joseph C. Salmons, whose guidance and advice have shaped not only my thesis but the present and future trajectory of my academic career. In addition, I must offer enormous gratitude to my peer, Taha Husain, whose dedication to and enthusiasm for empirical linguistic research is unparalleled and helped make this thesis possible.

Further thanks go to Gašper Beguš, Joe Eska, David Goldstein, Dieter Gunkel, and Jamison Nielsen. Without the conversation and assistance provided by these individuals, my work would be flat and easily rejected.

Newton wrote, "If I have seen further it is by standing on the shoulders of Giants." But even that was taken from a predecessor, Bernard of Chartres. Thank you to all the giants on whose shoulders I am standing and to the village that helped me get up here.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction[1]

Proto-Indo-European(PIE) is the theoretical ancestral language of all Indo-European languages. While a multitude of the world's languages, from Latin to English to Persian to Greek, descend directly from PIE, the proto-language itself is unattested; we have no documentation of PIE. For over a century, scholars have taken to trying to reconstruct the phonology, morphology, syntax, semantics, etc. of PIE. Within the domain of its reconstructed phonology, a survey of the literature shows that there are a number of differing traditions. This chapter aims to present a survey of these traditions while providing the current evidence in favor of each viewpoint as well the drawbacks of each theoretical framework. More specifically, I will be dissecting the various reconstructions of the PIE phonological stop system. Further, I introduce the problem on which I will be focused for the duration of this thesis, the paucity of */b/ in the Standard Method's conception of the phonology of PIE. By investigating this issue with modern, empirical methods, I hope to, first and foremost, demonstrate the utility of such methodology for diachronic investigation.

1.1 The Tools of Reconstruction

In order to discuss the various reconstructions of PIE phonology, we must first take a look at the method by which comparative linguists have arrived at the conclusions that form the basis of the various theories of reconstruction. The field of Indo-European linguistics has, since its beginnings, relied heavily on what is known as the comparative method. Campbell 1998 describes the comparative method thusly:

> "The aim of reconstruction by the comparative method is to recover as much as possible of the ancestor language (the proto-language)[...] The work of reconstruction usually begins with phonology, with an attempt to reconstruct the sound system; this leads in turn to reconstruction of the vocabulary and grammar of the proto-language" (pp. 122-3).

The comparative method, at its most basic form, involves lining up possible phonological or lexical cognates from related languages and making a phonologically-informed guess as to the common source.

[1]**Portions of this thesis are adapted from a forthcoming paper authored by myself and Dr. Andrew Miles Byrd, Barnett and Byrd under consideration. The Methods and Results sections for the first two experiments of Chapter 2 were taken directly from the paper with minor changes. Chapter 1 here is an expanded overview of the state of the field of Indo-European linguistics and the current treatments of the problem of interest and is therefore an elaboration on the brief literature review section of Barnett and Byrd, under consideration.**

Table 1.1: Germanic Cognates for ‘night’ in Native Orthography.

English	German	Dutch	Swedish	Danish	Afrikaans	Scots	Gothic
night	Nacht	nacht	natt	nat	nag	nicht	nahts

Using the cognates from Germanic languages for night, as shown in Table 1.1, scholars reconstruct the Proto-Germanic form **naxts*[2]. The */n/ is obvious, as each form begins with /n/. The comparative method dictates that a majority usually rules, hence the reconstructed */a/ for the vowel and the final */t/. The */x/ is less obvious however, as the daughter languages contain /j, ç, x, ∅/. This is where the comparative method relies on cross-linguistic data and phonological theory to determine the most likely phonological changes. Finally, you will notice that the proto-Germanic form ends in */s/ while only one daughter language listed here does. This is reconstructed as languages are significantly more likely to lose a final /s/ than a language is to randomly insert one. Furthermore, */s/ was the nominative singular animate marker in Proto-Germanic and is present in non-Germanic languages, adding morphological and cross-linguistic evidence for its existence. For a more comprehensive look at the comparative method see the first chapter of Fortson's *Indo-European Language and Culture* (2010) or the fifth chapter of Campbell's *Historical Linguistics: An Introduction* (1998).

1.2 The Standard Model (SM)

Scholars most commonly reconstruct the PIE stops with a three-way voicing contrast at each of five places of articulation. The voicing contrast consists of voiceless unaspirated (*/t/), voiced unaspirated (*/d/), and voiced aspirated (*/d^h/), which should be more precisely viewed as murmured.[3] I will henceforth refer to this as the Standard Model (SM).

Using the comparative method, scholars have reconstructed a vocabulary for PIE from cognates in the oldest attested languages in each branch of the Indo-European tree. The SM of PIE phonology (e.g., Byrd 2018, Fortson 2010), presents a system of plosives that contains five places of articulation and a three-way voicing distinction. The places of articulation are bilabial, dental, palatal, velar, and labiovelar while the voicing modes are voiceless, voiced, and murmured. This system is presented in Table 1.2 in an orthography common to the field.

[2]An asterisk in Indo-European linguistics indicates that the form or phoneme is reconstructed(i.e., unattested), rather than ungrammatical, as the symbol indicates in other fields of linguistics. To indicate an ungrammatical form, I have adopted the practice of Byrd 2018, which places a superscripted *X* before the form, as in ^{X}/DeD/.

[3]Phonetic murmured stops feature voiced aspiration, most accurately represented as seen in [$b^{ɦ}$]. As the exact phonetic realization of this series of PIE stops is unknown, they are often interchangeably referred to as voiced aspirated and murmured stops. Murmured stops are found in the Indic languages, suggesting this manner of articulation.

Table 1.2: The traditional reconstruction of the PIE stop system.

	bilabial	dental	palatal	velar	labiovelar
voiceless	*p	*t	*ḱ	*k	$*k^w$
voiced	*b	*d	*ǵ	*g	$*g^w$
murmured	$*b^h$	$*d^h$	$*ǵ^h$	$*g^h$	$*g^{wh}$

Problems with the Standard Model

The SM is not free of problems. Jakobson 1957 notes:

> "To my knowledge, no language adds to the pair /t/ – /d/ a voiced aspirate $/d^h/$ without having its voiceless counterpart $/t^h/$, while /t/, /d/, $/t^h/$ frequently occur without the comparatively rare $/d^h/$, and such a stratification is easily explainable (cf. Jakobson-Halle); therefore theories operating with the three phonemes /t/ – /d/ – $/d^h/$ in Proto-Indo-European must reconsider the question of their phonemic essence" (p. 528).

Indeed, the traditionally reconstructed system is marked in that no attested stop system is congruent. There are a few similar systems, which will be discussed in Section 1.4, but none are identical in voicing mode distinctions. Furthermore, the root structure */DeD/, where D represents any unaspirated voiced oral stop, is apparently non-existent in the reconstructed forms of PIE despite the presence of such forms as */*ped-*/ 'foot' and */*steig-*/ 'mount' (Byrd 2018).

The unaspirated voiced bilabial stop **b* is also strikingly absent in the forms of PIE according to the SM. However, while **b* is rare, it is not entirely absent as in the reconstructed root **bel-* 'strength.' The literature has struggled to explain this paucity. Joseph (1985:5), for instance, has suggested that */b/ may have held special semantic or pragmatic significance, resulting in its scarcity. While Joseph is undoubtedly correct that certain phonemes and phonemic sequences may be sequestered for affective usage, this seems to push the mystery back to an earlier stage of PIE and we must still address what caused */b/ to enter this function within the language.

Both Jakobson's objections to the SM and the aforementioned phonological peculiarities remain, though some have been addressed in the recent decades. For instance, Iverson and Salmons (1992:295) and Barrack (2003:12) argue that the absence of ^{X}/DeD/ is to be expected, as double-stop roots of any shape are infrequent; in addition, stops occur more often in syllable onsets than in coda position (Barrack 2002:82–4). Cooper (2009:63) also points out that PIE roots tended to avoid roots beginning and ending with any segment of like manner, and therefore the absence of ^{X}/DeD/ should not come as a surprise.

Jucquois 1966

Jucquois 1966 Guy Jucquois, *La structure des racines en indo-européen envisagée d'un point de vue statistique* (Universa Wetteren, 1966) presented a statistical look at PIE consonantism under the SM. Many scholars have since noted problems with Jucquois' tabulations of PIE consonants (e.g., Vine 1988:397), the study seems to confirm the notion that */b/ is virtually absent in the most salient of reconstructions in PIE. Furthermore, Jucquois 1966 presents an abundance of the other labials, */p/, */b^h/, and */m/ compared to other consonants within their series of manner of articulation (e.g., */m/ is more abundant than */n/). These counts have led many scholars to propose a merger between scarce */b/ and another abundant labial. For instance, Vine (1988:397) proposes that */b/ merged with */m/. Jucquois 1966 will be discussed further in Chapters 2 and 3.

1.3 The Glottalic Theory

Hopper 1973 and Gamkrelidze and Ivanov 1973 independently proposed new systems to resolve the issues of the SM, also containing three-way voicing contrasts, known collectively as the Glottalic Theory (GT). Unlike the SM, the GT proposed the inclusion of a set of voiceless ejectives to replace the SM's voiced stops. Meanwhile, the murmured stops of the SM were replaced by voiced stops that could be allophonically aspirated. For a visualization of the proposed change to the standard model, see Table 1.3.

Table 1.3: The Glottalic Theory

Manner	Stop Type		Manner	Stop Type
voiceless	*/T/	→	voiceless	*/T/
voiced	*/D/	→	voiceless ejective	*/T'/
breathy	*/D^h/	→	voiced	*/D/

Proponents claim the GT provides an explanation for the scarcity of */b/ in PIE, as the bilabial ejective stop */p'/ is less common than many other ejectives, */k'/ in particular, and also addresses the absence of ^{X}/DeD/ roots, as many languages with ejective phonemes ban roots of the shape /T'eT'/ (Gamkrelidze and Ivanov 1973:153). However, most scholars continue to follow the SM as it does a more succinct job of predicting daughter forms and relies on more phonologically viable sound changes.[4]

Table 1.4: The GT reconstruction of the PIE stop system.

	bilabial	dental	palatal	velar	labiovelar
voiceless	*$p^{(h)}$	*$t^{(h)}$	*$\acute{k}^{(h)}$	*$k^{(h)}$	*$k^{w(h)}$
ejectives	*p'	*t'	*$\acute{k}$'	*k'	*k'^{w}
voiced aspirates	*$b^{(h)}$	*$d^{(h)}$	*$\acute{g}^{(h)}$	*$g^{(h)}$	*$g^{w(h)}$

[4]For an in-depth treatment of the Glottalic Theory, see Salmons 1993.

1.4 Cross-linguistic evidence

Cross-linguistic evidence is of the utmost importance in the field of Indo-European linguistics and historical linguistics in general (e.g., Byrd 2018; Weiss 2009; Fortson 2010). As it is cross-linguistic evidence that raises many of the issues with the traditional reconstruction of the PIE stop system, it is natural that this problem be tackled in-kind.

Cao Bang Theory (CBT)

The Cao Bang Theory (CBT), first proposed by Haider 1983 and followed up and given this name by Weiss 2009b and again followed up by Kümmel 2012, uses cross-linguistic evidence taken from the Central Tai language Cao Bang, spoken in northern Vietnam. The language's recent history provides a clear path by which a system of implosive stops and voiced stops could become a system of voiced and murmured stops. Weiss posits that this could have been the case with an earlier stage of PIE, making the three-way distinction seen in the SM more plausible. See Table 1.5 for the specific path Weiss proposes as a possibility.

Table 1.5: The Cao Bang Theory

Manner	Stop Type		Manner	Stop Type
voiceless	*/t/	→	voiceless	*/t/
voiced implosive	*/ɗ/	→	voiced	*/d/
voiced	*/d/	→	breathy	*/d^h/

While Weiss did not propose that the CBT could be responsible for the paucity of PIE */b/, the possibility that an earlier form of PIE may have included implosives offers another perceptual and articulatory pathway to a merger between /b/ and some other segment. Haider (1983:12) noted, based on cross-linguistic evidence, the possibility of a merger between */ɓ/ and */m/ at an early stage of the language, which would result in a lack of */b/ and an abundance of */m/, as mentioned in Section 1.2.

Table 1.6: The CBT reconstruction of the PIE stop system.

	bilabial	dental	palatal	velar	labiovelar
voiceless	*$p^{(h)}$	*$t^{(h)}$	*$\acute{k}^{(h)}$	*$k^{(h)}$	*$k^{w(h)}$
implosives	*ɓ	*ɗ	*ʄ	*ɠ	*$ɠ^w$
voiced egressive	*b	*d	*ǵ	*g	*g^w

!Xóõ

The Southern Khoisan language !Xóõ, also known as Taa, contains a plosive series[5] with four distinctions: voiceless unaspirated, voiceless aspirated, voiced unaspirated, and voiced aspirated. Notably, the system actually lacks /b^h/ (Traill 1994).[6] Comparative scholarship on the Khoisan languages is scarce and muddled with controversy and most scholarship that exists focuses on the clicks present in the languages (see Sarostin 2006).

Thus, it is not presently definitive that Proto-Khoisan included this four-way distinction, but other Khoisan languages include this system such as !Xun, a Central Khoisan language that does include /b^h/ (König and Heine 2008). Because of the cross-linguistic rarity of voiced aspirated stops, it is unlikely that they each developed these systems independently and thus it is likely !Xóõ previously contained */b^h/ and eventually merged the segment with another sound or lost it entirely. This lends to the possibility that early PIE may have contained */b/, but it eventually merged with */b^h/ in many or all contexts. In fact, phonetically, this merge makes even more sense than the merge that likely occurred in !Xóõ as /b^h/ and /p/ are more acoustically and articulatorily distinct than /b/ and /p/ (Ladefoged and Johnson 2011), especially considering the traditional view that PIE did not contain /p^h/.[7]

1.5 Empirical Methods of Investigation in the Study of PIE

Hock (1986:625) points out that PIE is not unique in having a gap or near-gap in the voiced labial stops, with /b/ being absent in Dargwa (Northeast Caucasian) and being unexpectedly rare in Dehu (Southern Oceanic). Still yet, it would be preferable to identify a reason for the rarity of */b/ in PIE, if at all possible. In this thesis, while there are further implications of my work and findings, I focus solely on the rarity of */b/ in PIE. Is it possible to arrive at an explanation within the SM, or must we assume an alternative consonantism to come up with a sensible solution?

Traditionally grouped amongst the humanities, the study of Indo-European linguistics tends to rely on comparative methods that have been in use for centuries. Meanwhile, other fields of linguistics, such as phonology and morphology and sociolinguistics, have been quick to adapt to new technologies and methods of empirical investigation.

Byrd 2017 argues against certain traditional reconstruction methods in Indo-European studies while simultaneously advocating for previously underused methods. He proposes three specific rules to follow when performing phonemic reconstructions: "absence of evidence is evidence of absence" (p. 33), "reconstructions don't exist in a vacuum" (p. 37), and "proto-languages are languages, too" (p. 38). This last rule

[5] Here we are only referring to pulmonic egressive speech sounds as !Xóõ contains a large inventory of clicks and click clusters.

[6] It is important to note here that the /b^h/ found in !Xóõ contains voiceless aspiration, making it distinct from a murmured stop, which features voiced aspiration.

[7] It is important to note that while the Glottalic theory posits that /p^h/ was present in the PIE stop system (see Figure 1.4), it was not phonemic.

is especially valuable for constructing an argument in favor of using empirical methods using living participants to investigate historical questions. More specifically, Byrd argues that reconstructions must "be grounded in the universal properties of all languages" (p. 38). While this is not a new concept to the field, it is clear that using this as a guiding principal, an experiment constructed to uncover such universal properties could be helpful to any investigation of PIE.

Salmons Forthcoming presents a brief yet blistering critique of the existing reconstructions of PIE obstruents. The general conclusion he draws is that none of the existing theories actually connect the concepts of phonology to phonetics. Salmons highlights the phonetic features that are used cross-linguistically to distinguish stops:

> "Only three contrastive features are needed to capture laryngeal features in the world's languages (assuming 'privative' features and using the label from Iverson & Salmons 2011): [spread] (associated with aspiration), [constricted] (glottalics) and [voice] (voicing). To capture a three-way system, at least two features are needed" (p. 2).

It is widely believed that */b/ was a rare phoneme by late PIE. This does not mean, however, that such was the case at earlier stages of the proto-language. In fact, one may reasonably conclude that the rarity of */b/ in late PIE derives from the fact that it was completely absent at an earlier stage (henceforth Middle PIE), a sound that was slowly reintroduced to the phonemic inventory through borrowings (**bak*- 'staff'), onomatopoeia (**baba*- 'gibberish'), and phonologically derived segments (*/pd-/ > *-bd- 'foot', in Ved. upa-bd-á-, Gk. ἔπι-βδ-ᾱ, etc.). If */b/ were completely absent in Middle PIE, it is not unreasonable to assume that it was present at an even earlier stage of PIE (henceforth Early PIE), having undergone a phonetic shift to an entirely different sound or having merged with another phoneme in the language.

The evolution of /p/ in the history of Proto-Celtic (Russell 1995:11–2) provides a parallel of the proposed sequence of events. The phoneme */p/ was certainly present in PIE and was inherited by Proto-Celtic, at which stage the phoneme itself was not lost, but rather shifted to an alternate pronunciation (likely */ɸ/). It was then lost entirely as a phoneme in Celtic, with only minor traces in certain environments. The phoneme /p/ was eventually reintroduced into the Celtic languages through loanwords (e.g., Mod.Ir. póg 'kiss' < Lat. (osculum) pacis, onomatopoeia (Mod.Ir. plimp 'boom'), and later sound laws (e.g. *penkʷe > *kʷenkʷe > W. pimp 'five'). Thus, it is entirely possible that */b/ had existed as a phoneme in Early PIE and then either shifted to an alternate pronunciation or merged together with an already existing phoneme in Middle PIE. To my knowledge, there is no evidence of a shift from */b/ to an alternate, yet-still-contrastive sound in PIE – i.e., there is no reason to assume that */h_1/,[8] for example, derives from */b/, especially given the fact

[8]By most modern theories of PIE consonantism, the language is said to have contained what are known as the laryngeals, represented as */h_1/, */h_2/, and */h_3/. A communis opinio on the phonetic realization of the laryngeals has yet to be reached, but many scholars posit that they were glottal and/or pharyngeal fricatives due to their inheritance as such into Hittite.

that visual stimuli would cause these to be less susceptible to perceptual merger (see McGurk and MacDonald 1976). However, there are hints that */b/ merged with another segment in PIE.[9]

[9] The final two paragraphs of Chapter 1 were taken from Barnett and Byrd under consideration with moderate modification.

Chapter 2 Perceptual Confusability Experiments

2.1 Experiment One: Modal-Murmur Confusability

Table 2.1: PIE Stops in Root-Initial Position(taken form Jucquois 1966:59)

Modal	b	25(1.2%)	129(6.4%)	b^h	Murmured
	d	53(2.7%)	67(3.3%)	d^h	
	ǵ	22(1.1%)	52(2.3%)	$ǵ^h$	
	g	50(2.5%)	47(2.6%)	g^h	
	g^w	37(1.8%)	12(0.6%)	g^{wh}	

As seen in Table 2.1, the number of */b^h/ tokens in root-initial position tallied by Jucquois 1966 is quite striking, with nearly double the number of dental roots, and more than twice as many as the dorsal roots combined. The merger of plain voiced and voiced murmured stops is widespread, being the most common outcome of the two series within the Indo-European daughter languages, occurring in Albanian, Anatolian, Balto-Slavic, Celtic, Iranian, and in many dialects of Indic, though the merger is always in the direction of a voiced stop, and not a murmured one. Moreover, the mergers are never restricted to a specific place of articulation, and certainly do not seem to target the labial node alone. It is thus incumbent upon us to find additional evidence that such a diachronic development is a possible one. As existing cross-linguistic and philological evidence falls short of accomplishing this goal, we may turn to experimental phonology for answers. Scholars in nearly all subfields of linguistics are increasingly using experimentation to investigate theoretical claims; for the use of experimentation in historical phonology, see Yu 2015.

Methods

In 1955, Miller and Nicely conducted an experiment to determine which English consonants were most easily confusable to an English-speaking listener. Their experiment involved playing for participants pairs of nonce words[1] of the shape CV and then asking the participants to identify whether the two words were the same or different. As all of the consonants used in their experiment were phonemic in English, and were therefore naturally easy for listeners to distinguish, the researchers added varying amounts of noise to the audio recordings to increase confusability across the board. I have used their methodology as a model for the perceptual confusability experiments contained herein. Miller and Niceley (1955:340) found that, with high amounts of noise, [b] and [m] were no more likely to be confused than many other pairs including [m] and [s], to give an example, and lower amounts of noise produced virtually no confusability between [m] and any other phoneme. However, with an

[1] Nonce, here, indicates that the words do not hold any semantic or pragmatic value in the listener's language.

intermediate amount of noise, a slightly larger portion of participants indicated that [b] and [m] were the same than [d] and [n]. More specifically, when the stimuli were 12db lower than the volume of the noise, of the 125 times [b] and [m] were paired together, participants indicated they were the same sound 11 times. In comparison, given the same conditions, [d] and [n] were confused only twice. However, in this same trial, [m] and [ð] were also confused 11 times, so it is likely this confusability is due to chance. These results suggest a merger between a nasal and a non-nasal voiced stop is unlikely to occur exclusively at the bilabial node using English-speaking participants, so I did not investigate this merger in my own experiments below, but plan to explore this avenue using participants of a non-English L1 moving forward.

In Experiment One, we investigate the perceptual confusability of voiced murmured stops and their modal[2] counterparts. Participants (n = 23) were young adults enrolled at the University of Kentucky, a mid-sized university located in Lexington, Kentucky, at the boundary between the Midwestern United States and the Southern United States. All participants were natively fluent in English[3]. Participants were asked to sit in a sound-attenuated booth while wearing noise-canceling over-the-ear headphones. All of the nonce words followed a consonant-vowel pattern. Each consonant used for the experiment appeared with each of five vowels. The critical sounds used in the experiment are represented by the matrix in Table 2.2 below.

Table 2.2: Experiment One Critical Sound Matrix

unaspirated stops	b	d	ɟ	g	gʷ
aspirated stops	bʰ	dʰ	ɟʰ	gʰ	gʷʰ
vowels	æ	e	i	o	u

Through the headphones, participants heard pairs of nonce words with a one second interstimulus interval[4]. As voiced aspirates are non-phonemic in English, we did not include noise as a variable in these experiments as did Miller and Nicely (1955); as we suspected, English speakers found the task sufficiently difficult without the complication of adding noise to the stimuli. The experiment also featured a series of filler pairs, which consisted of non-plosive consonant-vowel pairings, replacing stops here with fricatives to avoid extraneous variables caused by similarity to critical pairs. The consultant who recorded the nonce words used in the experiment is a young adult male and a fluent heritage speaker of Urdu, which features a set of murmured stops. The consultant was instructed to aim for Urdu consonant and vowel targets when recording the stimuli. See Table 2.3 for examples of nonce words used in the experiment as derived from the critical sound matrix in Table 2.2.

[2]Modal voicing, most often used with vowels to describe optimal vocal fold tension and airflow creating maximal vibration is also used to describe "standard" pulmonic egressive vocal register used for obstruents. For a full treatment of modal and other voicing registers, see Ladefoged and Maddieson (1996).

[3]One participant was a heritage speaker of Spanish. This participant's results were not exceptional, and so their responses were included in the data.

[4]The interstimulus interval is the time between the beginning of a stimulus, here the first word, and the beginning of the next stimulus, the second word.

Table 2.3: Examples of the Nonce Pair Creation

Consonant 1	Consonant 2	Vowel	Word Pair
/b/	/b/	/u/	/bu/ - /bu/
/ɟ/	/ɟ^{h}/	/æ/	/ɟæ/ - /ɟ^{h}æ/
/g^{w}/	/g^{wh}/	/o/	/g^{w}o/ - /g^{wh}o/

I used the OpenSesame[5] experiment creation software package to construct and execute the experiment, which enabled the logging of response accuracy and precise response time. As they listened to these pairs, using a Black Box Toolkit response box[6], the participants indicated whether the two words in each pair were homophonous or heterophonous, or "SAME" and "DIFFERENT," respectively, as they were instructed (see Figure 2.1 below)[7].

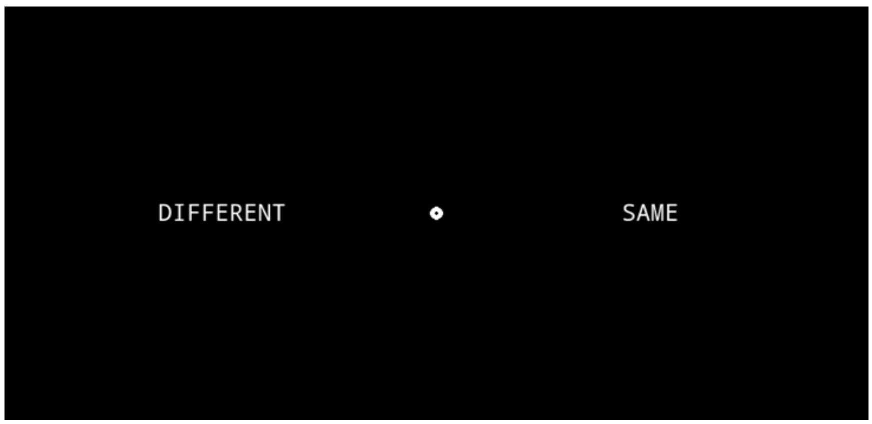

Figure 2.1: Screenshot of OpenSesame During Experiment

Given that my study investigates the perceptual confusability of /b/ and /b^{h}/ in comparison to other homorganic modal-murmur pairs, I was most interested in the accuracy of participants when given a heterophonous pair. When a participant rates two different sounds as the same, it indicates the participant had trouble distinguishing between the sounds. Pairs were played in semi-random order[8]. If neither

[5]OpenSesame is a drag-and-drop interactive experiment creation application that allows the experimenter to create, from scratch or using a template, a fully automated behavioral experiment using audio and visual stimuli on a standard computer. For more information on OpenSesame, see the software's website and official documentation at http://osdoc.cogsci.nl/.

[6]A response box is a handheld plastic box that has multiple buttons and connects to a computer, usually via USB port. The box functions in much the same way as a computer keyboard while making the buttons larger and fewer, making them the perfect tool for collecting quick responses from participants. Furthermore, the Black Box Toolkit Response Box is able to measure reaction time to the nearest millisecond, unlike most computer keyboards.

[7]Participants were asked to press the response box's rightmost button when the sounds were the same and the box's leftmost button when the sounds were different. The words "DIFFERENT" and "SAME" were left on the screen as seen here as a reminder of which button indicated which response

[8]After randomly sorting the list of pairs, I ensured that any given critical sound did not occur in two pairs in a row to prevent confounding recognition variables.

button was pressed within two seconds of the beginning of the second word in a pair, the program went on to the next pair. This measure was put in place in order to ensure that participants reacted as quickly as possible. Recording reaction time in an experiment such as this is important, as the time it takes a respondent to react may be more crucial to understanding the underlying linguistic and psychological processes than the actual response of the participant to the stimuli. In other words, if a participant takes a long time to respond, this indicates that they have trouble differentiating the two sounds.

Results

In this experiment, I investigated the relationship of place of articulation and confusability of homorganic consonants with a murmur distinction, which we will term here as "modal-murmur confusability." As such, this relationship may be represented with the following abstract linear equation, which will serve as our model[9]:

$$\text{modal-murmur confusability} \sim \text{place node} + \varepsilon^{10}$$

I chose a linear effects model to analyze the data because such models can be used to find clusters (such as place node) within the data when there are independent and dependent variables. *Mixed* linear effects models allow for both fixed and random variation. In essence, in the above equation, modal-murmur confusability is a function of place of articulation and random variability between participants. In other words, place node is an independent variable that, when changed, predicts the dependent variable, how confusable a modal-murmur pair is. Thus, we can use this mixed linear effects model[11] to determine how salient place of articulation is in predicting modal-murmur confusability. Let us now look at the results. Hypothesis One predicts that modal-murmur confusability is highest at the bilabial node, which could explain the merger of */b/ and */bh/ in PIE. Figure 2.2 below illustrates the accuracy of participants when given a homorganic modal-murmur word pair.

[9]A model such as this is a mathematical representation of the phenomenon/phenomena being studied. Models are useful to the researcher as they allow a clear path to understanding the situation via determining the values of the variables contained within the model.

[10]Epsilon here represents "error," or the unknown variable(s) that influences modal-murmur confusability that are not explicitly accounted for in our model.

[11]A mixed linear effects model is a model that uses both fixed and random variables. The fixed variables in our experiment here include the place of articulation and the perceptual confusability of the consonants. The random effects are the individual differences in participants and other unknown variables. Coincidentally, all of the random variables are represented here by ε, though ε may include fixed variables such as age, gender, etc.

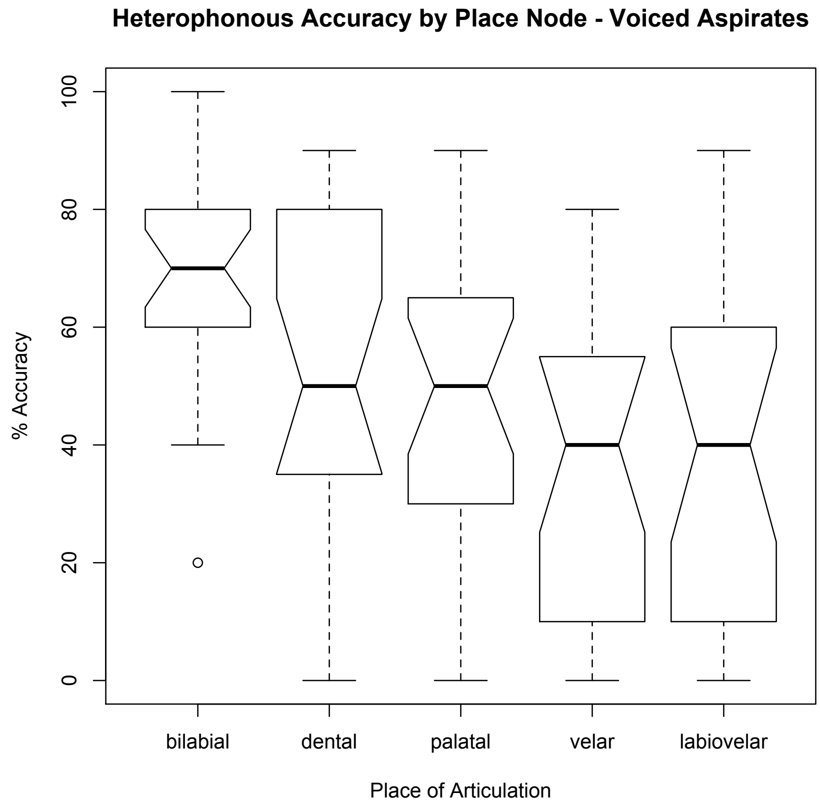

Figure 2.2: Experiment One Heterophonous Accuracy Results (n=23)

In homophonous non-critical pairs[12], participants exhibited 92.58% accuracy on average and the median was 94.12%, with all but two participants yielding at least 88% accuracy. Both of the participants who scored below this level scored 58.82% accuracy. Their responses were excluded in this analysis as this low level of accuracy suggests that something undesirable may have occurred during their trials. This same metric was used to exclude participants in all three experiments described in this chapter as the filler pairs were identical between experiments. Figure 2.2 above shows the accuracy of participants during critical heterophonous pairs by place node. When given a heterophonous bilabial pair, a /b/ and a /b^h/, in either order with a matching vowel, participant accuracy averaged higher than any other place node. The mean and median accuracy at the bilabial node were both 70%. Compare this with the next highest accuracy, at the dental node, with a mean of 50.95% and a median of 50%. Note also that the only perfect score from any of the participants in heterophonous pairs was at the bilabial node and that no participants scored 0% accuracy at this node, unlike the other four nodes. Using this data, place node explains a significant proportion of the variance in modal-murmur confusability (R^2

[12]A non-critical pair refers to a pair of sounds that does not involve the stop consonants investigated here. For this experiment, all non-critical consonants were fricatives. Non-critical pairs included such words as /fo/ and /zo/.

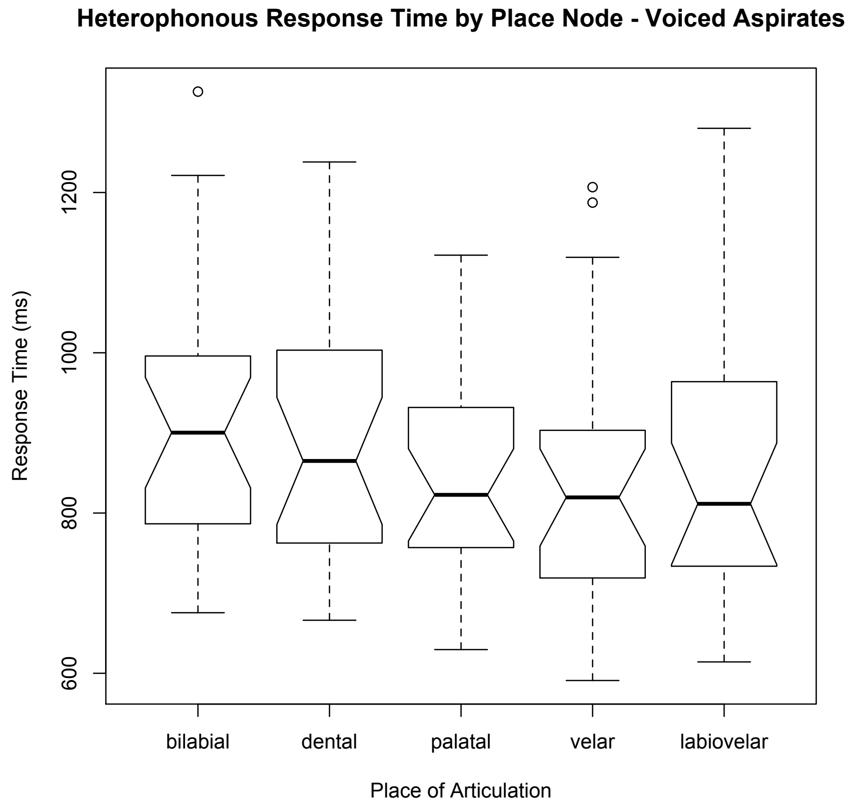

Figure 2.3: Experiment One Heterophonous Response Time Results (n = 23)

= 0.1675, $F(4, 110) = 5.535$, $p < 0.01$). However, this value for R^2 means only 16.75% of the variance in modal-murmur confusability is explained by our model. Thus, other unknown variables ε must account for the rest of the variance.

Finally, turning our attention to Figure 3, the response time at the bilabial node parallels accuracy as the highest place node, yielding a mean of 910.7 milliseconds(ms) and a median of 900.3 ms. Unlike our accuracy results, the response time averages and medians are clustered more tightly and, in our linear mixed model, do not reach statistical significance ($F(4, 100) = 0.5293$, $p = 0.7144$). However, as both accuracy of response and response time are measures of perceptual confusability, we can substitute both measures into the model as follows, using accuracy and response time as our dependent variables:

$$\text{modal-murmur accuracy} + \text{modal-murmur response time} \sim \text{place node} + \varepsilon$$

In other words, place node and unknown variables predict accuracy and response time in a modal-murmur pairing. Taking our two dependent variables into account, the model does not achieve statistical significance ($F(4, 100) = 1.036$, $p = 0.3925$). This result gives us no reason to believe place of articulation had any significant effect on response time. These results present a compelling picture. Because participants had higher accuracy and no significant variation in response time when judging heterophonous bilabial pairs than at other place nodes, we can conclude that in this experiment, the participants could more accurately perceive the difference between /b/ and /bʰ/ than /d/ and /dʰ/, /g/ and /gʰ/, etc. It is also worth stressing that the effect that place node has is much weaker than our unknown ε variables.

2.2 Experiment Two: Modal-Implosive Confusability

The results of the experiment discussed in Section 2.1 are inconsistent with a model in which a merger of PIE */b/ and */bʰ/ had occurred as the result of perceptual confusability. For this reason, we turn to the Cao Bang Theory, as discussed in Section 1.4. Unlike the GT, this view does not deny the SM, but rather assumes that the SM represents a very late stage of PIE consonantism. The CBT proposes a stop system that could have plausibly evolved into the SM, yet still provides a typologically viable system, one which we hope offers explanations for the numerous phonological peculiarities of the stop system discussed above, the rarity of */b/ included. The CBT proposes that the original PIE stop system consisted of voiceless stops, voiced implosives, and voiced stops, with the latter two having undergone a chain shift to voiced stops and voiced murmured stops, respectively.

Methods

Participants (n = 31) were young adults enrolled at the University of Kentucky. All participants were native speakers of English[13]. The design for Experiment Two featured the following critical sounds:

Table 2.4: Experiment Two Critical Sound Matrix

egressive stops	b	d	ɟ	g	gʷ
implosive stops	ɓ	ɗ	ʄ	ɠ	ɠʷ
vowels	æ	e	i	o	u

The consultant whose voice was used to record the stimuli for this experiment is a young adult male and a cultural Sindhi, a group which speaks an Indian/Pakistani language containing implosives, who grew up with a fluent Sindhi-speaking parent but is not himself fluent. I asked the consultant to aim for Sindhi targets while producing the implosives. Otherwise, the experimental design was identical to that used in the modal-murmur experiment.

Results

In Experiment Two, I investigated the relationship of place of articulation and confusability of homorganic consonants with an implosive distinction, which we will term

[13] One participant was also natively fluent in Gujarati, which features a series of murmured stops, but still contains no implosives. This participant's results were not exceptional, so they were included in our analysis.

here as "modal-implosive confusability." Using a linear mixed effects model as before, we can represent this relationship with the following abstract linear equation:

$$\text{modal-implosive confusability} \sim \text{place node} + \varepsilon$$

Once again, place node is the independent variable that, along with ε, predicts the dependent variable; in this experiment, modal-implosive confusability is the dependent variable. We will answer the question: how well does place of articulation predict modal-implosive confusability? Hypothesis Two predicts that modal-implosive confusability is highest at the bilabial node. Figure 2.4[14] below illustrates the accuracy of participants when given a homorganic modal-implosive word pair.

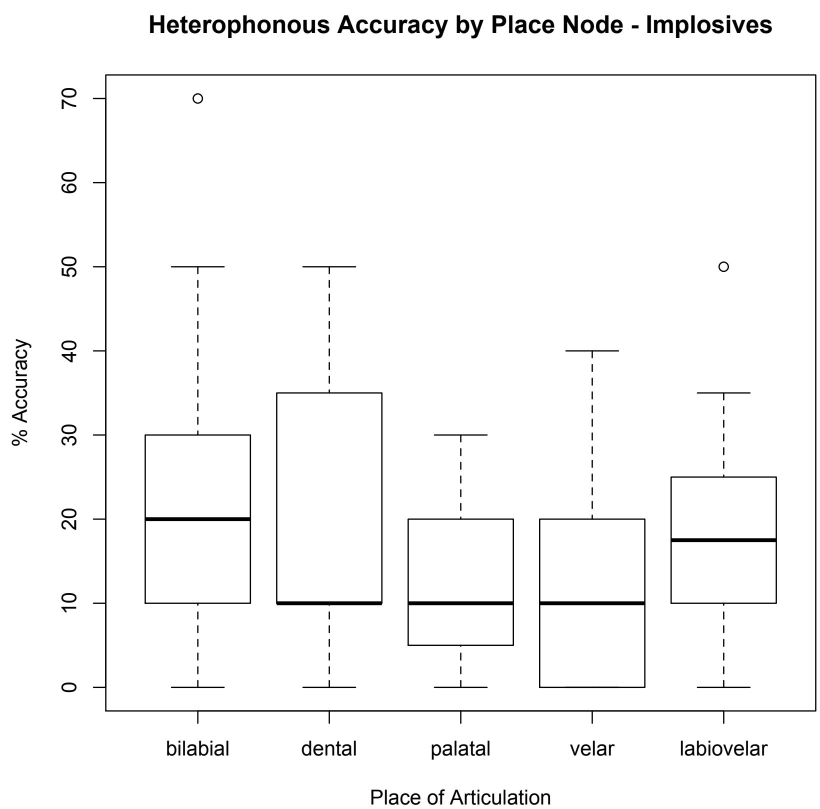

Figure 2.4: Experiment Two Heterophonous Accuracy Results (n = 31)

You will notice that, while the average accuracy of participants at the bilabial node is higher than the other nodes, 20% compared to 10%, the interquartile range[15]

[14] Note that Figure 2.2 does not include notches (the inward sloping sections cut from the other results graphs). Notches represent the confidence interval around the median and were excluded here due to the fact that both quartiles were not possible at the dental and palatal nodes.

[15] The interquartile range is a measure of the dispersion of the data and is comprised of the middle 50% of data points, represented as boxes in Figure 4.

overlaps in all cases. Using this data, place node explains a significant proportion of the variance in modal-murmur confusability ($R^2 = 0.09504$, $F(4, 150) = 3.938$, $p < 0.01$). The R^2 value is relatively low and, as we established above, translates to place node accounting for approximately 9.504% of the variation present in the data. Comparing Figure 2.4 to Figure 2.2, the accuracy data from Experiment One, it is not surprising that a weaker effect is predicted by the model in Experiment Two. Figure 2.5, however, paints a slightly different picture.

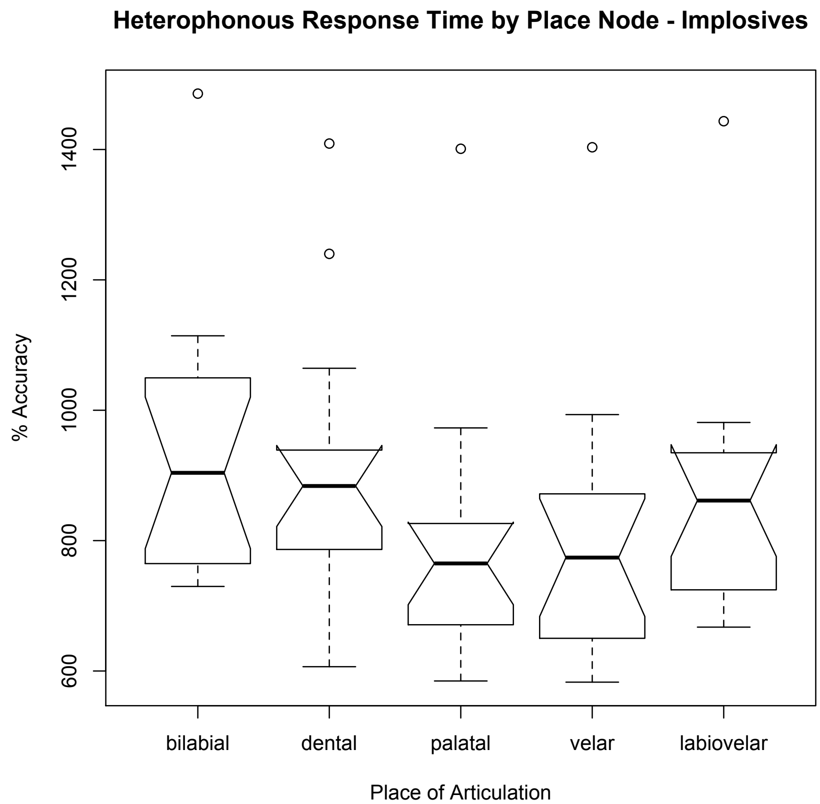

Figure 2.5: Experiment Two Heterophonous Response Time Results (n = 31)

Though the interquartile areas overlap again, the bilabial node, followed closely by the dental node, took participants longer, on average, to respond to. This could be an indication that participants were slowed down by the difficulty of discerning between /b/ and /ɓ/. Using this data, place node once again explains a significant proportion of the variance in modal-murmur confusability ($R^2 = 0.07981$, $F(4, 150) = 3.252$, $p < 0.05$). Thus only 7.981% of the variation in the response time data can be accounted for by place node. When both accuracy and reaction time are used in the model as before, once again, place node explains a significant portion of variance ($F(4, 150) = 3.525$, $p < 0.01$), and once again the R^2 value is low at 0.08592. The implications of this experiment are not quite as clear as those of the modal-murmur experiment. Not only is the trend in the data not as strong, but also, the meaning of the trend is less obvious. It seems that both bilabial and dental implosives are more closely clustered, differing on average by only 5%. In comparison, the other place nodes differ from the bilabial node by at least 10%. Even if the model were able to account for a greater share of the variation, the fact that the dental node also exhibits an increase in perceptual confusability would make the path to a bilabial merger more complicated as the same motivation would have been present at the dental node, for which we do not posit a merger.

2.3 Experiment Three: Voiced-Voiceless Confusability

As neither Experiment One or Experiment Two are able to account for the posited merger, I constructed a third and final experiment to test the confusability of unaspirated voiced stops and their unaspirated voiceless counterparts, but not to investigate a possible merger between */b/ and */p/, but rather to test the possibility of dissimilation between the two sounds due to similarity, which could have pushed /b/ to another phonetic realization.

Methods

Participants (n = 9) were young adults enrolled at the University of Kentucky. All participants were native speakers of English. The design for Experiment Three featured the following critical sounds:

Table 2.5: Experiment Three Critical Sound Matrix

voiced stops	b	d	ɟ	g	gʷ
voiceless stops	p	t	c	k	kʷ
vowels	æ	e	i	o	u

The consultant who recorded the nonce words used in the experiment is a young adult male and a fluent heritage speaker of Urdu, which features a set of voiceless unaspirated stops. The consultant was instructed to aim for Urdu consonant and vowel targets when recording the stimuli. Otherwise, the experimental design was identical to that used in the modal-murmur experiment.

Results

In Experiment Three, we investigate the relationship of place of articulation and confusability of homorganic consonants with a voicing distinction, which we will term here as "voiced-voiceless confusability." Using a linear mixed effects model as before, we can represent this relationship with the following abstract linear equation:

$$\text{voiced-voiceless confusability} \sim \text{place node} + \varepsilon$$

Once again, place node is the independent variable that, along with ε, predicts the dependent variable; in this experiment, voiced-voiceless confusability is the dependent variable. We will answer the question: how well does place of articulation predict voiced-voiceless confusability? Hypothesis Two predicts that voiced-voiceless confusability is highest at the bilabial node. Figure 2.6 below illustrates the accuracy of participants when given a homorganic modal-implosive word pair.

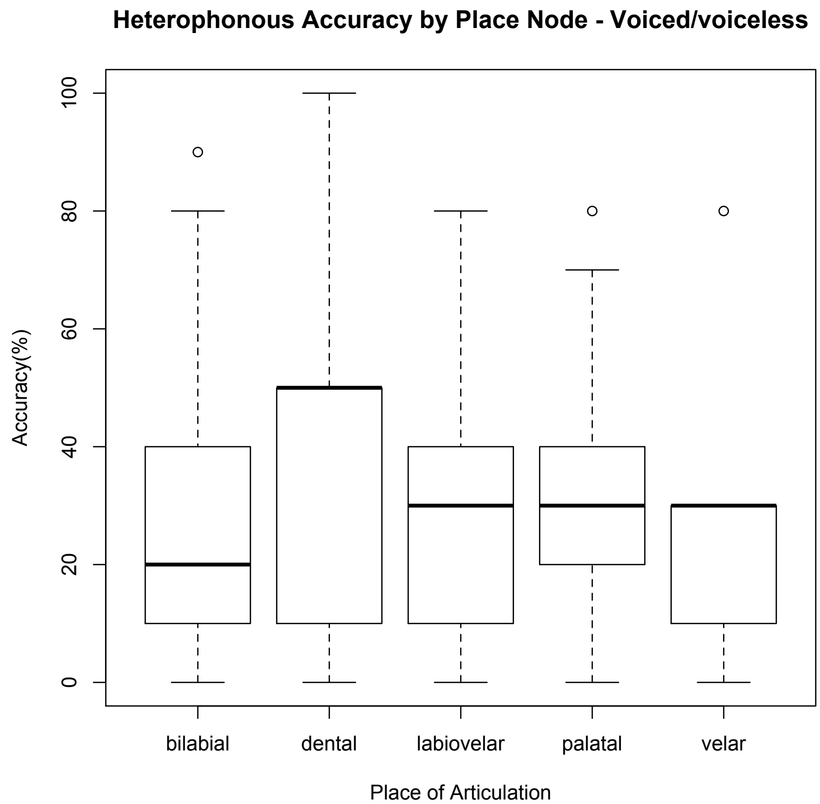

Figure 2.6: Experiment Three Heterophonous Accuracy Results (n = 9)

It is immediately clear that there is no clear trend in the accuracy data. Using this data, place node does not explain a significant proportion of the variance in voice-voiceless confusability ($R^2 = 0.04463$, $F(4, 40) = 0.4671$, $p = 0.7595$).

The data shown in Figure 2.3 demonstrates that place node explains even less of the response time data from the experiment ($R^2 = 0.001004$, $F(4, 40) = 0.01005$, $p = 0.9998$). It is clear that some other factor is affecting the voiced-voiceless confusability of the voiced-voiceless pairs. If we run a linear mixed effects model using individual participants as predictors of accuracy, however this variable seems to account for only a fraction of the variation ($R^2 = 0.1185$, $F(1, 43) = 5.781$, $p < 0.05$). Further, this significance goes away when response time is added into the mix ($R^2 = 0.07408$, $F(1, 43) = 3.44$, $p = 0.07049$).

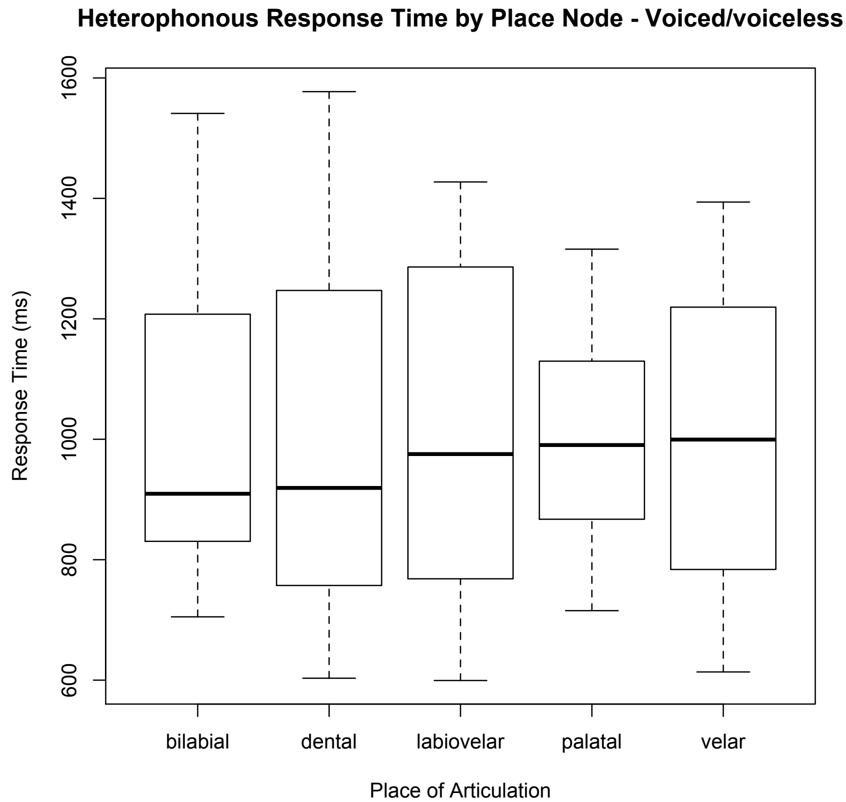

Figure 2.7: Experiment Three Heterophonous Response Time Results (n = 9)

2.4 Discussion

From the experimental data presented above, it is unlikely that perceptual confusability is a primary motivator for PIE */b/ merging with another segment. In all three experiments, place of articulation was able to account for only a small fraction of the variation in perceptual confusability if any at all. This indicates that if the perceptual confusability of modal-murmur, modal-implosive, or voice-voiceless pairs were to motivate a merger, it would be unlikely to occur at only a single node. As my hypotheses were not supported, the motivation for such a merger is still yet unknown. Thus we turn to Chapter 3 in search of further answers.

Chapter 3 DERBiPIE[1]

This chapter outlines the intended capabilities of an in-progress reference work, which I call the Database of Etymological Reconstructions Beginning in Proto-Indo-European (DERBiPIE). DERBiPIE, while currently in its developmental stages, promises to be a useful tool for Indo-Europeanists and historical linguists alike.

As mentioned in Chapter 1, Jucquois (1966) published a chart (see Figure 3) in 1966 that gave the counts for consonants in the forms of PIE as well as their occurrences in root-intial and root-final position. What is unclear is the source of his numbers. It is unclear what reference work, if any, Jucquois used to tally these numbers. One of the goals of this section is to recreate the stops portion of Jucquois' chart using DERBiPIE as a reference.

Consonne initiale de racine		Consonne finale de racine																								Totaux
		p	t	k	k̂	k^w	b	d	g	ĝ	g^w	bh	dh	gh	ĝh	gh^w	l	m	n	r	y	w	s	ə$_1$	ə$_2$	ə$_3$	ə?	
A	p	0	4	0	2	1	0	2	1	0	0	0	0	0	0	0	35	0	8	33	19	14	5	5	11	2	1	143
	t	3	0	3	1	1	0	0	4	0	0	0	0	0	0	0	9	4	10	35	7	18	0	2	9	0	0	106
	k	2	1	2	2	0	1	1	1	1	0	0	1(?)	1	0	0	30	6	15	43	8	19	6	2	48	0	0	190
	k̂	1(?)	0	1	0	2	0	0	0	0	0	0	1(?)	0	0	0	14	5	6	19	10	18	3	5	10	0	0	95
	k^w	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	3	1	2	5	1	0	1	0	3	0	0	18
B	b	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	4	0	1	3	1	0	0	1	13	0	0	25
	d	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	6	3	4	12	6	10	1	2	5	3	0	53
	g	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	1	4	14	5	7	0	1	7	0	0	50
	ĝ	0	0	0	0	0	0	0	0	0	0	2	0	1	1(?)	0	3	2	4	2	3	3	0	0	1	0	0	22
	g^w	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	6	0	1	11	9	2	1	2	2	0	0	37
C	bh	0	0	1	0	0	0	0	1	0	1	0	2	1	1	0	33	0	4	42	8	10	3	3	18	1	0	129
	dh	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	8	2	5	21	1	13	0	7	5	2	0	67
	gh	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	11	1	0	19	4	6	2	0	5	0	0	52
	ĝh	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	4	1	1	10	2	9	3	2	10	1	0	47
	gh^w	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	1	1	2	3	1	0	0	1	0	0	0	12
D	l	3	2	2	0	0	0	0	2	2	0	0	0	1	0	1	0	2	7	3	22	18	1	17	20	0	0	103
	m	0	0	1	1	0	0	2	0	1	0	0	2	1	0	0	18	1(?)	15	21	15	11	4	10	24	5	3	135
	n	1	0	0	1	1	0	3	0	0	2	1	1	0	1	0	0	2	0	6	9	8	3	5	9	1	0	54
	r	1	1	3	3	0	0	0	2	2	1	2	0	0	0	0	0	2	2	0	13	11	1	7	3	0	0	54
	y	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	2	1	1	1	1	12	1	6	7	0	0	36
	w	4	1	2	1	1	1	2	1	1	1	2	3	0	1	0	14	1	6	31	20	0	11	1	17	0	4	126
	s	18	29	13	2	2	0	1	1	0	0	0	0	0	1	0	16	15	17	17	6	46	0	8	13	1	0	206
E	ə$_1$	4	6	0	1	0	0	1	2	2	0	0	1	0	3	0	13	2	14	21	8	8	5	0	0	0	0	91
	ə$_2$	7	4	0	3	2	3	4	3	2	0	2	0	3	2	1	16	6	19	17	22	20	2	0	0	0	0	138
	ə$_3$	2	0	1	2	1	0	3	1	0	0	1	0	0	0	0	1	3	6	5	1	5	4	0	0	0	0	36
Totaux		46	52	30	22	11	6	28	20	11	5	13	15	9	10	3	258	62	154	394	202	268	57	87	240	16	8	2.027

N.B. : Dans ce tableau ne figurent ni les onomatopées, ni les mots du langage enfantin, ni les exclamations, ni les particules adverbiales ou

Figure 3.1: Jucquois' Chart of PIE Root Structure Counts (Jucquois 1966)

3.1 What is DERBiPIE?

DERBiPIE, like the other work in this thesis, straddles a boundary between two scholarly worlds. In this case, the two worlds are computational linguistics and com-

[1]**Special thanks for this chapter go first and foremost to Mark Richard Lauersdorf, who can turn a yes/no question into a full-scale epistemological debate that leaves you feeling like your idea is the best in the world, and whose conversation led to the creation of the database discussed herein.**

parative philology. Like its material and spiritual predecessor, Julius Pokorny's *Indogermanisches etymologisches Wörterbuch* (1959), DERBiPIE is intended to represent the complete lexicon of Proto-Indo-European. The ultimate goal of DERBiPIE is to be an easily accessible and searchable digital collection of the PIE language. While it is currently a set of searchable XML files, my hope is that DERBiPIE will soon evolve into a user-friendly database complete with a graphical user interface and a set of versatile search functions.

3.2 Source Material

In 1959, Julius Pokorny published what he believed to be a complete and accurate collection of the reconstructed forms of PIE. While many modern scholars are weary of the forms presented in the *IEW* due in part to his non-acceptance of laryngeal theory, the volume's completeness results in its tenacity as a commonly used reference work. In fact, Millar's revised version of *Trask's Historical Linguistics* (2007) lists the *IEW* as the go-to reference for Proto-Indo-European (p. x)[2]. Its completeness is also the reason I have selected it as a starting point for DERBiPIE. In its current form, the database still reflects its roots in Pokorny, and should thus be thought of as a "sandbox proof-of-concept"[3] rather than a fully matured source of linguistic information.

Creation of the Database

To begin work on DERBiPIE, I needed to convert the *IEW* into digital form. There were a few key pieces of information I knew I wanted to retain in my database. I needed each form as Pokorny reconstructed it, as well as the philological information included in each entry, namely the languages in which each form was attested. This information's usefulness will become clearer in Section 3.3 below. Luckily, the *IEW* exists in digital form hosted by the website indo-european.info. More specifically, the digital version can be found at the following web address:

`https://indo-european.info/pokorny-etymological-dictionary/contents.htm`

A quick glance at this page shows that it is a series of links, each to a different entry from Pokorny. Clicking any of these links will take you to the full text entry for the form, taken verbatim from the original Pokorny with the minor addition of an English gloss provided alongside the original German.

Since the website was written in basic HTML markup, I was able to easily write a Python program that scraped all of the desired information from this site, converting it directly into a useable XML database. The program followed the following process:

1. Search the contents page for the next entry link based on HTML markup and navigate to this page.

[2]Note that this is not a typographical error, the page on which this statement is made is within the front matter of the book.

[3]Phrasing à la Mark Richard Lauersdorf.

2. Locate the gloss and store this information.
3. Search the page for all mentions of daughter languages using a simple RegEx[4]; store which languages have an attested descendant of the form.
4. Locate and store the actual form itself using HTML markup and RegEx.
5. Finally, print all of this information in standard XML markup.

The final print step brings everything together and produces a single line of the database and numbers it, which, after manual modifications, looks something like this:[5]

```
<F ENTRY="75_3" ABG="N" ABR="N" ACORN="N" ACYMR="N" ACHECH="N" ABRIT="
    ↪ N" AFRIES="N" AFRZ="N" AGS="N" AHD="Y" AI="Y" AIR="N" AISL="Y"
    ↪ AKSL="N" ALAT="N" ALB="Y" AN="N" AOL="N" APERS="N" APREUSS="N"
    ↪ ARM="N" ASA="N" ASCHWED="N" ATT="N" AV="N" BRET="N" BSL="N"
    ↪ BULG="N" CECH="N" DAN="N" DOR="N" DRAVID="N" EL="N" ENGL="N"
    ↪ GAL="N" GALL="N" AOL="N" GEG="N" GERMAN="N" GOT="N" GR="Y" HITT
    ↪ ="N" HOLL="N" HOM="N" IDG="N" ILLYR="N" IR="N" ITAL="N" JAV="N"
    ↪  KELT="N" KELTIBER="N" LAT="Y" LETT="N" LIT="Y" LYK="N" MAK="N"
    ↪  MBRET="N" MENGL="N" MESSAP="N" MHD="Y" MIND="N" MIR="N" MLT="N
    ↪ " MND="N" MNDL="N" MPERS="N" NHD="Y" NIR="N" NORDILL="N" NORW="
    ↪ N" NPERS="N" OSK="Y" OSSET="N" PAL="N" PAMIR="N" PEHL="N" PHRYG
    ↪ ="N" POLN="N" RAET="N" RHOD="N" ROMAN="Y" RUSS="N" SAK="N" SBKR
    ↪ ="N" SCHWED="N" SEM="N" SERB="N" SLOV="N" SPATLAT="N" SUM="N"
    ↪ THRAK="N" TOCH="Y" TOSK="N" UMBR="N" URIND="N" URNORD="N" URSPR
    ↪ ="N" VED="N" VEN="N" VENILL="N" VGL="N" VLT="N" WRUSS="N" ZEM="
    ↪ N" GLOSS="mother" FA="Y" SA="Y">amï</F>
```

There are a few things about the code above that I would like to point out. First, you will notice that corner brackets encompass everything except the form *amï*. This information is referred to as metadata, or data regarding your data. The metadata begins with *F*, which indicates the data is a *form*. The *F*, along with each of the items color-coded in blue above, is an *attribute*. The information given in prime quotes in purple immediately following each attribute is called a *value*. Further, you will notice that the first part of the metadata ends with closed corner bracket immediately before the form, given in black. This form is the actual data and is followed in every instance by </*F*>. This lets programs analyzing the database know that that whatever follows is part of a different data point. Effectively, it ends the data point.

[4] RegEx refers to the *Reg*ular *Ex*pressions used by most programming languages that are used to define search patterns. To illustrate the abilities of RegEx, consider the following *string* of regular expressions: '\bwork\n'. This string will match the word 'work' when it is preceded by a word boundary(e.g. a space, punctuation, or beginning of a line) and followed immediately by a new line.

[5] Note that while presented here in multiple lines for space purposes, this would all be on the same line in the actual database file.

Drawing attention now to more specific features of this line of XML, you will notice that the ENTRY attribute has the value "75_3". This indicates that this form is the third form listed in the heading of the 75th entry of the *IEW*. This distinction was made using a text editor and RegEx find-and-replace strings after the database had been scraped from the web. Another artifact of post facto changes seen here is the character <ï>. Most scholars who have studied PIE will notice that this is not a typical character found in any mainstream orthography for the language. I intentionally inserted this character as a replacement for <ī̆> as it occupies more than one Unicode space. In other words, <ī̆> is actually two or three characters depending on the way it is typewritten. In order to make searching the database easier and more accurate, I replaced all instances wherein a segment was represented by more than one Unicode character with a single character.

3.3 The Current State of DERBiPIE

Querying the Database

To echo that which was stated above, any information derived from the database in its current form should be thought of as preliminary and a mere proof-of-concept. Nonetheless, we can begin to gain some insight into the phonology and phonotactics of PIE through careful, directed analysis of the database. Perhaps we may even gain some insight into the paucity of */b/. As this is the goal of this thesis, I will be looking specifically at the plosives of PIE.

Pokorny included a fourth manner of articulation, voiceless aspirates, as a way of adjusting for the lack of laryngeals. However, these voiceless aspirates occur so rarely[6] in the database that I have disregarded them in the charts and analysis in this section, focusing only on the stops present in the Traditional Model.

With disclaimers out of the way, we may begin to pick apart DERBiPIE. The database currently contains 3,801 forms, each tagged for languages in which its descendants are attested, as well as gloss. Using Python and RegEx, I constructed a series of short programs that searched the database for stops in various contexts, checking the counts manually using an advanced text editor.

[6] Total occurences – */$\mathrm{p^h}$/: 15; */$\mathrm{t^h}$/: 17; */$\mathrm{\acute{k}^h}$/: 0; */$\mathrm{k^h}$/: 9; */$\mathrm{k^{wh}}$/: 0

Full Database Results

Table 3.1: Frequency and Manner Ratio of PIE Stops for Full Database

	p		t		ḱ		k		k^w	
Total	496	23.6%	523	24.9%	277	13.2%	730	34.8%	73	3.5%
Form-initial	246	24.9%	192	19.4%	160	16.2%	351	35.5%	39	3.9%
Form-final	35	18.6%	57	30.3%	22	11.7%	64	34.0%	10	5.3%
Intervocalic	19	21.3%	47	52.8%	5	5.6%	15	16.9%	3	3.4%
	b		d		ǵ		g		g^w	
Total	146	12.8%	343	30.0%	123	10.8%	434	38.0%	96	8.4%
Form-initial	46	10.4%	126	28.4%	41	9.3%	168	37.9%	62	14.0%
Form-final	19	10.9%	58	33.1%	24	13.7%	59	33.7%	15	8.6%
Intervocalic	8	19.5%	16	39.0%	5	12.2%	6	14.6%	6	14.6%
	b^h		d^h		$ǵ^h$		g^h		g^{wh}	
Total	310	34.0%	237	26.0%	152	16.6%	181	19.8%	33	3.6%
Form-initial	204	40.3%	115	22.7%	82	16.2%	88	17.4%	17	3.4%
Form-final	23	18.4%	34	27.2%	26	20.8%	30	24.0%	12	9.6%
Intervocalic	9	30.0%	4	13.3%	10	33.3%	5	16.7%	2	6.7%
	w		y		m		n			
Total	518	82.2%	112	17.8%	427	44.2%	538	55.8%		
Form-initial	247	81.8%	55	18.2%	228	71.9%	89	28.1%		
Form-final	6	66.7%	3	33.3%	28	35.9%	50	64.1%		
Intervocalic	39	75%	13	25%	17	21.8%	61	78.2%		

Table 3.1 shows the distribution of the stops in PIE as found in the *IEW*, including what I have termed the manner ratio, or the percentage that each sound occupies within its manner of articulation series. You will notice that */b/ appears a total of 146 (12.8% of voiced unaspirated stops) times in the database. This makes */b/ only the third least common among the voiced stops. These numbers allow us to more accurately pin down the nature of the scarcity of */b/. Notice that while within the voiced stops, */ǵ/ and */g^w/ are also rare, so are each of their voiceless and murmured counterparts. Meanwhile, the other labials, */p/ and */b^h/, are found in high distribution within their manner series. Notice also that */w/ is by far more common than the palatal glide */y/, which may suggest a possible merger between */b/ and */w/. However, a common PIE suffix *-ye/o- would have made the segment much more common in actual speech.

Table 3.2: A Partial Recreation of Jucquois 1966 Using Data from DERBiPIE

		Root-final Position															
		p	k	ḱ	k	k^w	b	d	ǵ	g	g^w	b^h	d^h	$ǵ^h$	g^h	g^{wh}	Total
Root-initial Position	p	3	6	4	7	0	0	3	0	0	0	0	0	0	0	0	23
	k	7	4	1	6	4	2	0	0	0	2	1	0	0	1	0	28
	ḱ	8	1	3	4	3	0	9	0	0	0	0	0	0	0	0	28
	k	5	9	13	7	0	1	1	0	0	0	4	0	0	0	0	40
	k^w	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	4
	b	0	0	1	1	0	2	0	1	0	0	0	1	0	0	0	6
	d	5	2	3	4	2	2	0	1	1	0	2	0	1	3	1	27
	ǵ	3	0	0	2	0	0	0	0	1	0	2	0	0	0	0	8
	g	5	10	0	1	0	0	1	0	2	0	9	0	0	0	0	28
	g^w	0	1	0	0	0	0	0	0	0	0	2	3	0	0	0	6
	b^h	0	0	0	0	0	0	0	2	0	3	4	6	0	4	0	19
	d^h	0	0	1	1	0	0	0	0	0	3	8	0	0	4	1	18
	$ǵ^h$	0	0	0	0	0	0	1	0	0	0	1	1	1	3	0	7
	g^h	1	0	0	3	0	0	1	2	0	0	8	2	0	4	0	21
	g^{wh}	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2
	Total	38	34	27	36	9	7	16	6	4	8	42	14	3	19	2	265

Turning our attention to Table 3.2, a redux of the stops section of Jucquois' chart, at first glance, we once again see that */b/ does not seem unusually sparse. While it occurs less than */d/ and */g/ in both the root-initial and root-final position, it is on par with */ǵ/ and */g^w/. However, tabulation of each chart reveals a striking similarity. In Jucquois' count, */b/ accounts for approximately 13.4% (25/187) of voiced stops while */b^h/ accounts for a whopping 42.0% (129/307) of all murmured stops. In DERBiPIE, */b/ accounts for 11.2%(13/116) of voiced stops contained in roots, while */b^h/ accounts for 41.5% (61/147) of murmured stops contained within roots. The percentages here suggest that Jucquois methods were not far off from my own in result, regardless of methodology.

Partial Database Results

To further demonstrate the capabilities of a tool like DERBiPIE and further investigate the paucity of */b/, I created a sub-database using only the forms whose descendant forms are attested in all three, Latin, Greek, and Sanskrit. These three languages were selected as they are the earliest well-attested[7] descendants of PIE with the exception of Hittite (Pollock 2003; Olivier 1986), whose forms Pokorny unfortunately largely overlooked. Furthermore, these three languages, coming from entirely separate branches of Indo-European, give us a diverse look at their ancestral lan-

[7] A few Italic languages, Indo-Iranian languages, Celtic languages, and Phrygian are attested before one or more of these three, but none with a substantial corpus as Latin, Greek, and Sanskrit are. For a more comprehensive discussion, see Fortson's textbook on Indo-European language and culture (Fortson 2010).

guage. In short, I selected these three languages to build a sub-database because theoretically, these forms are most likely to have actually been present in the proto-language.

Table 3.3: Frequency and Manner Ratio of PIE Stops for Partial Database

	p		t		ḱ		k		k^w	
Total	162	28.1%	159	27.6%	87	15.1%	142	24.7%	26	4.5%
Form-initial	87	31.4%	61	22.0%	42	15.2%	75	27.1%	12	4.3%
Form-final	6	17.1%	12	34.3%	8	22.9%	3	8.6%	6	17.1%
Intervocalic	7	25.9%	16	59.3%	2	7.4%	2	7.4%	0	0.0%
	b		d		ǵ		g		g^w	
Total	29	10.0%	97	33.3%	39	13.4%	98	33.7%	28	9.6%
Form-initial	12	7.5%	61	38.1%	18	11.3%	49	30.6%	20	12.5%
Form-final	2	11.1%	7	38.9%	2	11.1%	6	33.3%	1	5.6%
Intervocalic	3	37.5%	3	37.5%	1	12.5%	0	0.0%	1	12.5%
	b^h		d^h		$ǵ^h$		g^h		g^{wh}	
Total	62	34.8%	50	28.1%	43	24.2%	15	8.4%	8	4.5%
Form-initial	43	39.4%	27	24.8%	30	27.5%	6	5.5%	3	2.8%
Form-final	1	5.9%	4	23.5%	6	35.3%	1	5.9%	5	29.4%
Intervocalic	0	0.0%	1	100.0%	0	0.0%	0	0.0%	0	0.0%
	w		y		m		n			
Total	176	83.0%	36	17.0%	143	45.4%	172	54.6%		
Form-initial	68	85.0%	12	15.0%	79	69.9%	34	30.1%		
Form-final	3	100.0%	0	0.0%	7	35.0%	13	65.0%		
Intervocalic	19	90.5%	2	9.5%	4	12.1%	29	87.9%		

Changing perspectives with the partial corpus paints a very different picture in the story of the paucity of PIE */b/. In the forms attested in Latin, Greek, and Sanskrit, */b/ is the second least common of the voiced stops, making up only 10% of the series, greatly outnumbered by */d/ and */g/. Meanwhile, the other bilabials are the most frequent among their manner of articulation. I did not construct a chart of roots in resemblance of Jucquois as the counts for this partial database are too small to draw any conclusion, but nonetheless, between the root counts for the full database and the total counts for the partial database, a scarcity of */b/ as noted in extant literature is evident.

3.4 The Future of DERBiPIE

As I have stated, DERBiPIE will be undergoing major changes in the near future of the project. This section outlines several of the intended changes and features. In an ideal situation, DERBiPIE will be accessible by a diverse group of Indo-Europeanists, linguists, and scholars in general. In order to make it as accessible as possible, I will create a graphic user interface (GUI) that will automate all of the search functions and other analysis tools, including a visualization module. As for updating the data itself, since Pokorny's *IEW* is regarded by the field as inaccurate, the first step will be to update the forms to reflect a modern conception of the language. With Andrew Miles Byrd's help, I will do so, using Byrd's working reconstruction of the language. However, unlike how Pokorny believed his own reconstruction to be the best available, only to have time and research wear away at its viability due to the stagnant nature of print, I am not naïve enough to think that the same will never happen to any given modern reconstruction of the language. Thus, DERBiPIE, in its ultimate stages, will be a living database that is constantly updated by researchers spanning the discipline. Not only will it represent the most modernly informed reconstruction at all times, but also it will be capable of representing multiple, competing reconstructions at once. To accomplish this, there will be drop-down menus in the GUI that allow you to select and compare counts based on different reconstructions. Thus, as the current iteration of DERBiPIE stores only one data point per form, it will one day store several data points per form. Ultimately, DERBiPIE has the potential to be the next *IEW*, serving as a catch-all reference for PIE.

Chapter 4 Conclusion

In my introduction, I promised to investigate the paucity of PIE */b/. While I have accomplished this goal, I have not actually answered any questions. Instead, I have only ruled out possible answers, which is nonetheless valuable for progress toward an answer. Through laboratory phonology and computational database construction and parsing, I have ensured a thorough treatment of the issue.

DERBiPIE in its current form is unable to suggest a different picture of the problem than what has been suggested. This is good, because it means that the field hasn't been throwing time and energy into a bottomless pit of a non-issue. The database suggests that there is indeed an unusual paucity of */b/ in PIE when compared to the other labials.

The experiments in Chapter 2 do not definitively show that a merger between PIE */b/ and some other segment was or was not possible. They do however show that if such a merger occurred, it is unlikely to have been motivated solely by perception. In other words, if PIE */b/ merged with another labial consonant, it was not solely because listeners misheard it. Another goal of this thesis was to determine whether it is possible to explain the paucity of */b/ using the Standard Model of PIE reconstruction or if a new model of PIE consonantism is required. While this was not conclusively answered, I hope to conduct further research that will continue to seek an answer to this question.

One future experiment that I would like to conduct in search of this answer is that of a production experiment. I would like to recruit informants of various first languages, all of which will contain one of the stop distinctions focused on in the experiments herein, and elicit their consonants. I hope to use acoustic and articulatory measures to determine if there is anything about the labial node that might make it particularly susceptible to a merger. Combining this data with the data I have already gathered (and more like it), hopefully I will be able to reach a conclusion.

While it is clear that there is much more ink to be spilled on the topic of PIE */b/, this thesis has accomplished a number of things. First and foremost, the experiments presented in Chapter 2 show that it is unlikely that perception could have been a sole motivator in a merger between PIE */b/ and another segment. Another future direction of my research will be to extrapolate this to sound change in general, investigating whether sound change is always multifaceted or if many factors are required to make sound change possible. Furthermore, this thesis has demonstrated the usefulness of computational methods in the investigation of large-scale data based projects involving language. I hope to continue this path of research as I continue my academic career, expanding my horizons to larger and more general questions of historical change, all-the-while continuing investigation in my first true academic love, Indo-European linguistics.

Bibliography

Abramson, Arthur S. "Laryngeal timing in consonant distinctions." *Phonetica* 34 (1977): 295–303.

Bailey, Todd, and Ulrike Hahn. "Phoneme Similarity and Confusability." *Journal of Memory and Language* 52 (April 2005): 339–362.

Barnett, Phillip, and Andrew Miles Byrd. "A Markedly Different Approach: An Experimental Look at the Rarity of PIE */b/." *Proceedings of the 29th Annual UCLA Indo European Conference*, under consideration.

Barrack, Charles M. "The Glottalic Theory Revisited: A Negative Appraisal." *Indogermanische Forschungen* 107 (2002): 76–95.

———. "The Glottalic Theory Revisited: A Negative Appraisal. Part II: The Typological Fallacy Underlying the Glottalic Theory." *Indogermanische Forschungen* 108 (2003): 1–16.

———. "The Labial Gap in Proto-Indo-European." *Interdisciplinary journal for Germanic linguistics and semiotic analysis* 108 (2018): 231–41.

Beddor, Patrice Speeter. "A Coarticulatory Path to Sound Change." *Language* 85, no. 4 (2009): 785–821.

Byrd, Andrew Miles. *The Indo-European Syllable.* Leiden: Brill, 2015.

———. "The Phonology of Proto-Indo-European." In *Comparative Indo-European Linguistics. An International Handbook of Language Comparison and the Reconstruction of Indo-European*, edited by Brian Joseph Jared S. Klein and Matthias Fritz. Mouton de Gruyter, 2018.

———. "The rules of reconstruction: Making our etymologies more grounded." *Etymology and the European Lexicon, Proceedings of the 14th Fachtagung of the Indogermanische Gesellschaft* 14 (2017): 17–22.

Campbell, Lyle. *Historical Linguistics: An Introduction.* Second. Cambridge, MA: The MIT Press, 1998.

Cooper, Adam. "Similarity Avoidance in the PIE Root." *Proceedings of the 32nd Annual Penn Linguistics Colloquium* 15, no. 1 (2009): 55–64.

Crowley, Terry, and Claire Bowern. *An Introduction to Historical Linguistics.* Oxford: Oxford University Press, 2010.

Drager, Katie. "Sociophonetic Variation in Speech Perception." *Language and Linguistic Compass* 4, no. 7 (2010): 473–480.

Dresher, Elan B. "The contrastive hierarchy in phonology." *Phonology* 27, no. 3 (2009): 532–536. doi:`10.1017/S0952675710000278`.

Elbourne, Paul. "Proto-Indo-European Voiceless Aspirates." *Historische Sprachforschung* 111, no. 1 (1998): 1–30.

Fortson, Benjamin W., IV. *Indo-European Language and Culture: an Introduction.* Wiley-Blackwell, 2010.

Gamkrelidze, Tamas. "On the Reconstruction of the Proto-Indo-European Stops." In *Typological Relationship and Time: A Collection of Papers on Language Change and Relationship by Soviet Linguists*, edited by Vitalij V. Shevoroshkin and T. L. Markey. Ann Arbor: Karoma, 1986.

Gamkrelidze, Tamas, and Vyacheslav Vsevolodovich Ivanov. "Sprachtypologie und die Rekonstruktion der gemeinindogermanischen Verschlüsse." *Phonetica* 27, nos. 150-56 (1973).

Haider, Hubert. "Der Fehlschluß der Typologie." In *Philologie und Sprachwissenschaft: Akten der 10. Österreichischen Linguisten-Tagung Innsbruck, 23. - 26. Oktober 1982*, edited by Wolfgang Meid, 79–92. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 1983.

Hopper, Paul J. "Glottalized and Murmured Occlusives in Indo-European." *Glossa* 7, no. 2 (1973): 141–66.

Iverson, Gregory K, and Joseph Salmons. "The Phonology of the Proto-Indo-European Root Structure Constraints." *Lingua* 87 (1992): 293–320.

Jakobson, Roman. "Typological studies and their contribution to historical comparative linguistics." In *Proceedings of the Eighth International Congress of Linguistics*, edited by Eva Siversten, 17–25. Oslo: University Press, 1957.

Joseph, Brian D. "Proto-Indo-European Consonantism: Methodological and Further Typological Concerns." Edited by Jacek Fisiak. *Papers from the 6th International Conference on Historical Linguistics*, 1985, 313–321.

Jucquois, Guy. *La structure des racines en indo-européen envisagée d'un point de vue statistique*. Universa Wetteren, 1966.

König, Christa, and Bernd Heine. *A concise dictionary of Northwestern !Xun*. Rüdiger Köppe, 2008.

Kümmel, Martin Joachim. "The Role of Typology in Historical Phonology." In *The Oxford Handbook of Historical Phonology*, edited by Patrick Honeybone and Joseph C. Salmons, 121–32. Oxford: Oxford University Press, 2015.

———. "Typology and Reconstruction: The Consonants and Vowels of Proto-Indo-European." In *The Sound of Indo-European: Phonetics, Phonemics and Morphophonemics*, edited by Benedicte Nielsen Whitehead, Thomas Olander, Birgit Anette Olsen, and Jens Elmegård Rasmussen, 291–329. Copenhagen: Museum Tusculanum, 2012.

Ladefoged, Peter, and Keith Johnson. *A Course in Phonetics*. 6th ed. Wadsworth, Cengage Learning, 2011.

Ladefoged, Peter, and Ian Maddieson. *Sounds of the World's Languages*. Oxford: Blackwell, 1996.

Lisker, Leigh, and Arthur S. Abramson. "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements." *Word* 20 (3 1964): 384–422.

McGurk, H., and J. MacDonald. "Hearing lips and seeing voices." *Nature* 264, no. 5588 (1976): 746–8.

Millar, Robert McColl. *Trask's historical linguistics*. 2nd. New York: Oxford University Press, 2007.

Miller, Brett. "Feature Patterns: Their Sources and Status in Grammar and Reconstruction." PhD diss., University of Cambridge, 2012.

Miller, George A., and Patricia E. Nicely. "An analysis of perceptual confusions among some english consonants." *Journal of the Acoustical Society of America* 27 (1955): 338–352.

Ohala, John J. "The Listener as a Source of Sound Change," edited by Carrie S. Masek, Roberta A. Hendrick, and Mary Frances Miller, 178–203. Chicago Linguistics Society, 1981.

Ohala, John J., and Brian W. Eukel. "Explanations for the instrinsic pitch of vowels." In *In honor of Ilse Lehiste*, edited by R. Channon and L. Shockey, 207–15. Foris: Dordrecht, 1987.

Olivier, J.P. "Cretan Writing in the Second Millineum B.C." *World Archaeology* 17, no. 3 (1986): 377–89.

Perey, Alan J. "Discrimination of Voice Onset Time by Human Infants: New Findings and Implications for the Effects of Early Experience." *Child Development* 52 (4 1981): 1135–1145.

Pokorny, Julius. *Indogermanisches etymologisches Wörterbuch*. Vol. 1. Berne: Francke, 1959.

Pollock, Sheldon. *The Language of the Gods in the World of Men: Sanskrit, Culture, and Power in Premodern India*. University of California Press, 2003.

Principles of Historical Linguistics. Berlin: Mouton de Gruyter, 1986.

Pulju, Timothy James. "Problems in the Reconstruction of Indo-European Stop Consonants." PhD diss., Rice University, 1995.

Russell, Paul. *An Introduction to Celtic Languages*. London and New York: Longman, 1995.

Salmons, Joe. "The larynx and Indo-European obstruent phonology and phonetics." *Historische Sprachforschung / Historical Linguistics*, Forthcoming.

Salmons, Joseph C. *The Glottalic Theory: Survey and Synthesis*. Journal of Indo-European Studies monograph series 10. Washington, D.C.: Institute for the Study of Man, 1993.

Sarostin, George. "From modern Khoisan languages to Proto-Khoisan: The value of intermediate reconstructions." *Mother Tongue: Journal of the Association for the Study of Language in Prehistory 11*, 2006, 337–470.

Shimizu, Katsumasa. "A cross-language study of voicing contrasts of stops." *Studia Phonologica* 23 (1989): 1–12.

Traill, Anthony. *A !Xoo Dictionary*. First. Rudiger Koppe, 1994.

Vine, Brent. "Review of Gamkrelidze & Ivanov, Indoevropejskij jazyk i indoevropejcy (1984)." *Language* 64, no. 2 (1988): 396–402.

Weiss, Michael. *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor: Beech Stave, 2009.

———. "The Cao Bang theory." *28th East Coast Indo-European Conference* (Reykjavik, Iceland), 13 June 2009.

Wodtko, Dagmar S., Britta Irslinger, and Carolin Schneider. *NIL: Nomina im Indogermanischen Lexikon*. Heidelberg: Winter, 2008.

Yu, Alan C. L. "The Role of Experimental Investigation in Understanding Sound Change." In *The Oxford Handbook of Historical Phonology*, edited by Patrick Honeybone and Joseph C. Salmons, 410–30. Oxford: Oxford University Press, 2015.

Vita

Bachelor of Arts; Theatre, University of Kentucky
Bachelor of Arts; Communication, University of Kentucky
Master of Arts (expected); Linguistic Theory and Typology, University of Kentucky

Phillip Barnett