This page intentionally left blank

Properties of the Anterior and Posterior Click Closures in N|uu

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophischen Fakultät
der Universität zu Köln
im Fach Allgemeine Sprachwissenschaft

vorgelegt von

Mats Exter

aus Ebersberg

Köln, 19. November 2008

Erster Referent: Prof. Dr. Hans-Jürgen Sasse
Zweite Referentin: Prof. Dr. Martine Grice

Datum der letzten Prüfung: 30. Januar 2009

Contents

Acknowledgments

There are many people who have contributed, directly or indirectly, to this work. First and foremost, I am deeply indebted to my N|uu consultants, Ouma Katrina Esau, Ouma Anna Kassie, Ouma Hanna Koper, and Ouma Griet Seekoei, for inviting me into their homes and sharing their knowledge of this beautiful and complex language with me.

Next, I would like to thank Nigel Crawhall for first introducing me to the N|uu-speaking community. I am also very grateful to the ǂKhomani San Communal Property Association for giving me permission to work on the language, and to the South African San Institute (SASI) for organizational support. Especially Grace Humphreys has always been a tremendous help during my stays in Upington.

On the academic side, I owe very much to Amanda Miller and Bonny Sands. Their friendship, cooperative spirit, generosity in sharing their data, and linguistic insight have always made me feel welcome in the community of 'Khoisan' researchers. I also thank Johanna Brugman, Chris Collins, and Levi Namaseb for support and helpful discussions.

Thanks are also due to Tom Güldemann for his openness and support in initiating me into the study of N|uu during our first joint fieldwork on the language, and for sharing his profound knowledge of all things 'Khoisan'. Many of my views on the topic have been shaped by the challenging discussions with him. Also, I am grateful to Martina Ernszt and Sven Siegmund for their help with collecting missing pieces of data, as well as to Alena Witzlack-Makarevich for helpful discussions.

I thank my supervisors, Hans-Jürgen Sasse and Martine Grice, for their insightful comments, friendly support, and patience. Hans-Jürgen, my teacher for many years, has had a profound influence on what I think linguistics is all about. I am very grateful for that.

Moreover, I thank the University of Cologne and the German Academic Exchange Service (DAAD) for their financial support. I thank Bernd Heine and the colleagues at the Khoisan Forum for the discussions and the inspiring atmosphere, and Rainer Voßen and my fellow Khoisanists of the International Symposium on Khoisan Languages and Linguistics in Riezlern, where I have always been overwhelmed by the friendly, enthusiastic, and supportive atmosphere.

My family and friends have always been interested, helpful and supportive of my research at home and abroad, for which I am very grateful. And there are almost certainly numerous other colleagues whom I have failed to mentioned here but who have nevertheless had their share in this work. I trust they do not take this oversight as a sign of lacking gratitude.

Last but not least, I wish to thank my wife Tamara for her love, patience, endurance, and support throughout this often difficult project.

1 Introduction

The N|uu language is of special interest for a number of different reasons. On the one hand, its speakers are among only a handful of people who still have some knowledge of one of the languages that once used to be spoken all over southern Africa. As such, they have become something of an emblem of the so-called New South Africa, a post-apartheid South Africa that is increasingly aware, and proud, of its historical roots. A strong symptom of this development is the fact that the motto on the new South African coat of arms is written in (reconstructed) |Xam, an extinct close relative of N|uu that was once spoken in the Karoo area of South Africa.

Bearing this in mind, it is maybe not quite as surprising that the local newspaper in my northern German hometown recently titled, “N|u-Sprache bleibt erhalten [N|u language is preserved]” (2005, May 30, p. 9). While it may be overly optimistic to think that a language with less than ten elderly speakers, who themselves seldom use the language in daily discourse because they live far from each other, is saved from extinction only through linguistic documentation, this example serves to illustrate the extent of public interest in the community, the speakers, and their language.

From a diachronic linguistic point of view, the N|uu language is of particular interest because it represents one of the few language families that (to the best of our current knowledge) have been spoken in situ in Southern Africa for as long as the prehistoric evidence goes. And because all those languages and language families are small and endangered (though not all as highly endangered as N|uu) and many are also poorly documented, reliable data on N|uu is invaluable for reconstructing the linguistic prehistory of the region.

Finally, from a synchronic linguistic point of view, N|uu is particularly challenging because it has one of the most complex sound systems in the world, rivaled only by some of its neighboring languages, like e.g. !Xóõ. At the heart of its sound system is its large inventory of contrastive click consonants, which is the main focus of the present work.

Our discussion will precede as follows: In Chapter 2, some basic background information on the N|uu language will be given. Chapter 3 will then present in detail the entire phoneme inventory of the language, before we will procede to discuss the phonological structure in Chapter 4. Chapter 5 contains two detailed experimental case studies (one on the anterior click closure, one on the posterior click closure) that help to shed some light on our understanding of the phonological representation of clicks. The information gathered in the preceding chapters will then lead us to Chapter 6, where some concrete proposals will be made concerning the categorization and, consequently, the transcription of clicks. Chapter 7 concludes this study.

2 N|uu: A Disappearing Language

N|uu is the sole surviving language of the !Ui branch of the Tuu language family, formerly known as ‘South Khoisan’. The Tuu family as a whole has no known relatives, and its second branch, named Taa, also only contains a single extant language (or dialect continuum), !Xóõ (Hastings, 2001; Güldemann, 2005). Traditionally, both languages are subsumed under a hypothetical ‘Khoisan’ language family (Greenberg, 1963); however, most specialists at present agree that no genealogical relationship among the supposed subgroups of ‘Khoisan’ can be convincingly established; see e.g. Güldemann and Vossen (2000).

The term *N|uu* is not really a glossonym; rather, *ᵑ|ùú* is a verb meaning ‘to speak the N|uu language’. From this, a verbal noun *ᵑ|ùú-cì* can be formed which literally means ‘(the act of) speaking N|uu’. In various orthographic variants, this verbal noun has been a common way of referring to the language in the literature. Another common term of reference for the language is *ᵑǁŋ̀ŋ̀ ǂé* (or any of its many orthographic renderings), literally meaning ‘people’. The most common way that the language and especially the community is referred to nowadays is by the term *ǂKhomani*. In a linguistic context, however, this term should be avoided since it is probably originally an exonym, used by the neighboring |’Auni tribe to refer to the N|uu (Bleek, 1937, p. 219). More importantly, none of the remaining N|uu speakers recognise the term *ǂKhomani* or identify with it in any way.

Formerly (before around 1870, when the colonization of the Southern Kalahari set in), N|uu was spoken in a fairly large but sparsely populated area of what is now South Africa, bounded approximately by the Orange river in the south, the Langeberg mountains in the east, the Molopo river in the north, and the Namibian border in the west (cf. Papst, 1895, 1895–1896; Herbst, 1908; Pöch, 1909a, 1909b, 1910; Bleek, 1927, 1929, 1939–1940, 2000; Dart, 1937;

Marais, 1939). Figure 2.1 shows the approximate extent of the traditional language area of N|uu.[1]

Figure 2.1 Topographic map of southern Africa The box indicates the approximate traditional language area of N|uu.

Due to forced relocation in the 1930s and social and economic subjugation and marginalization during the apartheid era, however, the state of the language quickly declined, and at the time of writing only eight elderly speakers of N|uu remain who live in three widely separated areas: Upington in the south, Olifantshoek in the east, and Andriesvale in the north.

The remaining speakers represent two original dialectal forms of the language which, however, are not extremely divergent (i.e., in all situations that I could

1 The map was produced using the mapping software GMT (Wessel & Smith, 2010) with the ETOPO2 (National Geophysical Data Center, 2006) and SRTM30 (Jet Propulsion Laboratory, 2005) data sets.

witness, mutual comprehension across the dialects was almost complete). All of my consultants represent what I for convenience will call the 'Western' dialect.

I conducted phonetic, lexicographic, and basic morphosyntactic studies of the language during three extended field trips of three months each between 2003 and 2005. I worked regularly with four speakers (all of them women and closely related), Ouma Katrina Esau, Ouma Anna Kassie, Ouma Hanna Koper, and Ouma Griet Seekoei. All of them explicitly consented (and indeed expressedly wished) to be mentioned by name.

The earliest more or less substantial linguistic data on N|uu date back to the 1930s, when an expedition from the University of the Witwatersrand in Johannesburg set out into the southern Kalahari to conduct linguistic and ethnographic studies on the communities that lived in or close to the area at the time. Several different language communities used to live in fairly close proximity in the area that is now the Kgalagadi Transfrontier Park; one of these groups consisted of N|uu speakers. The linguistic studies regarding N|uu (at that time referred to by the probable exonym, *ǂKhomani*) from this expedition were all published in a lengthy collection of articles, Rheinallt Jones and Doke (1937). Among these, Doke (1937) is the first article-length treatment of N|uu phonetics. After that, nothing was published for a long time, and the language was actually thought to be extinct until it was 'rediscovered' by human rights activists in the late 1990s. Nigel Crawhall, a South African-based human rights activist and sociolinguist, played a major role in making the discovery known, and he published a series of articles and a dissertation on the language (Crawhall, 2001, 2002, 2003, 2004, 2005a, 2005b).

Then, starting in 2003, a group of American and Namibian linguists around Amanda Miller, Bonny Sands, and Chris Collins started conducting in-depth phonetic (Miller, Brugman, Sands, Namaseb, Exter, & Collins, 2007, 2009;[2] Sands, Brugman, Exter, Namaseb, & Miller, 2007), lexicographic (Sands, Miller, & Brugman, 2007), and syntactic (Collins 2005) studies on N|uu within a large-scale National Science Foundation (NSF) project (*Collaborative Research: Descriptive and Theoretical Studies of N|u*). Miller et al. (2007, 2009) are the first modern in-depth descriptions of the phonetic structure of N|uu and will therefore serve as a reference point throughout the present work.

2 For convenience and better legibility, these two works will henceforth be referred to as Miller et al. (2007) and Miller et al. (2009).

Finally, another large-scale project, the Hans Rausing Endangered Languages Project (HRELP) grant *A text documentation of N|uu* under the direction of Tom Güldemann has set out to document the grammatical structure of N|uu by eliciting and analyzing a large number of original texts.

3 The Phoneme Inventory of N|uu

This chapter provides an in-depth study of the phoneme inventory of N|uu. The presentation will proceed as follows: First, descriptions of the non-click consonant, click consonant, vowel, and lexical tone inventories will in turn be given in separate sections. To do so, the individual phonemes (represented by their main allophones) will be organized in tables in the familiar IPA format (International Phonetic Association, 1999), and the descriptive parameters relevant for the respective classes of sounds will be presented. Finally, the allophonic realizations of the phonemes and their most salient phonetic properties will be discussed.

The phonological inventory of N|uu (with the exception of lexical tones) has been presented before by Miller et al. (2007, 2009). However, based on my own fieldwork data, I will present a slightly modified account here: Some phonemes reported earlier will be argued not to be part of the system, while other phonemes were only recently discovered and will therefore be added to the system; still others will be retained while being analyzed in a slightly different way. The majority of segments, however, will remain unchanged.

3.1 Non-Click Consonants

As is the case with all 'Khoisan' languages, the consonant system of N|uu is very large, involving many dimensions of contrast. Most conspicuous is the presence of clicks, which are for convenience presented here separately from the non-click consonants in Section 3.2. This should not be taken to imply that clicks are totally detached, phonetically or phonologically, from the other consonants. While I will argue later that clicks *are* indeed 'special' in a number of ways, they do of course share certain phonetic and phonological characteristics with the other consonants, and the separate presentation here is only due to the fact that it allows for a clearer arrangement of the many phonemes of N|uu in a 'classical' IPA manner.

The phonological analysis presented here is based on a lexical database containing 701 roots (cf. Chapter 4). Assuming a total number of 94 attested segmental phonemes in N|uu (25 non-click consonants, 52 click consonants, and 17 vowels) plus 2 lexical tones, it is obvious that there is a rather high probability that some phonemes might not yet have been discovered due to the limited size of the available lexicon. The segment */ʘ͡q$^{\mathrm{h}}$/ may serve as an example: It is not attested in my data but should be expected, given the otherwise strong tendency towards symmetry in the click subsystem, and it is therefore represented by a dotted circle in Table 3.2. As will be shown in Section 4.2.2, bilabial clicks are generally of low lexical frequency in N|uu; specifically, only 16 bilabial clicks are attested in my 701-root corpus.[3] On the other hand, aspirated simultaneously released linguo-pulmonic stops are represented 54 times in the corpus. Together, this results in an expected frequency of aspirated bilabial simultaneously released linguo-pulmonic stops (i.e. */ʘ͡q$^{\mathrm{h}}$/) of (16 × 54) / 701 = 1.23. This means that given a corpus of this size, only about one root containing */ʘ͡q$^{\mathrm{h}}$/ would be expected a priori, so it should not be too surprising that none has been attested so far.

As Maddieson (1997, p. 636) points out, “inventories tend to be built by the intersection of repeated characteristics”. Such considerations of symmetry, together with areal-typological arguments, lead me to assume that some of the more striking gaps in the system are in fact accidental gaps that might be filled if more lexical data were available. Such presumed accidental gaps are marked by dotted circles in the respective tables.

Table 3.1 shows the non-click consonants of N|uu. As is customary, columns correspond to place of articulation and rows correspond to manner of articulation. Within each cell, four glottal states are distinguished by the order in which the segments appear (this is more apparent in the click subsystem as given in Table 3.2 below); they are, from left to right: voiceless, voiceless aspirated, voiceless ‘glottalized’ (i.e. produced with a closed glottis, whether ejective or not), and voiced. Airstream contrasts, finally, are represented by the order of rows: Pulmonic segments are represented in the first eight rows, while glottalic egressive segments (ejectives) are found in the last two rows.

3 As will be shown in Section 4.2.1, clicks in N|uu are confined to a single prosodic position, called C_1 here; this position is obligatorily present in every root.

Table 3.1 The non-click consonant phonemes of N|uu. The velar nasal /ŋ̍/ is in parentheses because it only occurs as a syllable *nucleus*. Dotted circles indicate presumed accidental gaps.

	Bilabial	Alveolar	Prepalatal	Velar	Uvular	Glottal
Plosive	p ◌ b		c cʰ ɟ	k kʰ g		ʔ
Affricate		t͡s				
Heterorganic Affricate	◌		c͡χ			
Nasal	m	n	ɲ	(ŋ̍)		
Tap		ɾ				
Fricative		s z			χ	h
Approximant	β					
Lateral Approximant		l				
Ejective Stop	◌		c’	◌		
Ejective Affricate		t͡s’		k͡χ’		

The inventory of non-click consonants in N|uu as given in Table 3.1 differs from the one given by Miller et al. (2007, 2009) in several respects, which will now in turn be discussed.

Recent loans from Afrikaans. Miller et al. (2007, 2009) list /t/, /d/, and /f/ among the phonemes of N|uu, acknowledging that they only occur in loan words (mostly from Afrikaans). In the present work, they are not included because the loan words in question are hardly integrated into the phonological structure of N|uu. Rather, given that the N|uu speakers are isolated from each other and, at the same time, fully bilingual in N|uu and Afrikaans, those words are better regarded as nonce borrowings (‘single-word code-switching’) that are becoming increasingly necessary because N|uu is no longer supported by a productive speech community that would be able to coin new lexical items or to produce conventionalized phonologically integrated loan words through frequent usage. Therefore, any lexical item from Afrikaans can potentially be the target of such nonce borrowings, and, consequently, any Afrikaans phoneme could be regarded as a ‘loan phoneme’ of N|uu with the same justification.

Uvular non-click stop and affricate. *Pace* Miller et al. (2007, 2009), I do not include /q/ and /q͡χ’/ in the phoneme inventory of N|uu. Judging from my own data and from the data that was kindly made available to me by A. Miller and B. Sands, I do not see any compelling evidence that there is in fact a phonological contrast between */q/ and /k/ on the one hand, and between */q͡χ’/ and /k͡χ’/ on the other, at least for the variety of N|uu represented here. Instead, I argue that the lexical items that have been transcribed with the uvular segments (*/q/, */q͡χ’/) are really velar (/k/, /k͡χ’/).[4] A certain ‘back’ auditory impression could have been caused by the fact that in most of the lexical items in question, the initial (velar) consonant is followed by a rounded back vowel, which has the effect of lowering the resonance frequencies associated with the consonant due to anticipatory lip rounding.

Prepalatal ejective. One non-click consonant phoneme, a prepalatal ejective /c’/, was only recently discovered in the data (in the word *c’áˤè* ‘be in pieces’) and is therefore added to the inventory.

A minor terminological difference is that I use the term *prepalatal* for the categories of N|uu where Miller et al. (2007, 2009) use *palatal*. This choice will be motivated in Section 3.1.1.

Glottal stop. Miller et al. (2007, 2009) argue that the phonemic status of the glottal stop /ʔ/ in N|uu is doubtful and that it might be preferable to analyze it as a predictable surface segment whose function it is to fill an otherwise empty syllable onset. In the present work, however, I do attribute phonemic status to the glottal stop, on the following basis:

Consider first a lexical root like *‖͡qʰàná* ‘scorpion’, with an underlying intervocalic nasal. In a case like this, the nasal is regularly assigned to the onset of the second syllable: [‖͡qʰà.ná]. Now let us turn to a root like *ᵑ̊|ˀáǹʔá* ‘lick’ (syllabified as [ᵑ̊|ˀáǹ.ʔá]). In this root, there is what I take to be an underlying glottal stop that is assigned to the onset of the second syllable, while the nasal is assigned to the coda of the first syllable. Now, it would be theoretically possible to argue that the glottal stop is not yet present at syllabification (although it would not be clear why the two words would then be syllabified differently in the first place), and that a constraint militating against onsetless syllables then leads to the insertion of [ʔ].

4 To be sure, the affricate, /k͡χ’/, is uvular at *release*, as indicated by the transcription.

But now consider a nasal-final root like *ᵑ̊!ˀáń* ‘be old’. When the stative suffix *-à* is suffixed to such a root, the resulting form *ᵑ̊!ˀáń-à* is syllabified as [ᵑ̊!ˀáń.nà], with a geminate nasal. Obviously the strategy to ‘repair’ onsetless syllables in this case is gemination, not glottal stop insertion, and it is not obvious why a word like *ᵑ̊|ˀánʔá* ‘lick’ should represent a different repair strategy.

Let us assume for a moment that the glottal stop in ‘lick’ is indeed not present in the underlying representation. As far as I could ascertain, the form is not morphologically complex (at least not synchronically). But even if that possibility cannot be ruled out completely, there would be only two possibilities: Either the form is morphologically simple; then it should behave like ‘scorpion’; or it is morphologically complex; then it should behave like ‘be old’. Neither possibility is born out by the data.

As a last resort, one could assume different repair strategies for different situations. But it seems to me that the more natural and elegant solution is to assume that the glottal stop is indeed present underlyingly in forms like *ᵑ̊|ˀánʔá* ‘lick’, and many others.

Glottal fricative. In Miller et al. (2007, 2009), it is proposed that the glottal fricative in N|uu should most adequately be analyzed as *voiced*. While I do agree that there are contexts where it is indeed voiced, I assume that such voicing is not an invariant property of the phoneme, but rather that it is contextually determined (by coarticulation, e.g. in intervocalic contexts). This is in contrast to Afrikaans, where glottal fricatives are clearly voiced, even in utterance-initial position. An illustration of a typical voiceless realization in N|uu will be given in Section 3.1.2. Consequently, I will transcribe the glottal fricative phoneme in N|uu as /h/, not */ɦ/.

Bilabial approximant. The voiced bilabial stop /b/ is assumed by Miller et al. (2007, 2009) to have two allophones: [b] (which occurs root-initially) and [β] (which occurs root-medially, i.e. intervocalically). Such an analysis is perfectly legitimate in N|uu, since both realizations are phonetically similar and in complementary distribution. However, it is precisely because of such distributional considerations that I prefer to analyze /b/ and /β/ as independent phonemes of N|uu in the present work. As will be shown in Section 4.2.1, there is a strong correlation between obstruent vs. sonorant status and prosodic position in N|uu:

While obstruents occur only foot-initially, sonorants are excluded from that position.[5] Therefore, distributional generalizations can be captured more easily if one assumes that /b/ is an underlying obstruent, while /β/ is an underlying sonorant (which is in agreement with the phonetic facts). Also, there is no morphophonological connection (alternation) between [b] and [β] in N|uu that would argue compellingly for an analysis in terms of an allophonic relationship. And finally, there are no other obvious corresponding obstruent–sonorant pairs with a distribution parallel to that of [b] and [β] that would set a precedence for an allophone analysis.

To be precise, the bilabial *approximant* phoneme should be transcribed as /β̞/, not /β/ (which, strictly speaking, represents a voiced bilabial *fricative*), but in the interest of legibility I will use the simpler form /β/ for the bilabial approximant in the present work.

3.1.1 Places of Articulation

Within the class of non-click consonants in N|uu, six places of articulation can be distinguished in surface forms: bilabial, alveolar, prepalatal, velar, uvular, and glottal.

Bilabial. Non-click consonants with a bilabial place of articulation in N|uu are, expectedly, produced with the lower and upper lip (i.e. with a 'labio-labial' articulation).

Alveolar. In N|uu, alveolar non-click consonants are generally *lamino-alveolar*, i.e., they are produced with the *blade* of the tongue against the alveolar ridge. The distinction between click and non-click consonants is important in this context in that alveolar *clicks* are produced with an *apical* anterior closure (cf. Section 3.2.1).

Prepalatal. Before discussing the prepalatal place of articulation in N|uu in more detail, the term *prepalatal* itself merits some discussion. No general consensus exists in the literature on the most adequate description of consonants articulated in the area between that used by prototypically coronal segments (e.g. [t]) and prototypically dorsal segments (e.g. [k]). Most often, such sounds re-

5 Nasals are an exception to this generalization as they can occur in all positions.

ceive the label *palatal* (or variations of it), but it is not quite clear where the boundaries of this category are, whether there is a linguistically motivated need for further subdivisions within it, and, indeed, what the phonetic and phonological status of the category as such is (e.g. its status as coronal or dorsal). The description of sounds on the boundary between the palatal place of articulation and the postalveolar and velar places of articulation, respectively, is even more controversial.

First of all, it is necessary to make explicit the anatomical landmarks that are referred to when describing the relevant articulations. As for the passive places of articulation along the roof of the mouth, a major division into alveolar, postalveolar, palatal, and velar places seems to be generally accepted (although further subdivisions within the palatal and velar places are more contentious). As for the active articulators along the surface of the tongue, less consensus is found in the literature. Broadly speaking, two traditions can be discerned: On the one hand, some authors take the tongue blade (lamina) to be the relatively mobile part of the tongue that is roughly opposite the alveolar ridge when the tongue is at rest. Thus, e.g. Catford (1968, 1977) uses a definition of the lamina as extending about 10 to 15 mm back from the apex (tongue tip). Similarly, Ladefoged (1989) takes the lamina to extend for 8–10 mm behind the apex (which is assumed to be about 2 mm wide). Keating (1991), on the other hand, uses a figure of 1.5–2 cm for anterior coronals and maximally 3–4 cm for non-anterior coronals. It would seem that the shorter definitions of the lamina are primarily based on anatomical landmarks, whereas the longer definitions are rather based on typical phonological patterns (since the articulatory boundary between laminal and dorsal is usually taken to correspond to the phonological boundary between coronal and dorsal).

Secondly, authors differ in the degree to which they subdivide the dorsum (or tongue body, defined as the part of the tongue bounded by the lamina and the radix, or tongue root) and the palatal place of articulation (identified on anatomical grounds as the hard palate, bounded by the alveolar ridge and the soft palate). Ladefoged and Maddieson (1996), e.g., do not set up any subcategories within the dorsum or the palatal place at all, whereas Catford (1968, 1977) divides his dorsum more or less equally into *anterodorsum* and *posterodorsum* and his palatal zone (again, equally and somewhat arbitrarily) into *prepalatal* and *palatal proper*. Recasens (1990), finally, argues on the basis of X-ray, palatographic, linguographic, and EPG evidence that palatal consonants may involve a higher degree of articulatory precision than previously thought; he divides the dorsum into *predorsal*, *mediodorsal*, and *postdorsal* articulators and the palatal region into *prepalatal*, *mediopalatal*, and *postpalatal* places of articulation. Stone, Epstein, and Iskarous (2004), in a study based on ultrasound and tagged cine-MRI data, suggest that the tongue can be modeled computationally as consisting of five independent *functional* segments (i.e. segments that act as functional units). I would hypothesize that in traditional phonetic terms, those functional segments can be broadly translated into lamina plus apex, predorsum, mediodorsum, postdorsum, and radix, which is compatible with the descriptive framework suggested by Recasens (1990).[6]

Thirdly, there is some controversy in the literature as to whether segments produced in the broader palatal area are simple or complex. Three positions are found in the literature: Catford (1968, 1977) regards such sounds as simple segments; he distinguishes between *lamino- or [antero]dorso-prepalatal* sounds and *[antero]dorso-palatal* sounds. Likewise, Recasens (1990) distinguishes between *predorso-postalveolo-prepalatal* ('alveolo-palatal'), *predorso-prepalatal* ('front palatal'), *mediodorso-mediopalatal* ('mid palatal'), and *postdorso-postpalatal* ('back palatal') sounds, claiming that none of these are complex segments. Ladefoged and Maddieson (1996), on the other hand, analyze 'alveolo-palatal' sounds (of the [ɕ] type) as coronal segments with a secondary articulation, more precisely as *palatalized lamino-postalveolar* (in contrast to sounds of the [ç] type, which are analyzed as simple *dorso-palatal* segments). An even more extreme approach is taken by Keating (1988): Based on X-ray evidence from Czech, she concludes that palatals are complex, multiply articulated (coronal-dorsal) segments.

As for the situation in N|uu, I will adopt the descriptive framework put forward by Recasens (1990), both because it is largely in accordance the results of Stone, Epstein, and Iskarous (2004) and because it is very well suited for the description of clicks. As the linguogram of [c] in *cáà* 'lie (recline)' in Figure 3.1 (left) shows, the articulation is mainly predorsal; and while there may be some

6 As for the *velar* region, see Keating and Lahiri (1993) for arguments (based on X-ray, palatographic, and acoustic evidence) that [back] velars, fronted velars, palatalized velars, and palatals all need to be distinguished. In the framework of Recasens (1990), fronted velars and palatalized velars would be called *back palatals*, with palatalized velars being characterized by some *additional* mediopalatal involvement, resulting in a *longer* contiguous constriction. This parallels the distinction between his front palatals and alveolo-palatals: The latter have *additional* postalveolar involvement and thus a longer constriction.

mediodorsal contact as well, it is clearly not laminal (in the more restricted sense of the term).[7]

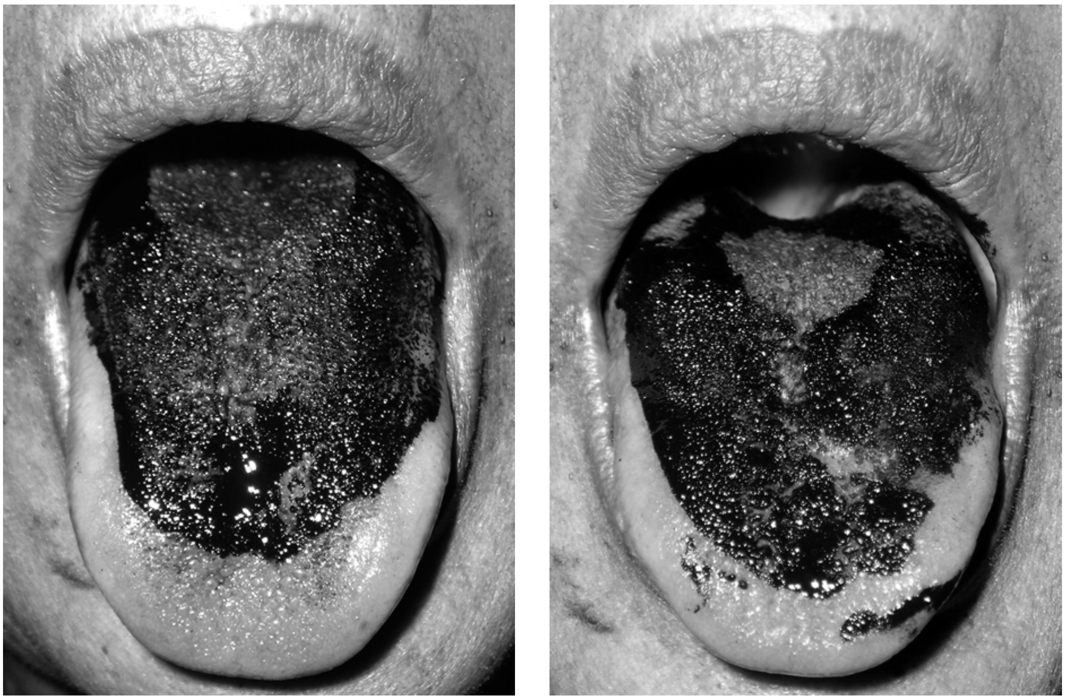

Figure 3.1 Linguograms of /c/ in *cáà* 'lie (recline)' (left) and /ɟ/ in *ɟá* 'someone' (right; both by speaker KE). The posterior border of the contact area in /c/ is somewhat obscured by traces of a previous linguogram.

As ist to be expected when a convex and rather inflexible part of the tongue (such as the predorsum) makes contact with a concave surface (such as the prepalatal area), the articulation is characterized by a very large contact area. Unfortunately, my data does not include a matching palatogram of sufficiently high quality, but the prepalatal area is definitely involved in the articulation of [c], although there seem to be tokens with additional postalveolar contact as well. I am not sure how consistent that additional contact is and whether it is part of the articulatory target or just an incidental consequence of the particular geometrical conditions of the vocal tract referred to above. Therefore, I cannot at present decide whether [c] in my data corresponds to the 'alveolo-palatal' or

7 See Ladefoged (2003) for a description of the procedure for static palatography and linguography.

rather to the 'front palatal' category suggested by Recasens (1990). In the absence of conclusive data on this point, I will use the simpler label, *(predorso-)prepalatal*, for the category instantiated by [c] in N|uu. (To avoid a proliferation of diacritics, the simple transcription symbol [c] rather than the more precise symbol [c̟] will be used here, since no ambiguity can arise from this.)

Figure 3.1 (right) shows a linguogram of [ǂ] in ǂá 'someone'. It is quite obvious that the area of contact corresponding to the *anterior* closure of the click is very similar to the one in [c]; in fact, the only relevant difference is that in [ǂ], there is a posterior closure present in addition to the anterior one, resulting in the characteristic lingual cavity that is clearly visible in the linguogram. Therefore, I will use the same articulatory label (*prepalatal*) for clicks of the [ǂ] type as I use for non-click consonants of the [c] type.

The large extent of the contact area is probably also responsible for two other characteristic features of prepalatal segments in N|uu. Firstly, at the release of prepalatal stops, the tongue cannot be moved away from the hard palate exactly simultaneously at all points of contact. This lack of precision results in a situation where the constriction at some parts of the articulator is already wide enough to preclude turbulent airflow, while other parts of the constriction are not yet quite wide enough, resulting in a short yet noticeable period of frication noise (affrication). And secondly, because the tongue dorsum moves relatively slowly (e.g. in contrast to the lamina), there is typically a certain amount of diphthongization in neighboring vowels, with a short high front vowel component characterizing the movement into and out of prepalatal consonants (the latter especially in nasals, which lack the masking effect of the affricated release). Thus, a word like *ᵑ|ùcú* 'nose' would be realized as [ᵑ|ùⁱcᶝú][8]. Figure 3.2 shows a spectrogram of that word.

Despite the fact that [c] is predorsal and *not* laminal, it is still phonologically coronal in N|uu, as will be shown in Section 4.2.1 (the same applies to the anterior closure of [ǂ]). Such ambiguous behavior, however, should not be too surprising for segments that are situated in between typical 'focal' categories; this is in accordance to the theory of emergent features as proposed e.g. by Mielke (2008).

8 The superscript symbols in this narrow transcription are meant to indicate that the amount of affrication and diphthongization is noticeably less extensive than in 'true' affricates and diphthongs, respectively.

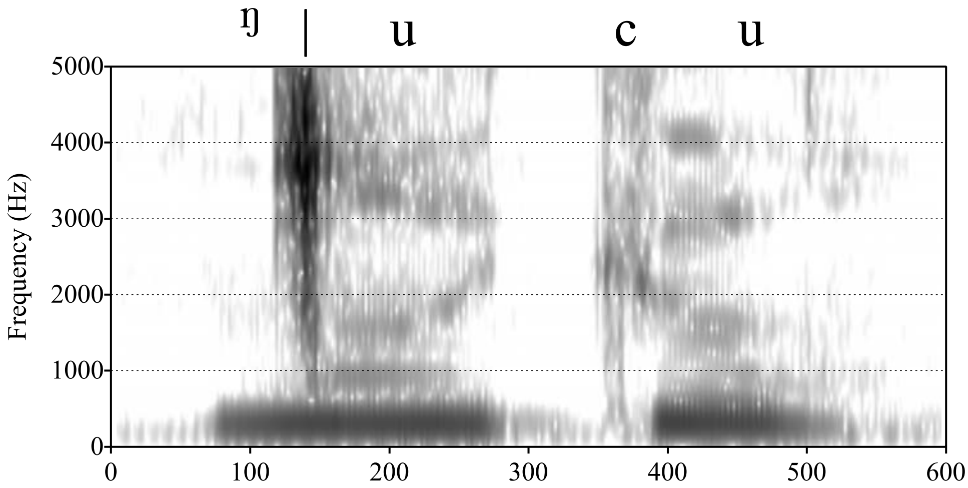

Figure 3.2 Spectrogram of *ᵑ|ùcú* ‘nose’ (speaker KE). Note the affrication of the palatal plosive and the diphthongization of the vowel preceding it.

As has been described above, alveolar and prepalatal segments both occur in *surface* forms in N|uu; the two places of articulation are, however, only very marginally *contrastive* (cf. Section 4.2.1). It is not very common typologically for a language to lack alveolar plosives altogether while possessing a whole set of prepalatal plosives (cf. Maddieson 1984), as N|uu does. But in fact there seems to be an ongoing sound change in the language that results in the unmarked alveolars shifting to the more marked prepalatals. This assumed sound change has almost completely annihilated the underlying contrast between alveolars and prepalatals, and it is interesting to note that in the eastern dialect of N|uu (which is not covered in depth in the present work), the change seems to have progressed even further in that the only remaining marginally contrasting context (word-initial nasals), the contrast has been removed in favor of the prepalatal (western N|uu *ná* ‘I’ corresponds to eastern N|uu *ɲá*).

Velar. Underlyingly velar non-click consonants in N|uu are invariably produced at the soft palate (velum). While this seems tautological at first, it is actually quite unusual: Velar segments are only minimally, if at all, affected coarticulatorily by the tongue position of a following vowel.[9] Instead, in se-

9 Anticipatory lip rounding of consonants preceding rounded vowels, however, does occur.

quences of a velar consonant and a following front vowel, retraction of the vowel occurs, as in *kéré* [kə́ré] 'lightning'.

One can hypothesize that this is in analogy to the coarticulatory relationship between *clicks* (which also have a dorsal posterior closure but a much higher lexical and text frequency than 'plain', non-click, velar consonants) and following front vowels: Like non-click velars, clicks remain largely unaffected by such coarticulatory influences while at the same time exerting a very strong influence on the articulation of following vowels. Another possible explanation is that coarticulatory fronting of a velar before a front vowel would entail the risk of losing the distinction between velars and prepalatals, e.g. between [k] and [c].

Uvular. There is only one underlyingly uvular non-click segment in N|uu, the fricative [χ]. The spectrogram of [χ] in *χàá* 'scratch' given in Figure 3.3 shows the irregular fluctuations in intensity characteristic of voiceless uvular fricatives (caused by involuntary, more or less periodic movements of the uvula in the airstream).

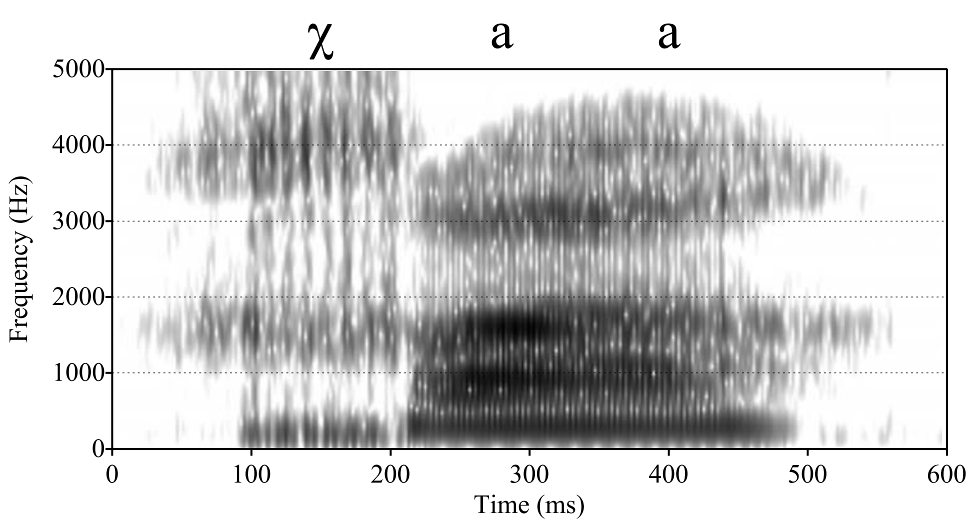

Figure 3.3 Spectrogram of *χàá* 'scratch' (speaker KE). The uvular fricative /χ/ shows characteristic irregular fluctuations in intensity caused by involuntary, more or less periodic movements of the uvula in the airstream.

It should be noted that while both velar and uvular non-click consonants are clearly present on the *surface* in N|uu, they are not in phonological *contrast* to each other (cf. Section 4.2.1).

Glottal. Both the glottal stop [ʔ] and the glottal fricative [h] have the phonetic values suggested by their transcription, i.e., they are produced at the *glottal* place of articulation.

3.1.2 Manners of Articulation and Their Phonetic Realization

In surface forms in N|uu, ten manners of articulation have to be distinguished within the class of non-click consonants, which will now be taken up in turn.[10]

Plosive. The class of plosives (i.e. pulmonic consonants produced with a complete closure of the vocal tract) in N|uu includes the voiceless unaspirated segments /p/, /c/, /k/, and /ʔ/, which are bilabial, prepalatal, velar, and glottal, respectively (3.1a); the voiceless aspirated segments /c^h/ and /k^h/, which are prepalatal and velar (3.1b); and the voiced segments /b/, /ɟ/, and /g/, which are bilabial, prepalatal, and velar (3.1c).

(3.1) a. *píɾí* 'goat'
cú 'mouth'
ká^{n}ì^n 'be startled'
ʔá 'give'

b. *c^{h}óè* 'be naked'
k^{h}ànà 'be wide'

c. *báˤùkè* 'bark (verb)'
ɟó^n 'skin'
gáò 'thing'

Voiceless unaspirated segments have a very short voice onset time (VOT) of approximately 0 ms. Phonologically aspirated segments have a comparatively short aspiration with a VOT on the order of only 50 ms (although no quantitative study on this has been undertaken yet). Phonologically voiced segments, finally, have comparatively weak voicing that is frequently absent in utterance-initial

10 As will be noted, allophonic variation is very limited in N|uu. Unless noted otherwise, the surface realization of the phonemes is as suggested by the choice of the respective transcription symbols.

position (i.e., when the preceding context is *not* voiced). Also, as noted by Miller et al. (2009), the degree (likelihood and duration) of voicing in utterance-internal segments depends on the presence and strength of any prosodic boundary preceding it: The stronger the prosodic boundary, the weaker the voicing in the segment that follows it. Figure 3.4 shows spectrograms of the words *ká^{n}ì^n* 'be startled', *k^{h}ànà* 'be wide', and *gáò* 'thing', realized in the context *ná ká* __ 'I say __'.

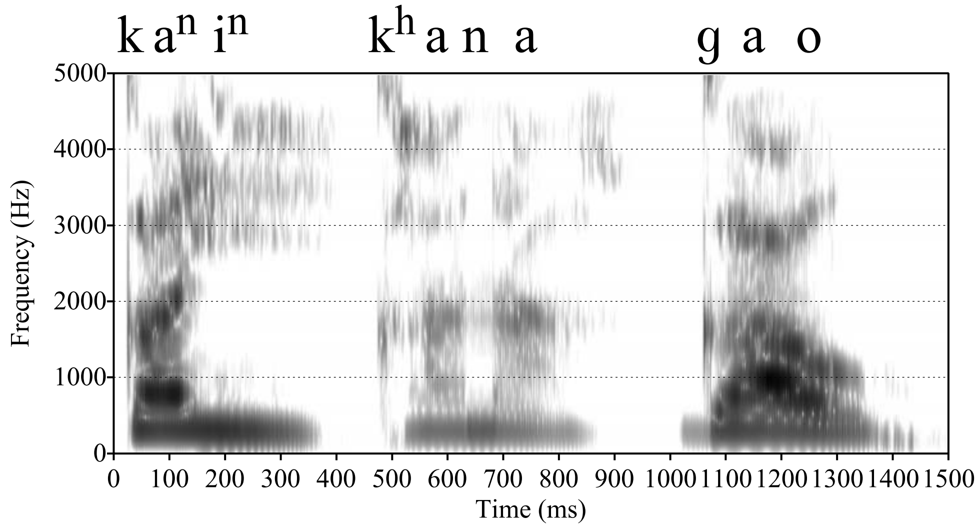

Figure 3.4 Spectrograms of *ká^{n}ì^n* 'be startled', *k^{h}ànà* 'be wide', and *gáò* 'thing', realized in the context *ná ká* __ 'I say __' (speaker KE)

Affricate. The phoneme inventory of N|uu includes only one pulmonic (homorganic) affricate, namely the *voiceless lamino-alveolar sibilant affricate* /t͡s/ (3.2). It consists, as the transcription suggests, of a lamino-alveolar stop portion released into a lamino-alveolar sibilant fricative portion.

(3.2) *t͡sá^{n}á^n* 'buchu'

Heterorganic affricate. In addition to the homorganic affricate /t͡s/, N|uu, like many other languages in the Khoisan area, has what is often referred to as a *heterorganic affricate*. This term is sometimes applied to cases where the stop portion and the fricative portion of the affricate are are not produced *exactly* at the same place of articulation. This is e.g. the case in German /p͡f/, where the stop portion is bilabial whereas the fricative portion is labiodental; i.e., the two places of articulation are contiguous, with only a slight adjustment in the position of the active articulator taking place.

In N|uu (and other Khoisan languages), however, the two portions are truly heterorganic: While the stop portion is *prepalatal*, the fricative portion is *uvular*, resulting in the *voiceless prepalatal-uvular heterorganic affricate* /c͡χ/ (3.3).

(3.3) *c͡χáá* 'tear up'

Such heterorganic affricates have been described for languages outside the Khoisan linguistic area before, e.g. by McDonough and Ladefoged (1993), who describe the heterorganic affricate [t͡χ] in Navajo, although phonologically, the authors prefer to treat it as /tʰ/ in that language. Such an analysis is not possible in N|uu, because, as we have seen, /cʰ/ is an independent and contrasting phoneme.

There are three reasons for still treating /c͡χ/ as a unitary phoneme and not as a sequence of phonemes such as */cχ/ in N|uu. The first is that /c͡χ/ has the *phonetic* characteristics of an affricate. As Howell and Rosen (1983) have shown, *rise time* (the duration between the start of the frication noise and the point where it reaches its maximum amplitude) is a parameter that reliably distinguishes affricates from fricatives. Mitani, Kitama, and Sato (2006) distinguish between affricates and fricatives by the parameters *frication duration* and *amplitude rise slope*. Because (I assume) affricates and fricatives in my data do not differ substantially in any systematic way in terms of maximum amplitude values, the parameter *amplitude rise slope* should be inversely proportional to the parameter *rise time*. Therefore, in the present work, I will assume that the duration and rise time of the frication noise constitute the two main parameters that distinguish between affricates (low values for both parameters) and fricatives (high values for both parameters). No quantitative study has yet been undertaken on /c͡χ/ with regard to those parameters (this has so far only been done for clicks; cf. Section 5.2.2), but qualitatively, it seems quite clear that the fricative portion of /c͡χ/ has a considerably shorter duration and rise time than the fricative /χ/. Figure 3.5 shows representative waveforms of *χàá* 'scratch' and *c͡χáá* 'tear up'.

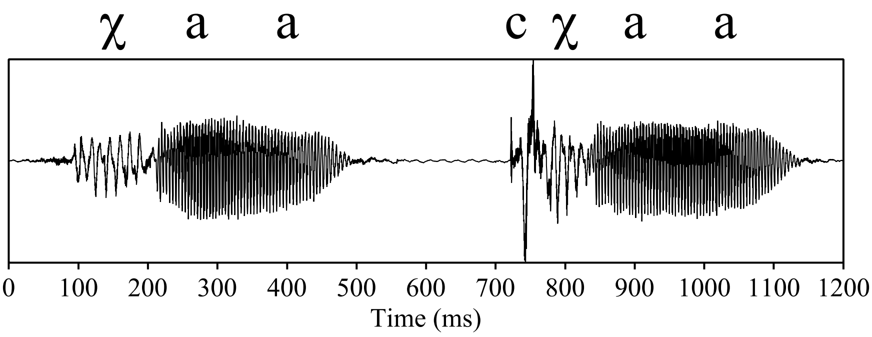

Figure 3.5 Waveforms of *χàá* 'scratch' and *c͡χáá* 'tear up' (speaker KE)

The second reason for treating /c͡χ/ as a unitary phoneme is its phonological distribution. As will be shown in Section 4.2.1, there are certain strong, foot-initial prosodic positions (called C_0, C_1, and C_3 in the present work) to which obstruents are confined; no obstruents can occur in any other position. These positions are either word-initial or bounded by vowels or sonorants, implying that obstruent clusters cannot occur in the language. If one were to analyze [c͡χ] as an underlying *sequence* */cχ/, then that would constitute the only exception to this generalization, leading to a considerably more complicated prosodic and phonotactic structure.

The third reason, finally, is the parallelism to the click system: I hypothesize that languages in the Khoisan area have a systematic tendency towards including heterorganic affricates such as /c͡χ/ in their phoneme inventories because it provides the possibility to build up parallel series of (complex) segments between their non-click and click systems. Consider e.g. the two phonemes /c͡χ/ and /ǂ͡χ/ in N|uu: Both are characterized auditorily by a coronal stop burst followed by dorsal frication noise. While the articulatory details of their production are quite different, the auditory similarities are close enough to be a stabilizing factor in the establishment of such parallel series. This is not so obvious in N|uu because only one heterorganic affricate has so far been found, but other languages like e.g. !Xóõ (Traill 1985, 2009) have much more elaborate systems of heterorganic affricates, including both labial and coronal stop portions and both pulmonic and glottalic airstream mechanisms.

Nasal. N|uu has four contrastive nasals: bilabial /m/, lamino-alveolar /n/, prepalatal /ɲ/, and velar /ŋ̍/; these are all voiced (3.4).

(3.4) *|ʰàmà* 'sneeze'
ᵑ‖ò ˤnó 'maxilla'
ɲéβécé 'greet'
ᵑ‖ŋ̀ŋ̀ 'house'

It is worth noting in this context that the velar nasal /ŋ̍/ differs from the other three nasals in that it occurs exclusively as a syllable *nucleus* (i.e., with a *vocalic* function), while the others occur only as syllable *margins* (onsets or codas; /m/, /n/, and /ɲ/ show different distributions in this regard, cf. Section 4.2.1). To reflect that distribution, the syllabicity diacritic is used with /ŋ̍/ here (unless it carries a tone mark).

Tap. There is one rhotic in N|uu, the *lamino-alveolar tap* /ɾ/ (3.5).

(3.5) *ᵍ!ùɾù* 'thunder'

Sometimes, /ɾ/ is realized as the corresponding *flap* instead; this seems to be speaker-dependent, but no systematic study on this has been carried out yet. Very occasionally, tokens of /ɾ/ realized as a *trill* [r] are encountered, although that that is not a consistent feature in any of the speakers I worked with. I hypothesize that it is due to momentary influence from Afrikaans.

Fricative. The class of fricatives contains the *voiceless lamino-alveolar sibilant fricative* /s/, the *voiceless uvular fricative* /χ/, and the voiceless glottal fricative /h/ (3.6a); it also contains the *voiced lamino-alveolar sibilant fricative* /z/ (3.6b).

(3.6) a. *sàá* 'give'
χú 'face'
hùí 'help'

b. *zéˤé* 'fly (verb)'

The latter is of particular interest because it is the only voiced fricative in the phoneme system of N|uu and, in my data, only occurs in a single word, *zéˤé* 'fly (verb)', so it is several ways isolated within the system. However, the realization of /z/ is always unambiguously voiced, 'fly' is clearly part of the basic vocabulary, and there is no obvious source that the word could have been borrowed from, so the status of /z/ as a phoneme of N|uu (and, with it, the status of voicing as a contrastive feature of fricatives in N|uu) is quite firmly established.

As was noted before, I analyze /h/ as underlyingly *voiceless* here, despite the fact that it can acquire voicing from the surrounding vocalic context. However, especially (but not exclusively) in utterance-initial position, the realization of /h/ in my data is predominantly voiceless, which is in contrast to the typical *voiced* realization [ɦ] in Afrikaans in all positions. Since all N|uu speakers are fluent in Afrikaans, the consistently different realization of the segments in the two languages provides further evidence that /h/ in N|uu is indeed underlyingly voiceless. Figure 3.6 shows a typical voiceless realization of /h/ in the word *hárú* 'be far', realized in utterance-initial position.

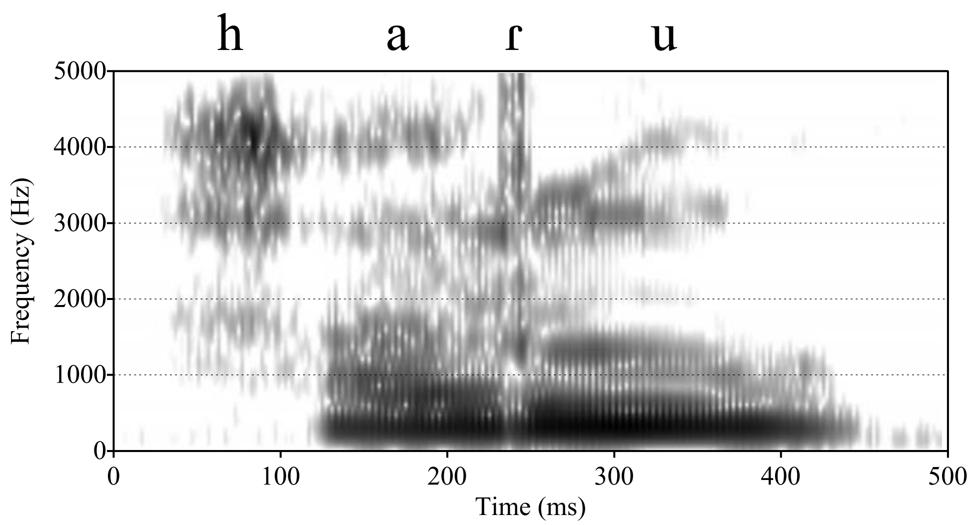

Figure 3.6 Spectrogram of *hárú* 'be far', realized in utterance-initial position (speaker KE)

Approximant. As argued above, there is one (central) approximant in N|uu, namely the *voiced bilabial approximant* /β/ (3.7).

(3.7) *|óβà* 'child'

Lateral approximant. In addition to the *central* approximant /β/, there is also a *voiced lamino-alveolar lateral approximant* /l/ in the N|uu phoneme inventory (3.8).

(3.8) *sòˁlè* 'grab'

This finishes the description of the *pulmonic* non-click consonants in the language and leads us to the glottalic segments, which are (as is typical of languages in the Khoisan area) exclusively *ejective* (i.e., there are no *implosive* consonants).

Ejective stop. As was noted above, there is one ejective stop in N|uu (recently discovered in one word, *c’áˤè* ‘be in pieces’), namely the *voiceless prepalatal ejective stop* /c’/ (3.9).

(3.9) *c’áˤè* ‘be in pieces’

As for the VOT of aspirated plosives, no quantitative study has yet been conducted on the *glottal lag* (the duration between the release of the oral closure and the release of the glottal closure, i.e., the onset of voicing) of ejective stops in N|uu. From qualitative inspection of the tokens that are found in my data, an order of magnitude for the glottal lag of around 50 ms (similar to the estimate given for the VOT above) seems reasonable, though. Figure 3.7 shows spectrograms of typical realizations of *cáà* ‘lie (recline)’ and *c’áˤè* ‘be in pieces’. Both /c/ and /c’/ show the slight affrication typical of prepalatal stops.

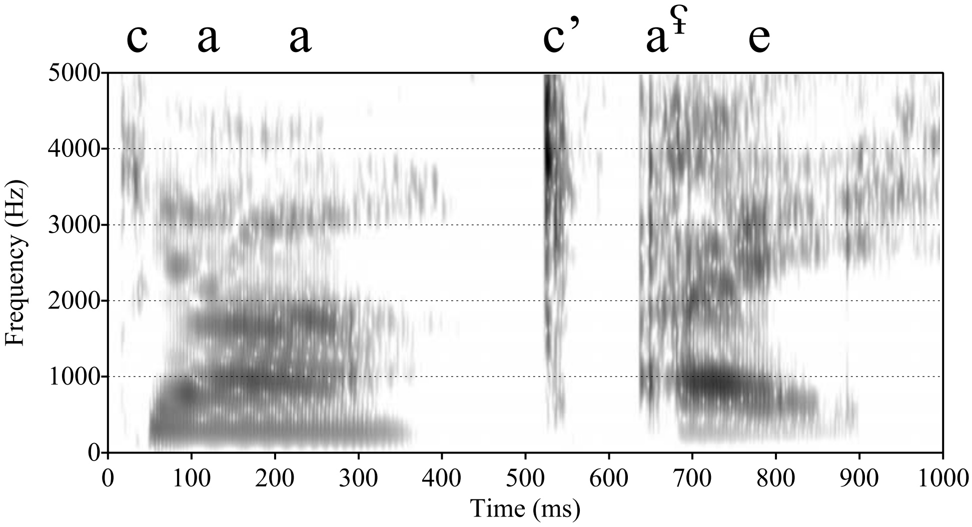

Figure 3.7 Spectrograms of *cáà* ‘lie (recline)’ and *c’áˤè* ‘be in pieces’ (speaker KE). Note the affrication typical of prepalatal stops.

Ejective affricate. In addition to the ejective stop, N|uu also has two *voiceless ejective affricates*: lamino-alveolar sibilant /t͡s’/ and velar-uvular /k͡χ’/ (3.10).

(3.10) *t͡s'áá* 'squeeze'
k͡χ'áà 'weep'

Both segments have a glottal lag similar to the one described for /c'/ above (roughly 50 ms; again, quantitative data are not yet available). While the realization of /t͡s'/ is more or less the same as that of /t͡s/ (except for the airstream mechanism), /k͡χ'/ is what I would call a *quasi-homorganic affricate*: In contrast to /c͡χ/, which is a true heterorganic affricate (as discussed above), /k͡χ'/ shows a slight articulatory accomodation between the stop portion (which is velar) and the fricative portion (which is uvular).

3.2 Click Consonants

Having discussed the non-click consonant inventory of N|uu, we now turn to the click consonant system, which is even more extensive. Many aspects of the click system have been covered by Miller et al. (2007, 2009), so this section will concentrate on giving a short overview as well as focus on a few points that merit further discussion.

To set the grounds for this and the following chapters, Figure 3.8 shows schematic sagittal sections of a dental click [|] and an alveolar click [!] at different stages of their production. Both click types display what is the defining feature of all clicks: First, an *anterior* and a *posterior closure* is formed in the mouth. Then, the portion of the tongue between the two closures is lowered (or, rather, retracted and lowered), thereby rarefying the air in the cavity between the two closures (called the *lingual cavity* or simply *click cavity*). Finally, the anterior closure is released, and there is a brief period of *ingressive airflow* due to the fact that the air in the lingual cavity has negative pressure relative to the atmospheric pressure. It is this ingressive airflow that gives rise to the characteristic 'clicking' sound of a click.

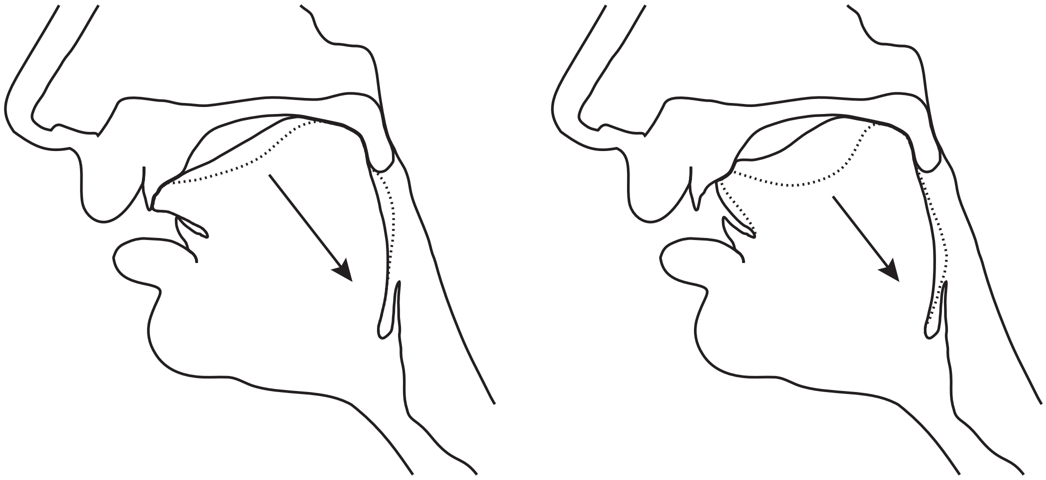

Figure 3.8 Schematic sagittal sections of the production of a dental click /|/ (left) and an alveolar click /!/ (right). Solid tongue contours correspond to the configuration at the beginning of the rarefaction process; dotted tongue contours correspond to the configuration just prior to release. Arrows indicate the direction of active tongue displacement.

These two click types were chosen because they represent two different classes of clicks, as will be discussed in detail later. The dental click [|] (which is typically laminal in N|uu) has a relatively *small* click cavity and a *shallow concave* shape of the tongue center at release. The alveolar click [!], on the other hand (which is typically apical in N|uu) has a *large* click cavity and a *deep concave* shape of the tongue center at release.

Table 3.2 presents the inventory of contrastive click consonants in N|uu. The arrangement is precisely the same as the one used in Table 3.1 above for the non-click consonants. Places of articulation are given in the columns; in the case of clicks, the place of articulation of the *posterior* constriction is not contrastive but predictable and changes dynamically from velar to uvular in the course of the production of each click. Because it is uvular at release, that label was chosen to represent the posterior place of articulation for expository purposes here. The dynamic nature of the posterior articulation should nevertheless be kept in mind.[11]

11 As shown by Miller et al. (2007, 2009), the posterior place of articulation (at release) does differ slightly but systematically between click types, though.

Table 3.2 The click consonant phonemes of N|uu. Dotted circles indicate presumed accidental gaps.

	Bilabial-Uvular	Dental-Uvular	Alveolar-Uvular	Lateral Alveolar-Uvular	Prepalatal-Uvular
Linguo-Pulmonic Stop	⊙ ◌ ᶢ⊙	ǀ ǀʰ ᶢǀ	ǃ ǃʰ ᶢǃ	ǁ ǁʰ ᶢǁ	ǂ ǂʰ ᶢǂ
Linguo-Pulmonic Sequential Stop	⊙͡q ◌ ◌	ǀ͡q ǀ͡qʰ ᶢǀ͡ɢ	ǃ͡q ǃ͡qʰ ◌	ǁ͡q ǁ͡qʰ ◌	ǂ͡q ǂ͡qʰ ◌
Linguo-Pulmonic Affricate	⊙͡χ	ǀ͡χ	ǃ͡χ	ǁ͡χ	ǂ͡χ
Linguo-Pulmonic Nasal	◌ ŋ̊⊙ˀ ŋ⊙	ŋ̊ǀʰ ŋ̊ǀˀ ŋǀ	ŋ̊ǃʰ ŋ̊ǃˀ ŋǃ	ŋ̊ǁʰ ŋ̊ǁˀ ŋǁ	ŋ̊ǂʰ ŋ̊ǂˀ ŋǂ
Linguo-Glottalic Sequential Stop	⊙͡q’	ǀ͡q’	ǃ͡q’	ǁ͡q’	ǂ͡q’
Linguo-Glottalic Affricate	◌	ǀ͡χ’	ǃ͡χ’	ǁ͡χ’	ǂ͡χ’

As was the case with the non-click consonants, there are some differences from the analysis by Miller et al. (2007, 2009) evident in Table 3.2, which will now be taken up.

Terminology for places and manners of articulation. Apart from the use of the term prepalatal, which has already been discussed in Section 3.1.1, the terminological framework used for the description of clicks in the present work departs slightly from the one introduced by Miller et al. (2007, 2009). The segment /ǀ͡q/, e.g., is called a *linguo-pulmonic sequential stop* here instead of a *linguo-pulmonic stop*. The motivation for this terminological change is given in Section 3.2.2.

Linguo-pulmonic affricates. The linguo-pulmonic affricates (i.e. /⊙͡χ/, /ǀ͡χ/, /ǃ͡χ/, /ǁ͡χ/, /ǂ͡χ/), listed in the *aspirated* column by Miller et al. (2007, 2009), are instead grouped as *voiceless (unaspirated)* here. The reason is that in N|uu, there are no phonological processes, constraints or the like that would speak in favor of classifying these segments as aspirated phonologically. Therefore, in the absence of such evidence, and since the segments do not have any phonetic aspiration in the sense of breathy phonation following the end of the fricative portion, they are classified as voiceless, along with all other affricates in N|uu.

Voiced bilabial click. The voiced bilabial click /ᶢ⊙/ was only recently discovered (in the word *ᶢʘóé* ‘dried food’) and is therefore included in the inventory here.

Linguo-glottalic sequential stops. Miller et al. (2007, 2009) assume only a single ejective series, the *linguo-glottalic ejective affricate* series (/͡|χʼ/, /͡!χʼ/, /͡‖χʼ/, /͡ǂχʼ/). In the present work, however, I argue that there is an additional, contrasting ejective series, the *linguo-glottalic ejective stop* series (/͡ʘqʼ/, /͡|qʼ/, /͡!qʼ/, /͡‖qʼ/, /͡ǂqʼ/).

Voiced linguo-pulmonic sequential stops. An additional series that was only recently discovered in the data is the *voiced linguo-pulmonic sequential stop* series. So far, only one member of this series has been discovered, the *dental(-uvular)* one, /ᶢ͡|ɢ/ (in the words *ᶢ͡|ɢànà* 'spleen' and *ᶢ͡|ɢùú* 'be constipated'). However, given the very strong tendency towards symmetry in the click system, it is reasonable to assume that the remaining, 'missing' members of the series are accidental gaps due to the limited size of the data set.

3.2.1 Click Types

As can be seen in Table 3.2, there are five *click types*, i.e. five ways in which the *anterior* closure of click consonants can be realized, in N|uu. They are, in traditional IPA terms, the *bilabial*, *dental*, *alveolar*, *lateral alveolar*, and *prepalatal* click types.[12]

As was explained above, the term *click type* as used in the present work refers to the entirety of the articulatory characteristics of the anterior closure of a click, including its formation and release (the latter is, of course, of particular importance because it is what makes a click so auditorily salient). Many aspects of the articulatory, acoustic, and auditory characteristics of the click types of N|uu have been described before by Miller, Brugman, and Sands (2007), Miller et al. (2007, 2009), and Sands, Brugman, Exter, Namaseb, and Miller (2007); those will shortly be reviewed below.

Bilabial. The bilabial click type (represented by the symbol 'ʘ') in N|uu is, expectedly, produced with a *bilabial* anterior closure.[13] Figure 3.9 shows photo-

12 The composite labels given in Table 3.2, like *bilabial-uvular* etc., refer to the places of articulation of the *entire click segment*. Since the posterior place of articulation is not contrastive, there is no need to specify it in the click type labels.

13 As mentioned before, the *posterior* closure in all click types in N|uu is assumed to be *velar* at the formation of the closure and *uvular* at the release of the closure. This is taken graphs of frontal and lateral views of typical realizations of /ʘ/ in the word *ʘáàχè* ‘daughter’ (thus, without any contextually induced lip rounding) just prior to the release of the bilabial closure. It is obvious that the labial gesture is quite similar to the equivalent gesture in the realization of a pulmonic segment like /p/; in particular, there is no appreciable protrusion of the lips.

Figure 3.9 Photographs (frontal view, left; lateral view, right) of /ʘ/ in *ʘáàχè* ‘daughter’ (speaker GS). The photographs are extracted from video films, each illustrating the frame immediately prior to the release of the bilabial closure.

The release of bilabial clicks in N|uu is auditorily quite weak (i.e. acoustically not very intense) and associated with a certain amount of noisiness. Two distinct types can be distinguished: Depending on the exact trajectory of the lower lip following the release of the closure (and maybe also influenced by individual anatomical differences, like differences in dentition), the release can be either strictly *bilabial* or *labiodental* (although the closure itself is always bilabial). These variants are neither contrastive nor (as far as I could ascertain) systematic, i.e., there do not seem to be any contextual conditions, consistent individual preferences, or the like. Figure 3.10 shows the waveform of a typical token of [ʘ] in *ʘáàχè* ‘daughter’.[14]

here to be an invariant feature of all clicks and will therefore not be specifically mentioned in this section any more.

14 Here and below, the different click *types* are illustrated by *waveforms* because they highlight the most salient acoustic properties of the *anterior* release, namely relative duration, rise time, and intensity.

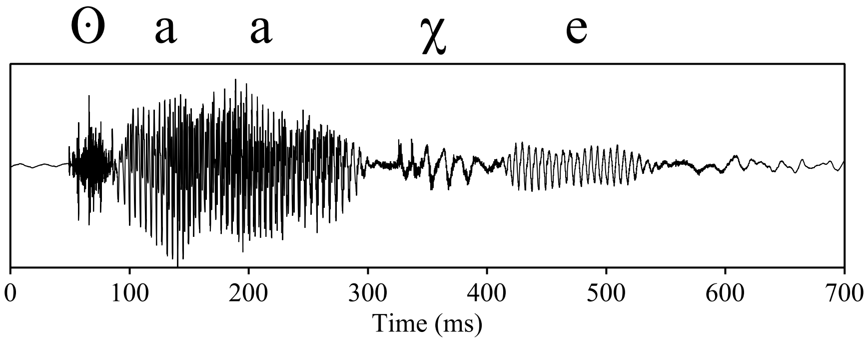

Figure 3.10 Waveform of /ʘ/ in *Θáàχè* 'daughter' (speaker KE)

Dental. The dental click type (represented by the symbol '|') in N|uu is realized with a *lamino-dental* anterior closure (see Sands, Brugman, Exter, Namaseb, & Miller, 2007 for palatographic and linguographic evidence). Like with bilabial clicks, the release of dental clicks in N|uu is noisy and relatively weak. Figure 3.11 shows a typical example of [|] in the word *|áàχè* 'female cousin'.

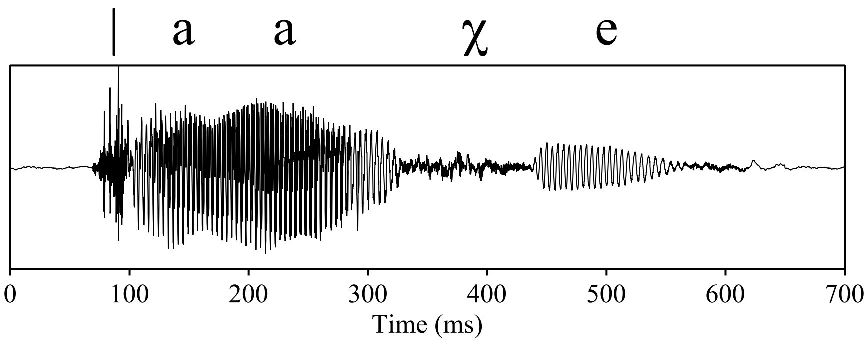

Figure 3.11 Waveform of /|/ in *|áàχè* 'female cousin' (speaker KE)

Alveolar. Clicks produced with the alveolar click type (represented by the symbol '!') in N|uu have an *apico-alveolar* anterior closure (cf. Sands, Brugman, Exter, Namaseb, & Miller, 2007). The release is not noisy but *abrupt* and very intense (auditorily salient). Figure 3.12 shows an example of [!] in the word *!ámà-sí* 'kidney' (*-sí* 'singulative').

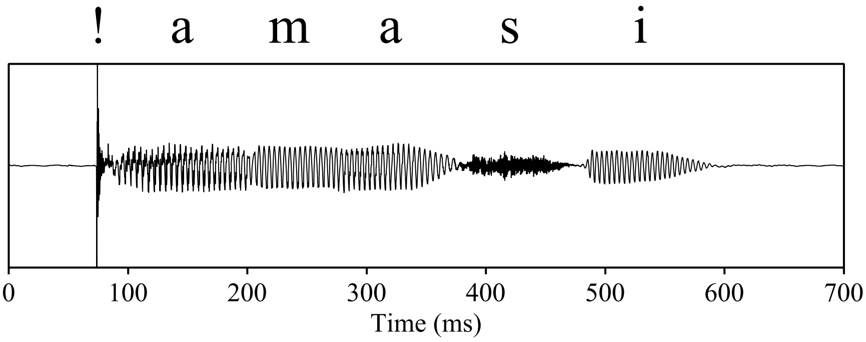

Figure 3.12 Waveform of /!/ in *!ámà-sí* 'kidney' (*-sí* 'singulative'; speaker KE)

Lateral alveolar. As explained above, the term *click type* refers to the articulatory characteristics of the anterior closure of a click; as such, it is a classificatory term that includes more information than just the anterior place of articulation. The lateral alveolar click type (represented by the symbol '‖') is a good example of this: The term *lateral* here refers to the fact that the *release* is not central but lateral. Lateral clicks are thus not 'lateral' segments in the narrow sense of the word; rather, they are laterally released stops.

As for their articulation, Sands, Brugman, Exter, Namaseb, and Miller (2007) have shown that in lateral clicks in N|uu there is less consistency than in clicks produced with other click types: While the place of articulation is quite consistently alveolar, the active articulator was found to vary between apical and laminal, usually depending on the speaker. Lateral alveolar clicks in N|uu can therefore be said to be either *apico-alveolar* or *lamino-alveolar*.[15]

The release of clicks with the lateral alveolar click type is noisy and intense (auditorily salient). In Figure 3.13, a representative waveform of [‖] in the word *‖áàχè* 'sister' is given.

15 But note that this variation has practically no consequences for the auditory quality of the clicks in question.

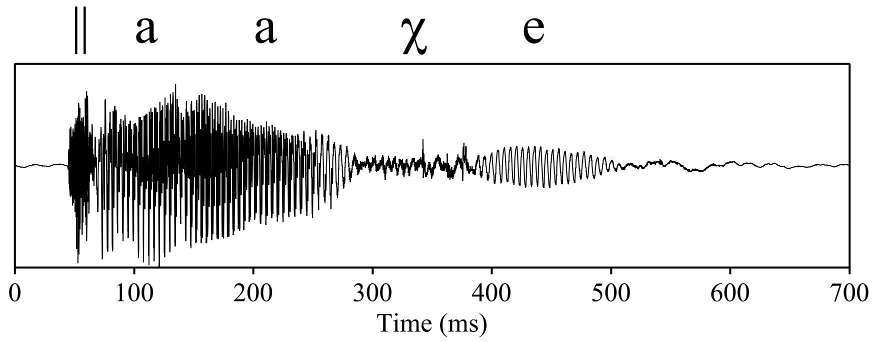

Figure 3.13 Waveform of /ǁ/ in *ǁáàχè* ‘sister’ (speaker KE)

Prepalatal. Clicks with the prepalatal click type (represented by the symbol ‘ǂ’) are produced with a *predorso-prepalatal* anterior constriction (commonly with some involvement of the postalveolar region as well), as was discussed in the context of the prepalatal place of articulation in *non-click* consonants in Section 3.1.1 (cf. Figure 3.1 for a linguogram of [ǂ], and Sands, Brugman, Exter, Namaseb, and Miller, 2007, for further information, including palatograms and linguograms).

The release of prepalatal clicks in N|uu is abrupt and quite intense, although not as intense as the release of alveolar and lateral alveolar clicks. Figure 3.14 shows the waveform of a token of [ǂ] in the word *ǂám̋ʔá* ‘hit’.

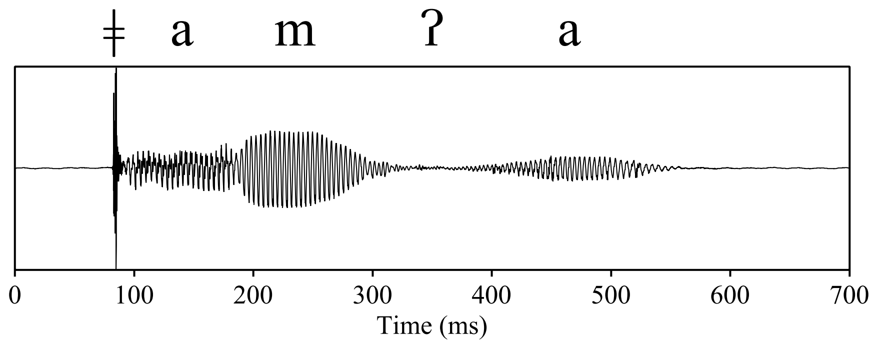

Figure 3.14 Waveform of /ǂ/ in *ǂám̋ʔá* ‘hit’ (speaker KE)

3.2.2 Click Series and Their Phonetic Realization

As explained above, clicks with any of the five click types can be produced in a number of ways. The entirety of phonetic characteristics of a click consonant *except* those associated with the click *type* itself have traditionally been called click *accompaniment*. Miller et al. (2007, 2009) have argued that this term is unfortunate in that it implies some coherent phonetic content while it is really only defined negatively, including such heterogeneous aspects as articulatory properties of the posterior closure and release, nasality, voicing etc.

In the present work, the validity of that criticism is in principle recognized. Yet, I think that for expository purposes, a term that captures the essence of the traditional term *accompaniment* is very helpful, in the informal sense of 'ways in which different click consonants can be formed with any given click type'. To capture that information without implying that it stands for any independent or coherent phonetic or phonological concept, click *series* is used here as a maximally neutral term. In that sense, click *type* and click *series* are complementary terms that together provide an exhaustive description of all aspects of the production of a click consonant.

As explained above, the posterior place of articulation of clicks in N|uu is velar at the formation of the closure and uvular at release (the latter with minor differences that depend on the particular click type). Since this is valid for all clicks, it will not be specifically mentioned any more in the description of click series that follows.

Linguo-pulmonic stop. In the linguo-pulmonic clicks in N|uu, the posterior closure is released just after the anterior closure. Auditorily, the two releases take place practically at the same time, which is the reason why I propose the term *simultaneously released* stop (or *simultaneous* stop for short). While the *anterior* closure is released with a *lingual* (i.e. 'velaric') airstream[16], the *posterior* closure is released in the manner of a *pulmonic stop*; however, because of the quasi-simultaneous nature of the two releases and because the anterior release is usually auditorily much more salient than the posterior one, the latter is typically inaudible even though it is technically a stop.

There are three series of (simultaneously released) linguo-pulmonic stops in N|uu, paralleling the non-click stop system: *voiceless unaspirated*, *voiceless as-*

16 Cf. Miller et al. (2007, 2009) for arguments in support of replacing the misleading term *velaric airstream* by *lingual airstream*. The term was originally introduced by Laver (1994, p. 180) for situations with an *incidental* ingressive airstream initiated by the tongue, as in labial-velar double articulations, thus reserving the term *velaric airstream* for clicks proper.

pirated, and *voiced*. First of all, the *voiceless unaspirated linguo-pulmonic stop* series contains the segments /ʘ/, /ǀ/, /!/, /ǁ/, and /ǂ/ (3.11).

(3.11) *ʘòˤná* 'horned adder (*Bitis caudalis*)'
ǀèré 'mane'
!áì 'run'
ǁáβé 'leopard (*Panthera pardus*)'
ǂàβú 'be dull'

Example waveforms of the linguo-pulmonic stop series for all five click types have been given in Section 3.2.1 above. Additionally, Figure 3.15 shows the spectrogram of [!] in the word *!áì* 'run'.[17]

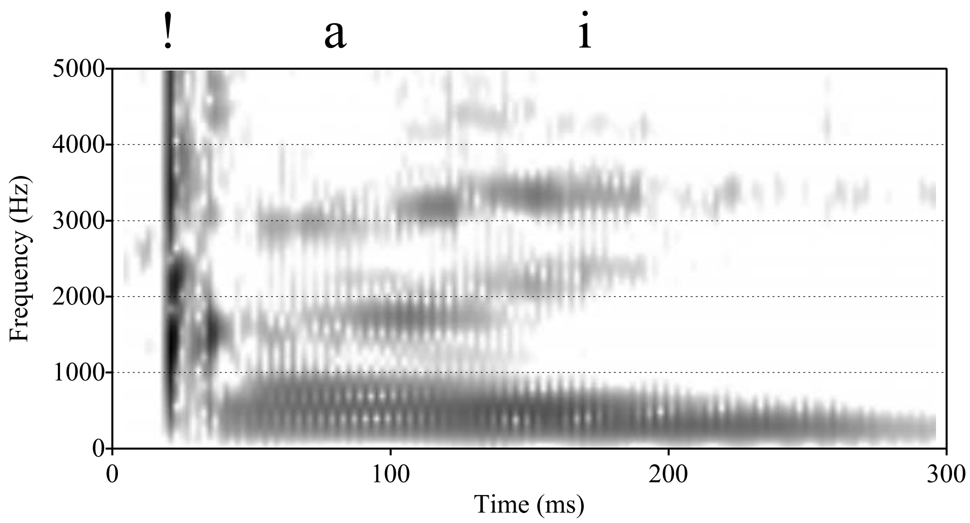

Figure 3.15 Spectrogram of /!/ in *!áì* 'run' (speaker KE)

Secondly, the *voiceless aspirated linguo-pulmonic stop* series contains the segments /ǀʰ/, /!ʰ/, /ǁʰ/, and /ǂʰ/ (3.12).

17 Here and below, the different click *series* are illustrated by *spectrograms* because they show the different *posterior* release properties of the segments more clearly. Wherever possible, clicks with the *alveolar* click type have been used for illustration, because the short duration and great intensity associated with it make it easy to distinguish between the anterior and posterior releases.

(3.12) *|ʰùìcé* ‘be quiet’
!ʰòò ‘be awake’
‖ʰàˤβà ‘whitequilled korhaan (*Eupodotis afroides*)’
ǂʰúì ‘brother-in-law’

This series differs from the voiceless *unaspirated* series in that there is a longer delay (VOT) between the *posterior* release and the onset of voicing for the following vowel. Figure 3.16 shows the spectrogram of a typical production of [|ʰ] in the word *!ʰòò* ‘be awake’.

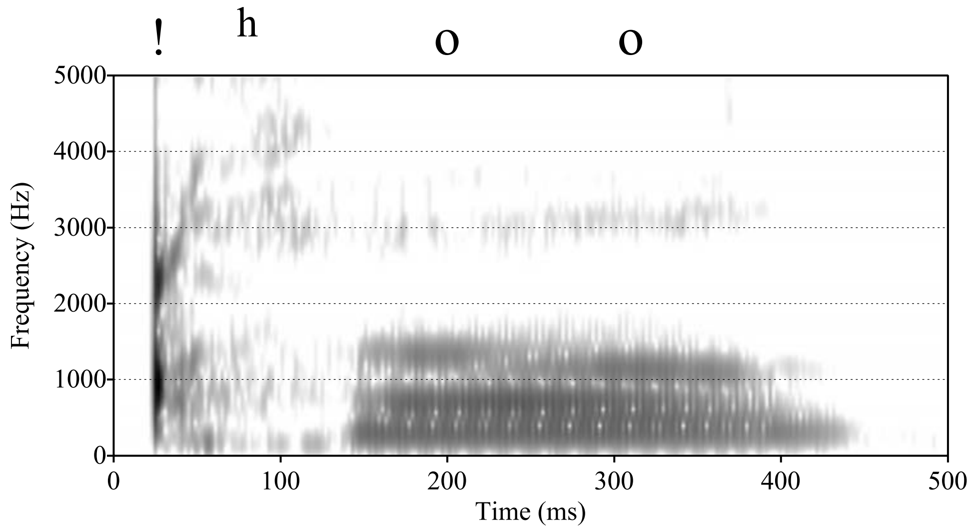

Figure 3.16 Spectrogram of /!ʰ/ in *!ʰòò* ‘be awake’ (speaker KE)

Thirdly, there is a *voiced linguo-pulmonic stop* series which contains the segments /ᵍʘ/, /ᵍ|/, /ᵍ!/, /ᵍ‖/, and /ᵍǂ/ (3.13).

(3.13) *ᵍʘóé* ‘dried food’
ᵍ|òá ‘bay (verb)’
ᵍ!àìχà ‘healer’
ᵍ‖óà ‘spoon’
ᵍǂàrù ‘sheep’

As has been shown to be the case with *pulmonic* stops (cf. Section 3.1.2), underlyingly voiced *linguo-pulmonic* stops are frequently devoiced following a prosodic boundary, with the likelihood of voicing decreasing with an increase in boundary strength. (A similar phenomenon of contextually dependent realization of voicing in clicks has been reported by Jessen 2002 for Xhosa, where voicing in underlyingly voiced clicks is phonetically present only in post-nasal position). Figure 3.17 shows the spectrogram of a production of [ᶢ!] in the word *ᶢ!àìχà* 'healer' as realized in the context *ná ká* __ 'I say __' (with voicing present).

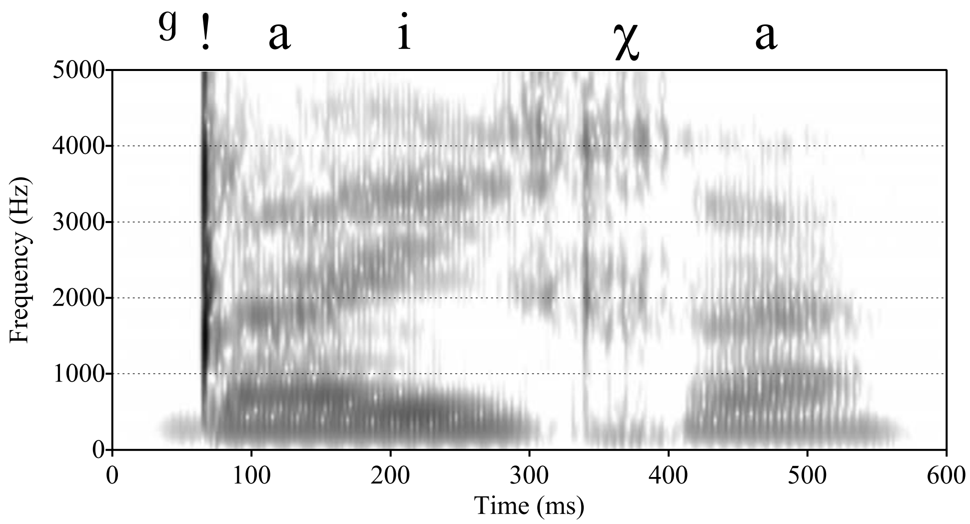

Figure 3.17 Spectrogram of /ᶢ!/ in *ᶢ!àìχà* 'healer', realized in the context *ná ká* __ 'I say __' (speaker KE)

As was mentioned before, one phoneme from the voiced linguo-pulmonic stop series, the bilabial /ᶢʘ/, was only recently discovered and is therefore not given by Miller et al. (2007, 2009). Figure 3.18 shows the spectrogram of a production of the word in which it was so far found to occur, *ᶢʘóé* 'dried food', again realized in the context *ná ká* __ 'I say __'.

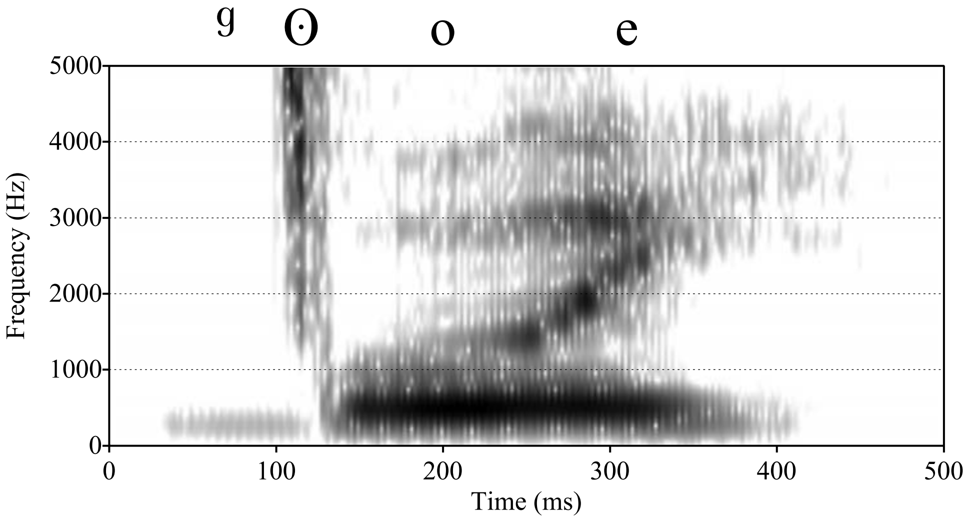

Figure 3.18 Spectrogram of /^{g}ʘ/ in *^{g}ʘóé* 'dried food', realized in the context *ná ká* __ 'I say __' (speaker KE)

Linguo-pulmonic sequential stop. Having illustrated the voiceless unaspirated, voiceless aspirated, and voiced linguo-pulmonic *simultaneously released* stop series, we now turn to the three linguo-pulmonic *sequentially released* stop series, or *sequential* stop series for short.

The first of those three series comprises the *voiceless unaspirated linguo-pulmonic sequential stops*, /ʘ͡q/, /|͡q/, /!͡q/, /‖͡q/, and /ǂ͡q/ (3.14).

(3.14) *ʘ͡qùì* 'sweat (verb)'
|͡qàá 'shine'
!͡qà^{n}á^{n} 'rope'
‖͡qóé 'salt pan'
ǂ͡qò^{n}è^{n} 'be short'

They differ from the corresponding linguo-pulmonic *simultaneously released* stops in that the posterior release occurs some time after the anterior release and is therefore clearly audible as a separate event (recall that in simultaneously released stops, the two releases occur so close together that they are not usually perceived as separate events; rather, only the anterior closure is audible due to its much greater intensity). Other than this difference in timing, the two classes of clicks are very similar. Figure 3.19 shows an example of a spectrogram of /!͡q/ in the word *!͡qà^{n}á^{n}* 'rope'.

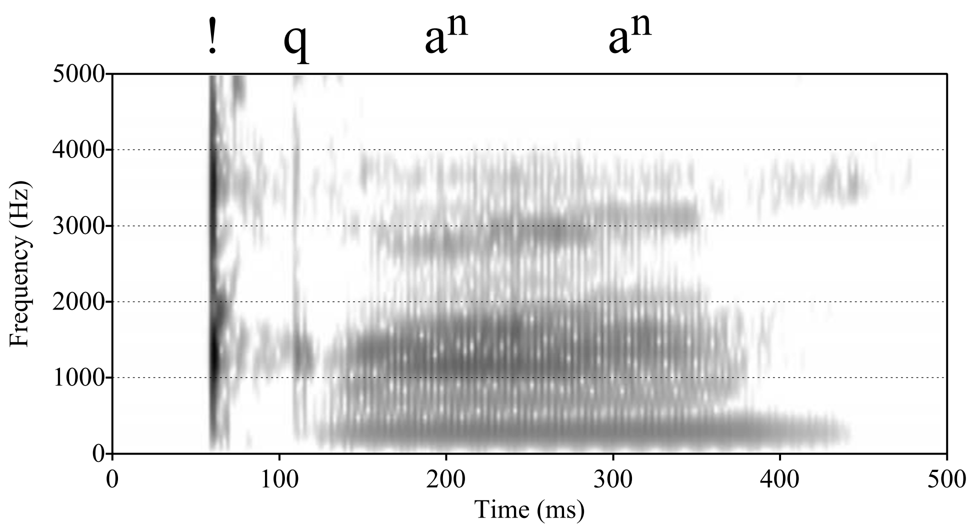

Figure 3.19 Spectrogram of /!͡q/ in *!͡qà*^{n}*á*^{n} ‘rope’ (speaker KE)

Secondly, the *voiceless aspirated linguo-pulmonic sequential stop* series comprises the segments /|͡q^{h}/, /!͡q^{h}/, /‖͡q^{h}/, and /ǂ͡q^{h}/ (3.15).

(3.15) *|͡q*^{h}*èè* ‘common duiker (*Sylvicapra grimmia*)’
!͡q^{h}*àà* ‘water’
‖͡q^{h}*àmà* ‘aardvark (*Orycteropus afer*)’
ǂ͡q^{h}*áà* ‘light’

As was the case with the *pulmonic* plosives and the linguo-pulmonic *simultaneously released* stops, the voiceless *aspirated* linguo-pulmonic sequential stops differ from the corresponding *unaspirated* segments in that they have a longer VOT (defined here as the time interval between the release of the *posterior* closure and the onset of the following vowel), the duration of which is approximately of the same order as that found in the other two classes of stops described before. Figure 3.20 shows a spectrogram of /!͡q^{h}/ in the word *!͡q*^{h}*àà* ‘water’.

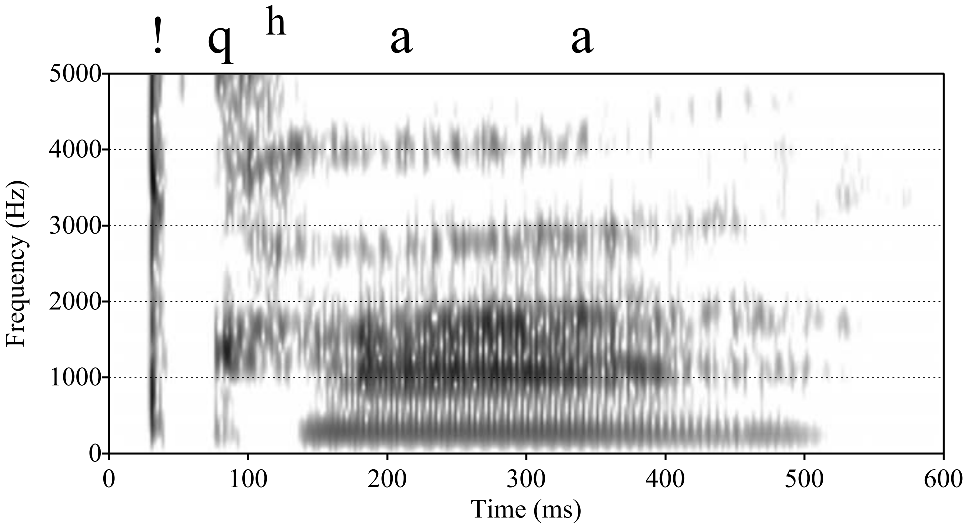

Figure 3.20 Spectrogram of /!͡q^{h}/ in *!͡q^{h}àà* ‘water’ (speaker KE)

Thirdly, there is also a *voiced linguo-pulmonic sequential stop* series in N|uu that was not included in the phoneme system by Miller et al. (2007, 2009) because it was only recently found in the data. So far, only the segment produced with the *dental* click type, /^{g}|͡ɢ/ (3.16), has been found (in the words *^{g}|͡ɢàná* ‘spleen’ and *^{g}|͡ɢùú* ‘be constipated’).

(3.16) *^{g}|͡ɢùú* ‘be constipated’

The symbol used here for the phonemic transcription, /^{g}|͡ɢ/, was chosen to reflect the facts that the segment is underlyingly both *voiced* (symbolized by ‘^{g}’, paralleling e.g. /^{g}|/) and *sequentially released* (the release being voiced; symbolized by ‘ɢ’, paralleling e.g. /|͡q/).It should be noted, however, that *phonetically*, voicing is consistently present only during the *posterior release* and not, as e.g. in /^{g}|/, during the closure phase as well. While this is at present only a qualitative observation made on the basis of very limited data, one can speculate that it is a strategy to overcome the problem of maintaining voicing in a uvular stop for an extended period of time (due to the small volume of the cavity between the glottis and the uvular closure, subglottal and supraglottal pressure values are usually equalized very quickly in uvular stops, making it difficult to maintain a prolonged airflow through the glottis for voicing). As reported by Traill (1985), !Xóõ uses a different strategy to reach the same goal: There, underlyingly voiced sequentially released stops are phonetically *prenasalized* (i.e., there is voiced nasal airflow during the closure phase, thus e.g. [ᵑǀ͡ɢ] in the transcription system used here); thereby the time interval needed to produce an audible *voiced* uvular burst is greatly reduced. One could therefore argue that there is a (phonetic) *voicing contour* (a change in voicing from voiceless to voiced) within the segments of this click series in N|uu.

A second characteristic of the segment /ᵑǀ͡ɢ/ in N|uu is worth mentioning (again, keeping in mind the rather limited data the observation is based on): Unlike the other click series, the phonetic realization of /ᵑǀ͡ɢ/ seems to show quite large inter-speaker variation as far as the posterior release is concerned: While some speakers produce the segment as [ǀ͡ɢ], with a *voiced uvular stop* posterior release, others produce it as [ǀ͡g] (i.e., the posterior place of articulation seems to be further to the front than in the corresponding voiceless segment, resembling a *voiced velar stop*) and others as [ǀ͡ʁ], where the posterior articulation resembles a *voiced uvular affricate* (as noted above, lack of voicing during the *closure* phase seems to be a consistent feature of the surface realization of /ᵑǀ͡ɢ/). What all these realizations have in common, though, is that (1) *voicing* is present during the realization of the segment (more specifically, during the posterior release), and that (2) the anterior and posterior releases occur *sequentially*. Therefore, systematically, it is justified to classify them as *voiced linguo-pulmonic sequential stops*.[18] Figure 3.21 shows a spectrogram of /ᵑǀ͡ɢ/ in the word *ᵑǀ͡ɢùú* 'be constipated', realized as [ǀ͡ɢ], [ǀ͡g], and [ǀ͡ʁ].

18 To be precise, this is not quite true of the realization [ǀ͡ʁ], which in the framework proposed in the present work would be called a *voiced linguo-pulmonic (simultaneously released) affricate* (i.e., the voiced equivalent of [ǀ͡χ], except that in that case, voicing would be expected to be present throughout the closure phase as well, at least in 'favorable' phonological contexts). However, considering the other, stop-like, realizations, it seems more advantageous in this case to explain [ǀ͡ʁ] as a case of lenition of the posterior articulation.

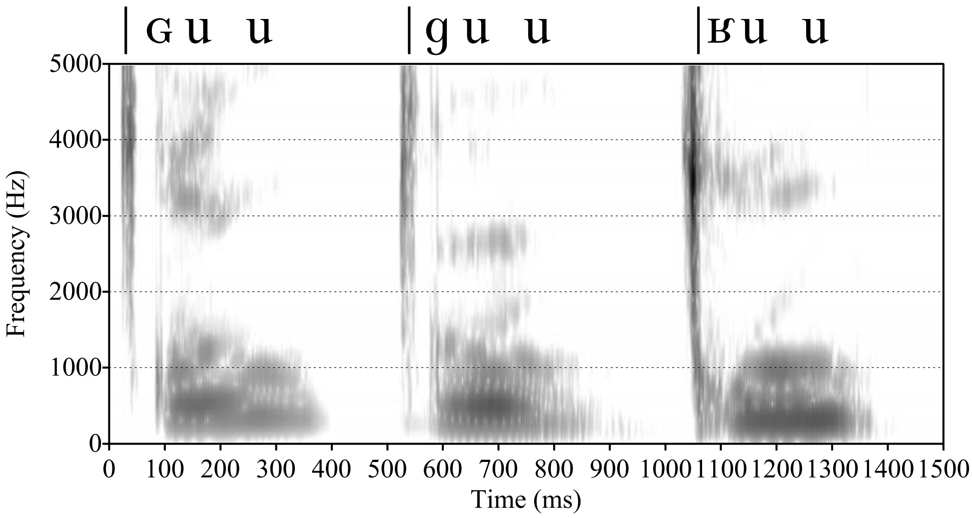

Figure 3.21 Spectrogram of /ᶢǀ͡ɢ/ in *ᶢǀ͡ɢùú* ‘be constipated’, realized as [ǀ͡ɢ] (speaker GS), [ǀ͡ɡ] (speaker HK), and [ǀ͡ʁ] (speaker KE)

Linguo-pulmonic affricate. The *voiceless linguo-pulmonic (simultaneously released) affricate* series comprises the segments /ʘ͡χ/, /ǀ͡χ/, /ǃ͡χ/, /ǁ͡χ/, and /ǂ͡χ/ (3.17).

(3.17) *ʘ͡χúú* ‘rub in’
ǀ͡χùɾú ‘juice’
ǃ͡χàɾú ‘snore’
ǁ͡χà^{n}*á*^{n} ‘sack’
ǂ͡χóá ‘speak’

This series is akin to the linguo-pulmonic (simultaneously released) *stop* series, from which it differs in that the posterior closure is released not in a stop-like but in an *affricate*-like manner; i.e., the posterior closure is not released abruptly but *gradually*. Keeping in mind that the posterior place of articulation is velar at the beginning of the click articulation but uvular at the end, the frication noise that is created by the gradual release of the posterior closure is therefore *uvular*.

It should be kept in mind, however, that this fricated interval is not a *fricative* but the fricated portion of an *affricate*-like articulation, the stop portion of which is the posterior articulation of the click. The *burst* of this affricate-like articulation occurs immediately after the anterior release (i.e., they occur *quasi-simultaneously*) and is therefore usually inaudible, the salient anterior burst masking the posterior one auditorily. Clicks of this series are thus called *simultaneously re-*

leased affricates here, differing from simultaneously released *stops* only in the manner of the posterior release, not in the timing between the anterior and posterior releases. Figure 3.22 shows a spectrogram of /!͡χ/ in the word *!͡χàɾú* ‘snore’.

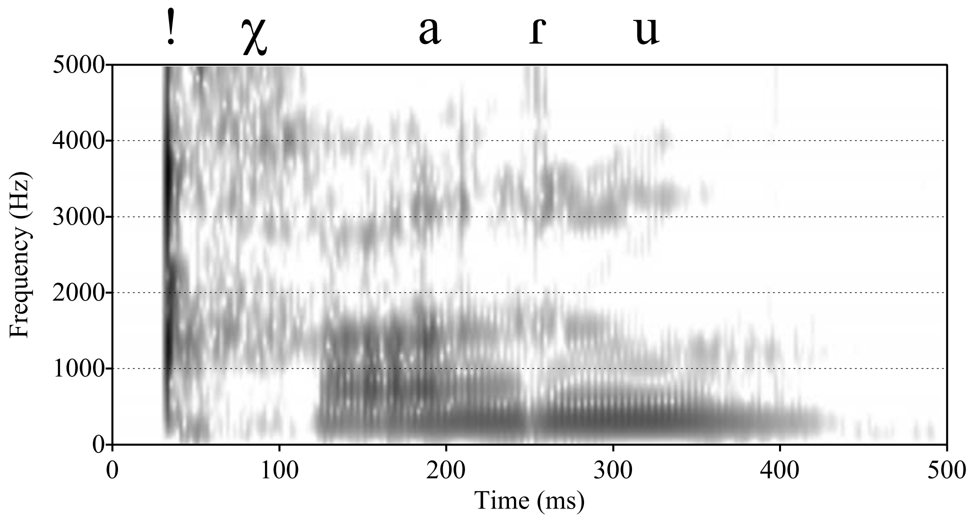

Figure 3.22 Spectrogram of /!͡χ/ in *!͡χàɾú* ‘snore’ (speaker KE)

Linguo-pulmonic nasal. All of the click series discussed so far involved a complete closure at the velum, so that as long as the interval when the anterior or the posterior click closure (or both) are in place, there is no pulmonic egressive airflow. In contrast, in the three series that are subsumed under the term *linguo-pulmonic nasal*, the velum is held in a lowered position so that there is pulmonic egressive nasal airflow throughout the production of the segments. At the moment of release of the anterior closure, there is also lingual ingressive oral airflow. In other words, the two directions of airflow (lingual ingressive and pulmonic egressive) occur simultaneously, or more precisely, a short (but intense) pulse of lingual ingressive oral airflow is *superimposed* on a continuous stream of pulmonic egressive nasal airflow.

The term *linguo-pulmonic nasal* should not be taken to imply that these sounds behave phonologically like nasals (it will be seen in Section 4.2.1 that they pattern with the other click consonants with respect to their phonotactic distribution). It is a purely descriptive articulatory label that expresses the fact that there is continuous nasal airflow throughout the production.

The first of these three series is the *voiceless aspirated linguo-pulmonic (simultaneously released) nasal* series, which comprises the segments /ŋ̊|ʰ/, /ŋ̊!ʰ/, /ŋ̊‖ʰ/, and /ŋ̊ǂʰ/ (3.18).

(3.18) *ŋ̊|ʰùú* 'be cool'
ŋ̊!ʰáì 'bone pipe'
ŋ̊‖ʰàní 'carrying kaross'
ŋ̊ǂʰàˤá 'push'

These clicks are characterized by a voiceless nasal airstream (which is usually inaudible) which, following the release of the posterior closure, is gradually replaced by a voiceless oral airstream (aspiration). Auditorily, this gives rise to a 'crescendo' effect of the aspiration noise, which has led earlier researchers like Snyman (1978) to call these clicks 'delayed aspirated'. Traill (1991, 1992), however, demonstrated by means of aerodynamic investigations that there is indeed a 'trade-off', as it were, between a decreasing (but inaudible) nasal airflow and an increasing (and audible) oral airflow.

For the present work, no aerodynamic investigations could be made, but the pulmonic egressive nasal airflow could be verified clearly by means of a mirror that was held under the nostrils during the production of these sounds. The nasality of the sounds is also apparent in the fact that the interval before the release of the anterior closure usually assimilates in voicing to a preceding vowel, so that the surface realization most often includes a *voiced* nasal pulmonic egressive airstream. The anterior release and the following aspiration, however, are always voiceless, so that in a sequence like *ná ká* __ 'I say __', *ŋ̊!ʰ* would normally be produced as [ᵑ!ʰ], not [ŋ̊!ʰ].

Figure 3.23 shows a spectrogram of /ŋ̊!ʰ/ in the word *ŋ̊!ʰáì* 'bone pipe'.

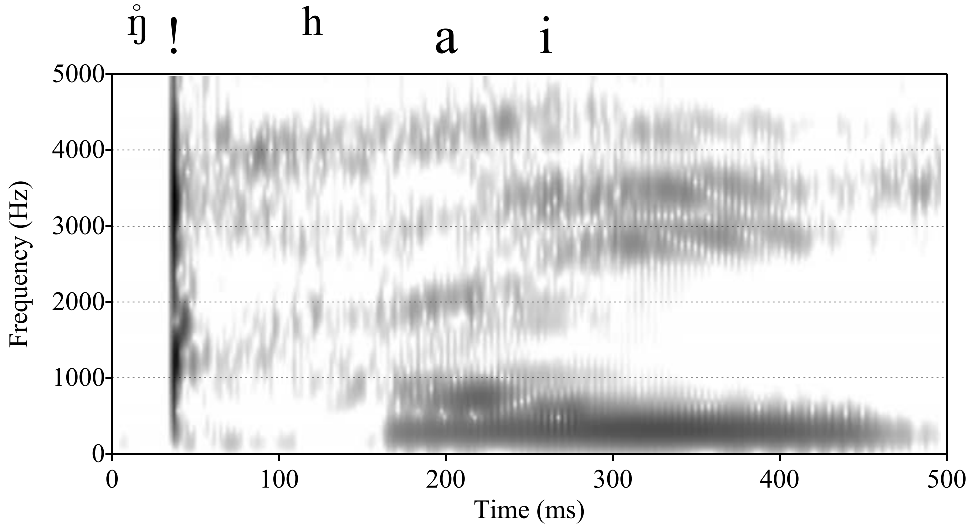

Figure 3.23 Spectrogram of /ŋ̊!ʰ/ in *ŋ̊!ʰáì* 'bone pipe' (speaker KE)

The difference between 'regular aspiration' and 'delayed aspiration' in the words *!ʰàí* 'tail' and *ŋ̊!ʰáì* 'bone pipe' can be seen in Figure 3.24.

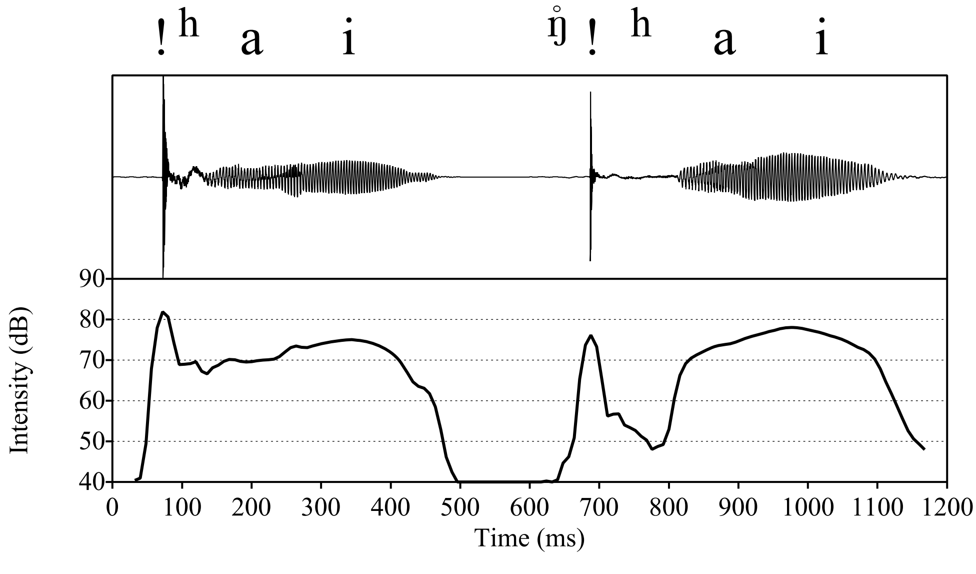

Figure 3.24 Waveforms and intensity tracks of /!ʰ/ in *!ʰàí* 'tail' (left) and /ŋ̊!ʰ/ in *ŋ̊!ʰáì* 'bone pipe' (right; speaker KE)

The second linguo-pulmonic nasal series is the *voiceless glottalized linguo-pulmonic (simultaneously released) nasal* series, comprising the segments /ŋ̊⊙ˀ/, /ŋ̊|ˀ/, /ŋ̊!ˀ/, /ŋ̊‖ˀ/, and /ŋ̊ǂˀ/ (3.19).

(3.19) *ᵑ̊ʘˀúíʔí* ‘be ill’
ᵑ̊|ˀàⁿàⁿ ‘be long’
ᵑ̊!ˀúβí ‘small-spotted genet (*Genetta genetta*)’
ᵑ̊‖ˀáú ‘dig’
ᵑ̊ǂˀáú ‘be narrow’

This series is characterized by a complete closure of the glottis which occurs after the formation of the anterior and posterior closures and is also released after their release. This results in the audible sequence of events *nasal airflow – glottal closure – anterior release – glottal release*. As with the corresponding aspriated (‘delayed aspirated’) segments, the nasal airflow at the onset is only audible in a voiced (vocalic) preceding context. Figure 3.25 shows a spectrogram of /ᵑ̊!ˀ/ in the word *ᵑ̊!ˀúβí* ‘small-spotted genet (*Genetta genetta*)’.

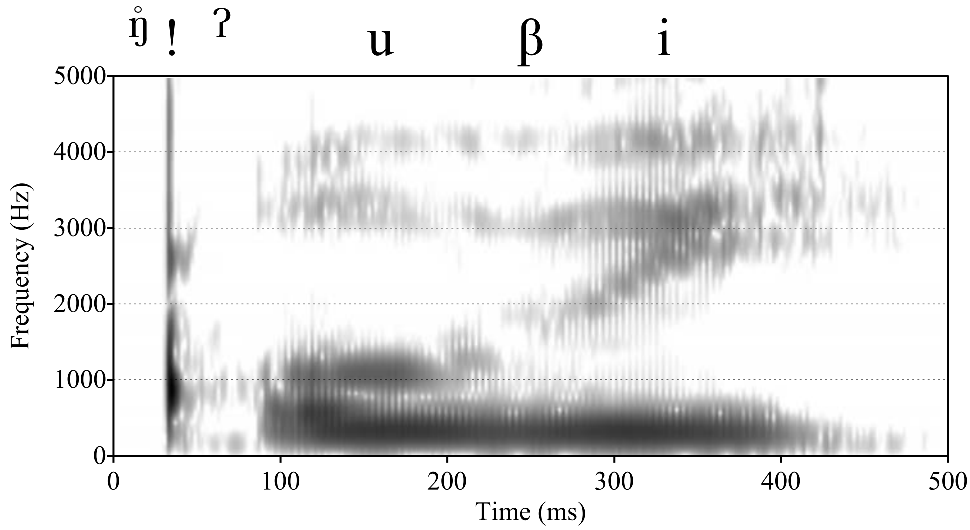

Figure 3.25 Spectrogram of /ᵑ̊!ˀ/ in *ᵑ̊!ˀúβí* ‘small-spotted genet (*Genetta genetta*)’ (speaker KE)

The third, and final, linguo-pulmonic nasal series is the *voiced linguo-pulmonic (simultaneously released) nasal* series, which contains the segments /ᵑʘ/, /ᵑ|/, /ᵑ!/, /ᵑ‖/, and /ᵑǂ/ (3.20).

(3.20) *^{ŋ}ʘà^{n}à^{n}* 'African wild cat (*Felis lybica*)'
^{ŋ}ǀáù 'sweep'
^{ŋ}!ò^{n}ó^{n} 'dune'
^{ŋ}ǁàá 'stay'
^{ŋ}ǂàˤà 'kick'

This series is similar to the two aforementioned ones in that it is characterized by the superimposition of a lingual ingressive airpulse on a nasal pulmonic egressive airstream, but in contrast to the other two series, voicing is maintained throughout the producion of the segment. In Figure 3.26, a spectrogram of /^{ŋ}!/ in the word *^{ŋ}!ò^{n}ó^{n}* 'dune' is shown.

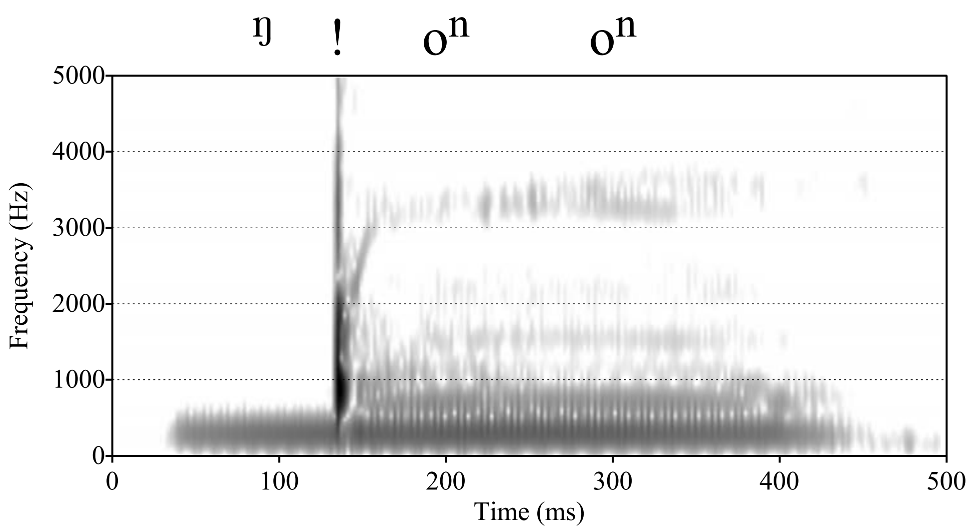

Figure 3.26 Spectrogram of /^{ŋ}!/ in *^{ŋ}!ò^{n}ó^{n}* 'dune' (speaker KE)

Linguo-glottalic sequential stop. The final category of click consonants in the phoneme system of N|uu that remains to be discussed is the category of clicks involving (alongside the lingual airstream mechanism) *glottalic* airflow. Glottalic airflow in N|uu, be it in clicks or non-click consonants, is always *egressive*, thus involving an *ejective* articulation. In clicks, since the anterior closure is by definition released with a lingual (ingressive) airstream, ejection is confined to the release of the *posterior* closure. That is, instead of releasing the posterior closure on a pulmonic egressive airstream as in the examples discussed so far, the volume of air between the closed glottis and the posterior closure is compressed by raising the larynx. Then, after the anterior release, the posterior closure is released, and the compressed air in the pharyngeal cavity results in a short (but intense) egressive air pulse.

Two series of clicks in N|uu employ this mechanism. The first one is the *voiceless linguo-glottalic sequential stop* series, comprising the segments /ʘ͡q'/, /|͡q'/, /!͡q'/, /‖͡q'/, and /ǂ͡q'/ (3.21).

(3.21) *ʘ͡q'úβúkà-sí* 'stink bug (*Pentatomidae* spp.)' (*-sí* 'singulative')
|͡q'ùríχà 'be dirty'
!͡q'òˤò 'stretch'
‖͡q'àˤàʔí 'be bitter'
ǂ͡q'ání 'twist off'

In this series, the two releases are audibly realized *in sequence*, just like in the corresponding linguo-pulmonic series; consequently, both are called *sequential stops* here. The sounds are voiceless throughout. Figure 3.27 shows a spectrogram of /!͡q'/ in the word *!͡q'òˤò* 'stretch'.

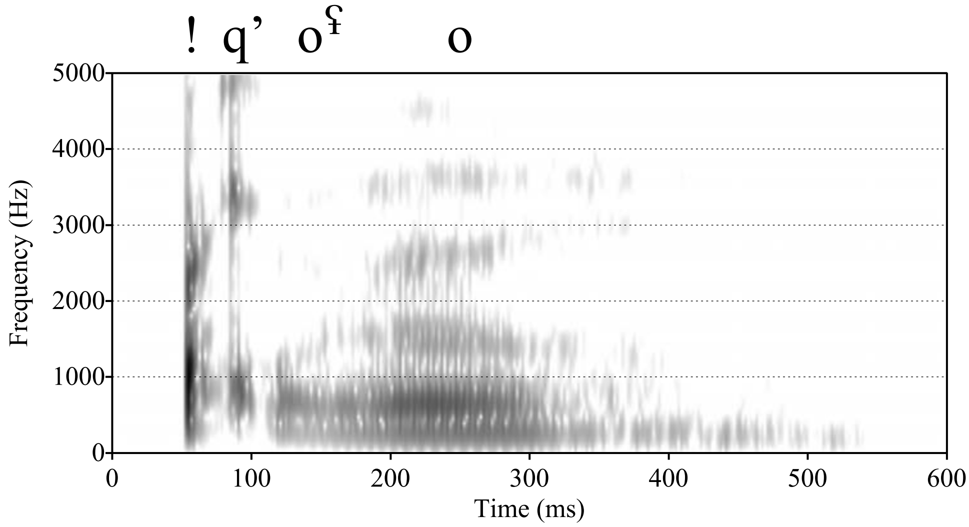

Figure 3.27 Spectrogram of /!͡q'/ in *!͡q'òˤò* 'stretch' (speaker KE)

Linguo-glottalic affricate. The second series involving ejection in N|uu is the *voiceless linguo-glottalic (simultaneously released) affricate* series, which contains the segments /|͡χ'/, /!͡χ'/, /‖͡χ'/, and /ǂ͡χ'/ (3.22).

(3.22) *ǀ͡χ’á* ‘arm’
!͡χ’ùú ‘foot’
ǁ͡χ’àà ‘scold’
ǂ͡χ’ùákà ‘be rotten’

Like the preceding (*voiceless linguo-glottalic sequential stop*) series, the posterior closure is released on a glottalic airstream. In contrast, however, the posterior release occurs immediately after the anterior release (i.e., ‘quasi-simultaneously’), and is characterized by affrication. It can therefore be said to be the glottalic counterpart of the *voiceless linguo-pulmonic (simultaneously released) affricate* series described above. Figure 3.28 shows a spectrogram of /!͡χ’/ in the word *!͡χ’ùú* ‘foot’.

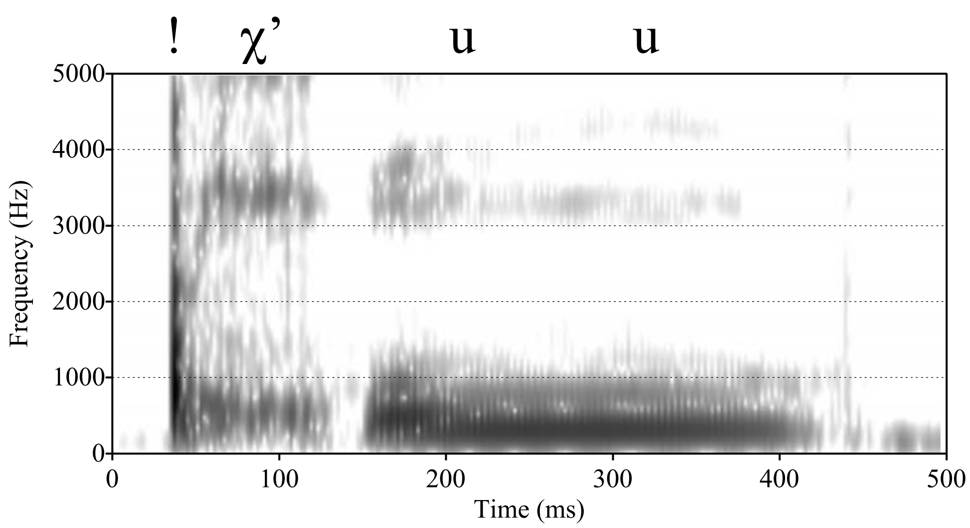

Figure 3.28 Spectrogram of /!͡χ’/ in *!͡χ’ùú* ‘foot’ (speaker KE)

Note that Miller et al. (2007, 2009) did not distinguish between /!͡q’/ etc. on the one hand and /!͡χ’/ etc. on the other hand. Comparison of Figure 3.27 with Figure 3.28, however, shows the two distinguishing characteristics, namely, (1) one vs. two (audible) releases, and (2) presence vs. absence of affrication at the posterior release.

3.3 Vowels and Diphthongs

Having described the consonantal segment inventory of N|uu, we will now turn to the vowel system. N|uu has a comparatively large inventory of vocalic segments, which is not primarily due to a large number of 'basic' vowels (in terms of height, backness, or roundedness), but rather to the presence of a nasality contrast and a phonation contrast and the resulting combinations.

3.3.1 Descriptive Dimensions

Vowels in N|uu contrast along five dimensions: Firstly, a system of 5 'basic' vowels can be identified that contrast along the dimensions of height (high vs. mid vs. low), backness (front vs. central vs. back) and roundedness (unrounded vs. rounded). Secondly, all vowels are contrastively non-nasalized ('plain') vs. nasalized. And thirdly, all vowels *except high front vowels* can be non-epiglottalized ('plain') vs. epiglottalized. This results in a system of 17 contrastive vowels in total (one combination, the nasalized epiglottalized mid front vowel */$e^{ʕn}$/, is not attested in the data). For convenience, Figure 3.29 schematically shows the non-epiglottalized vowel system.

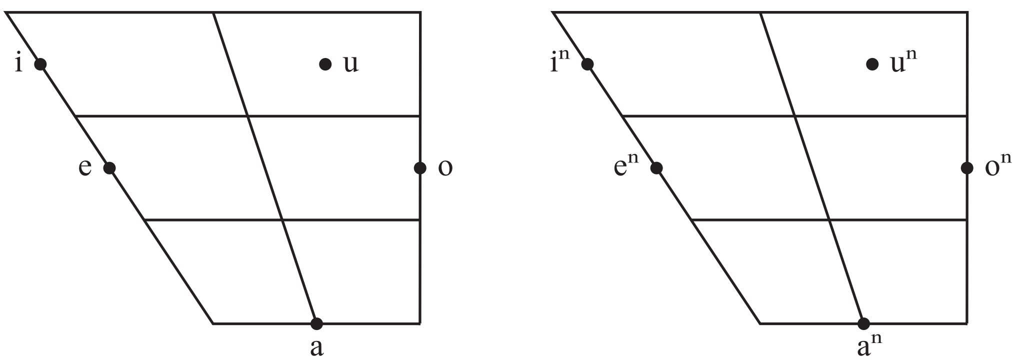

Figure 3.29 Schematic diagrams of the 'plain' (left) and nasalized (right) non-epiglottalized vowel inventories of N|uu

The epiglottalized vowel system of N|uu is summarized in Figure 3.30. As was the case with the consonant system, there are certain differences from the vowel inventory as presented by Miller et al. (2007, 2009), which will shortly be discussed.

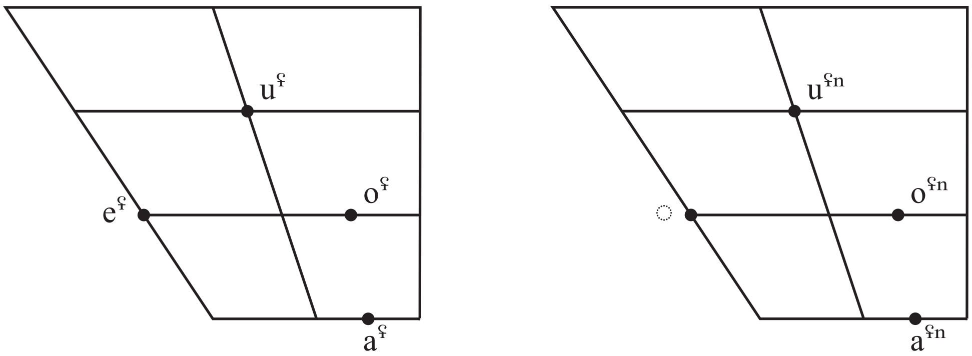

Figure 3.30 Schematic diagrams of the ‘plain’ (left) and nasalized (right) epiglottalized vowel inventories of N|uu. The dotted circle indicates a presumed accidental gap.

Non-low nasalized vowels. In the present work, it is argued on the basis of the analysis of the data available to me that *pace* Miller et al. (2007, 2009), there is in fact a phonological contrast between /i^{n}/ and /e^{n}/ as well as between /u^{n}/ and /o^{n}/ and between /$u^{ʕn}$/ and /$o^{ʕn}$/; that is, I assume that there is a height distinction among non-low nasalized vowels in N|uu.

Nasalized epiglottalized high back rounded vowel. The vowel /$u^{ʕn}$/ wa only recently discovered in the data (in the word *^{ŋ}ǂù^{ʕn}à^{n}* ‘common duiker calf, steenbok calf’).

Diphthongs. Phonetic diphthongs are treated here phonologically as underlying sequences of individual vowels. Their phonotactic structure and distribution can be explained by more general principles that are independent of whether or not the vowels are adjacent or not; therefore, diphthongs will be treated entirely in Section 4.3 and not listed in the present chapter.

3.3.2 Phonetic Realization of Vowels

We will now turn to a description of the phonetic realization of the individual underlying vowels, organized in terms of the four vowel categories already familiar from Figure 3.29 and Figure 3.30.

‘Plain’ vowels. The ‘plain’ (i.e. non-nasalized, non-epiglottalized) vowel system of N|uu contains the segments /i/, /e/, /a/, /o/, and /u/ (3.23).

(3.23) *ᵍ|ìí* ‘candle thorn (*Acacia hebeclada*)’
|͡q’éé ‘tell’
sáá ‘eland (*Taurotragus oryx*)’
ᵑ̊!ˀóò ‘axe’
‖ʰúù ‘be excited’

These are realized in their canonical form (i.e. in unmarked contexts) as a *lowered high front vowel* [i̞], *lowered high-mid front vowel* [e̞], *low central vowel* [ä], *lowered rounded high-mid back vowel* [o̞], and *mid-centralized rounded high back vowel* [ŭ], respectively. For reasons of convenience and in order not to overload the transcription with diacritics, however, the phonetic realizations are transcribed in the present work as [i], [e], [a], [o], and [u].

Nasalized vowels. The nasalized vowel system consists of the segments /i^{n}/, /e^{n}/, /a^{n}/, /o^{n}/, and /u^{n}/ (3.24).

(3.24) *cʰì*^{n}*ì*^{n} ‘leg’
kà|é^{n} ‘name’
k͡χ’á^{n} ‘drink’
ɟó^{n} ‘skin’
!͡qú^{n}*ú*^{n} ‘flour’

The phonetic realization is by and large the same as that of the ‘plain’ vowels (but of course with accompanying nasalization), without very large differences in tongue position or lip rounding (although a quantitative investigation of this has not yet been undertaken). That is, the segments are realized phonetically as [i̞^{n}], [e̞^{n}], [ä^{n}], [o̞^{n}], and [ŭ^{n}] but transcribed phonetically here as [i^{n}], [e^{n}], [a^{n}], [o^{n}], and [u^{n}].

Epiglottalized vowels. The epiglottalized vowel system consists of the segments /eˤ/, /aˤ/, /oˤ/, and /uˤ/ (3.25).

(3.25) *zéˤé* ‘fly (verb)’
ᵑǂàˤà ‘kick’
kóˤó ‘swell’
|ʰùˤá ‘arrow poison’

The phonetic realization in terms of tongue position, however, is radically different from that of the corresponding ‘plain’ vowels: The segments are realized, in turn, as an *epiglottalized low-mid front vowel* [ɛˤ], *epiglottalized fronted low back vowel* [ɑ̟ˤ], *epiglottalized fronted rounded low-mid back vowel* [ɔ̟ˤ], and *epiglottalized rounded high-mid central vowel* [ɵˤ]. That means that in comparison with the ‘plain’ vowels, the corresponding epiglottalized vowels are ‘modified’ in the following ways: /e/ is lowered, /a/ is retracted, and /o/ and /u/ are both lowered and fronted.

The term *epiglottalization* requires some elaboration. In his work on !Xóõ, Traill (1986) presented instrumental (e.g. cineradiographic) evidence that in certain (phonologically contrastive) vocalic segments in that language, a certain ‘harsh’ sound quality was achieved by an extreme contraction of the epilaryngeal structures in the region of the epiglottis. Later on, John Esling and colleagues (Esling, 1996, 1999, 2005; Edmondson, Ziwo, Esling, Harris, & Li, 2001; Esling & Edmondson, 2002; Esling & Harris, 2005; Edmondson & Esling, 2006) initiated a research program to investigate the phenomenon, which was found to occur in more languages than previously thought. By using laryngoscopic methods, they found that the main articulatory gesture in the formation of what they call *harsh voice* is an activation of the so-called *aryepiglottic sphincter*, which results in an approximation of the arytenoid cartilages and the epiglottis.

No instrumental investigation of the articulation of these sounds has yet been undertaken in N|uu, but the close proximity to !Xóõ and the auditory similarity between the respective sounds in the two languages leads me to hypothesize that what are called *epiglottalized vowels* here are indeed instances of Esling’s *harsh voice*. The auditory quality of this phonation type in N|uu is somewhat speaker-dependent: For some speakers, it is characterized by a strongly pharyngealized quality accompanied by a very characteristic low-frequency periodic vibration (trilling), presumably of the aryepiglottic folds. For other speakers, the periodic, ‘growling’ quality is absent, the sounds sounding essentially pharyngealized. Again, for reasons of convenience, they are transcribed phonetically as [ɛˤ], [ɑˤ], [ɔˤ], and [ɵˤ] in the present work. Figure 3.31 shows a spectrogram of /aˤ/ in the word *ᵑǂàˤà* ‘kick’.

Nasalized epiglottalized vowels. Finally, the nasalized epiglottalized vowel system of N|uu consists of the segments /a^{ˤn}/, /o^{ˤn}/, and /u^{ˤn}/ (3.26).

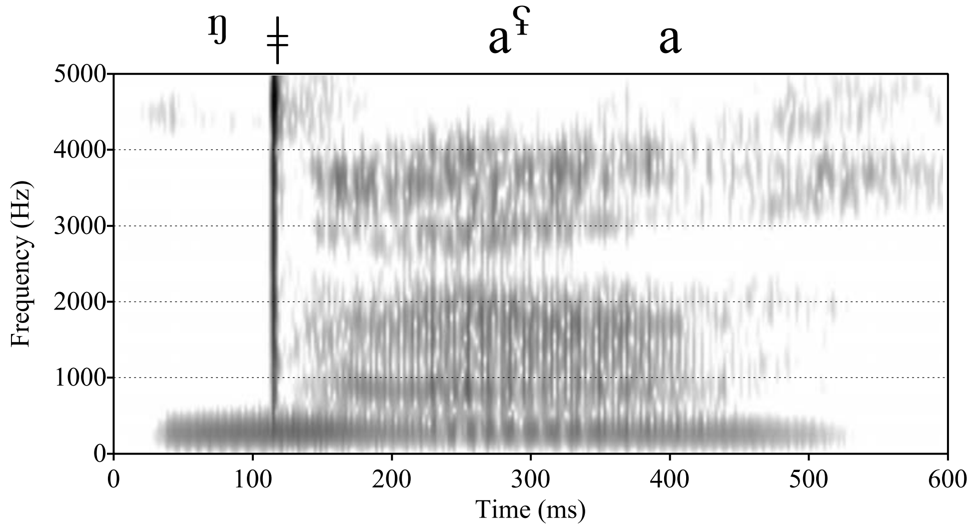

Figure 3.31 Spectrogram of /aˤ/ in *ᵑǂàˤà* 'kick' (speaker KE). Note the quasi-periodic vibration (trilling) towards the middle of the vocalic interval.

(3.26) *kàˤ*^{n}*á*^{n} 'breathe'
‖ʰòˤ^{n}*ó*^{n} 'cough'
ᵑǂùˤ^{n}*à*^{n} 'calf of common duiker (Sylvicapra grimmia) or steenbok (Raphicerus campestris)'

As was the case with the 'plain' vs. nasalized non-epiglottalized vowels, the phonetic realization of these segments is largely the same as that of the corresponding non-nasalized epiglottalized vowels, of course with the addition of epiglottalization. Also, as before, [ɑ̟ˤ^{n}], [ɔ̟ˤ^{n}], [ɵˤ^{n}] are transcribed here as [ɑˤ^{n}], [ɔˤ^{n}], [ɵˤ^{n}].

3.4 Lexical Tones

While it has been acknowledged before by Miller et al. (2007, 2009) that N|uu is indeed a tone language (like all other 'Khoisan' languages), the tonal structure of the language has not been described in detail so far. A preliminary investigation of the data available to me has shown that there seem to be two underlying lexical tones, *high* (H) and *low* (L). High tones are transcribed with an acute accent /◌́/, low tones with a grave accent /◌̀/ in the present work (3.27).

(3.27) *ǀ͡χ’á* ‘arm’
ǀʰà ‘kill’

Like in !Xóõ (Naumann 2008, *pace* Traill 1977), The *mora* is the tone bearing unit in N|uu; this means that within any given syllable, each nuclear element (i.e. vowel or syllabic nasal) and each coda element (i.e. syllable-final nasal, if present) bears exactly one tone.

My preliminary study suggests that *underlying* lexical tones in N|uu seem to be stable accross contexts, but it should be noted that no in-depth investigation has been undertaken yet that systematically takes many different phonological, morphological, and syntactic environments into account to search for possible morphophonological or grammatical tonal phenomena. On the phonetic *surface*, however, coarticulation between adjacent tones is widespread (cf. Xu 1994). Figure 3.32 shows averaged pitch tracks (f_0 contours) of ‘canonical’ realizations of the four two-moraic tone sequences found in N|uu on the words *ǀ͡χáá* ‘edge’ (HH), *!áà* ‘red hartebeest (*Alcelaphus caama*)’ (HL), *ǀàá* ‘hold’ (LH), and *ǁ͡χ’àà* ‘scold’ (LL).

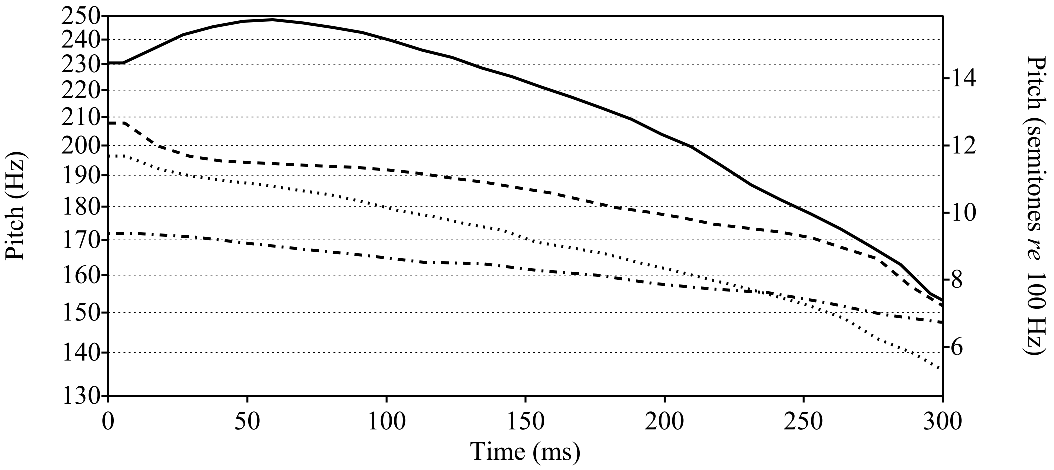

Figure 3.32 Averaged pitch tracks of the tones on *ǀ͡χáá* ‘edge’ (HH, solid line), *!áà* ‘red hartebeest (*Alcelaphus caama*)’ (HL, dotted line), *ǀàá* ‘hold’ (LH, dashed line), and *ǁ͡χ’àà* ‘scold’ (LL, dashed-dotted line). Each pitch track was ensemble-averaged over 30 tokens.

What is of crucial importance in understanding the appearance of the pitch tracks given above is the *elicitation context* of the words used to illustrate the tonal patterns. Like most of the words used in the present work, they were elicited in a controlled context with the structure *ná ká* __ ‘I say __’. That means that *preceding* every (two-moraic) example word, there is a *high* lexical tone H (on *ká*), and *following* every example word (which is utterance-final in this context), there is what I postulate to be a *low* boundary tone L-%.[19] This results in the following four sequences (lexical tones on example words in **bold** type):

- H **HH** L-%
- H **HL** L-%
- H **LH** L-%
- H **LL** L-%

Looking at the four cases, it seems reasonable to assume with Xu (1994) that the coarticulatory 'pressure' in the given context is strongest on words with a low–high (LH) tonal structure because context conflicts with that structure both at the left and at the right edge. And indeed, words with an underlying LH structure in my data are in the majority of cases realized phonetically not with a *low–high rising* tone but with a *mid level* tone in this context, as formalized in (3.28).

(3.28) /LH/ → [MM] / H __ L-%

This effect can also be observed in Figure 3.32 above, where the pitch contour of the word *|àá* 'hold' has a (more or less) mid level trajectory on the phonetic surface.

19 It must be noted, though, that no systematic study of the intonation system of the language has been carried out yet.

4 The Phonological Structure of N|uu

In the preceding chapter, the *phonetic* realization of the segmental inventory of N|uu has been discussed. In the present chapter, an overview of the *phonological* system of the language will be given. For that purpose, the prosodic structure of N|uu (syllables, morae, feet, and prosodic words as well as their relation to roots and stems) will first be investigated in Section 4.1. Following that, the distributional properties of the phonemes in terms of prosodic positions are discussed in Section 4.2. Finally, Section 4.3 deals with phonotactic co-occurrence restrictions between individual phonemes.

4.1 The Prosodic Structure

In a top-down approach to the prosodic structure of N|uu, lexical roots are best described by reference to a *foot* structure: Roots in N|uu consist minimally of one foot.[20] However, the constituent feet of any given root, if it contains more than one foot, do not have equal hierarchical status. Adopting a terminology originally used by Van Valin and LaPolla (1997) in a syntactic context, a *nucleus* can be distinguished from the *periphery*. Within the nucleus, in turn, a *core* can be distinguished from the *non-core* nucleus. To the entire complex, finally, a prefix and/or a suffix can be added; however, these do not enter into the foot structure in any obvious way. In phonological words in N|uu, the core is the only obligatory constituent.

The decisive argument for such a layered prosodic structure of the root in N|uu comes from distributional restricions on segmental phonemes. As will be seen below, the distribution of consonantal as well as vocalic phonemes is best

20 *Prosodic structure* is restricted to the structure of phonological words and their constituent units in the present context. No systematic study of larger prosodic units has been undertaken yet.

described by reference to their respective positions *within the foot* (e.g., obstruents only occur in the initial position of any foot). On the other hand, as was mentioned above, the feet do not have equal status. The peripheral foot is strongly reduced in comparison with the nuclear feet in that it is shorter (the periphery is always monomoraic, as opposed to the mono- or bimoraic nuclear feet) and can only contain a severely reduced subset of segments. Within the nucleus, the segmental structure of the core and the non-core closely parallel each other, but again, the non-core is reduced relative to the core in that it only shows a reduced subset of phonological oppositions (e.g., while obstruents do occur in the initial position of both, clicks are only found in the initial position of the core foot). It is argued here that the distributional differences between the constituents are systematic and clear-cut enough to justify an analysis in terms of a *categorical* and *hierarchical* difference.

To sum up, the structure of phonological words in N|uu can be described schematically as in (4.1):

(4.1) $[(CV\text{-})\ ([C_0V_0]_{\phi_0})\ [C_1V_1(C_2)(V_2)]_{\phi_1}\ ([C_3V_3(C_4)(V_4)]_{\phi_2})\ (\text{-}CV)]_{\omega}$

The phonological word is marked as ω. The peripheral foot is marked as ϕ_0, the core nuclear foot as ϕ_1, and the non-core nuclear foot as ϕ_2. CV- denotes a prefix, -CV a suffix; i.e., only root segments receive an index. Elements in parentheses are not obligatory. The hierarchical structure between the levels of the phonological word and the foot have been omitted for better legibility. Of all possible foot-internal combinations within this template, the structure $*[C_3V_3V_4]_{\phi_2}$ for the non-core nuclear foot (ϕ_2) is not attested in my data; I assume that this is an accidental gap.

It can well be hypothesized that this layered, hierarchical structure has historical origins, e.g. in a scenario where formerly independent elements in a more isolating stage of the language gradually developed greater phonological coherence between each other until they finally formed a single phonological word. That way, the distributional patterns of different foot positions could reflect former differences between words of different status. However, as far as I am aware, there are no convincing arguments that words with a complex foot structure in N|uu could be analyzed as sequences of independent words or even single but morphologically complex words *synchronically*. To take an example where such an analysis might at first seem attractive, consider (4.2):

(4.2) *Θáàχè* ‘daughter’
ǀáàχè ‘female cousin’
ǁáàχè ‘sister’

One could theoretically argue that there is a (derivational) suffix **-χè* involved here, with a meaning along the lines of ‘female relative’. But such a putative ‘suffix’, however plausible it might be historically, has no synchronic justification in N|uu: Not only would it be completely unproductive, but also would the remainder of the respective words, the putative ‘stems’, have no discernible meaning. The situation is quite parallel to that of a number of Indo-European terms of relationship, given in (4.3) in their reconstructed forms according to Wodtko, Irslinger, and Schneider (2008):

(4.3) **b^hráh$_2$ter-* ‘brother’ (pp. 38–41)
**máh$_2$ter-* ‘mother’ (pp. 457–461)
**ph$_2$tér-* ‘father’ (pp. 554–562)

On the face of it, one could argue that the three English words *brother*, *mother*, and *father* are all really morphologically complex, containing a ‘suffix’ **-ðə*. But just like in N|uu, such a suffix would be unproductive, and the remaining ‘stems’ would have no meaning. In fact, this situation can be taken as far back as reconstructed proto-Indo-European, where no completely convincing hypothesis for a putative suffix **-ter* has yet been presented.

The point of the argument is that under the analysis assumed in the present work, N|uu is a language with a rather elaborate phonological word structure in a substantial part of its lexicon, a fact that makes it stand out from many, if not most, other languages of the ‘Khoisan’ linguistic area.

We will now in (4.4) turn to examples of all attested prosodic word types in N|uu. All examples are taken from my 703-root lexical database, from which two items, *púɾúkú͡tsì* ‘butterfly’ and *ɟùɾùkúɟúí-sí* ‘Namaqua sandgrouse (*Pterocles namaqua*)’ (*-sí* ‘singulative’), had to be excluded. (These are the only items in the database that do not fit into the analysis presented above; it can be hypothesized that this is due to the ideophonic character of the two words.) An expression like “1 + 2 morae per syllable” means that in the words with the respective structural template, the first syllable has one mora and the second syllable has two morae; i.e., such a word would have two syllables and three morae (not counting any possible affixes). Numbers in parentheses denote the absolute number of occurrences of the respective structural type in the database (*N* = 701).

(4.4) a. 1 mora per syllable:
[CV]$_{\phi_1}$: *cú* ‘mouth’ (42)

b. 2 morae per syllable:
[CVC]$_{\phi_1}$: *!àń* ‘brain’ (49)
[CVV]$_{\phi_1}$: *ǂúí* ‘egg’ (331)

c. 1 + 1 morae per syllable:
[CV]$_{\phi_0}$ [CV]$_{\phi_1}$: *kà|é*^{n} ‘name’ (1)
[CV]$_{\phi_1}$ [CV]$_{\phi_2}$: ^{ŋ}*ǁùcú* ‘nose’ (12)
[CV.CV]$_{\phi_1}$: *kéɾé* ‘lightning’ (149)

d. 1 + 2 morae per syllable:
[CV]$_{\phi_0}$ [CVC]$_{\phi_1}$: *kàǁúǹ-sí* ‘sociable weaver (*Philetairus socius*)’ (*-sí* ‘singulative’) (1)
[CV]$_{\phi_1}$ [CVC]$_{\phi_2}$: *t͡s’ík*^{h}*ùm̀* ‘common mole rat (*Cryptomys hottentotus*)’ (2)
[CV]$_{\phi_0}$ [CVV]$_{\phi_1}$: *kú!*^{h}*òè* ‘be black’ (4)
*[CV]$_{\phi_1}$ [CVV]$_{\phi_2}$: not attested

e. 2 + 1 morae per syllable:
[CVC]$_{\phi_1}$[CV]$_{\phi_2}$: *χáǹcì* ‘mother’ (8)
[CVV]$_{\phi_1}$ [CV]$_{\phi_2}$: *bá*^{ʕ}*ùkè* ‘bark (verb)’ (58)

f. 2 + 2 morae per syllable:
[CVC]$_{\phi_1}$ [CVC]$_{\phi_2}$: *ǁ͡χúḿʔúm̀* ‘spread out’ (2)
[CVV]$_{\phi_1}$ [CVC]$_{\phi_2}$: $^{ŋ̊}$*!*^{h}*ò*^{n}*è*^{n}*cíń* ‘evening’ (2)

g. 1 + 1 + 1 morae per syllable:
[CV]$_{\phi_0}$ [CV]$_{\phi_1}$ [CV]$_{\phi_2}$: *sí!úχú* ‘Burchell’s zebra (*Equus burchelli*)’ (1)
[CV]$_{\phi_0}$ [CV.CV]$_{\phi_1}$: *kàɟámá* ‘show’ (4)
*[CV]$_{\phi_1}$ [CV.CV]$_{\phi_2}$: not attested
[CV.CV]$_{\phi_1}$ [CV]$_{\phi_2}$: *!͡χánáχà* ‘hate’ (31)

h. *1 + 1 + 2 morae per syllable: not attested

i. 1 + 2 + 1 morae per syllable:
[CV]$_{\phi 0}$ [CVC]$_{\phi 1}$ [CV]$_{\phi 2}$: *kàcúm̀ʔí* ‘listen’ (1)

j. 2 + 1 + 1 morae per syllable:
[CVV]$_{\phi 1}$ [CV.CV]$_{\phi 2}$: *$^{\text{ŋ}}$ǂù$^{\text{n}}$ì$^{\text{n}}$cùɾú-sí* ‘caterpillar’ (*-sí* ‘singulative’) (2)

k. 1 + 1 + 1 + 1 morae per syllable:
[CV]$_{\phi 0}$ [CV.CV]$_{\phi 1}$ [CV]$_{\phi 2}$: *kàgìɾìcè* ‘tickle’ (1)

Roots in N|uu thus have 1–3 feet, 1–4 syllables, and 1–4 morae. Table 4.1 shows the lexical frequency of the possible foot templates (CV structures) in the lexical database over the three hierarchical foot positions (core, non-core nucleus, and periphery).

Table 4.1 Lexical frequency of different foot templates by hierarchical status in N|uu (N = 701). The missing non-core nuclear template *$C_3V_3V_4$ is presumed to be an accidental gap.

Periphery	Nucleus	
	Core	Non-Core
C_0V_0 (13)	C_1V_1 (58)	C_3V_3 (112)
—	$C_1V_1C_2$ (61)	$C_3V_3C_4$ (6)
—	$C_1V_1V_2$ (397)	*$C_3V_3V_4$ (0)
—	$C_1V_1C_2V_2$ (185)	$C_3V_3C_4V_4$ (2)

Of course, the numbers in Table 4.1 do not add up to N = 701 (except for those in the *Core* column, which is the only obligatory position in a word), because many lexical items have more than just a core foot. From this perspective, the attested combinatory possibilities (with their absolute lexical frequencies) are as follows:

- Core only (571)
- Core + non-core nucleus (117)
- Periphery + core (10)
- Periphery + core + non-core nucleus (3)

4.2 The Distribution of Phonemes

Having established the basic prosodic structure of N|uu (at the level of the phonological word), we will now focus on the statistics of the distribution of the phonologically contrastive segments in terms of their prosodic position.

4.2.1 The Phoneme Inventory in Terms of Prosodic Position

In Chapter 3 above, the phonological inventory of the language was presented in a format that corresponded as closely as possible to the conventional format of the International Phonetic Association (IPA). For the discussion at hand, however, it will be more convenient to organize this information in a way that is phonologically less redundant. In other words, while still being grounded in articulatory categories, now only *phonologically contrastive* categorical dimensions will be used. As will be seen, this results in a considerably more 'compact' presentation of the same data. The discussion will proceed linearly along the individual C and V positions, each indexed as defined in Section 4.1.

Segments in position C_0. In position C_0, i.e., initially in the peripheral foot, only two segments are attested: /k/ and /s/. Given the small number of items in the database with that position, any further generalization at this point runs the risk of overgeneralizing; but a sensible hypothesis seems to be that only *non-click obstruents* occur in this prosodic position.

Segments in position V_0. The peripheral vocalic position is similarly restricted as the corresponding consonantal position, containing only the segments /i/, /a/, and /u/. Thus, again at the risk of overgeneralizing slightly, only *plain (non-nasalized, non-epiglottalized) vowels* occur in this position.

Segments in position C_1. In stark contrast to the set of segments in position C_0, the set occuring in position C_1 is extremely large, containg the bulk of the consonantal segments found in the language. Table 4.2 presents the click segments that are attested in the database for the position C_1.

Table 4.2 Click segments occuring in prosodic position C_1 in N|uu. Dotted circles indicate presumed accidental gaps.

	Labial	Tense Deep Coronal	Tense Shallow Coronal	Lax Shallow Coronal	Lateral Coronal
Stop	ʘ ◌ ᶢʘ	ǃ ǃʰ ᶢǃ	ǂ ǂʰ ᶢǂ	ǀ ǀʰ ᶢǀ	ǁ ǁʰ ᶢǁ
Sequential Stop	ʘ͡q ◌ ʘ͡q' ◌	ǃ͡q ǃ͡qʰ ǃ͡q' ◌	ǂ͡q ǂ͡qʰ ǂ͡q' ◌	ǀ͡q ǀ͡qʰ ǀ͡q' ᶢǀ͡ɢ	ǁ͡q ǁ͡qʰ ǁ͡q' ◌
Affricate	ʘ͡χ ◌	ǃ͡χ ǃ͡χ'	ǂ͡χ ǂ͡χ'	ǀ͡χ ǀ͡χ'	ǁ͡χ ǁ͡χ'
Nasal	◌ ŋ̊ʘˀ ŋʘ	ŋ̊ǃʰ ŋ̊ǃˀ ŋǃ	ŋ̊ǂʰ ŋ̊ǂˀ ŋǂ	ŋ̊ǀʰ ŋ̊ǀˀ ŋǀ	ŋ̊ǁʰ ŋ̊ǁˀ ŋǁ

Cf. Chapter 5 for the terminology used in this table. The term ‘deep’ is short for *deep concave* (i.e., large click cavity), ‘shallow’ for *shallow concave* (i.e., small click cavity); ‘stop’ stands for *simultaneously released stop*, whereas ‘sequential stop’ is short for for *sequentially released stop*.

Likewise, most of the non-click consonants can occur in position C_1, as is shown in Table 4.3.

Table 4.3 Non-click segments occuring in prosodic position C_1 in N|uu. Dotted circles indicate presumed accidental gaps.

	Labial	Coronal	Dorsal	Laryngeal
Stop	p ◌ ◌ b	c cʰ c' ɟ	k kʰ ◌ g	ʔ
Affricate		t͡s t͡s'	k͡χ'	
Heterorganic Affricate	◌	c͡χ		
Fricative		s z	χ	h
Nasal	m	ɲ		

In the variety on N|uu investigated in the present study, there is only a single item in the database that would seem to justify an underlying contrast between *coronal* segments at two *different* places of articulation (the pronoun *ná* ‘I’, with lamino-alveolar /n/). Other than in this word, coronal nasals in C_1 position are prepalatal, as in *ɲéβécé* ‘greet’ (with prepalatal /ɲ/). More generally, for any coronal category in any given prosodic position, only one place of articulation can occur (e.g., coronal *stops* in C_1 position are always *prepalatal*, whereas coronal *fricatives* in C_1 position are always *lamino-alveolar*). The same applies to dorsals, where velar and uvular segments are never in contrast with each other *within the same prosodic position*.

Summing up the information that can be gathered from Table 4.2 and Table 4.3, the segments that can occur in C_1 position include *obstruents* and *nasal sonorants*.[21] *Non-nasal sonorants* are excluded from this position.

Segments in position V_1. Table 4.4 shows the inventory of segments that are attested in position V_1 in the database.

Table 4.4 Segments occuring in prosodic position V_1 in N|uu. The dotted circle indicates a presumed accidental gap.

	Front	Central	Back
High Vowel	i i^n		u u^n $u^ʕ$ $u^{ʕn}$
Mid Vowel	e e^n $e^ʕ$ ◌		o o^n $o^ʕ$ $o^{ʕn}$
Low Vowel		a a^n $a^ʕ$ $a^{ʕn}$	

The inventory of segments in position V_1 includes all *vowels* found in the language. Additionally, the *velar nasal* /ŋ̍/ (which only occurs as a syllable nucleus) also occurs in this position.

Segments in position C_2. Table 4.5 summarizes the set of segments in position C_2.

Table 4.5 Segments occuring in prosodic position C_2 in N|uu

	Labial	Coronal
Nasal	m	n
Oral Sonorant	β	ɾ
Lateral Oral Sonorant		l

This set includes all *sonorants* (except /ŋ̍/). It should be noted at this point that *non-nasal* sonorants can only occur in this position if the V_2 position is filled (as

21 Cf. Miller (2011) for evidence that all clicks, including nasal clicks, are obstruents.

e.g. in *ǁùrú* ‘quartz’); in other words, non-nasal sonorants cannot occur as a syllable coda.

Segments in position V_2. Table 4.6 shows the segments that are attested in position V_2.

Table 4.6 Segments occuring in prosodic position V_2 in N|uu

	Front		Central		Back	
High Vowel	i	i^n			u	u^n
Mid Vowel	e	e^n			o	o^n
Low Vowel			a	a^n		

This set of segments can be characterized as the entire set of *non-epiglottalized vowels*. As for position V_1 above, the syllabic *velar nasal* /ŋ̍/ also occurs in this position, but only if V_1 is also /ŋ̍/ and C_2 is not filled (as e.g. in *ᵑǂŋ̀ŋ̀* ‘house’).

Segments in position C_3. In the non-core nuclear foot, the set of possible segments is again reduced as compared to the one in the peripheral foot, though not quite as severely. The following segments are attested in the database: /b/, /c/, /k/, /kʰ/, /ʔ/, /t͡s/, /k͡χʼ/, /s/, and /χ/. Although the set is not complete (due to the limited size of the database), it is plausible to hypothesize that this is the set of *non-click obstruents*.

Segments in position V_3. The situation is similar in position V_3, which is reduced as compared to the core, but not as strongly reduced as in the periphery. The following segments are attested: /i/, /e/, /a/, /u/, and /ŋ̍/, which I hypothesize is the set of *plain (non-nasalized, non-epiglottalized) vowels* plus the syllabic *velar nasal*.

Nasalized vowels also do occur in this position, but only under very restricted circumstances, namely if V_1 (and V_2 if it is present) is also nasalized *and* C_3 is the glottal stop /ʔ/ (as e.g. in *cúⁿáⁿʔáⁿ* ‘have nothing’). This can be described as a form of vowel harmony targeting nasalization; interestingly, the process operates across the foot boundary. In this specific context, /iⁿ/, /eⁿ/, and /aⁿ/ are attested in V_3 position.

Segments in position C_4. Just like C_3 in the non-core nucleus can be said in a way to parallel C_1 in the nucleus, C_4 parallels C_2. The following segments are attested: /m/, /n/, and /ɾ/. They are hyphothesized here to constitute the (incomplete) set of *sonorants*. As for C_2 above, *non-nasal* sonorants can only occur in the C_4 position if the V_4 position is filled (as e.g. in *ŋ̊!ˀóécàɾá* 'female adolescent').

Segments in position V_4. Finally, in position V_4, only the segments /a/ and /u/ are attested. Because of the limited size of the database, it seems not unreasonable to hyphothesize that they constitute the (incomplete) set of *plain (non-nasalized, non-epiglottalized) vowels*. The syllabic velar nasal */ŋ̍/, which is not attested in this context, could be assumed to be an accidental gap.

One of the main results of the preceding discussion is the fundamental distributional difference between obstruents, sonorants, and vowels in the phonology of N|uu, with clicks as the phonologically strongest segments (cf. Traill, 1979) occupying the (arguably) strongest prosodic position, namely the initial position (C_1) in the core foot, in the majority of lexical items (cf. Section 4.2.2).

Interestingly, (non-click) *nasals* are the only segments in the phoneme inventory that show an ambiguous behavior with respect to the prosodic positions. For instance, the labial nasal /m/ occurs both in C_1 and in C_2 position, patterning with obstruents in the first case but with sonorants in the second. The same applies in principle to the coronal nasal, although the effect is not as readily visible there because of the position-dependent split into /ɲ/ (which patterns with obstruents) and /n/ (which patterns with sonorants, except for the word *ná* 'I'). The ambiguous picture is completed by the syllabic velar nasal /ŋ̍/, which patterns *neither* with obstruents *nor* with sonorants but exclusively with *vowels*.

4.2.2 The Statistical Distribution of Phonemes

In the present chapter, the statistical distribution of the phonemes of N|uu is presented in terms of the prosodic positions discussed in the previous section. Numbers refer to absolute counts of occurrence in a given prosodic position and refer to the same 701-item lexical database as described above.

Segments in position C_0. The segments in position C_0 occur in the database with the following absolute frequencies (n = 13): /k/ (11), /s/ (2).

Segments in position V_0. In position V_0, the frequencies are as follows (n = 13): /i/ (1), /a/ (10), /u/ (2).

Segments in position C_1. Table 4.7 summarizes the absolute lexical frequencies of all click segments that are attested in prosodic position C_1 in the database.

Table 4.7 Click segments occuring in prosodic position C_1 in N|uu with cell counts (in parentheses), row counts, and column counts (N = 701). Dotted circles indicate presumed accidental gaps.

⊙ (9)				! (44)				ǂ (31)				| (26)				‖ (34)				144
	◌ (0)				!ʰ (14)				ǂʰ (11)				|ʰ (11)				‖ʰ (18)			54
																				—
			ᶢ⊙ (1)				ᶢ! (14)				ᶢǂ (5)				ᶢ| (8)				ᶢ‖ (9)	37
⊙͡q (1)				!͡q (12)				ǂ͡q (8)				|͡q (13)				‖͡q (7)				41
	◌ (0)				!͡qʰ (6)				ǂ͡qʰ (11)				|͡qʰ (10)				‖͡qʰ (9)			36
		⊙͡q’ (1)				!͡q’ (2)				ǂ͡q’ (3)				|͡q’ (3)				‖͡q’ (5)		14
			◌ (0)				◌ (0)				◌ (0)				ᶢ|͡G (2)				◌ (0)	2
⊙͡χ (1)				!͡χ (13)				ǂ͡χ (11)				|͡χ (6)				‖͡χ (12)				43
																				—
		◌ (0)				!͡χ’ (5)				ǂ͡χ’ (5)				|͡χ’ (14)				‖͡χ’ (12)		36
																				—
																				—
	◌ (0)				ŋ̊!ʰ (8)				ŋ̊ǂʰ (7)				ŋ̊|ʰ (11)				ŋ̊‖ʰ (3)			29
		ŋ̊⊙ˀ (1)				ŋ̊!ˀ (24)				ŋ̊ǂˀ (11)				ŋ̊|ˀ (17)				ŋ̊‖ˀ (22)		75
			ᵑ⊙ (2)				ᵑ! (11)				ᵑǂ (12)				ᵑ| (19)				ᵑ‖ (13)	57
11	0	2	3	69	28	31	25	50	29	19	17	45	32	34	29	53	30	39	22	568

Again, for reasons of convenience and legibility, absolute frequencies of non-click segments in C_1 position are given separately in Table 4.8.

Table 4.8 Non-click segments occuring in prosodic position C_1 in N|uu with cell counts (in parentheses), row counts, and column counts (N = 701). Dotted circles indicate presumed accidental gaps.

p (1)				c (16)				k (16)				ʔ (7)		40
	◌ (0)				c^h (5)				k^h (5)					10
		◌ (0)				c’ (1)				◌ (0)				1
			b (2)				ɟ (6)				g (10)			18
				t͡s (2)										2
														—
						t͡s’ (10)				k͡χ’ (7)				17
														—
◌ (0)				c͡χ (5)										5
														—
														—
														—
				s (14)				χ (12)				h (7)		33
														—
														—
							z (1)							1
														—
														—
														—
			m (3)				ɲ (3)							6
1	0	0	5	37	5	11	10	28	5	7	10	14	—	133

As in the previous section, a note on the coronal nasal is in order. As was explained above, there is a single attested item in N|uu with a *lamino-alveolar nasal* /n/ in C1 position (the pronoun *ná* ‘I’). However, this item would not appear in the table anyway, because only *lexical roots* were included in the database, not grammatical morphemes. Therefore, this isolated instance of /n/ does not influence the statistical distribution of segments as given in the table, where only the count for /ɲ/ is included in the cell corresponding to the coronal nasal.

Segments in position V_1. In Table 4.9, the lexical frequencies of the segments occurring in position V_1 are presented.

Table 4.9 Segments occuring in prosodic position V_1 in N|uu with cell counts (in parentheses), row counts, and column counts (N = 701). The dotted circle indicates a presumed accidental gap.

i (23)								u (138)				161
	i^{n} (5)								u^{n} (11)			16
										$u^{ʕ}$ (2)		2
											$u^{ʕn}$ (1)	1
e (21)								o (99)				120
	e^{n} (5)								o^{n} (19)			24
		$e^{ʕ}$ (2)								$o^{ʕ}$ (28)		30
			◌ (0)								$o^{ʕn}$ (5)	5
				a (252)								252
					a^{n} (49)							49
						$a^{ʕ}$ (25)						25
							$a^{ʕn}$ (8)					8
44	10	2	0	252	49	25	8	237	30	30	6	693

In addition, as discussed above, a further segment that occurs in position V_1 is the syllabic *velar nasal* /ŋ̍/ (8).

Segments in position C_2. Table 4.10 summarizes the statistical distribution of segments in prosodic position C_2.

Table 4.10 Segments occuring in prosodic position C_2 in N|uu with cell counts (in parentheses), row counts, and column counts (N = 701)

m (42)	n (70)	112
β (52)	ɾ (81)	133
	l (3)	3
94	154	248

The conspicuously low frequency of /l/ could lead one to hyphothesize that the present situation could be the result of an (almost completed) historical merger of /l/ with /ɾ/.

Segments in position V_2. In Table 4.11, lexical counts are given for all segments attested in position V_2 in the database.

Table 4.11 Segments occuring in prosodic position V_2 in N|uu with cell counts (in parentheses), row counts, and column counts (N = 701)

i (94)				u (105)		199
	i^n (26)				u^n (16)	42
e (81)				o (63)		144
	e^n (13)				o^n (15)	28
		a (138)				138
			a^n (25)			25
175	39	138	25	168	31	576

As was the case with position V_1 above, the inventory of segments in position V_2 is augmented by the syllabic *velar nasal* /ŋ̍/ (6).

Segments in position C_3. The segments that are attested in position C_3 occur in the database with the following frequencies (n = 120): /b/ (1), /c/ (32), /k/ (27), /k^h/ (1), /ʔ/ (41), /t͡s/ (1), /k͡χʼ/ (4), /s/ (2), /χ/ (11).

Segments in position V_3. For position V_3, the counts are as follows (n = 120): /i/ (25), /i^n/ (2), /e/ (40), /e^n/ (1), /a/ (31), /a^n/ (3), /u/ (17), /ŋ̍/ (1).

Segments in position C_4. In prosodic position C_4, the following frequencies are observed (n = 8): /m/ (4), /n/ (2), /ɾ/ (2).

Segments in position V_4. Finally, position V_4, the final position in the root template of N|uu, shows the following absolute frequencies in the database (n = 2): /a/ (1), /u/ (1).

4.2.3 The Statistical Distribution of Lexical Tones

Having discussed the lexical statistics of the segmental phonemes of N|uu, we will now investigate the statistical distribution of the lexical tones. The same 703-item database as before will be taken as a basis for the investigation. The two items *púɾúkút͡sì* ‘butterfly’ and *ɟùɾùkúɟúí-sí* ‘Namaqua sandgrouse (*Pterocles namaqua*)’ (*-sí* ‘singulative’) had already been excluded above, because as probable *ideophones*, they did not fit into the proposed root template.

Now, when discussing tone, three more items need to be excluded because they deviate from the general pattern. All three words also share a segmental peculiarity: Historically, they are clearly *reduplicated* forms.[22] The items in question are:

- *‖ám̍-‖àm̀* ‘talk’
- *‖͡χàá-‖͡χáà* ‘teach’
- *ᵑ̊ǂˀòˤé-ᵑ̊ǂˀòˤé* ‘titbabbler (*Parisoma subcaeruleum*)’

That means that, excluding the two ideophones and the three reduplicated forms, for the purpose of investigating the *tonal* structure of N|uu, the database has a size of N = 698.

22 Note, though, that to my knowledge there is no evidence of a *synchronically productive* reduplication mechanism in N|uu.

Table 4.12 summarizes the frequencies of tonal patterns in different segmental contexts found in *peripheral* feet.

Table 4.12 Tonal patterns in N|uu peripheral feet, by segmental context (N = 698)

	H	L
C_0V_0	3	10

Next, Table 4.13 gives the counts for the different tonal patterns attested in *core* feet.

Table 4.13 Tonal patterns in N|uu core feet, by segmental context (N = 698). Unexpected cell counts are **bold** (see text).

	H	L	HH	HL	LH	LL
C_1V_1	48	10	—	—	—	—
$C_1V_1C_2$	—	—	18	18	17	7
$C_1V_1V_2$	—	—	107	114	101	73
$C_1V_1C_2V_2$	—	—	50	**11**	**87**	37

Inspection of the table reveals that the occurrence of the HL tone pattern on $C_1V_1C_2V_2$ feet is conspicuously *low*, while the frequency of the LH tone pattern on the same foot type is correspondingly *high*. A χ^2 test performed on the bitonal patterns from Table 4.13 (i.e., HH, HL, LH, and LL) by segmental context was statistically significant, $\chi^2(6) = 51.65$, $p < .001$, implying that the tonal patterns of core feet in N|uu are indeed *not independent* of their segmental structure. At present I have no plausible hypothesis as to what the explanation for this interdependence could be.

Finally, Table 4.14 summarizes the statistical distribution of tonal patterns in *non-core nuclear* feet.

Table 4.14 Tonal patterns in N|uu non-core nuclear feet, by segmental context ($N = 698$). Missing *$C_3V_3V_4$ is presumably an accidental gap.

	H	L	HH	HL	LH	LL
C_3V_3	52	60	—	—	—	—
$C_3V_3C_4$	—	—	2	2	2	2
*$C_3V_3V_4$	—	—	0	0	0	0
$C_3V_3C_4V_4$	—	—	0	0	2	0

4.3 Phonotactic Restrictions

Having established the relevant prosodic domains as well as the distribution of segments within these domains, we are now in a position to investigate possible co-occurrence restrictions (constraints) that are active in the phonology of N|uu.

As we have seen in Section 4.1, there is considerable phonological coherence within the feet of a phonological word in N|uu, but much less coherence between the feet. It is therefore not surprising that to my knowledge, there are very few active phonotactic restrictions *between* the periphery and the nucleus or *between* the core and the non-core nucleus, but quite many *within* each of these constituents. We will now take up these foot-internal restrictions in turn.

4.3.1 Phonotactic restrictions involving vowels in general

The combinations of vowels in V_1 and V_2 positions that are well-formed, i.e. attested in the data, are given in (4.5); /…/ stands for an optional intervening element, which, in this case, would be a consonant in C_2 position. The vowel symbols here stand for the *vowel quality* as such; whether or not in any given case nasalization and/or epiglottalization is possible is determined by independent rules.

(4.5) a. /i…i/

b. /e…e/

c. /a…i/
/a…e/
/a…a/
/a…o/
/a…u/

d. /o…e/
/o…a/
/o…o/

e. /u…i/
/u…a/
/u…u/

f. /ŋ̇…ŋ̇/

Note that it is irrelevant in this context whether a consonant intervenes between the two vowels or not; if it does, the vowels are heterosyllabic, whereas if it does not, the vowels are tautosyllabic and form a long vowel or a diphthong.

Several generalizations can be formed on the basis of these admissible combinations:

1. Front vowels are only followed by *identical* front vowels.
2. /a/ is followed by *all* vowels (except /ŋ̇/).
3. Back vowels are either followed by /a/ or by non-open vowels of *identical height.*
4. /ŋ̇/ is only followed by /ŋ̇/.

This system clearly displays a kind of *vowel harmony* with respect to *height* (and, in the case of front vowels, *backness* as well); in this system, /a/ is a *neutral* vowel in that it is compatible with all vowels. The domain of vowel harmony in N|uu is the *foot*; the generalizations given above hold not just within the core, but within the non-core nucleus as well.

It was stated above that when no consonants intervenes between V_1 and V_2, the result is a long vowel or a diphthong. The respective frequencies of the two cases in the lexical database are as follows ($N = 701$):

- Long monophthongs ($V_1 = V_2$): 194

- Diphthongs ($V_1 \neq V_2$): 203

4.3.2 Phonotactic restrictions involving front vowels

In addition to the constrictions regarding vowels in general, there are several important constraints that target *front* vowels (/i/ and /e/) specifically.

1. Front vowels cannot be epiglottalized (2 exceptions for /e/: *zéˤé* ‘fly (verb)’, *ŋ̊ǂʰèˤβé* ‘loincloth’).
2. Front vowels cannot occur after the click types /ʘ/, /!/, or /ǁ/, i.e. after clicks with a *deep concave* tongue configuration, a phenomenon known as the *Back Vowel Constraint* (Traill, 1985) (1 exception for /e/: *ŋ̊ǁˀé* ‘go’).
3. Front vowels cannot occur after /c͡χ/, /χ/, /k͡χʼ/, /ǀ͡χ/, /ǀ͡χʼ/, /ǂ͡χ/, or /ǂ͡χʼ/, i.e. after segments with a *uvular fricated* release (3 exceptions for /e/: *ʘáàχè* ‘daughter’, *ǀáàχè* ‘female cousin’, *ǁáàχè* ‘sister’).

Note that by rule 2, front vowels are also excluded after other click types with a uvular fricated release, like /ʘ͡χ/ etc. It is also noteworthy that in all three cases, the constraints are exceptionless for /i/ but show a small number of exceptions for /e/ (which can be said to be a ‘less extreme’ front vowel).

4.3.3 Phonotactic restrictions involving nasalization

Nasalization provides the second example for a system resembling *vowel harmony* in N|uu. If present, nasalization applies to *all vowels of the core*; in that case, no consonant C_2 may intervene between V_1 and V_2. But interestingly, when C_3 is a *glottal stop* /ʔ/, nasalization spreads from the core to the non-core nuclear vowel or vowels (again, no consonant C_4 may then intervene). In other words, all consonants are *opaque* to the spread of nasalization, except for /ʔ/, which is *transparent*.

An interesting constraint regarding nasalization is that nasalized $V_1(V_2)$ cannot occur after a nasal C_1 consonant. This is an important piece of evidence to show that nasal clicks (at least in N|uu) are phonologically *not* nasals but *obstruents*.

4.3.4 Phonotactic restrictions involving epiglottalization

Apart from the constraint that epiglottalization may not occur on front vowels that was discussed in Section 4.3.2 above, there is also a constraint that bans epiglottalized V_1 vowels from occurring after C_1 consonants with a *uvular fricated release* (including uvular fricatives).

Note also that in contrast to nasalization, which, as we have seen, has certain suprasegmental properties in that it can spread across a wider domain, the domain of epiglottalization is strictly the V_1 vowel.

4.3.5 Phonotactic restrictions involving consonant clusters

Finally, there is at least one constraint on well-formed (root-internal) consonant clusters in N|uu. Due to the restrictive root structure rules as presented above, root-initial consonant clusters cannot easily arise in N|uu. In fact, there is only one admissible cluster structure that is attested in the database, namely one of the type $C_2.C_3$, where C_2 is a nasal sonorant and C_3 is a non-click obstruent, with a syllable boundary between the two.

In the database ($N = 701$), 11 items with such clusters are found, in 2 of which C_3 is /c/ (e.g. *χáǹcì* 'mother') and in 9 of which C_3 is /ʔ/ (e.g. *ǂáńʔám̀* 'cover (verb)').

5 Two Empirical Studies

In the present chapter, I will present additional evidence for the phonetic structure of clicks in N|uu. As will become clear, this evidence will also have a bearing on possible explanations for the phonological patterning of clicks and is thus an important complement to the previous studies mentioned in Chapter 2. The experimental studies undertaken are twofold: Firstly, in Section 5.1, I investigate the different realizations of vowels following the so-called ‘front’ vs. ‘back’ clicks; secondly, in Section 5.2, the acoustic properties of linguo-pulmonic stops vs. affricates are looked into more closely.

It has been known since Traill’s (1985) groundbreaking work that in !Xóõ, clicks (along with other consonants) can be classified phonologically into two major groups, based on whether or not they can co-occur with a following front vowel. The segments that do *not* co-occur with front vowels are said to be subject to the so-called Back Vowel Constraint (henceforth BVC) because they are constrained to co-occurring exclusively with back vowels (/a/, /o/, and /u/). Recently, in Miller et al. (2007, 2009), we have considerably enhanced the knowledge about the BVC, presenting evidence that seems to support the view that the BVC is grounded primarily in the *posterior* constriction of clicks. As we will see below, N|uu has a very similar phonotactic constraint that I will argue has a physiological basis in the production mechanism of those clicks that are subject to the constraint.

The BVC is of central importance for the understanding of click phonology (or more specifically, the *internal* phonological structure of clicks) because it is the only known phonological rule with respect to which different click *types* pattern differently: Whereas the so-called ‘front’ clicks (dental /|/ and prepalatal /ǂ/) co-occur freely with all vowels, the so-called ‘back’ clicks (bilabial /ʘ/, alveolar /!/, and lateral alveolar /ǁ/) do not. I will argue below that the terms ‘front’ and ‘back’ clicks are misleading and should better be avoided; an alternative terminology will be proposed.

5.1 Properties of the Anterior Constriction

As was just noted, the 'front' click types, /|/ and /ǂ/, do co-occur with front vowels (i.e. /i/ and /e/) in N|uu. However, there is a noticeable (phonetic) diphthongization, or formant movement, at the onset of those vowels. At present, this has not yet been studied quantitatively, and it is not clear whether the vowel onset is retracted, or lowered, or both, with respect to the vowel midpoint. This question has potential consequences for the phonological representation of the respective clicks: Do 'front' vs. 'back' clicks differ systematically in their coarticulatory influence on following vowels?

5.1.1 Method

The data for the present study consisted of targeted words uttered in a controlled frame sentence (*ná ká* __ 'I say __') by speakers AK, HK, and KE (all female). Each word was repeated 10 times by each speaker, which resulted in a total of 30 tokens (3 speakers × 10 repetitions) for any given word. Which words were used depended on which particular phenomenon was being investigated; details on the exact nature of the words used are given below.

An effort was made to find as quiet a recording environment as possible, either at the speakers' homes or at the office of the South African San Institute (SASI) in Upington. The setup used for the recordings was an AKG C420 head-mounted condenser microphone and a Sound Devices USBPre high-quality combined microphone preamplifier and analog-to-digital converter, recording onto an Acer TravelMate 230 laptop computer.[23] The recordings (WAV format) were originally made at a sampling rate of 48000 Hz and a bit depth of 24 bits but were downsampled to 22050 Hz / 16 bits for further acoustic analysis.

The recordings were then organized in such a way that all repetitions of a given word by a given speaker were contained in a single sound file, resulting in 3 sound files (one for each speaker) for any given word. In each such file, the relevant portions of the signal were subsequently segmented and labeled in Praat (Boersma & Weenink, 2010), thus resulting in 10 labels per file. Praat scripts

23 Note that with this setup, the entire digitization process took place *outside* the computer, thus bypassing the sound card. – A number of recordings that were necessary to complete the data set were kindly made available to me by Amanda Miller and Bonny Sands; those additional recordings were made using very similar procedures.

were then written to loop though each of the 30 labeled segments (3 sound files × 10 labels) for any given word in order to measure the acoustic parameters under investigation. The results were written to a tab-delimited text file. For the experiments described in this section, formant transitions were then analyzed in Praat.

Formant transitions. Frequency values for the first two formants, F_1 and F_2, were determined at two different points of measurement: (1) at the first identifiable glottal pulse and (2) at the midpoint of the vowel. In the case of diphthongs, steady states had already been roughly identified during the labeling process, and the onset and midpoint of the first part of the respective diphthong was chosen as a point of measurement.

The formant values were determined as follows: First, the vocalic interval was extracted from its context to make sure that no frequency components other than from the vowel would be used for spectral calculation. then, following Boersma and Weenink's (2010) recommendation regarding suitable shapes for the analysis window, the middle of a Gaussian window with an effective length of 512 samples (corresponding to 23.22 ms at a sampling rate of 22050 Hz) was aligned with the midpoint of the vowel. If the onset was investigated, the interval was multiplied with a Hanning window, and 512 zero samples (i.e. samples with a value of zero) were added to the beginning of the interval. Then the middle of the analysis window was aligned with the onset of the original vowel, so that one half of the analysis window spanned the zero samples and the other half spanned the onset of the vowel.[24] (This procedure was chosen because it was important for the following analyses to make sure that the very first glottal pulse was given as much weight as possible.) Then, formant values were extracted with Praat's formant tracking algorithm. Formants are notoriously difficult to determine automatically, so as soon as a formant value was identified, it was checked manually by the following procedure, which was implemented by a Praat script and run automatically for each interval: The signal in the interval was downsampled to 12000 Hz (an assumed formant spacing of 1200 Hz × 5 formants × 2), then an LPC spectrum with a prediction order of 12 (5 formants × 2 + 2) was displayed for the point of measurement, and the formant value found by the formant tracking algorithm was superimposed on the

24 The actual, 'physical' length of a Gaussian window is twice its effective length (cf. Boersma & Weenink, 2010).

graphical display of the LPC spectrum. A high degree of accuracy and robustness against errors could be achieved through combining these different methods into a system of ‘checks and balances’; in the case of discrepancies between the results of the two methods, formant values were identified manually in the spectrogram, which was routinely displayed as a fallback.

Having extracted the necessary acoustic parameters from the data, the results were finally read into R (R Development Core Team, 2010) for statistical analysis (again with the help of scripts written for that purpose). The data were quantitatively analyzed and visualized using locus equations (Sussman, McCaffrey, & Matthews, 1991; Sussman, Hoemeke, & Ahmed, 1993); in addition, two distinct classes of statistical procedures were employed:

Locus equations. These provide a way of characterizing a given consonant by its coarticulatory relationship with a following vowel. To this end, F_2 values at vowel onset ($F2_{\text{onset}}$) as a function of those at vowel midpoint ($F2_{\text{vowel}}$) were plotted for a full set of vowels following a given consonant. Then a regression line was fitted to the resulting scatter plot. Subsequently, slope and y-intercept of the regression line were determined for each speaker and consonant to find any systematic patterns.

ANOVA. As an obvious first approach, an ANOVA (*analysis of variance*) with center of gravity as dependent variable and consonant as independent variable was carried out in each case (thus, a *one-way* ANOVA, since only one factor – consonant – was included). Furthermore, since multiple repetitions from each speaker were included, a *repeated measures* ANOVA (with speaker as a random factor) was called for to account for the requirement of independence of cases (since observations *within* a speaker are likely to be more strongly correlated than observations *between* speakers).

Two complications soon became apparent, however: Firstly, due to the high amount of random fluctuations, center of gravity values for FFT spectra included a large number of *outliers*, which made the applicability of ANOVA in principle problematic (due to violation of the assumption of normality). This problem could be remedied by only including *auditory* center of gravity values in the analysis; for these, the values within the groups were more or less normally distributed. As mentioned above, auditory major peak values were used rather than LPC ones, even though the latter presented considerably less problems with outliers than the FFT spectra.

Secondly, however, calculation of Greenhouse-Geisser and Huynh-Feldt ε revealed that these values were close to the lower bound, indicating a serious violation of the assumption of sphericity (i.e. equality of the variances of the differences between levels of the repeated measures factor, consonant) for the data. Because sphericity is a necessary condition for the application of a repeated measures ANOVA, a solution had to be found, and two possibilities were tried to that end. The first is to correct the degrees of freedom for the *F* test by multiplying them with the epsilon values mentioned above. While this was feasible, it resulted in most of the tests yielding only marginally significant results, which seemed at odds with what visual inspection of the data revealed, namely, clearly separated groups. Therefore, the second alternative was generally chosen, namely, fitting a totally different model to the data, a so-called *linear mixed effects* (LME) model (also known as a *multilevel* model).

LME. LME models have only recently been introduced into linguistic research but have been strongly advocated as an alternative to traditional repeated measures ANOVAs, both because they offer the possibility of 'fine-tuning' the model to allow for an optimal fit to the data, and because they are not sensitive to the violation of many assumptions that can make the application of ANOVA-based methods problematic, such as sphericity (cf. Baayen, Davidson, & Bates, 2008 for an introduction to the method and justifications for its use). For that reason, LME models (with a random intercept for speaker) were generally fitted to the data to evaluate the statistical significance of differences between values.[25] The *lme4* (Bates & Maechler, 2010) and *languageR* (Baayen, 2009) packages for R were used to compute those models.

The corpus for the investigation consisted of three groups of words. The first group consisted of words beginning with a prepalatal plosive /c/, followed by the full set of (non-nasalized, modal) vowels. It was chosen as a vocalic frame of

25 Note that *p* values in LME models in R were computed in a way that is fundamentally different from the one in conventional *F* tests (i.e. ANOVAs), namely by so-called *Markov chain Monte Carlo* (MCMC) sampling. Such *p* values are given in the text as p_{MCMC}.

reference, so to speak, against which the articulations after clicks could be compared.[26]:

- *cí* ‘it’
- *céè* ‘identificational’ (grammatical morpheme)
- *cáà* ‘lie (recline)’
- *còó* ‘bend’
- *cú* ‘mouth’

The second group consists of word beginning with a dental click /|/, as a representative of the class of ‘front’ clicks:

- *|í* ‘take’
- *|éè* ‘blue wildebeest (*Connochaetes taurinus*)’
- *|àá* ‘hold’
- *|óβà* ‘child’
- ^{g}*|ùù* ‘lie’ (i.e. ‘tell a lie’)

Finally, the third group consists of words that begin with an alveolar click /!/ as a representative of a ‘back’ click:

- *!áì* ‘burp’
- ^{g}*!àé* ‘springbok (*Antidorcas marsupialis*)’
- *!áà* ‘red hartebeest (*Alcelaphus caama*)’
- *!óó* ‘aardwolf (*Proteles cristatus*)’
- *!ùú* ‘camel thorn (*Acacia erioloba*)’

A remark is in order at this point regarding the first two items in the third group, *!áì* and ^{g}*!àé*. Note that these two words contain the diphthongs, /ai/ and /ae/, instead of the expected monophthongs, /i/ and /e/. This is no accident but will rather play a major role in the discussion that follows. At this stage, it suffices to say that there are two very conspicuous irregularities, or gaps, in the distribution

26 Note that the context after [c] was a natural choice, for the following reasons: There are no vowel-initial words in N|uu; there are hardly any words with pulmonic labial stops; there are no alveolar stops; there are hardly any velar stops with following front vowels; there are no uvular stops; and there are too few words that start with a glottal stop.

of /i/ and /e/ in the lexicon of N|uu (taking /|/ and /!/ as cover symbols for 'front' and 'back' clicks, respectively): As we have already seen, the sequences */!i/ and */!e/ do not exist, whereas /!ai/ and /!ae/ do. On the other hand, the sequences /|i/ and /|e/ exist but, surprisingly, */|ai/ and */|ae/ do *not*. I interpret this distributional peculiarity as the result of a set of (articulatorily based) diachronic processes of the type seen in (5.1), operating on an original system that had only /i/ and /e/ (in all contexts), but neither /ai/ nor /ae/ (in any context):

(5.1) a. **i > ai / !__*

b. **e > ae / !__*

The decisive argument to assume a *diachronic* process here (besides its phonetic plausibility) is the otherwise mysterious absence of the sequences */|ai/ and */|ae/ in the lexicon, for which there is no plausible phonetic reason whatsoever. I will return to this issue later, when I to demonstrate the articulatory plausibility of such a development.

5.1.2 Results

As a first step, the vowel space of N|uu was visualized by plotting Bark-scaled F_1 vs. $F_2 - F_1$ values (as advocated e.g. by Ladefoged, 2003) of the 'neutral' set of vowels (in the context after /c/) in Praat. Figure 5.1 shows a representation of this vowel space as a scatter plot of the individual formant values, pooled from all 3 speakers.

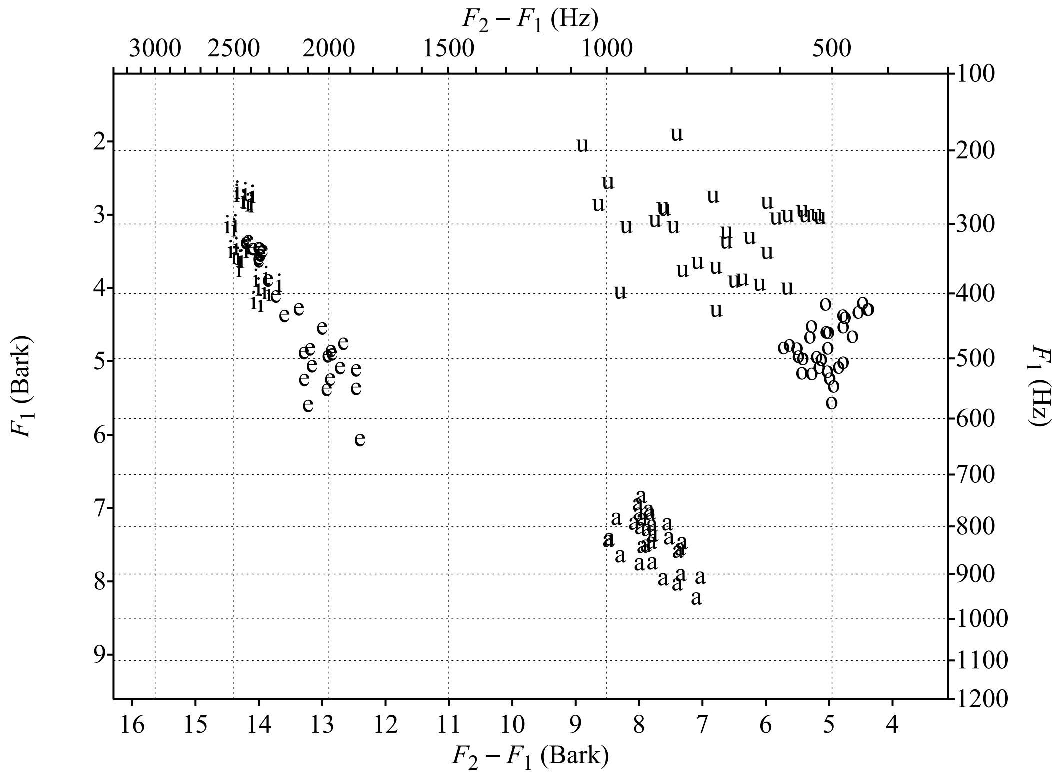

Figure 5.1 Bark-scaled scatter plot of F_1 vs. $F_2 - F_1$ values from the midpoint of 5 vowels in the context after /c/, pooled from 3 speakers (N = 150)

The individual formant values were seen to be fairly discretely distributed over the vowel space (with only a small amount of overlap between /i/ and /e/). It was therefore deemed legitimate to pool the data from the different speakers here and below. To summarize the formant data more clearly, the scatter plot was then replaced by a plot showing only the mean values and associated 95% confidence ellipses. Such a representation of the 5 vowels in the 'neutral' context is given in Figure 5.2.

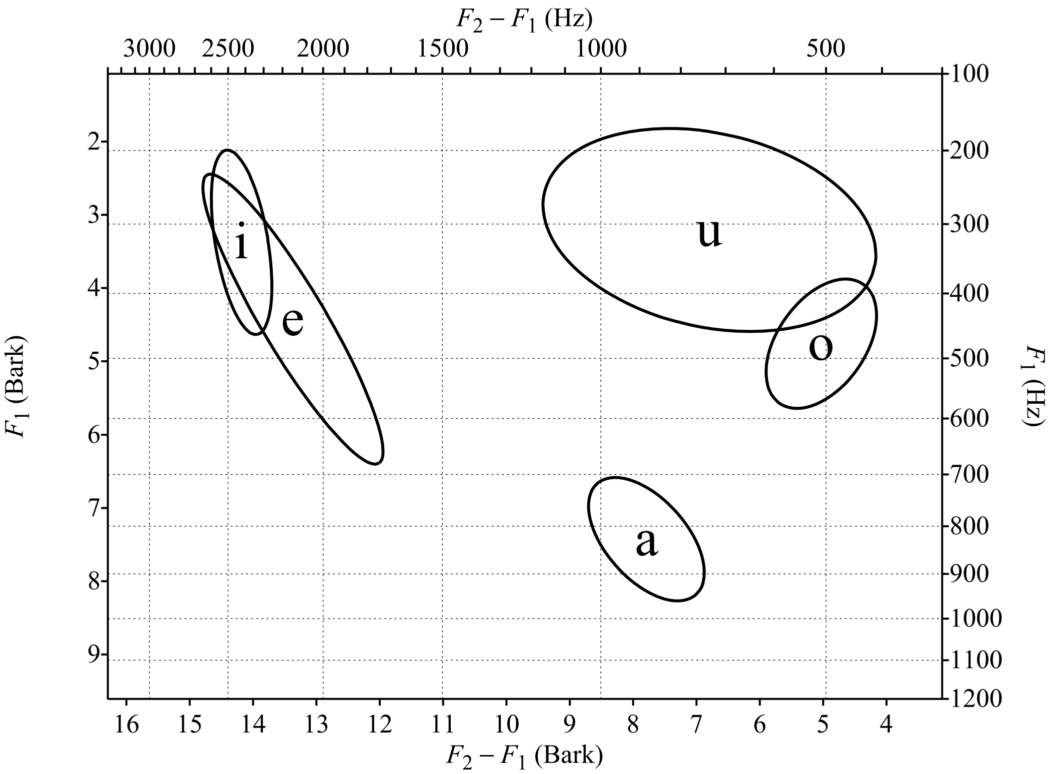

Figure 5.2 Bark-scaled plot of mean F_1 vs. $F_2 - F_1$ values and associated 95% confidence ellipses from the midpoint of 5 vowels in the context after /c/, pooled from 3 speakers (N = 150). The symbols are placed at the grand mean for each vowel.

These 'canonical' vowel ellipses were then used to provide a backdrop, a frame of orientation (represented with light dashed lines), in all specialized formant charts that were subsequently produced. This will not be further mentioned in what follows.

Next, the coarticulatory influence of the 'front' vs. 'back' clicks on a following vowel were investigated, starting with /|/. A preliminary check of all vowel contexts showed that such coarticulation was primarily present in the non-back vowels, /i/, /e/, and /a/ (the latter being a *mid* vowel in N|uu), and the study therefore concentrated on these. Figure 5.3 shows the formant movements (indicated by arrows) between the onset of the vowels (bold gray ellipses) and the vowel midpoint (bold black ellipses).

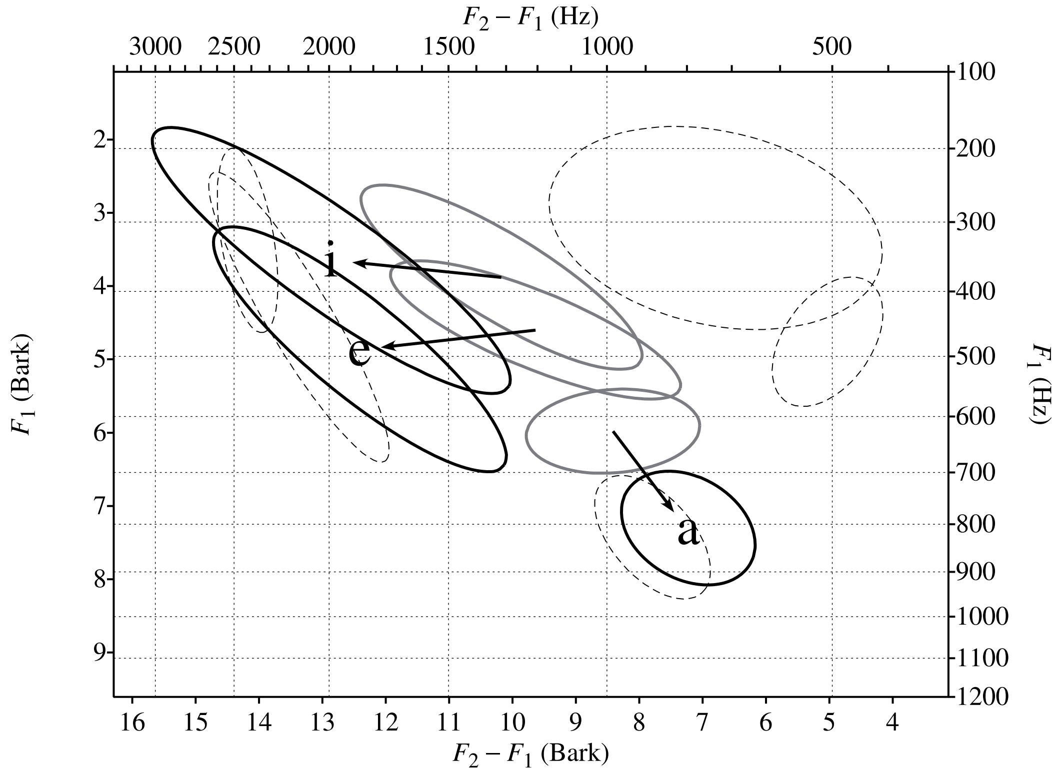

Figure 5.3 Bark-scaled plot of mean F_1 vs. $F_2 - F_1$ values and associated 95% confidence ellipses from the onset (bold gray ellipses) and the midpoint (bold black ellipses) of 3 vowels in the context after /ǀ/, pooled from 3 speakers (N = 90). Dashed ellipses represent vowels in unmarked contexts.

Two clearly distinct patterns emerged from this: (1) the *front* vowels, /i/ and /e/, are strongly *retracted* at onset; (2) the *low* vowel, /a/, is strongly *raised* at onset.

The patterns of formant movement following /!/ turned out to more complicated and will be presented for one vowel at a time. Figure 5.4 shows the formant movement between the onset and the midpoint of the first part of /ai/ after /!/.

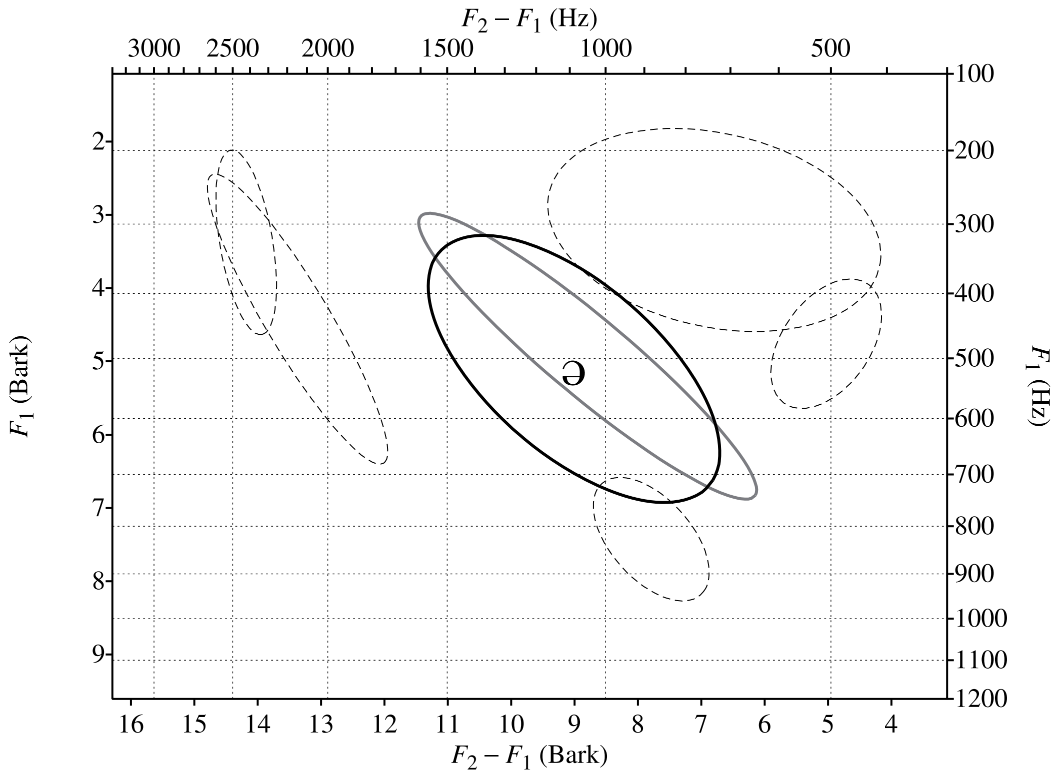

Figure 5.4 Bark-scaled plot of mean F_1 vs. $F_2 - F_1$ values and associated 95% confidence ellipses from the onset (bold gray ellipse) and the midpoint (bold black ellipse) of the first part of /ai/ in the context after /!/, pooled from 3 speakers (N = 30). Dashed ellipses represent vowels in unmarked contexts.

The first observation that was made is that there is no real formant *movement* at all, the values at onset and midpoint being centered around the *same* values, symbolized allophonically here as [ə]. Secondly, under the assumption detailed above that /ai/ (or [əi], allophonically) derives historically from /i/, [ə] is both *retracted* and *lowered* even more strongly at onset in comparison with /i/ in the context after /ǀ/.

Next, a corresponding chart for the formant movement between the onset and the midpoint of the first part of /ae/ after /!/ was plotted, as given in Figure 5.5.

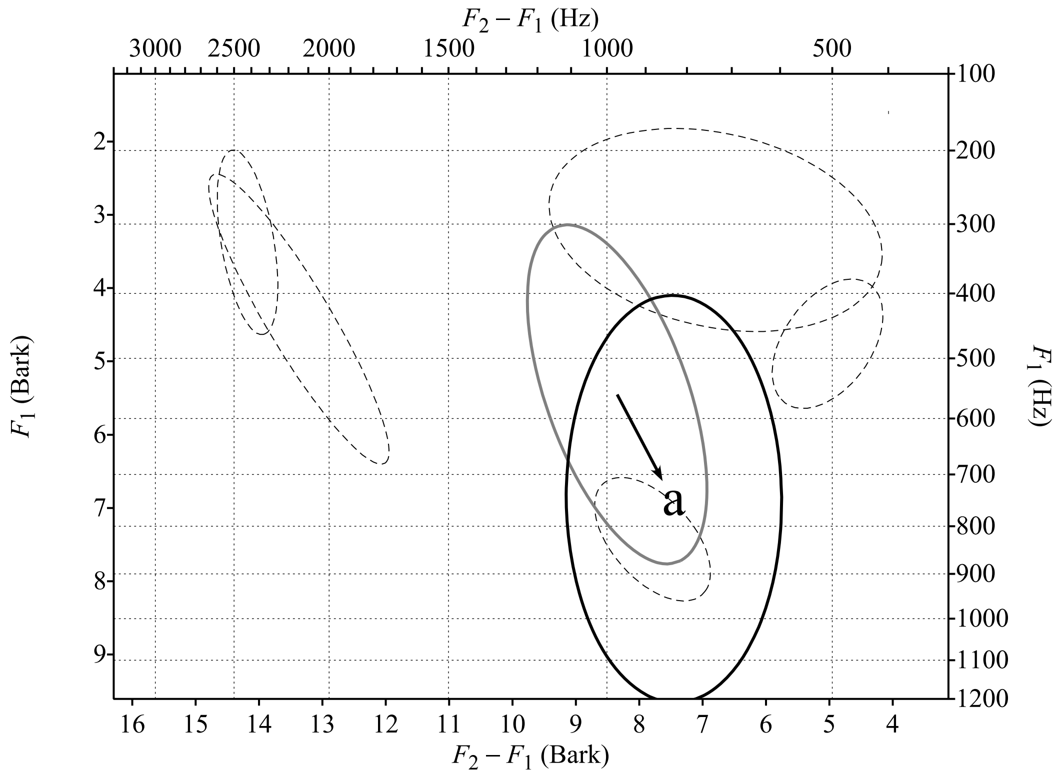

Figure 5.5 Bark-scaled plot of mean F_1 vs. $F_2 - F_1$ values and associated 95% confidence ellipses from the onset (bold gray ellipse) and the midpoint (bold black ellipse) of the first part of /ae/ in the context after /!/, pooled from 3 speakers (N = 30). Dashed ellipses represent vowels in unmarked contexts.

As compared with the plot for /ai/, Figure 5.5 shows the following differences: (1) There is considerable formant movement between onset and midpoint of the vowel; (2) the onset is only slightly lowered and retracted as compared to /ai/, whereas the midpoint is much lower and further back, being in the region of the 'canonical' vowel /a/; and (3) the formant values at midpoint (corresponding to the articulatory *target*) surprisingly display very large variability.

Finally, for the sake of completeness, Figure 5.6 shows the formant movements associated with /a/ in the context after /!/.

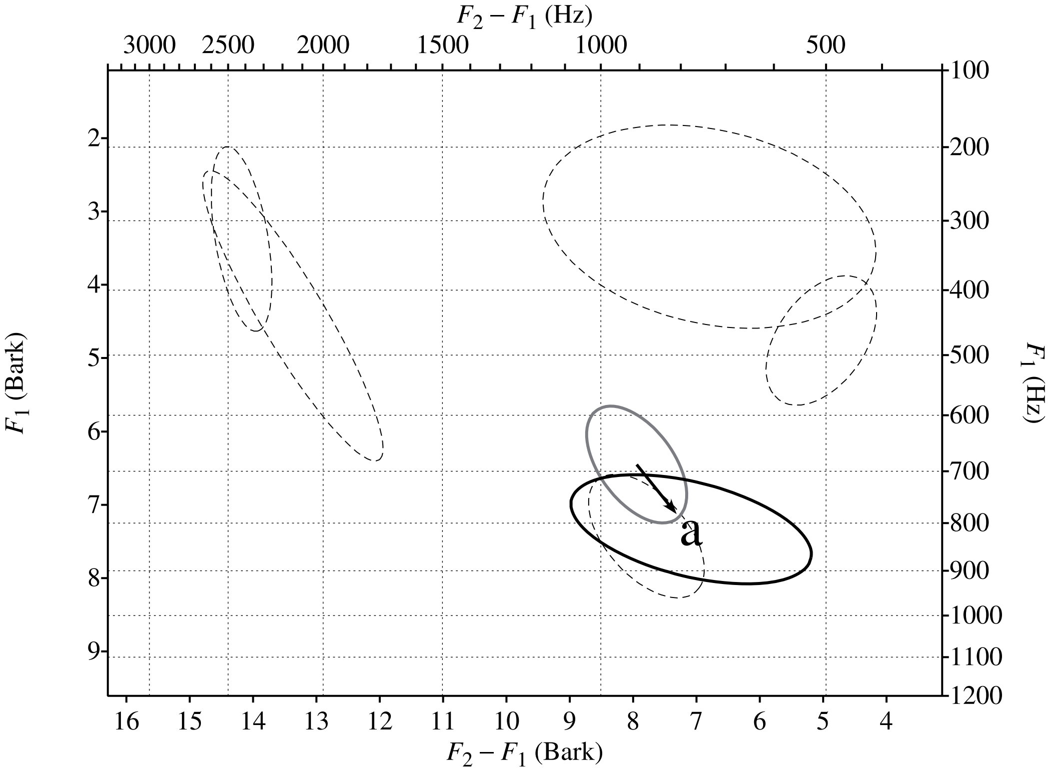

Figure 5.6 Bark-scaled plot of mean F_1 vs. $F_2 - F_1$ values and associated 95% confidence ellipses from the onset (bold gray ellipse) and the midpoint (bold black ellipse) of /a/ in the context after /!/, pooled from 3 speakers (N = 30). Dashed ellipses represent vowels in unmarked contexts.

Comparing this with /a/ in the context after /|/, it was observed that in the present context, the onset of /a/ is slightly lower, while the midpoint shows no major differences.

To quantify and assess statistically the observations made so far, an LME ANOVA for F_1 at onset vs. midpoint for /ai/, /ae/, and /a/ in the context after /!/ was conducted; corresponding box plots[27] are presented in Figure 5.7.

27 The bold band near the middle of the box represents the median (50th percentile) of the distribution; the lower and upper boundaries of the box represent the lower and upper quartile (25th and 75th percentile), respectively. The so-called 'whiskers' extend to the smallest and largest observations, respectively, unless these are more than 1.5 times the interquartile range from the box, in which case they are declared *outliers* and represented as small circles.

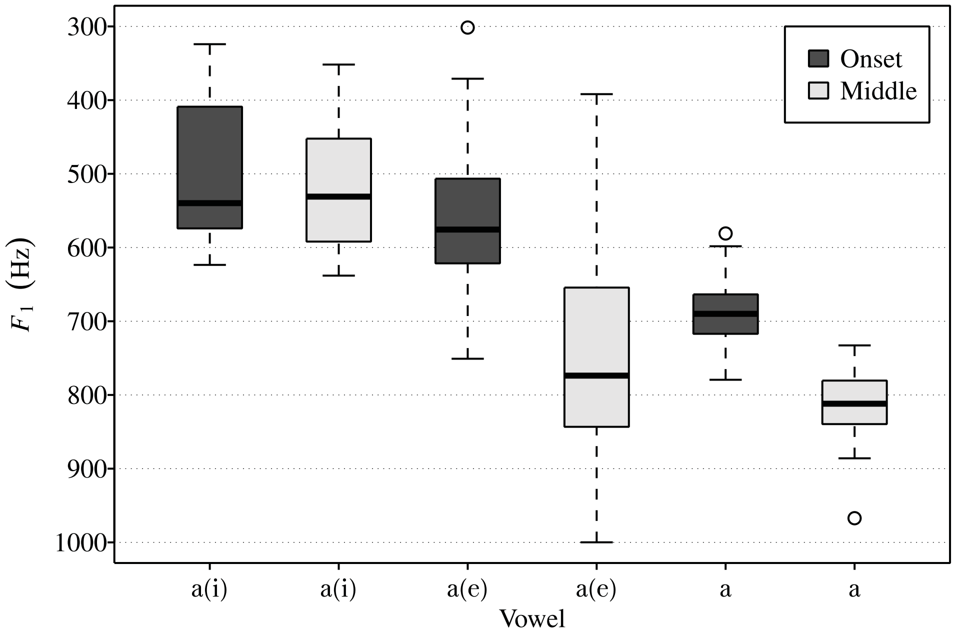

Figure 5.7 Box plots of F_1 at onset vs. midpoint for 3 vowels (N = 180)

The analysis revealed a significant main effect, $p_{\mathrm{MCMC}} < 0.001$. A one-way repeated measures ANOVA (without sphericity corrections) gave the same result, $F(5, 10) = 9.41$, $p = 0.002$. A pairwise comparison revealed that the only pair that was *not* significantly different was /ai/ at onset vs. midpoint, whereas all other pairs showed highly significant differences at $p < 0.001$.

Similarly, an LME ANOVA for F_2 at onset vs. midpoint for /ai/, /ae/, and /a/ (again in the context after /!/) was performed, as shown graphically in Figure 5.8.

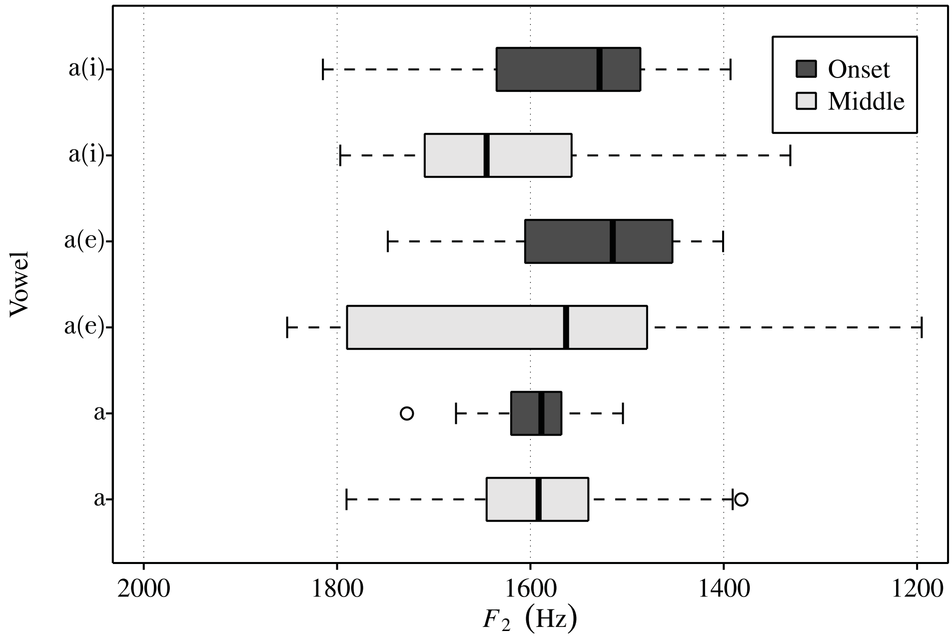

Figure 5.8 Box plots of F_2 at onset vs. midpoint for 3 vowels (N = 180)

Here, a radically different picture emerges: there is *no* significant main effect, and even the individual pairs do not show any significant differences.

Comparing all the formant charts presented above, one gets the impression that the movements ‘radiate’ from a certain (virtual) point in the vowel space, as it were, in analogy to the traditional concept of consonantal *locus*. Seeking to quantify this impression, I calculated the grand mean of all formant values (including all five vowels) at onset for each consonant, /|/ and /!/, and plotted the positions of the two resulting values in the vowel space, as shown in Figure 5.9.

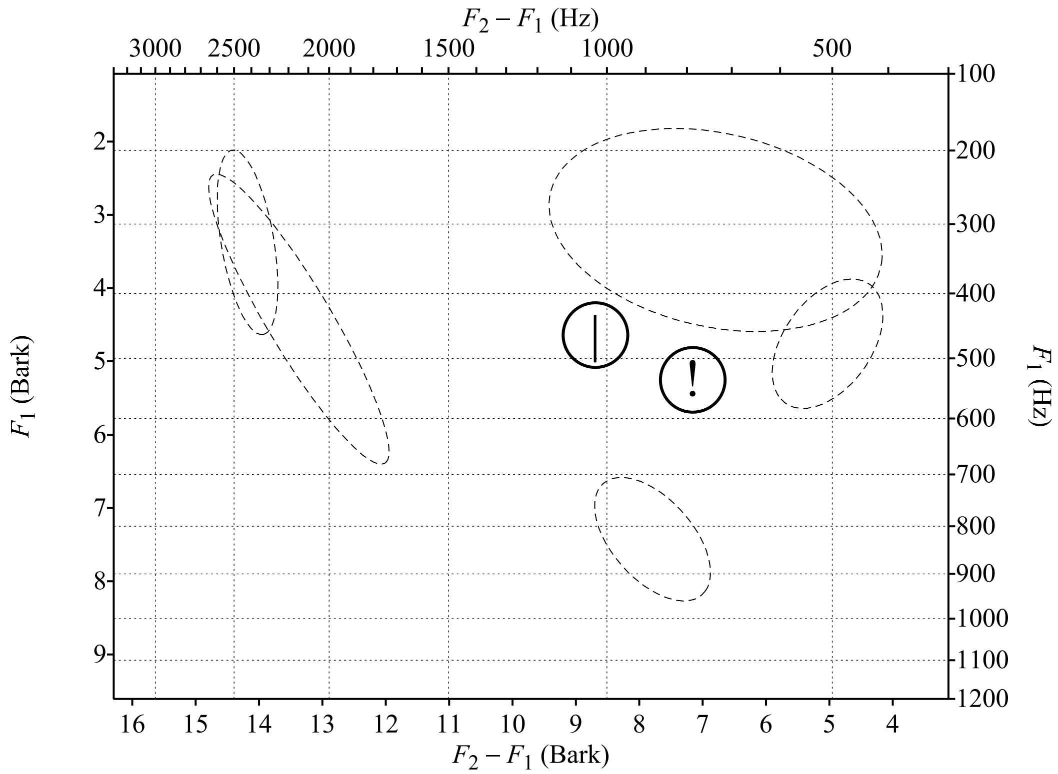

Figure 5.9 Bark-scaled plot of the grand mean of the F_1 vs. $F_2 - F_1$ values from the onset of 5 vowels in the contexts after the 'front click' /|/ and the 'back click' /!/, pooled from 3 speakers (N = 300). Symbols for the clicks are enclosed in circles for better legibility. Dashed ellipses represent vowels in unmarked contexts.

Clearly those values should be interpreted with caution because they do not take into account the possibility that coarticulatory patterns might vary not only between consonants, but also between vowels *within* consonants. Also, they are abstractions that do not *directly* correspond to anything in the actual data, as might be implied when placing these values in an actual formant chart. However, I take them to be a helpful heuristic tool to get a gross impression of the overall *direction* of coarticulatory adjustments in vowels following the respective consonants, and the results confirm the overall tendency observed so far: The onset of vowels following /!/ is *retracted* and *lowered* as compared to /|/. Recall that /!/ here stands for a 'back click' and /|/ for a 'front click'.

In order to investigate the coarticulatory relationship between different classes of clicks and the vowels that follow them more closely, *locus equations* for /|/ and /!/ were calculated. This method has been developed by Harvey Sussman and his colleagues and expounded in a series of papers (Sussman, McCaffrey, & Matthews, 1991; Sussman, Hoemeke, & Ahmed, 1993). As explained above, it involves plotting $F2_{\text{onset}}$ as a function of $F2_{\text{vowel}}$ for a given consonant over all vowel contexts. The regression line that is subsequently fitted to the data can be characterized in terms of just two values: (1) slope, and (2) *y*-intercept. Slope, being the more important of the two, is interpreted as a direct reflection of the degree of coarticulation between the consonant and its vocalic context. Two theoretical extreme cases can be distinguished, between which the actual values can range:

5. If slope = 0, it means that $F2_{\text{onset}}$ is *invariant* across all vowel contexts. In that case, there is *no* coarticulation (in the sense of regressive influence of the vowel on the preceding consonant).

6. If slope = 1, it means that $F2_{\text{onset}} = F2_{\text{vowel}}$ across all vowel contexts. If that is the case, there is *maximal* coarticulation (again, of the vowel on the consonant).

In the present case, however, I argue that due to the rigid requirements of the click articulation concerning the shape of the tongue, it is more fruitful to interpret the value of slope the other way around: A lesser slope means *stronger* coarticulatory influence of the *consonant* on the onset of the following *vowel*; a greater slope, conversely, is interpreted as *weaker* coarticulatory influence of the consonant on the vowel. That is, whereas the usual interpretation of the value is *regressive*, I will rather interpret it *progressively*.

Note that the values yielded by locus equations are *not* the same as the traditional, so-called *Haskins-type* locus (Delattre, Liberman, & Cooper, 1962), although the two quantities are not independent of each other. In the present study, I will note Haskins-type locus values only in passing.

Figure 5.10 presents the scatter plot and locus equation regression line that were calculated for /ǀ/.

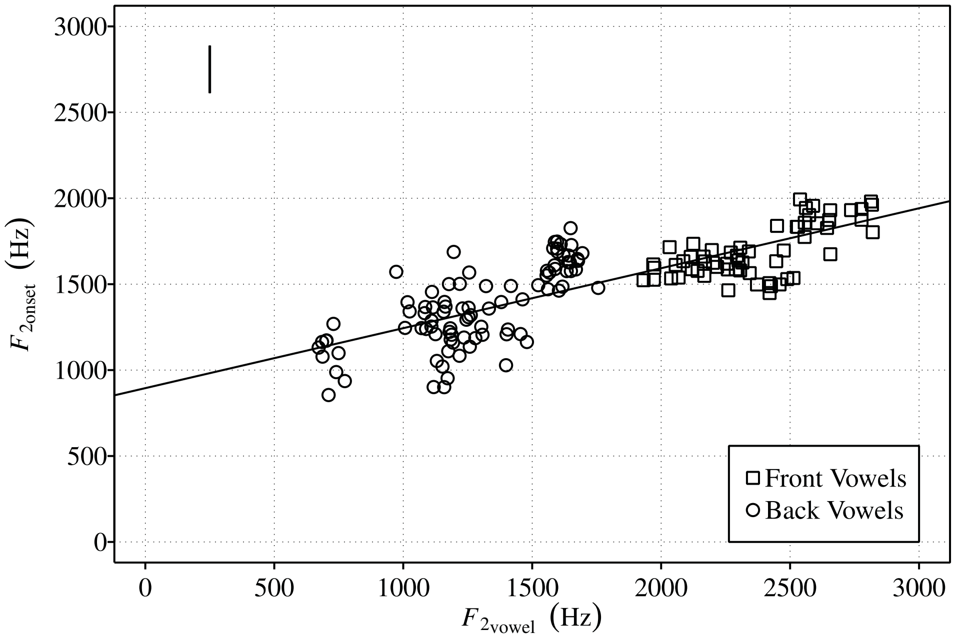

Figure 5.10 Scatter plot and regression line of F_2 at onset as a function of F_2 at midpoint for /ǀ/ over 5 vowels (N = 150)

By itself, the locus equation regression line had little informative value; rather, it was the *relative differences* between different regression lines that was the object of interest. Therefore, a second regression line was calculated for /!/, as shown in Figure 5.11. (One note is in order regarding this second plot: To facilitate comparison between the plots, data points for the different vowel contexts were categorized as either ‘front vowels’ or ‘back vowels’ for convenience. These labels are not to be taken literally: For /ǀ/, the former category comprises /i/ and /e/; for /!/, it comprises /ai/ and /ae/).

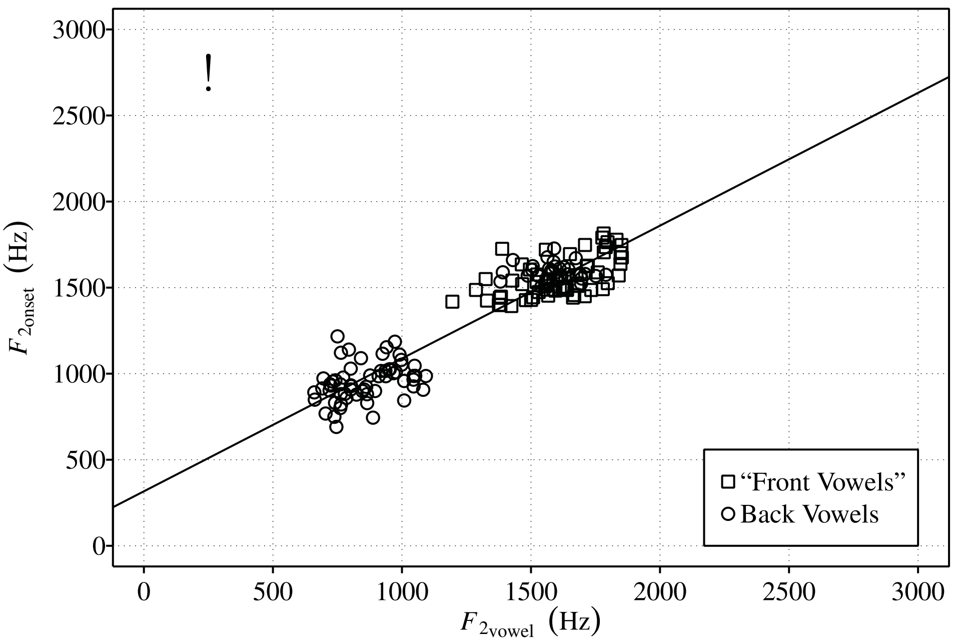

Figure 5.11 Scatter plot and regression line of F_2 at onset as a function of F_2 at midpoint for /!/ over 5 vowels (N = 150). The data points labeled "Front Vowels" represent /ai/ and /ae/, i.e. diphthongs moving towards a front target.

Two principal differences were noted between the plots: (1) The slope for /|/ was flatter than for /!/; and (2) for /!/, the data points for /ai/ and /ae/ were indistinguishable from one group of back vowel data points (actually /a/). For the sake of completeness, the mean Haskins-locus values obtained for /|/ was 1383.25 Hz (*SD* = 99.13 Hz); for /!/, the mean was very similar at 1398.10 Hz, but the standard deviation was much higher at 424.35 Hz.

Next, following the procedure in Sussman, McCaffrey, and Matthews (1991) and Sussman, Hoemeke, and Ahmed (1993), the *locus-equation-defined CV space* for /|/ and /!/ was constructed by plotting the *y*-intercept values as a function of the slope values for each consonant and speaker, as shown in Figure 5.12.

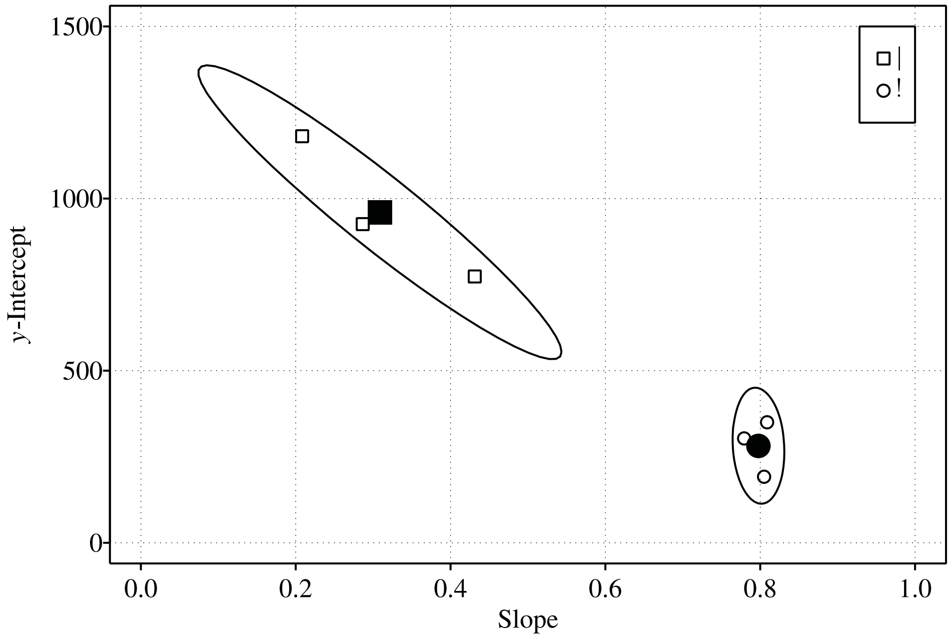

Figure 5.12 Locus-equation-defined CV space for /|/ and /!/ (N = 300). Data points represent speakers. The grand mean for each consonant is represented by a large symbol; data ellipses with a radius of one standard deviation from the mean enclose the respective categories.

Three facts were especially notable in this plot: (1) The relationship between slope and *y*-intercept is broadly linear: The greater the slope for a given speaker is, the lower the *y*-intercept will be; (2) as was already noted above, /|/ has a consistently lesser slope than /!/; and (3) /|/ shows a *much* higher variability in both parameters than /!/.

We will now proceed to look at possible explanations for the phenomena described so far.

5.1.3 Discussion

Putting together the pieces of evidence from the preceding section, we will review them one by one. As will become apparent, the seemingly disparate phenomena are really all reflections of the same underlying factor, namely the *size of the lingual cavity*, which typically relates directly to the apicality/laminality of the respective clicks.

Turning first to the *retraction* of vowels at onset as compared to their midpoint, recall that /i/ and /e/ were very strongly retracted following dental and prepalatal clicks /|/ and /ǂ/, whereas /ai/ and /ae/ were not at all retracted following bilabial, alveolar, and lateral alveolar clicks /ʘ/, /!/ and /ǁ/, just like /a/ following both clicks. However, plausible arguments have been brought forward in the previous discussion that /!ai/ and /!ae/ historically derive from */!i/ and */!e/. If one accepts that hypothesis, not only does the defective lexical distribution become understandable (the absence of */|ai/ and */|ae/ in the lexicon cannot plausibly be explained by any kind of co-occurrence constraint), but also the phonetic data under discussion suddenly becomes transparent: Instead of a situation where /|/ acts strongly retracting and /!/ hardly so, we have a situation where historical */!i/ and */!e/ have been retracted so far that they have been rephonologized as central vowels. Specifically, because the retraction is combined with lowering in the case of /!/, the retracted (and lowered) realizations of */!i/ and */!e/ at onset have been reanalyzed as realizations of /a/, resulting in /!ai/ and /!ae/.[28] In the case of /|i/ and /|e/, on the other hand, the retraction was (and is) slightly less extreme, and crucially, there is no lowering involved; therefore, there were no independent non-low central vowels in the inventory that could have served as targets of a reanalysis. In other words: all of /|i/, /|e/ (or [|ɨi], [|əe], allophonically), /!ai/, and /!ae/ are *phonetically* diphthongs; however, /!ai/, and /!ae/ have been *rephonologized* as diphthongs.

Let us now consider the *lowering* and *raising* phenomena discovered in the data. Firstly, looking at /|/, /i/ and /e/ did not show any signs of lowering, while /a/ even showed some *raising* at onset.

Again, the situation with /!/ is more complicated. I will summarize it vowel by vowel:

1. /ai/ starts off strongly lowered (with respect to its assumed original target, */i/) or strongly raised (with respect to 'canonical' /a/) and *remains* at that position.

28 Actually, it might be the case that the slightly stronger degree of retraction for /!/ is the result of enhancing (or stabilizing) a newly established contrast rather than the consequence of any articulatory mechanism. If that is so, then the additional lowering would have been the decisive factor historically.

2. /ae/ is slightly lower than /ai/ at onset (this relationship parallels that between /i/ and /e/), but in contrast to the latter, its midpoint target is radically lower, around the position of 'canonical' /a/. Moreover, there is very pronounced variability in its midpoint values.

7. /a/ is slightly raised at onset, but not as much as in the context following /|/.

With respect to /ai/, I interpret the combination of stability over onset and target together with lack of variability as evidence that the first part of this diphthong has not only historically been rephonologized, but that it implements an articulatory target that is distinct from 'canonical' /a/. In other words, it is encoded phonologically as a distinct *allophone*, [ə].

/ae/, on the other hand, shows the opposite behavior with respect to stability and variability. I assume that it realizes an articulatory target that is identical to 'canonical' /a/, but that it is subject to strong coarticulatory influence by the following part of the diphthong, the high-mid vowel /e/. The fact that its values show so much variability is a symptom of gradient, rather than categorical, influence, and hence *coarticulation* rather than allophony (A. Miller, personal communication, August 27, 2008).

So far, a residue of phenomena connected with raising, lowering, and retraction has remained unexplained. I believe that there is a unified explanation for these phenomena, namely the *degree of expansion of the lingual cavity*, reflecting the distinction between clicks with a typically *laminal* articulation (the so-called 'front' clicks) and those with a typically *apical* articulation (the so-called 'back' clicks). At the instant of release of the anterior closure, the lingual cavity is comparably small for laminal clicks but much larger for apical clicks (cf. Ladefoged & Maddieson 1996). Moreover, as evidenced by ultrasound images of click articulations (Miller et al., 2009), the *horizontal* movement of the tongue root (and consequently also the tongue body) also varies systematically with the type of cavity formed.

I argue that this parameter is sufficient to explain the patterns of variation left unaccounted for so far: Apical clicks show stronger retraction because of the more extreme downward-backward excursion of the tongue root that is facilitated (and indeed required) by the larger cavity. Furthermore, vowel height is broadly (and somewhat simplistically) speaking a function of the vertical position of the highest point of the tongue dorsum. But at the moment of release of a click, that point actually corresponds more or less to the *lowest* point of the lingual cavity. Therefore, the larger the cavity, the lower the onset of the vowel will be expected to be; hence the lower onset of /ai/ as compared to /i/ and of /ae/ as compared to /e/. This is true for *non*-low vowels. For low vowels, the opposite applies: There, the lowest point of the cavity is *higher* than the low-vowel target; hence the smaller the cavity, the higher the onset of the vowel will be. For that reason, /a/ at onset is raised more strongly for laminal than for apical clicks.

Analogously, vowel *backness* is in broad approximation a function of the *horizontal* position of the highest point of the tongue dorsum, hence in the case of clicks, of the lowest point of the lingual cavity. Because the cavity is always formed in a general downward-backward direction with only relatively slight horizontal leeway, vowels following *all* clicks display a more-or-less centralized articulation at onset, with much less variation than was seen for the vertical dimension.

The incompatibility of retroflex segments (in the broader sense of apical segments with displaying a sublingual cavity) with front vowels has been described in detail before by Hamann (2003). Based on the analysis of X-ray tracings, Traill (1985, pp. 115–116) previously discussed the incompatibility of 'back' clicks with high front vowels, but apparently did not relate this fact explicitly to the BVC. Miller (2010), on the basis of ultrasound evidence, discussed the possibility that in the context of the BVC, the shape of the tongue in 'back' clicks is what is responsible for the incompatibility with the tongue body shape found with high front vowels.

Because it is not the place of articulation of the anterior articulation (in the narrow, traditional sense of the term) but rather the overall *shape* of the anterior part of the tongue, I propose the terms *deep concave* and *shallow concave* clicks instead of the more traditional terms, apical and laminal.

A final question that arises in this context is why *sequentially* released clicks pattern just like plain, 'simultaneously' released clicks with respect to the BVC. I assume that this does not have articulatory reasons but rather arises through analogical extension of the constraint from the plain to the sequentially released clicks, leading to patterns such as /ǀ͡qi/ vs. */!͡qi/.

Summing up what we have established so far, the following hypotheses seem justified:

- The *degree of expansion of the lingual cavity* (correlating with shallow concave vs. deep concave shape of the front of the tongue) is mainly responsible for the different patterns of raising, lowering, and retraction at onset for vowels following /ǁ/ and /!/.

- Historically, the following processes can be postulated: */i/ > /ai/ and */e/ > /ae/ (in the context following /!/).

- Synchronically, (the first part of) /ai/ and /ae/ (as well as /a/) all share the same underlying target, /a/.

- Allophonically, however, (the first part of) /ai/ → [ə], whereas (the first part of) /ae/ as well as /a/ → [a]

- Coarticulation with the following /e/ part of the diphthong is responsible for the observed differences between /ae/ and /a/.

Finally, we will discuss the results from the locus equation study detailed above. As we have seen, /ǁ/ turned out to have a flatter slope than /!/, which, as I have explained above, I interpret as indicating a stronger coarticulatory influence of the consonant, /ǁ/, on the following vowels. However, I assume that a situation holds true here that is directly analogous to the one just discussed in connection with the formant charts: I take it that historically, the coarticulatory influence of /!/ on following front vowels was *so strong* that these were rephonologized as low central vowels. That is the situation that we see in the locus equation plots: For /!/, the 'front' vowels are indistinguishable in the plots from other tokens of /a/ (this is especially so because in these plots, only F_2 but not F_1 is being considered). Consequently, with /ai/ and /ae/ having merged with /a/ as far as F_2 is concerned, it is clear why /ǁ/ (with retraction of front vowel onsets still active synchronically) ends up with the flatter slope.

As for the locus-equation-defined CV space, two aspects were noted: (1) a broadly linear relationship between slope and *y*-intercept, and (2) more widely dispersed values for /ǁ/ than for /!/. Both can be explained quite straightforwardly:

The linear relation can be expressed in other words as: the steeper the slope, the lower the *y*-intercept. This means that for different speakers, the locus equation regression line (for /ǁ/) 'rotates' around a fixed point, so to speak; this point lies in the region of /a/. That is not surprising, given that /a/ as a low vowel shows hardly any F_2-related coarticulatory activity, as we have seen above. It is rather the *front* vowels (and obviously to some degree the back vowels, /o/ and

/u/) that display both strong coarticulatory retraction (manifested in differences in F_2) and large inter-speaker variation. Therefore, the wide dispersion of values for /ǀ/ in CV space can be translated as follows: If the degree of retraction of front vowels (and of fronting of back vowels) at onset between speakers gets higher, the slope gets flatter and the *y*-intercept higher, and vice versa.

5.2 Properties of the Posterior Constriction

In Section 5.1, experimental evidence has been presented that sheds some new light on the nature of the *anterior* constriction of clicks in Nǀuu. Next, we will turn to a different set of experiments that concentrates on the *posterior* constriction. As before, we will focus on the patterning of certain classes of clicks with respect to front vowels.

Linguo-pulmonic *sequential stops* like /ǀ͡q/ pattern just like 'plain', simultaneously released linguo-pulmonic stops (e.g. /ǀ/) with respect to the Back Vowel Constraint (BVC). Linguo-pulmonic *affricates* like /ǀ͡χ/, however, exhibit a different pattern: Such sounds never co-occur with front vowels, no matter what the click *type* is. In other words: In the sequential stops, the posterior release seems to be transparent with respect to the BVC, while in the affricates, it is obviously opaque, preventing the BVC from applying between the anterior release (determining the click type) and the following vowel. How does this difference arise? Several a priori possibilities come to mind:

- There might be a (very small and auditorily hardly perceptible) difference in place of articulation between the posterior constrictions of /ǀ͡q/ etc. and /ǀ͡χ/ etc. which could account for the different patterning of the two classes of sounds: Different places of articulation lead to different behavior.

- Under the assumptions (1) that the physiological basis of the BVC lies in the *posterior* rather than the anterior constriction (Miller et al., 2007, 2009) and (2) that the fricated phase in /ǀ͡χ/ etc. is the same for all click types, there might be coarticulatory adjustment between the place of articulation of the fricated phase in /ǀ͡χ/ etc. and the place of articulation of the posterior constriction during the lingual phase of the sound. In other words: The posterior *closure* is assimilated to the posterior *release* in terms of place of articulation, thus leveling the differences between the click types.

- Finally, it is conceivable that the reason for the different patterning is phonological: /͡ǀq/ etc. could be *units* and /͡ǀχ/ etc. could be *clusters* of click + pulmonic fricative, or possibly click + pulmonic affricate. While this could provide a convincing explanation for the fact that the fricated phase in /͡ǀχ/ is opaque with respect to the BVC, in my view the evidence suggests strongly that *all* clicks in N|uu as presented in Section 3.2 are indeed units. However, it has to be acknowledged that the evidence is not unambiguous regarding the patterning of /͡ǀχ/ etc.

To resolve this question, burst and frication properties of a suitably selected corpus of words were investigated acoustically.

5.2.1 Method

The data set for this second study was obtained in the same way as described in Section 5.1.1, thus consisting of sound files containing a total of 30 labeled segments for any given word. To test the a priori hypotheses outlined above, the posterior *stop burst* was segmented and labeled in plain and sequential linguo-pulmonic stops (e.g. /ǀ/ and /͡ǀq/); in linguo-pulmonic affricates (e.g. /͡ǀχ/), it was the posterior *affrication* that was segmented and labeled. Some discussion is in order to clarify my use of terminology, the labeling procedure, and the consequences that procedure has for the further analysis.

Following Stevens (1998), I take stop releases to be analyzable into a sequence as given in (5.2) (brackets denote an element that is not necessarily present):

(5.2) transient – frication phase – [aspiration]

The transient is the (very brief) impulse response of the vocal tract immediately following the release of the closure; the frication phase is the phase dominated acoustically by turbulent noise at the point of maximal constriction; and the aspiration (if present) is the phase where the main source of acoustic energy is (low-intensity) turbulence at the glottis.

Voice onset time (VOT) is calculated by taking all three phases together; what is generally referred to as the *burst* corresponds to the transient and frication phases taken together. I take affricate releases to be analyzable in the same way, except that I treat them as having a prolonged frication phase.[29]

Thus, returning to our segmentation problem, it is important to stress that in all three cases (plain and sequential linguo-pulmonic stops as well as linguo-pulmonic affricates), only the *frication phase* (and possibly the transient, which was rarely identifiable) of the posterior constriction release was labeled; great care was taken *not* to include any aspiration (if present). In other words, in all cases a portion of sound was labeled that is acoustically characterized solely by turbulent noise at the point of maximal constriction, namely, the posterior constriction.

One final point that merits discussion is the identification of the posterior burst in plain clicks. It has often been pointed out (cf. for instance Traill 1985) that the posterior constriction is released inaudibly, and in fact, that perceptual fact is the main justification for my term *simultaneously released* (as opposed to *sequentially released*) *linguo-pulmonic stops*. There is, however, a discrepancy here between the acoustic signal and the perceived event. Recall that the basic mechanism of click production (what I call the *click mechanism*) necessarily involves three aspects: (1) two closures, (2) expansion of the cavity between those two closures so that the air in it is rarefied below atmospheric air pressure, and (3) release of the anterior closure. It is a necessary condition for the production of a click that the anterior closure is released *before* the posterior one, because otherwise the difference between the air pressure in the cavity and the atmospheric air pressure[30] will be leveled before the anterior closure is released. If that happens, no appreciable release noise can be produced.

Therefore, what actually happens is that the two closures are released in very close sequence, so close that the posterior release is normally not perceived as a separate event. Acoustically, a large part of the posterior burst (the transient and part of the frication phase) is masked by the much more intense anterior burst, but the remaining, non-masked part of the posterior burst (on the order of 10 ms) is consistently identifiable in the acoustic signal.

29 Note, however, that neither of the terms *voice onset time* or *burst* is commonly used for affricates in the way defined here, so to avoid confusion, I will restrict the use of those terms to stops.

30 Strictly speaking, of course, the difference between the air pressure in the cavity and the *intraoral* air pressure would be leveled, which leads to the same result (because under the circumstances investigated here, the intraoral air pressure during speech is greater than or equal to the atmospheric air pressure).

Given the properly segmented sound files, Praat scripts were then written to obtain the following parameters for each of the 30 segmented and labeled frication phase tokens per word:

FFT spectra. FFT (*fast Fourier transform*) spectra provide one way to characterize and compare the spectral properties of sounds. In the present study, the overall shape of the spectra as well as the number, location, relative prominence, and relative bandwidth of major peaks was of potential interest. Primarily, however, FFT spectra were needed for the derivation of auditory spectra and for the calculation of spectral moments (see below). To ensure that only spectral information from the *frication phase* itself was used for the calculation of the spectra, the relevant stretches of speech (henceforth *intervals*) were extracted from their context (in other words, no information from preceding anterior bursts nor from following vowel onsets were included in the spectra). As before in the analysis of formant transitions, the middle of a Gaussian window with an effective length of 512 samples (corresponding to 23.22 ms at a sampling rate of 22050 Hz) was aligned with the middle of the interval, so that the samples in the middle of the interval were given more weight than those at the edges. Again, if the total duration of the interval was less than the time required to fit the analysis window to it (a situation commonly occurring with posterior bursts of simultaneously released clicks, as explained above), the extracted interval was multiplied with a Hanning window, and an equal number of zero samples (i.e. samples with a value of zero) were added to the beginning and end of the interval until the total duration of the interval equaled the analysis window length. Finally, an FFT spectrum was calculated from the interval.

LPC spectra. Secondly, LPC (*linear predictive coding*) spectra were computed for each interval using Praat's *autocorrelation* method. LPC spectra are better suited for gross visual inspection of turbulent sounds than FFT spectra because the overall *spectral envelope* is estimated, abstracting away from minor, sample-to-sample random fluctuations. The settings and general procedure for their calculation were generally the same as described above for the FFT spectra. A prediction order of 20 was found to give the best overall results for all three speakers.

Auditory spectra. Thirdly, auditory spectra were derived from the FFT spectra using Praat's *excitation* method (see Boersma, 1998, pp. 103–109) for a description and justification of the algorithm used). Auditory spectra attempt to model the representation of sounds in the human auditory system (with loudness [in phon] as a function of frequency [in Bark]) and therefore potentially offer better insights than acoustic spectra into which aspects of an acoustic signal are relevant, and which are not, from a human listener's perspective. Cf. Johnson (1993) for an example of a study that has advocated the use of auditory spectra, and successfully demonstrated their application.

Averaged spectra. To prevent individual differences in vocal tract shape between speakers from biasing the results of spectral investigation, spectra for all tokens of a given word were initially overlaid on each other, separated by speaker. Inspection of the resulting overlaid plots revealed, however, that the gross patterns were consistent across speakers; furthermore, the major spectral peaks of the individual tokens consistently tended to coincide, minor fluctuations canceling each other out. Therefore, it seemed justified to calculate averaged spectra (FFT, LPC, and auditory) over all 30 tokens for a given word. These have the distinct advantage of reducing the information in a spectrum in such a way that it is much easier to comprehend than in an overlaid plot of up to 30 individual spectra. The method is known as *ensemble averaging* (as opposed to *time averaging*) and is commonly applied to spectra containing a large amount of random fluctuation, typically fricative spectra. To arrive at such averaged spectra, I first calculated the individual spectra, then converted the magnitude of each spectrum to a linear scale, then added the spectra and divided the sum by the number of spectra, and finally converted the magnitude of the resulting averaged spectrum back to a logarithmic scale (decibels and phons, respectively). Figure 5.13 illustrates a typical ensemble-averaged LPC spectrum superimposed on the 30 individual spectra that it is based on.

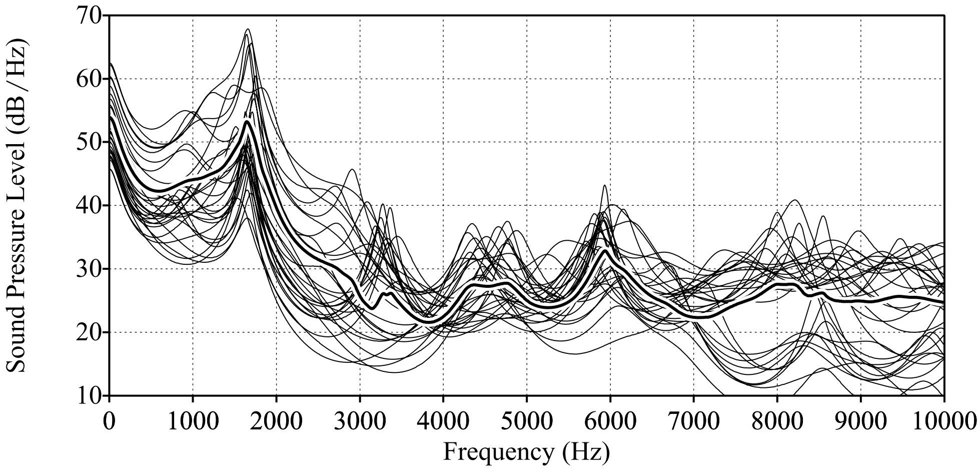

Figure 5.13 Ensemble-averaged LPC spectrum of the stop burst / frication phase midpoint for /k/ (bold line), superimposed on the 30 tokens that it is based on (light lines)

Waterfall spectra. While ensemble-averaged spectra represent a convenient way to reduce the information in a set of spectra, they do not provide any insights into changes in the spectra *over time*. For this purpose, so-called *waterfall spectra* – or more specifically, *ensemble-averaged* waterfall spectra – have been produced (again using the FFT, LPC, and auditory methods described above). For that purpose, a certain length of time, starting at the beginning of each labeled interval, and a certain time step (typically 5 ms) were chosen; then, ensemble-averaged spectra were computed at each time step within the predetermined length of time: For example, within a given frication interval, a 50 ms portion starting at the left edge of the interval would have been chosen and spectra computed every 5 ms within that portion, resulting in a total of 11 spectra. These spectra were then displayed graphically for inspection.

Spectral center of gravity. The spectral center of gravity, or first spectral moment[31] (Forrest, Weismer, Milenkovic, & Dougall, 1988), has gained importance in recent times as a parameter that characterizes the distribution of energy within

31 Higher spectral moments (particularly the second, third, and fourth moment, or quantities derived thereof) have been used in the literature, but their diagnostic value has not been as convincing as in the case of the first moment (cf. Fulop, Ladefoged, Liu, & Vossen, 2003, for an application of spectral moments to the acoustics of clicks).

a spectrum: A low center of gravity indicates that relatively more energy is concentrated in the lower frequency ranges (*grave* spectra in Jakobsonian terms), whereas a high value corresponds to a spectrum with more energy in the higher frequency ranges (*acute* spectra). Center of gravity is a very convenient spectral parameter in that it is *quantifiable*; that is to say, it captures an important spectral property that can be expressed as a single value and for instance be used as a basis for statistical analyses. Furthermore, in contrast with the major peak measure discussed above, it can be calculated automatically from any given spectrum. Care should be taken, however, not to rely on a numerical value such as center of gravity alone without comparing the actual *shapes* of the spectra that it describes: Center of gravity is reliable first and foremost if the corresponding spectrum has a single prominent peak (*compact* spectra). In such cases, it corresponds quite well to place of articulation. If, on the other hand, the spectrum is essentially flat with no prominent peaks (*diffuse* spectra), or if it has multiple peaks, the resulting center of gravity value is potentially hard to interpret or even misleading. For the present study, center of gravity values of FFT and auditory spectra have been calculated in Praat to complement the major peak measure as well as the various graphical displays described above, and to facilitate further statistical analysis.

Major spectral peak. As explained above, the spectra coincided quite consistently in terms of their major peaks, the most prominent of which was subsequently identified numerically in LPC and auditory spectra for statistical analysis. Due to the larger number of outliers in the LPC spectra, and because it was not always possible to decide unambiguously which peaks corresponded to each other in individual spectra, major peak values from auditory spectra were preferentially used.

Formant transitions. Finally, to determine whether any notable coarticulatory influence (particularly through retraction of the tongue body) is exerted by the different consonants on the following vocalic onset, F_2 values at the first glottal pulse after the consonantal release were measured, using the same method as detailed in Section 5.1.1 above.

The aim of the statistical analysis to which the major peak and the center of gravity values were subsequently subjected (over and above the descriptive statistics, such as mean and standard deviation) was to investigate whether the values differed between the different consonants in a *statistically significant* way; in other words, it was tested whether major peak or center of gravity as parameters could serve as reliable predictors for the identity of a given consonant. For this purpose, LME models were fitted to the data as explained in Section 5.1.1. For purposes of comparison, classical one-way repeated measures ANOVAs were also conducted.

A corpus of 8 words was selected for the study. Each word begins with a different targeted consonant, followed by the vowel /a/. (Note that due to the combination of an extremely high number of consonant, vowel, and tone contrasts on the one hand and the comparatively small-sized lexicon that was available to me on the other hand, it was practically impossible to find true minimal pairs for most contrasts. However, the words in the present corpus are sufficiently similar phonetically to provide a controlled context for the acoustic investigations at hand.) The following words were chosen:

- *ká* ‘say’
- *χàá* ‘scratch’
- *|àá* ‘hold’
- *!áà* ‘red hartebeest (*Alcelaphus caama*)’
- *|͡qàá* ‘shine’
- *!͡qá*^{n}*à*^{n} ‘migrate’
- *|͡χáá* ‘edge’
- *!͡χáá-kè* ‘stretch mark’ (*-kè* ‘plural’)

This corpus includes two items (one with a laminal and one with an apical click type) from each of the classes of (plain) linguo-pulmonic stops (dental /|/ and alveolar /!/), linguo-pulmonic sequential stops (/|͡q/, /!͡q/), and linguo-pulmonic affricates (/|͡χ/, /!͡χ/) as well as the pulmonic stop /k/ and the pulmonic fricative /χ/ for comparison.

The dental and alveolar click types are taken here as representative for the classes of laminal and apical clicks, respectively, because of their identical phonological patterning and on the grounds of the plausibility of the suggested physiological basis of that patterning. In any event, a complete set of contrasts for the present corpus that would represent all five click types in N|uu is at present not possible (due to the restrictions of my lexicon, especially for the bilabial click type).

A note on the points of measurement for the different words is in order here. For stops (/k/, /|/, /!/, /|͡q/, /!͡q/), measurements were taken at the midpoint of the frication phase of the burst (the posterior burst in the case of the clicks). For fricatives (/χ/) and affricates (/͡ǀχ/, /͡ǃχ/), the point of measurement was also the midpoint of the frication phase. In certain cases, however, the *onset* of the frication phase, immediately after the release, served as the point of measurement. If that is the case, it will be explicitly noted in the text.

5.2.2 Results

As a starting point, spectral center of gravity values were computed and visually inspected; an LME model was then fitted to those values to assess the statistical significance of the observed differences.

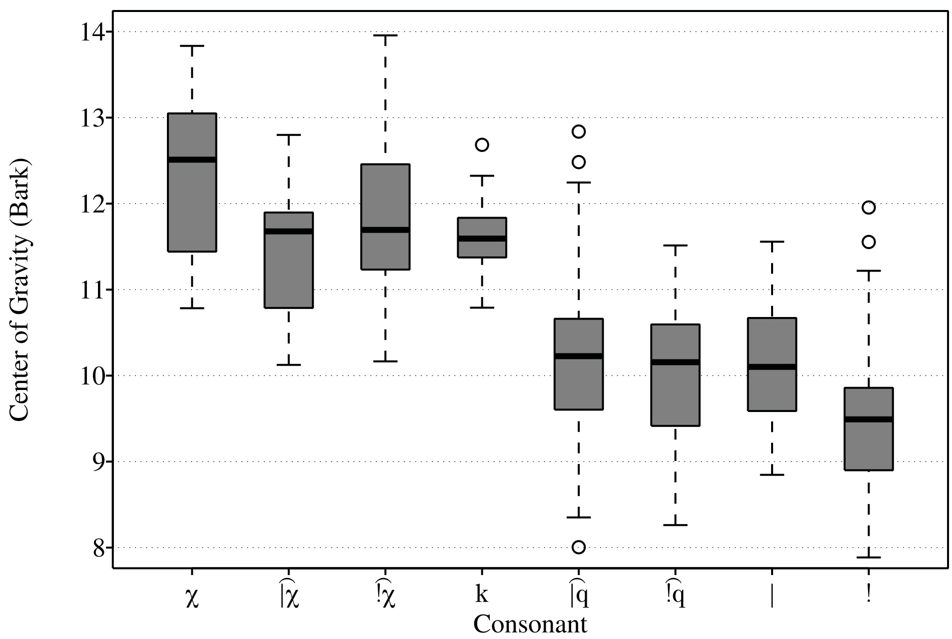

Figure 5.14 Box plots of the spectral center of gravity of the stop burst / frication phase midpoint for 8 consonants (N = 240)

Figure 5.14 shows box plots of the center of gravity values thus obtained. Four groups can be roughly identified (all values are in Bark):

1. high frequency: /χ/ (M = 12.34, SD = 0.93);

2. high-mid frequency: /ǀ͡χ/ (*M* = 11.45, *SD* = 0.69), /ǃ͡χ/ (*M* = 11.85, *SD* = 0.92), [k] (*M* = 11.56, *SD* = 0.45);
3. low-mid frequency: /ǀ͡q/ (*M* = 10.31, *SD* = 1.11), /ǃ͡q/ (*M* = 10.08, *SD* = 0.91), /ǀ/ (*M* = 10.16, *SD* = 0.73);
4. low frequency: /ǃ/ (*M* = 9.61, *SD* = 0.98).

An LME model that was fitted to the data (henceforth: LME ANOVA) revealed a significant main effect for consonant, $p_{\mathrm{MCMC}} < 0.001$. A one-way repeated measures ANOVA that was conducted for comparison yielded the same result, $F(7, 14) = 9.81$, $p < 0.001$; however, sphericity was not controlled, so the results of the latter test need to be interpreted with caution.

Pairwise comparison of the individual levels of the consonant factor resulted in the same four groupings as given above, with *p* values as summarized in Table 5.1.

Table 5.1 Pairwise comparison of differences between center of gravity values of individual levels of the consonant factor (based on p_{MCMC} values; ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, n.s.: not significant)

	χ	ǀ͡χ	ǃ͡χ	k	ǀ͡q	ǃ͡q	ǀ
ǀ͡χ	***	—	—	—	—	—	—
ǃ͡χ	*	***	—	—	—	—	—
k	***	n.s.	n.s.	—	—	—	—
ǀ͡q	***	***	***	***	—	—	—
ǃ͡q	***	***	***	***	n.s.	—	—
ǀ	***	***	***	***	n.s.	n.s.	—
ǃ	***	***	***	***	***	*	**

As explained above, center of gravity as a quantity should be complemented with other measures to compensate for the fact that center of gravity only yields gross results for the concentration of energy in a spectrum, abstracting away from the location of peaks and valleys and their relative positions and magnitudes. As we will see below, all consonant releases (frication phases) under investigation here show a single prominent peak in their spectra, as expected for segments produced in the velar–uvular area. Therefore, major spectral peak serves well as such a complementary quantity. Using the same target words as for the center of gravity measurements, the major peak in each auditory spectrum was identified in Praat; Figure 5.15 shows box plots of the distribution of those values by consonant type.

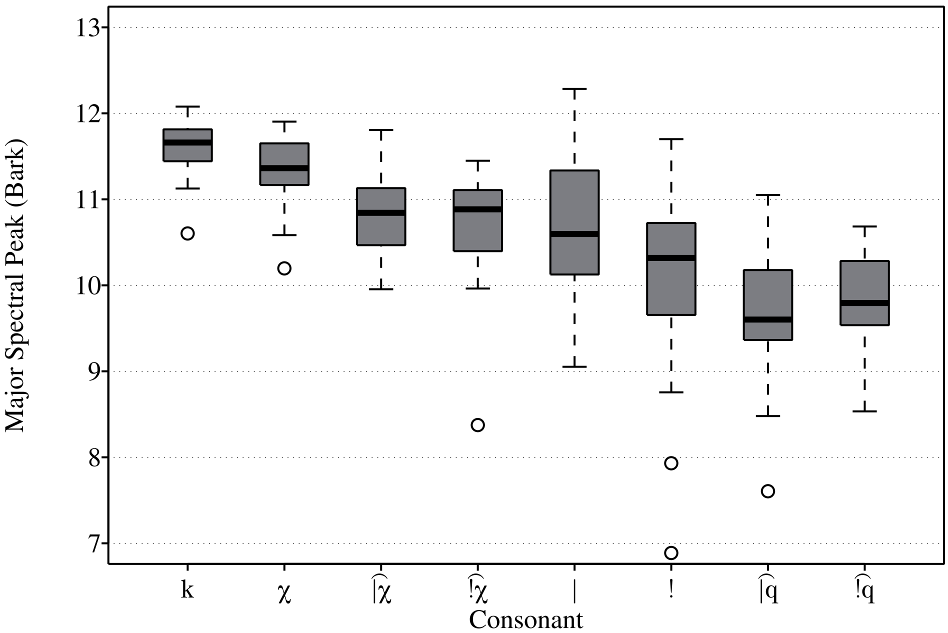

Figure 5.15 Box plots of the major spectral peak of the stop burst / frication phase midpoint for 8 consonants (N = 240)

As in the case of the center of gravity, four groups can again be identified; however, these groups are not the same as those seen before in terms of their membership (again, all values are in Bark):

1. high frequency: /k/ (M = 11.62, SD = 0.32), /χ/ (M = 11.33, SD = 0.41);
2. high-mid frequency: /|͡χ/ (M = 10.62, SD = 0.58), /!͡χ/ (M = 10.70, SD = 0.60), /|/ (M = 10.72, SD = 0.81);
3. low-mid frequency: /!/ (M = 10.10, SD = 1.02);
4. low frequency: /|͡q/ (M = 9.69, SD = 0.70), /!͡q/ (M = 9.85, SD = 0.55).

An LME ANOVA revealed a significant main effect for consonant, $p_{\mathrm{MCMC}} < 0.001$. A one-way repeated measures ANOVA (without sphericity corrections) yielded the same result, $F(7, 14) = 8.75$, $p < 0.001$.

The visual categorization of the consonants is again largely born out by pairwise comparison of the individual p values, as shown in Table 5.2. The only value that does not conform to this clear-cut categorization is the one for /ǃ/, which is intermediate between the high-mid frequency and the low frequency group; while /ǃ/ vs. /ǀ͡q/ is significantly different, /ǃ/ vs. /ǃ͡q/ unexpectedly does not reach significance (despite the significant difference between /ǀ͡q/ and /ǃ͡q/).

Table 5.2 Pairwise comparison of differences between major peak values of individual levels of the consonant factor (based on p_{MCMC} values; ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, n.s.: not significant)

	k	χ	ǀ͡χ	ǃ͡χ	ǀ	ǃ	ǀ͡q
χ	n.s.	—	—	—	—	—	—
ǀ͡χ	***	***	—	—	—	—	—
ǃ͡χ	***	***	n.s.	—	—	—	—
ǀ	***	***	n.s.	n.s.	—	—	—
ǃ	***	***	**	***	***	—	—
ǀ͡q	***	***	***	***	***	*	—
ǃ͡q	***	***	***	***	***	n.s.	n.s.

The appropriateness of comparing midpoints of (frication phases of) stop bursts with midpoints of frication phases of affricates merits some discussion at this point. Are those two points of measurement directly comparable at all? I take the position that they are. In terms of source–filter theory, in both cases the frication noise at (or shortly downstream from) the constriction is the main source of acoustic energy, with the cavities before and behind the constriction being insufficiently coupled for the back cavity to make any substantial spectral contribution to the resulting signal. (Note that this is only true when no aspiration noise is included in the interval used for spectral analysis, which, however, I took account of when labeling the data.) As Stevens (1998, p. 377) points out for voiceless unaspirated stop consonants, “the filtering of the transients and the frication sources is determined primarily by the dimensions of the cavity anterior to the constriction.” It becomes clear from the discussion in Stevens (1998, chap. 7–8) that a certain amount of acoustic coupling between the front and back cavities can occasionally be observed in the production of both voiceless unaspirated stops and fricatives (and, by extension, affricates) in real speech data. But by and large, the assumption appears justified that the source mechanism of the frication phases in stops, fricatives, and affricates is sufficiently similar to be subjected to direct comparison, as Johnson (2003, p. 141) states concisely:

> The burst is essentially a transient …. This noise source is then shaped by the resonances of the portion of the vocal tract in front of the stop closure, because the closure at the time of the release is still quite narrow, and thus the front and back cavities of the vocal tract are not acoustically coupled ….

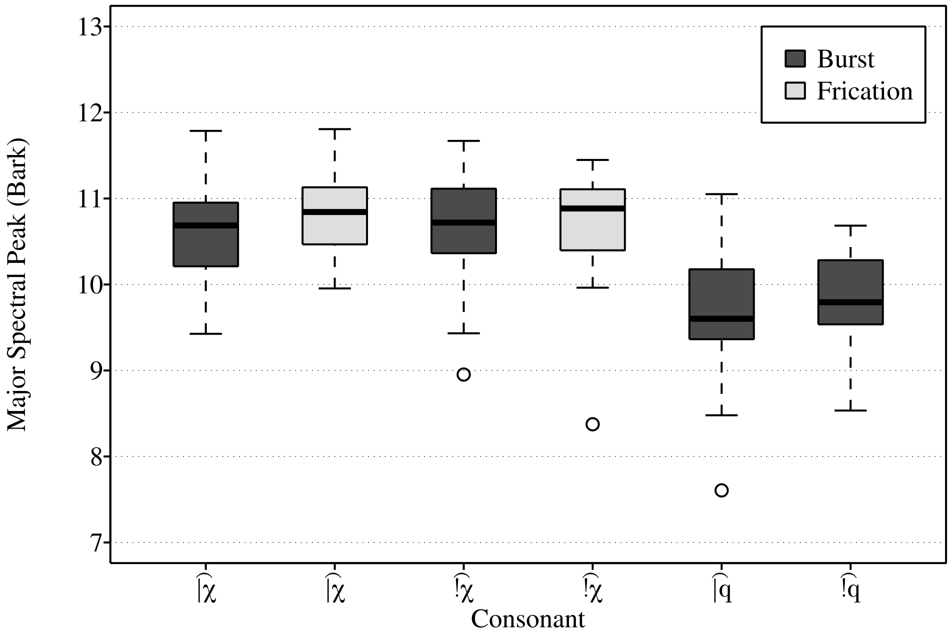

Figure 5.16 Box plots of the major spectral peak of the stop burst / frication phase onset (labeled 'Burst') vs. midpoint (labeled 'Frication') for 3 consonants (N = 180)

To consolidate this position, six classes of major spectral peak values were calculated and compared statistically: /ǀ͡χ/, /!͡χ/ at the onset of the frication phase (i.e. immediately following the release); /ǀ͡χ/, /!͡χ/ at the midpoint of the frication phase; and /ǀ͡q/, /!͡q/ at the onset of the (frication phase of the) burst. The prediction that major peak values would *not* differ significantly between onset and midpoint for the affricates, but that all four would be significantly different from the values for the stops, is confirmed by the data, as seen in Figure 5.16.

An LME ANOVA showed a significant main effect for consonant, $p_{\text{MCMC}} < 0.001$. A one-way repeated measures ANOVA (without sphericity corrections) gave the same result, $F(5, 10) = 18.78$, $p < 0.001$. A pairwise comparison revealed that *within* the groups of affricates and stops, respectively, no significant differences existed, whereas all pairs *between* the groups showed highly significant differences at $p < 0.001$. This shows very clearly that for the affricates, no significant spectral differences exists between the onset of the frication phase (which is temporally comparable to the burst of the corresponding stops) and the frication phase 'proper'. This has two important consequences: (1) The (linguo-pulmonic) stops and affricates can legitimately be directly compared with respect to their spectral properties; and (2) the linguo-pulmonic affricates are indeed true *homorganic* affricates in that the (posterior) place of articulation after the release remains the same from the release and throughout the frication phase.

To illustrate this further, ensemble-averaged LPC waterfall spectra of the first 50 ms (in steps of 5 ms) following the release of the posterior closure are shown in Figure 5.17 for /ǀ͡χ/.

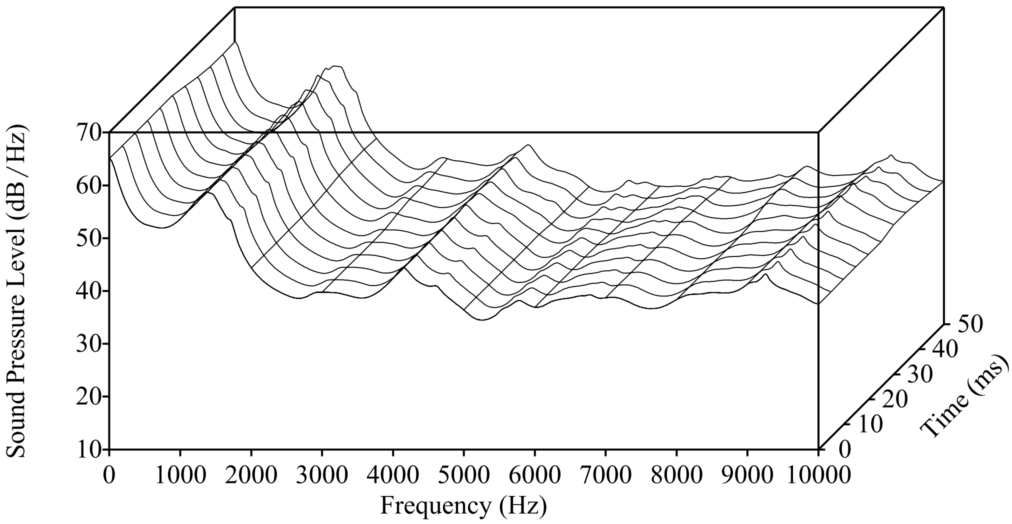

Figure 5.17 LPC waterfall spectrum of the first 50 ms of the frication phase following the release of the posterior closure in /ǀ͡χ/ (each slice was ensemble-averaged over 30 tokens)

As becomes apparent, the visual impression fully confirmed the results of the statistical tests above, namely that the major spectral peak, and indeed all other peaks, do not change appreciably over time between the moment of release and the last slice, 50 ms into the frication phase. Comparison with a corresponding waterfall spectrum for /ǃ͡χ/, as illustrated in Figure 5.18, yielded the same result.

Having shown the legitimacy of the comparisons conducted so far, I then proceeded to inspecting the graphical representations of the individual spectra to find out the reason for the discrepancies noted above between the categorizations based on center of gravity on the one hand and on major peak on the other hand.

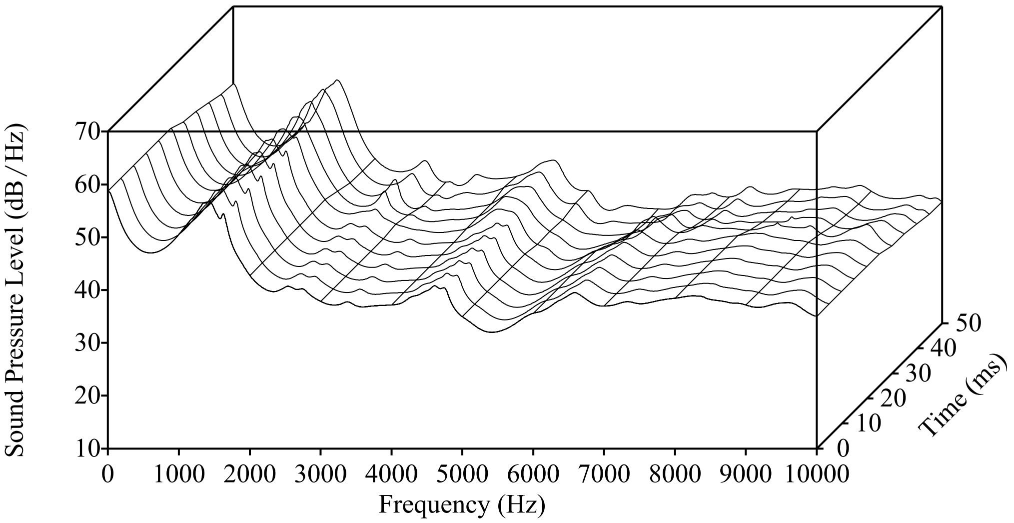

Figure 5.18 LPC waterfall spectrum of the first 50 ms of the frication phase following the release of the posterior closure in /ǃ͡χ/ (each slice was ensemble-averaged over 30 tokens)

The first object of interest were the (again, ensemble-averaged) LPC spectra of /ǀ/, /ǀ͡q/, and /ǀ͡χ/, which have been superimposed on each other in Figure 5.19 to facilitate comparison.

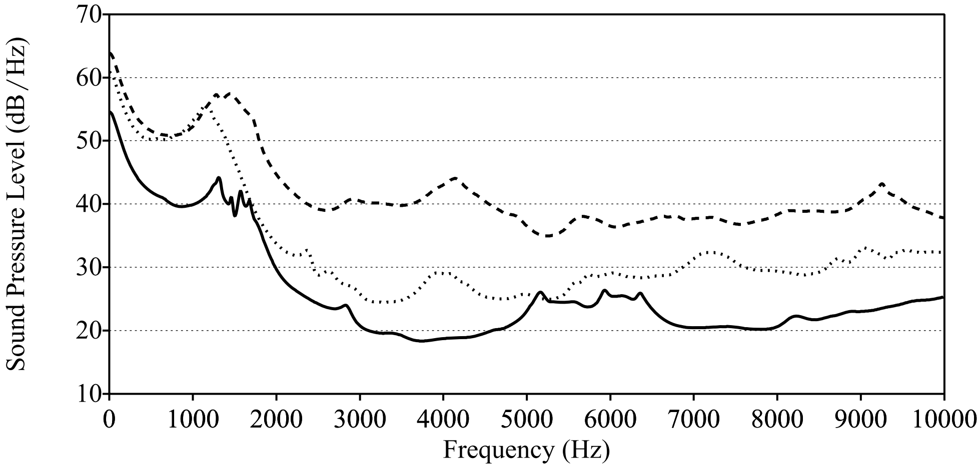

Figure 5.19 LPC spectra of the stop burst / frication phase midpoint for /ǀ/ (solid line), /ǀ͡q/ (dotted line), and /ǀ͡χ/ (dashed line). Each spectrum was ensemble-averaged over 30 tokens.

The spectra differed from each other with respect to two very different properties: (1) the location of the major peak (this is not quite as obviously visible at this frequency resolution as might have been expected, but it was, nevertheless, statistically significant, as we have seen above); and (2) the *overall magnitude* of the spectral values, particularly in the frequency region *above* the major peak. Table 5.3 provides a summary of the mean major spectral peak values and band energy differences for the three consonants under discussion. The former are based on the same values as used above for statistical analysis; thus, they were derived from auditory spectra but converted to Hertz for convenience. For the latter, the spectra were conveniently divided into three frequency bands: (1) a low-frequency band between 0 and 1000 Hz; (2) a mid-frequency band between 1000 and 2000 Hz, containing the major peak; and (3) a high-frequency band between 2000 and 10000 Hz. Band energies were then calculated for the high-frequency band for all three consonants. Taking the band energy of /ǀ/ (the lowest) as a reference value, the other two values were compared with it.

Table 5.3 Mean and standard deviation of major spectral peak values (derived from auditory spectra and converted to Hertz) and band energy differences (2000–10000 Hz) for ensemble-averaged LPC spectra of the stop burst / frication phase midpoint for /ǀ/, /ǀ͡q/, and /ǀ͡χ/ (*N* = 90)

	Major spectral peak (Hz)		Band energy difference (dB)
	Mean	Standard deviation	
ǀ	1433.69	75.13	0
ǀ͡q	1216.20	64.72	7.09
ǀ͡χ	1411.14	54.27	15.67

With reference to these two properties (major peak frequency and high-frequency band energy), two generalizations could be made:

1. /ǀ/ and /ǀ͡χ/ had *high* frequency peaks, while /ǀ͡q/ had a *low* frequency peak;
2. /ǀ/ had a *low* band energy above the peak, while /ǀ͡q/ had a *medium* and /ǀ͡χ/ had a *high* band energy above the peak.

These results were now compared to the situation with respect to the 'back' clicks, /ǃ/, /ǃ͡q/, and /ǃ͡χ/. Figure 5.20 shows the corresponding overlaid LPC spectra for those three consonants.

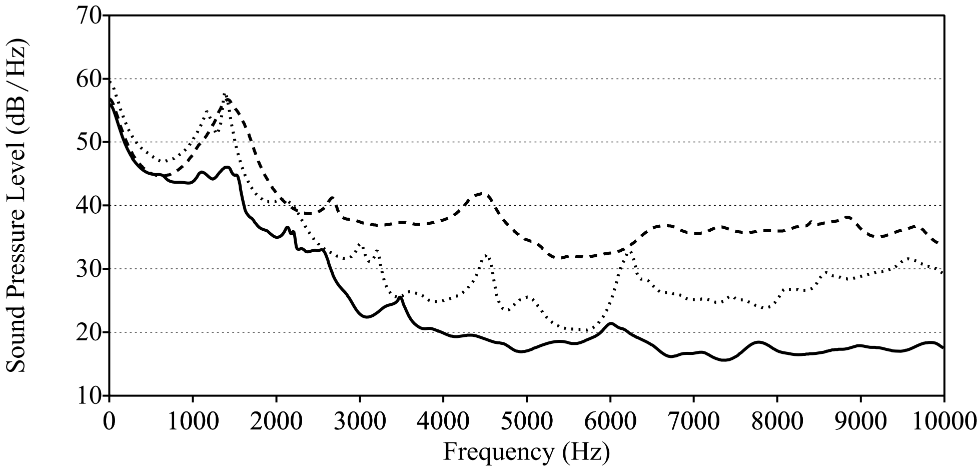

Figure 5.20 LPC spectra of the stop burst / frication phase midpoint for /!/ (solid line), /!͡q/ (dotted line), and /!͡χ/ (dashed line). Each spectrum was ensemble-averaged over 30 tokens.

As can be seen, the situation was found to be very similar to the one described above for the laminal clicks, /|/, /|͡q/, and /|͡χ/, the only obvious difference being in the spectrum for /!/; we will get to that below. Table 5.4 again summarizes the corresponding mean major peak and band energy difference values.

Table 5.4 Mean and standard deviation of major spectral peak values (derived from auditory spectra and converted to Hertz) and band energy differences for 2000–10000 Hz (BED) for ensemble-averaged LPC spectra of the stop burst / frication phase midpoint for /!/, /!͡q/, and /!͡χ/ (*N* = 90)

	Major spectral peak (Hz)		BED (dB)
	Mean	Standard deviation	
!	1299.34	94.63	0
!͡q	1248.23	51.16	5.33
!͡χ	1428.72	55.54	11.05

The next step was to compare the spectra for /|/ and /!/ directly with each other. Recall that the two seemed to show notably different spectra which departed from the pattern displayed by all other clicks investigated thus far, for which the click *type* did not seem to make any significant spectral difference. Figure 5.21 shows superimposed LPC spectra for /ǀ/ and /!/.

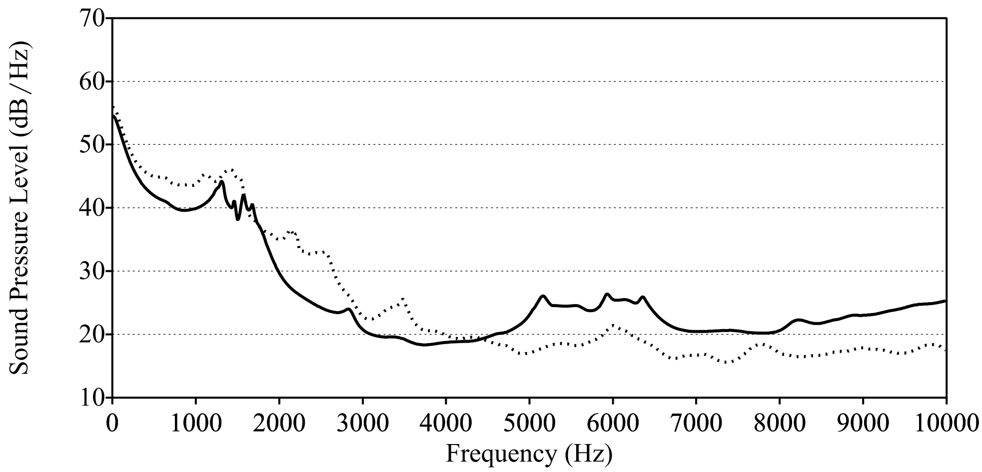

Figure 5.21 LPC spectra of the stop burst / frication phase midpoint for /ǀ/ (solid line) and /!/ (dotted line). Each spectrum was ensemble-averaged over 30 tokens.

It was noticable that the spectrum for /ǀ/, as compared with the one for /!/, showed slightly *less* energy in the region between 0 and 1000 Hz (auditory spectra showed a minor peak around 640 Hz for /!/), but slightly *more* energy in the region between 5000 and 6000 Hz (here, auditory spectra revealed a fairly pronounced minor peak around 5690 Hz for /ǀ/) and above.

To complete this study of spectral shape, the pulmonic consonants, /k/ and /χ/, were compared with the different linguo-pulmonic consonants investigated thus far. Cursory inspection showed that the spectral shape of /k/ corresponded closely to that of (the posterior burst of) /ǀ͡q/ and /!͡q/, but differed from the latter by having a distinctly higher spectral peak (*M* = 1646.26 Hz, *SD* = 29.85 Hz). Figure 5.22 shows overlaid LPC spectra for /k/, /ǀ͡q/, and /!͡q/.

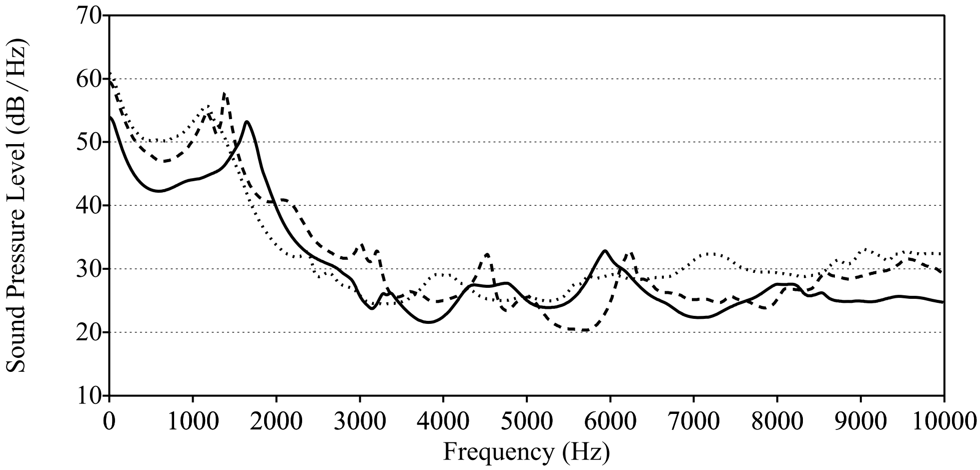

Figure 5.22 LPC spectra of the stop burst / frication phase midpoint for /k/ (solid line), /ǁ͡q/ (dotted line), and /!͡q/ (dashed line). Each spectrum was ensemble-averaged over 30 tokens.

Similarly, /χ/ corresponded to /ǁ͡χ/ and /!͡χ/ in its spectral shape, again differentiated from these by a higher peak (*M* = 1575.41 Hz, *SD* = 37.67 Hz), as shown in Figure 5.23.

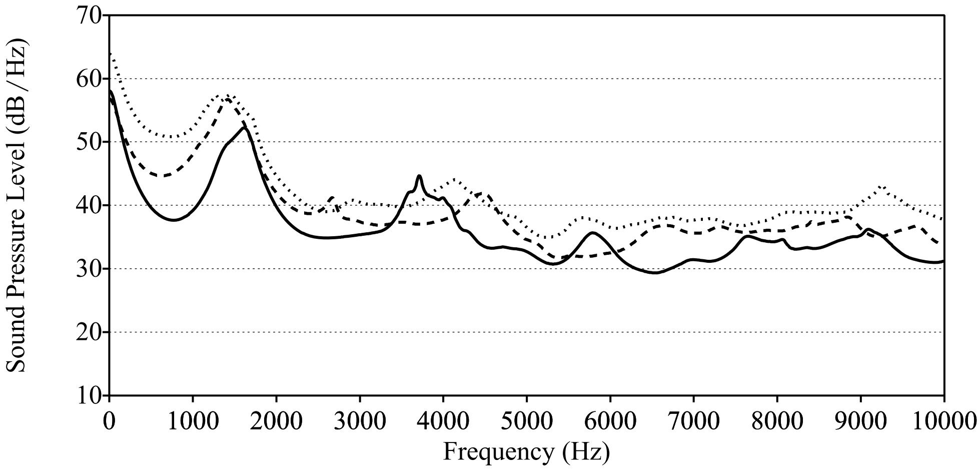

Figure 5.23 LPC spectra of the stop burst / frication phase midpoint for /χ/ (solid line), /ǁ͡χ/ (dotted line), and /!͡χ/ (dashed line). Each spectrum was ensemble-averaged over 30 tokens.

As a last source of evidence, F_2 at the first glottal pulse for the vowel /a/ was measured following /k/, /ǀ/, /!/, /ǁ͡q/, and /!͡q/ to find out whether any differences in place of articulation between the consonants were reflected in the formant transitions of a following vowel. A fairly clear pattern emerged, as evidenced in Figure 5.24.

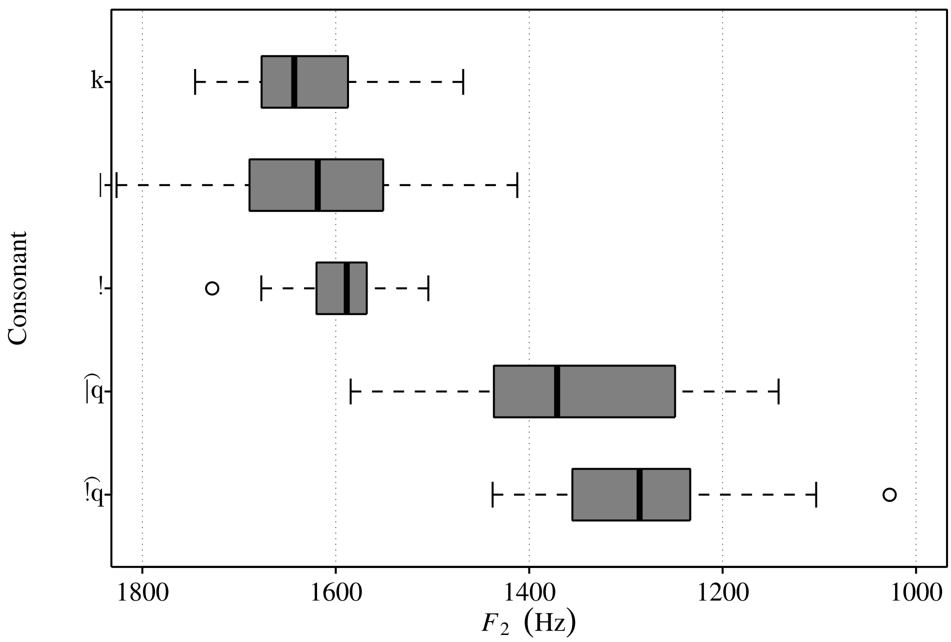

Figure 5.24 F_2 at the onset of vocal fold vibration for the vowel /a/ following 5 different consonants (N = 150)

Visual inspection led to the identification of 2 distinct groups, one with a higher F_2 (/k/, /ǀ/, and /!/), implying a more fronted articulation for the vowel, and one with a lower F_2 (/ǀ͡q/ and /!͡q/), implying a more retracted articulation. An LME ANOVA showed a significant main effect for consonant, $p_{\mathrm{MCMC}} < 0.001$. A one-way repeated measures ANOVA (without sphericity corrections) also indicated a significant main effect, $F(4, 8) = 25.00$, $p < 0.001$. A pairwise comparison revealed that *within* the more fronted group no significant differences existed, whereas all pairs *between* the groups showed highly significant differences at $p < 0.001$. Interestingly, however, despite the considerable overlap in the data as seen in the box plots, even /ǀ͡q/ and /!͡q/ differed significantly from each other at $p = 0.001$. This latter fact is at present unexplained.

5.2.3 Discussion

While the various pieces of evidence presented in the previous section might seem confusing at first, I think that several patterns emerge from the data that allow us to formulate a comprehensive account of the posterior articulation in click consonants that is consistent with the observed data. Table 5.5 summarizes the insights we have gained in the previous section, classifying the consonants under discussion by means of two acoustic parameters, (1) relative frequency of the major spectral peak and (2) relative energy in the high frequency band between 2000 and 10000 Hz, which was found to be the most robust of the parameters characterizing overall spectral shape for the present data.

Table 5.5 Classification of consonants by major peak frequency and high frequency band energy (HFBE)

HFBE	Major peak frequency			
	High	High-mid	Low-mid	Low
Low		ǀ	ǃ	
Mid	k			ǀ͡q, ǃ͡q
High	χ	ǀ͡χ, ǃ͡χ		

The interesting question now arises how those two acoustic parameters can be interpreted in articulatory terms. I think the following two assumptions are plausible and justified:

1. Major peak frequency is a fairly straightforward reflection of the length of the cavity anterior to the maximal constriction in the vocal tract, or, in other words, of the traditional *place of articulation* category (the *posterior* place of articulation in the case of clicks); cf. Stevens (1998, chap. 7–8) for a detailed discussion.

2. I take differences in the overall energy of major spectral regions (such as the high frequency band in this case) between two given spectra that otherwise share (more or less) the *same distribution of peaks and valleys* to be attributable, at least to a large degree, to acoustic damping by the vocal tract walls

(which in turn is determined by the mechanical compliance of the articulators). In more articulatory terms, this would translate to increased muscular *stiffness* (tenseness) of the articulators leading to less damping and hence a higher value for high frequency band energy.

Let us simplify the facts slightly for the moment and assume that /!/ is placed in the high-mid major peak frequency category along with /|/ (we will differentiate between the two again shortly). Looking at Table 5.5 again now, it becomes apparent that we can re-interpret our two acoustic parameters in a way that makes the distribution of the consonants in the table very plausible:

- As noted above, major peak frequency can be interpreted as a measure of *place* of articulation. In this case, *high* would correspond to velar, *high-mid* to back velar (or front uvular), and *low* to uvular. (Because the acoustic values only give us *relative* orientation, other sequences are conceivable, too, like e.g. back velar / front uvular – uvular – back uvular.)[32]

- As for the high frequency band energy parameter, there is a correspondence to *manner* of articulation in a broader sense: *low* corresponds to 'simultaneously' released clicks, *mid* to traditional stops (including the posterior release of sequentially released clicks), and *high* to traditional fricatives (and also including affricated clicks).

So how are the individual consonants placed in their particular categories? We will first take up the major spectral peak parameter, now reinterpreted as *place* parameter.

I postulate that on an underlying level only *two* articulatory targets are potentially involved in the production of the (posterior constrictions of the) consonants in question here: (1) velar (or back velar / front uvular), as for /k/ and arguably /χ/; and (2) uvular (or back uvular), as for /|͡q/ and /!͡q/, all other variation

32 One question that remains open for the time being is why [χ] is classified in the same place of articulation category as [k], judging from the criterion of major spectral peak. Auditorily, [k] is clearly *not* uvular, whereas [χ] frequently shows typically uvular properties, such as irregular vibrations in the signal likely to result from trilling of the uvula. My hypothesis is that [χ] has a *longer* constriction than [k], so that their anterior edges coincide but the constriction of [χ] extends far enough back to induce uvular trilling through the Bernoulli effect.

being explainable through coarticulatory processes of different kinds. In the light of the evidence brought forward in this section, I postulate the sequence of articulatory events for the different consonants to be as follows:

1. /k/ and /χ/ are straightforward implementations of a specific articulatory target, namely velar (or back velar / front uvular; see above for the problem of connecting the acoustic evidence with an *absolute* articulatory label).

2. For the dental click /|/, the situation is more complicated; this is due to what I have termed the *click mechanism*. At the moment when the two closures for the click are formed, two independent articulatory targets are realized: one for the front closure (which we will not discuss further in this section) and one for the back closure. Two pieces of evidence bear on the fact that, pace Miller et al. (2007, 2009), this initial articulatory target for the back closure is also *velar*. Firstly, in clicks involving nasality (like e.g. /$^{\text{ŋ}}$|/), the part at the onset of the segment which has (1) pulmonic nasal airflow and which shows (2) nasal resonances that are characterized by the posterior click closure is auditorily velar, not uvular. Even more clearly, in N|uu, clicks with voiceless nasality (like e.g. /$^{\text{ŋ̊}}$|$^{\text{h}}$/ or /$^{\text{ŋ̊}}$|$^{\text{ʔ}}$/) are frequently assimilated in voicing to a preceding vowel, leading to what has been described as an 'intrusive' nasal between the end of the vowel and the beginning of the click. Also, in languages that have both clicks and prenasalization, such as e.g. Zulu, the (by definition homorganic) nasal portion of prenasalized clicks is also described as velar. The second, and in my view even more important, piece of evidence is that there are several tokens in my data of what I would call *defective clicks*, namely clicks lacking the click mechanism, presumably due to fatigue of the articulators following larger numbers of repetitions of the same word. In those cases, the air pressure in the lingual cavity is not reduced sufficiently to produce a salient front burst; instead, the back closure is released audibly, so a word such as *|àá* 'hold' would, after numerous repetitions, occasionally be realized as [kàá]. There are examples of such defective clicks for all click types in my data. –Next, after the formation of the two closures, the cavity between them is expanded. As a result of this, the *posterior edge* of the cavity recedes slightly.[33] It is important to note that the articulatory target at this

33 At the same time, the *anterior* edge *advances* slightly, which, however, is only relevant for the spectral characteristics of the *anterior* burst and will not be further discussed here.

stage is to *maintain* the two existing closures while at the same time expanding the cavity by forcefully pulling the center of the tongue downwards and backwards (presumably through the action of the hyoglossus muscle). I see the recession of the posterior edge of the cavity as a more-or-less passive consequence of this cavity enlargement, and thus an example of coarticulatory adjustment. But because the burst noise at the release of the posterior closure is produced with a pulmonic *egressive* airstream, it is precisely that point – the posterior edge of the cavity or, in other words, the *anterior* edge of the back closure – that is responsible for the frequency location of the major peak in the spectrum.

3. As far as the alveolar click /!/ is concerned, the same line of argument as for the dental click /|/ applies in principle. However, as we have seen, the major spectral peak for /!/ is distinctly (and significantly) lower than for /|/. This is explained straightforwardly by looking at the fundamentally different shapes of the tongue involved in the production of /!/ (and other 'back' clicks) as opposed to /|/ (and other laminal clicks). Recall that in /!/ etc., the apical articulation leads to a *deep concave* tongue shape. This results in a much larger cavity than for the *shallow concave* tongue shape involved in /|/ etc. It is plausible to assume that the *size* of the cavity correlates directly with the *degree of recession* of the posterior edge of the cavity, resulting in a systematically lower frequency location for the major spectral peak in /!/. One decisive argument for the assumption that the same articulatory target is involved in the production of /k/, /|/, and /!/ (as opposed to /|͡q/ and /!͡q/, see below) could be seen above: I interpret the fact that all three show the same formant transition onset for a following vowel /a/ as reflecting that the gross position of the *center of the tongue* (meaning the center of the tongue body itself, not of the dorsum) is by and large the same for all three segments, which would not be the case if there were an active retraction gesture. Rather, the data imply that the tongue body aims at maintaining a constant position, so to speak, counteracting the forces that act upon the tongue tip/blade (by the formation of the anterior constriction) and the tongue root (by the downward/backward pull for the expansion of the lingual cavity).

4. Turning to /|͡χ/ and /!͡χ/, I have already shown that the release of the posterior closure in fact constitutes a true homorganic affricate. The obvious question of why the frication phase of these two segments differs significantly from the one seen in the production of /χ/ can be answered with reference to what

has been said about the production of /ǀ/ and /!/ above: First, the point of articulation recedes due to cavity expansion. Then, an affricated (instead of abrupt, as with /ǀ/ and /!/) release takes place at that ‘modified’ point of articulation. Thus, the lower major peak frequency of /ǀ͡χ/ and /!͡χ/ is also the result (although more indirect) of coarticulatory accomodation to the expanding lingual cavity.[34]

5. Finally, /ǀ͡q/ and /!͡q/: As we have seen, these two sequentially released clicks have a lower major peak than the corresponding simultaneously released clicks, /ǀ/ and /!/. Moreover, they do *not* differ from each other with respect to their posterior place of articulation. The two facts that (1) their posterior closure is released after a considerable temporal lag and that (2) they show a distinct place of articulation for that closure strongly suggest that a different articulatory target is implemented than in the segments discussed so far, namely a uvular (or back uvular) place of articulation. My hypothesis is that this is because of the requirement for the (hitherto *passive*) retraction of the posterior edge of the lingual cavity to be held over a longer period of time, namely until after the sequentially released posterior burst. This ‘stabilization’ of the retraction requires *active* muscular control and results not only in even further retraction of the point of articulation (as seen in the major spectral peak values) but also in actual *tongue body* retraction (As evidenced in the formant transitions at the onset of a following vowel).

In sum, my analysis agrees with the account given by us in Miller et al. (2007, 2009) in a number of important respects but also suggests some new interpretations that differ from her view. I share Miller’s view that for none of the clicks, the place of articulation of the back closure *at release* is velar; rather, it is in all cases further back than that. As argued in Miller et al. (2007, 2009), clicks like /ǀ͡q/ and /!͡q/ are most consistently described as *airstream* contour segments, *not* place contour segments.

34 It is not so obvious to me at present, however, why [ǀ͡χ] and [!͡χ] do not significantly differ from *each other*. My best assumption is that, frication noise in affricates being more prolonged and therefore much more salient than frication noise in stop bursts, a constant turbulent noise quality is in fact an *auditory* target. If that is so, it would follow that in [!͡χ], the posterior place of articulation at the release is coarticulatory accomodated to that of the following fricative.

The main difference between Miller et al. (2007, 2009) and the arguments put forward here is that that the former crucially attribute both (1) the different patterning of laminal and apical clicks with respect to the BVC and (2) the different posterior place of articulation of these two classes of clicks to observed differences in pharyngeal narrowing between the two classes. I fully agree that such a difference indeed exists (this is born out quite clearly by the ultrasound data presented in Miller et al., 2009). However, I disagree in my interpretation as to what is the cause and what is the effect.

As has been argued in this chapter, I attribute both sets of differences – the BVC patterning and the posterior place of articulation – to the shape of the tongue *in front of* the posterior constriction. While acknowledging the existence of different patterns of pharyngeal constriction associated with the different classes of clicks, I argue that it is more plausible to assume that those pharyngeal constriction patterns are the consequence of expanding a lingual cavity under the requirement of maintaining a shallow concave vs. deep concave shape for the front of the tongue.

For all clicks, three articulatory gestures (broadly corresponding to directions of muscular activity) can be identified: (1) an upward and forward push of the tip/blade of the tongue for the anterior closure; (2) an upward and backward pull of the back of the tongue for the posterior closure; and (3) a downward and backward pull of the root of the tongue to expand the lingual cavity. All three gestures are simultaneously active. I assume that due to the hydrostatic nature of the tongue (i.e. its property of maintaining a constant volume regardless of the forces that act upon it), with the two upward-directed forces in place, there is a compensatory deformation of the tongue in the direction of the pharyngeal wall as soon as the process of cavity expansion sets in. The exact combination of the muscular forces needed to maintain the shape of the front of the tongue (shallow concave or deep concave) on the one hand and the size of the lingual cavity on the other hand determine the exact nature (location and extent) of the pharyngeal excursion. Apparently (and plausibly) it varies systematically between the two classes of clicks, and it is this that can be identified in the ultrasound data. However, apart from the fact that I think a plausible account of the phenomena has been given in this chapter, I consider it unlikely that this additional pharyngeal constriction could play a central role in determining the phonological patterning of clicks by way of their *acoustic* properties because the cavities in front of and behind the posterior constriction are insufficiently coupled acoustically.

So, to highlight the main points of divergence between Miller et al. (2007, 2009) and the arguments put forward here, my central positions can be summarized as follows:

- The posterior closure of all clicks at *formation* is velar.
- The posterior closure of simultaneously released clicks at *release* is not (truly) uvular but rather back velar / front uvular; crucially, it does differ from the one found in sequentially released clicks, which is indeed uvular.
- 'Pharyngealization' (at whatever level of the pharynx) is present but not the decisive factor for the patterning of clicks with respect to the BVC; instead, the shape of the front of the tongue (shallow concave vs. deep concave) determines the patterning.

Turning now briefly to the second parameter, high frequency band energy (now reinterpreted as *manner* parameter), recall that a higher value for this parameter was attributed to higher stiffness (i.e. inverse compliance) of the walls of the cavity. Why should it be so that the posterior articulation of simulaneously released clicks (/ǀ/ and /!/) has low stiffness, while 'regular' stops (including the posterior articulation of /͡ǀq/ and /͡!q/) have medium stiffness and fricatives (including the posterior articulation of /͡ǀχ/ and /͡!χ/) have high stiffness?

1. Especially the first group (/ǀ/ and /!/) seems at first puzzling, given that we have just seen that during the formation of the lingual cavity, extremely high muscular tension acts upon the tongue from three opposing directions. However, it is important to keep in mind that we are discussing the *release of the posterior closure* here. I think it is plausible to assume that the phase of extreme tension in the tongue is followed by sudden relaxation at the moment of release of the anterior closure, which would still be in effect when the posterior closure is released, only some 10 ms or so later. In fact, this might contribute to the known perceptual weakness of the posterior burst (in addition to masking by the much more intense anterior burst).[35]

35 The additional differences between [ǀ] and [!] that were noted above (more energy between 0 and 1000 Hz for [!] vs. more engergy between 5000 and 6000 Hz for [ǀ]) remain more

2. As for /k/, /ǀ͡q/, and /ǃ͡q/, I take their high frequency band energy as a frame of reference, in that I think that there are plausible ways to explain why the other two groups have particularly low or particularly high values, respectively, compared with this ‘default’ group.

3. Thus, what could make /χ/, /ǀ͡χ/, and /ǃ͡χ/ as a group display such high stiffness? I assume that it is the high degree of precision necessary for maintaining just the critical cross-sectional area at a given volume flow rate to create a turbulent air stream that necessitates a comparatively high degree of muscular tension in the articulators (in this case, the tongue body). Cf. Kirchner (2001) for an argument that in precise articulations, like those required for (strident) fricatives, isometric tension (i.e. stiffness) is built up in the articulator involved by simultaneous activation of agonist and antagonist muscle groups to maintain the critical constriction degree for turbulent airflow; cf. also Laboissière, Lametti, and Ostry (2009), who describe the relationship of isometric agonist–antagonist muscle cocontraction, articulator stiffness, and movement accuracy (inversely related to movement variability) in speech.

To summarize: /ǀ/ and /ǃ/ have low stiffness because of sudden relaxation following an *extremely* stiff setting; /χ/, /ǀ͡χ/, and /ǃ͡χ/ have high stiffness because they need to maintain a certain position with a high degree of precision, and /k/, /ǀ͡q/, and /ǃ͡q/ have medium stiffness because there is no need for them to operate with such precision.

One thing that has become very apparent in the course of the discussion is that in this case (where the objects of comparison all have very similar, yet significantly and systematically different spectra), spectral center of gravity is neither a reliable measure nor one that is likely to lead to any meaningful insights. The reason is clear: The location of the major spectral peak on the one hand and the overall energy in different frequency bands on the other hand are *independent* parameters that both operate on the same quantities, namely the magnitudes of spectral components. Center of gravity summarizes the information in a spectrum in such a way that these two parameters are combined in a *single* value,

difficult to explain, however. I hypothesize that the former might be due to the lower (minor) pharyngeal constriction for [ǃ] as compared to a higher one for [ǀ] and that the latter could be the result of accidentally including a few spectral components from the preceding front burst in [ǀ].

leading to a situation where a spectrum with a slightly *lower* major peak but *more* high frequency band energy might be shown to have the same center of gravity value as a spectrum with a slightly *higher* peak but *less* high frequency energy. Therefore, in the given situation, the center of gravity values cannot receive any direct interpretation in articulatory terms.

What does all this tell us about the original question of why /|͡q/ does occur with front vowels, whereas /|͡χ/ does not? Unfortunately, the acoustic analysis has not yielded a conclusive answer to this question, but it has ruled out most of the potential explanations offered above: (1) There is indeed a difference in posterior place of articulation between the segments, but in such a way that /|͡χ/ is more *front* than /|͡q/, the opposite of what might have been expected; (2) an assimilation (most likely perceptually based) of the closure to the frication phase in /|͡χ/ and /!͡χ/ is actually likely, thus leading to a leveling between the click types in the case of the affricates, but the same is the case for the sequential clicks, /|͡q/, and /!͡q/. We are left with three different categories of explanations:

Articulatory reasons. Apart from the reasons just ruled out, another explanation could be hypothesized. As we have seen above, /|͡χ/ requires more *articulatory precision* than /|͡q/. The same could be said for a high front vowel like /i/, for the production of which a constriction is necessary that is just wide enough *not* to produce turbulent airflow.[36] It might be the case that a sequence like */|͡χi/ is disfavored because it involves two segments with very similar requirements of precision. The same might be said to apply for high *back* vowels, for which no co-occurrence constraint exists; however, I think it is plausible to assume that the homorganicity involved in a sequence like /|͡χu/ reduces the articulatory load because the required precision is concentrated at the same place of articulation; thus, its production is facilitated.

Auditory reasons. Apart from the BVC, there is another co-occurrence constraint in N|uu that /|͡χ/ (as well as /!͡χ/ and all other segments involving uvular frication) are subject to: They do not co-occur with a following *epiglottalized vowel* (thus */|͡χaˤ/ etc.). Epiglottalization, in turn, does not co-occur with *front vowels* either. Miller (2007) has shown for Ju|'hoansi that segments with an

36 The BVC is *exceptionless* only for *high* front vowels; in all areas where it applies, some (although very few) counterexamples with [e] are found.

ejected uvular affricated release form a phonological class with epiglottalized vowels. While on the basis of her data, she explicitly excludes the corresponding *non-ejected* segments from that class, It might be possible that in N|uu, the two do form a phonological class. If that were so, then the non-occurrence of */ǀ͡χi/ (given that neither uvular frication nor epiglottalization is compatible with front vowels and, in addition, the two are not compatible with each other) might be explainable by some auditory property (most probably connected with a measure of periodicity) that uvular frication in N|uu shares with epiglottalization and that makes it incompatible with front vowels. Further research is necessary in that direction, however.

Phonological reasons. It is still a possibility that /ǀ͡χ/ is analyzed as a cluster at some level of representation, whereas /ǀ͡q/ is a unit. Still, it is a puzzle why the posterior release should be transparent to the BVC in /ǀ͡q/ but not in /ǀ͡χ/, especially since we have seen that the *acoustic* evidence points towards a separate articulatory target for the posterior constriction of /ǀ͡q/ but not /ǀ͡χ/, and not vice versa, as would have been expected if the hypothesis were correct.

To further test this hypothesis, Praat scripts were written to measure duration and rise time to peak amplitude for /χ/, /ǀ͡χ/, and /ǃ͡χ/. Since these two quantities are commonly given as diagnostic of the fricative vs. affricate status, the expectation was that if /ǀ͡χ/ and /ǃ͡χ/ were consonant clusters rather than unit phonemes, their respective frication phases would resemble fricatives rather than affricates. The results of the duration measurements are given in Figure 5.25.

It was immediately obvious that the duration for /χ/ (M = 160.83 ms, SD = 29.44 ms) was much longer than for /ǀ͡χ/ (M = 73.23 ms, SD = 26.68 ms) and /ǃ͡χ/ (M = 66.54 ms, SD = 14.83 ms). An LME ANOVA showed a significant main effect for consonant, $p_{\mathrm{MCMC}} < 0.001$. A one-way repeated measures ANOVA (without sphericity corrections) yielded a similar result, $F(2, 4) = 12.11$, $p = 0.02$. A pairwise comparison revealed that *within* the group of affricates (/ǀ͡χ/ and /ǃ͡χ/), no significant differences existed, whereas all pairs *between* the affricates and /χ/ showed highly significant differences at $p < 0.001$.

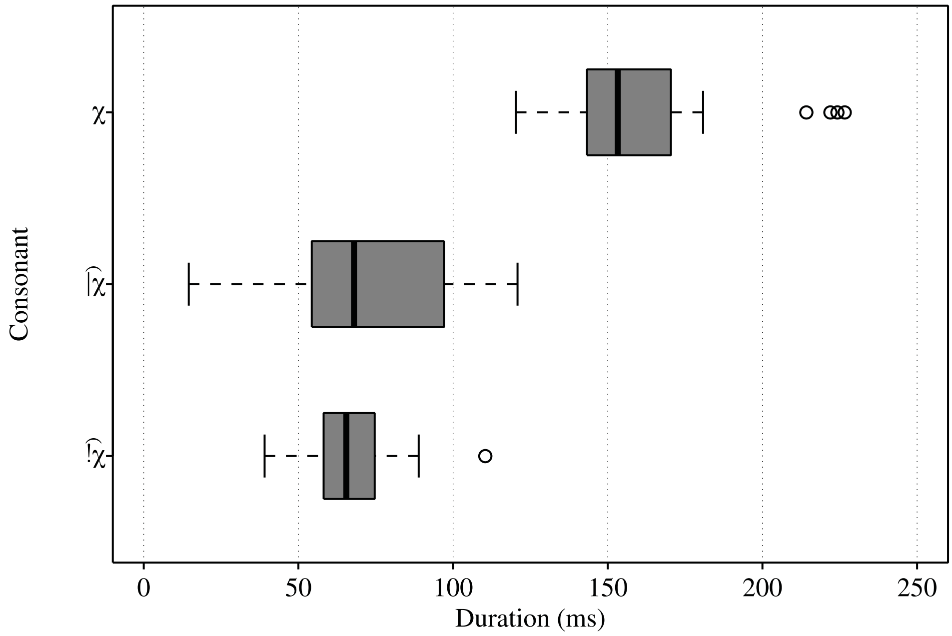

Figure 5.25 Box plots of the duration of the frication phase for /χ/, /ǀ͡χ/, and /!͡χ/ (N = 90)

Very much the same picture (if anything, even more extreme) emerged when comparing the rise time to peak amplitude for the same three consonants, as shown in Figure 5.26.

Again, the rise time for /χ/ (M = 109.27 ms, SD = 38.99 ms) was much longer than for /ǀ͡χ/ (M = 8.61 ms, SD = 9.79 ms) and /!͡χ/ (M = 6.02 ms, SD = 4.45 ms). An LME ANOVA showed a significant main effect for consonant, $p_{MCMC} < 0.001$. A one-way repeated measures ANOVA (without sphericity corrections) had a similar result, $F(2, 4) = 34.59$, $p = 0.003$. A pairwise comparison revealed that *within* the group of affricates (/ǀ͡χ/ and /!͡χ/), no significant differences existed, whereas all pairs *between* the affricates and /χ/ showed highly significant differences at $p < 0.001$.

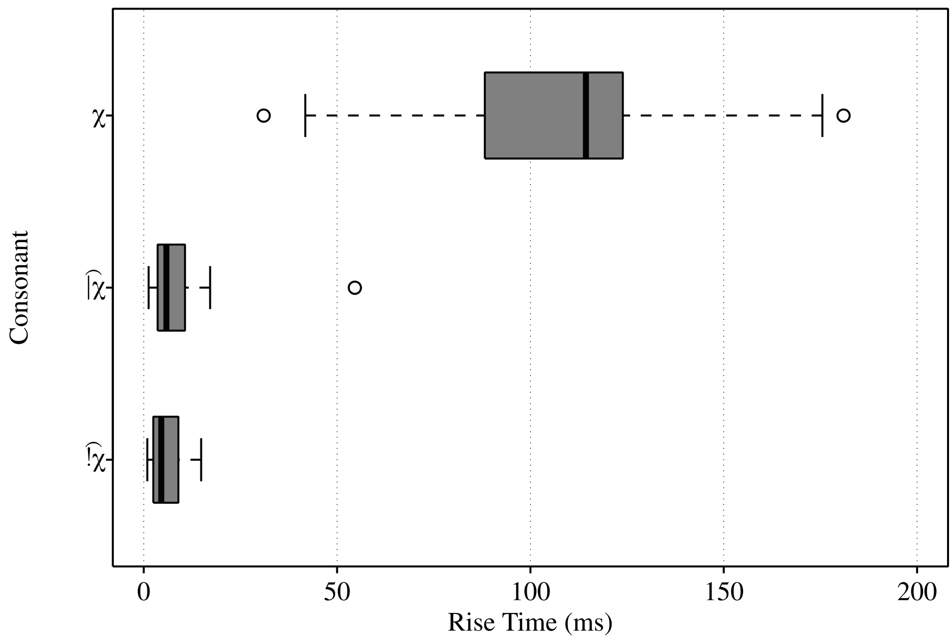

Figure 5.26 Box plots of the rise time to peak amplitude of the frication phase for /χ/, /ǀ͡χ/, and /ǃ͡χ/ (N = 90)

All in all, the phonetic evidence points very strongly against analyzing /ǀ͡χ/ and /ǃ͡χ/ as consonant clusters, and therefore either an explanation based on articulatory precision or an auditory explanation based on periodicity seem to be the most promising candidates for solving the problem of why */ǀ͡χi/ is an illicit combination in N|uu.

6 The Representation and Transcription of Clicks: A Revised Proposal

We are now in a position to formulate some proposals, or minimum requirements, for an adequate articulatory representation of clicks. The proposal that I would like to make for a revised categorization of click consonants here is consistently based on articulatory parameters (this should not be taken to imply, however, that a categorization in terms of auditory parameters is not equally called for).

Turning first to a categorization of click *types*, I think that all attested clicks types can exhaustively, and at the same time economically, be captured by the introduction (in addition to the established categories of *active articulator* and *central* vs. *lateral* articulation) of two additional articulatory parameters, namely (lingual) *cavity size* (cf. Exter, 2011), and *tenseness*, of the articulators involved, particularly the tongue (cf. Chapter 5).

Tenseness as an articulatory parameter has been discussed previously e.g. by Slis (1971), Cho and Keating (2001), and Cho, Jun, and Ladefoged (2002), who attribute characteristics like longer closure duration, greater linguopalatal contact, and greater muscle activity to tense ('fortis') segments. Similarly, in their Articulatory Phonology model, Browman and Goldstein (1989) describe tense ('stiff') segments as having greater constriction degree and greater articulator speed. Svirsky, Stevens, Matthies, Manzella, Perkell, and Wilhelms-Tricarico (1997) investigate tongue surface displacement (reflecting intraoral pressure and mechanical compliance of the tongue surface) during the production of bilabial stops. Compliance values were found to be much higher for voiced than for voiceless bilabial stops, which they interpret as evidence for active stiffening and/or relaxation of tongue muscles.

Finally, in connection with clicks, Traill (1985, based on the parameter 'cavity spread', i.e. comparing the length of the lingual cavity at formation and release), explicitly suggests that two classes of clicks can be distinguished: tense vs. non-tense, stating that the use of the feature tenseness "rests precisely on a hypothesized greater muscular tension for !, ǂ as opposed to | and ‖" (p. 111).

The terminology proposed her for the parameter *tenseness* is simply *tense* vs. *lax*, tense clicks being characterized, in comparison with their lax counterparts, by (at least) a broader contact area, a less noisy release (due to a faster release gesture), and more pronounced formant peaks (due to less damping by the vocal tract walls).

The new terminology that I propose for the parameter cavity size is *deep* vs. *shallow*, which is short for 'deep concave' and 'shallow concave', respectively, and refers to the shape of the center of the tongue at the moment of the anterior release.

A maximal system of click types based on this new combination of parameters is presented in Table 6.1.

Table 6.1 A maximally filled system of click *types*

	Labial	Coronal			
		Deep		Shallow	
		Tense	Lax	Tense	Lax
Central	ʘ	!	‼	ǂ	|
Lateral		‖			

A few notes on the table are in order. First of all, trivially, the distinction between central and lateral articulations is only relevant for *coronal*, not labial, clicks. Secondly, neither labial clicks nor lateral clicks show any contrasts in terms of tenseness or cavity size; at least, no such contrasts are to my knowledge attested in any language. If any such contrasts are later found to exist, they can easily be incorporated in the system. And thirdly, the symbol for a *lax deep coronal* click in this system, /‼/, is not an official IPA symbol. Such clicks have been described (under the term *retroflex* clicks) for a variety of languages, however (cf. Doke, 1923–1926; Traill & Vossen, 1997), and their reported properties (distinctively concave tongue shape, fricated release) make them a perfect candidate for filling the only vacant spot in the system.

Perhaps the most important consequence of this proposal is the claim that click types are not actually distinguished from one another by means of 'traditional' places of articulation, as the conventional IPA labels would suggest. This is born out by the fact that auditorily *identical* clicks can be produced at a variety of *different* places of articulation (cf. Sands, Brugman, Exter, Namaseb, & Miller, 2007), as long as the parameters (active) articulator, centrality, (lingual) cavity size, and tenseness remain unchanged; as Maddieson (2003, p. 37) puts it: “Such context-free liberty to vary place of articulation is rarely encountered with other classes of consonants.” I propose that the observed differences in ‘traditional’ place of articulation are best described as *enhancing features* in the sense of Stevens and Keyser (1989).

Table 6.2 A system for transcribing anterior place of articulation in clicks

ʘ	bilabial
ʘ̪	labiodental
ǀ̟	lamino-interdental
ǀ̺	apico-dental
ǀ̻	lamino-dental
ǀ̠	lamino-alveolar
ǃ̟	apico-alveolar
ǃ̺	apico-postalveolar
ǃ̻	lamino-postalveolar
ǃ̠	apico-prepalatal
‼	subapico-prepalatal
ǂ	predorso-prepalatal

Nevertheless, there can be contexts where the precise (anterior) place of articulation of a click is of importance, as e.g. in the analysis of inter-speaker variation or in historical or comparative descriptions, e.g. in studies like Traill and Vossen’s (1997) investigation of click weakening. In that case, I propose to use diacritics on the basic click symbols; this can be modeled closely on the proposal by Ladefoged and Maddieson (1996, p. 15) for transcribing fine-grained differences in coronal places of articulation. Table 6.2 summarizes my proposal for a transcriptional convention for (anterior) places of articulation in clicks. Again, some notes are in order:

- Place of articulation in *lateral* clicks of the /ǁ̺/ type can be modeled on the transcription of /!/.
- If a clearly *shallow* lateral click were to be attested, the symbol /⦀/ (as used e.g. by Miller-Ockhuizen & Sands, 2000) is proposed; its transcriptional variants could then be modeled on the transcription of /|/.
- A *flapped* click (i.e. a click of the /!/ type with an additional transient at release that is caused by the underside of the tongue hitting the floor of the mouth) has been described as a non-distinctive variant of /!/ in a number of languages (cf. Tucker, Bryan, & Woodburn, 1977; Wright, Maddieson, Ladefoged, & Sands, 1995). For such a click, the symbol /!¡/ has been proposed; its articulatory variants can, if needed, be modeled on /!/.

Turning now to what has traditionally been called click *effluxes*, or *accompaniments*, I propose an extension of the model introduced by Miller et al. (2007, 2009). As has been convincingly argued by Miller, the term *accompaniment* as traditionally used is unfortunate because it does not correspond to any coherent phonetic phenomenon (as opposed to the term *click type*). Nevertheless, I think that it is helpful to have a term of reference for the different kinds of realizations that a given click type may be involved in. I propose the term *click series* as a term of convenience that does not commit itself to any direct phonetic interpretation.

As for the proposed extension of Miller's innovative model of *airstream contours*, I think that two airstreams have to be identified in *every* click realization, not only in segments like e.g. /!͡q/. As we have seen in the empirical studies above, the posterior burst is always present (though not always *perceived*). Strong support for this view comes from the fact that in several languages, e.g. |Gui (Nakagawa, 2006), there are parallel realizations for linguo-pulmonic and linguo-glottalic clicks: In *both*, the two closures can be released either quasi-simultaneously or sequentially. For that reason, I propose the terms (plain) *linguo-pulmonic stop* (or linguo-pulmonic *simultaneously released* stop, if explicitness is required) for segments like /!/, and *linguo-pulmonic sequential stop* (or linguo-pulmonic *sequentially released* stop) for segments like /!͡q/. For the corresponding *affricates* and linguo-*glottalic* segments, the same applies mutatis mutandis. Table 6.3 presents a maximal system of click series for the /ǃ/ click type as an illustration of this new system of categorization.

Table 6.3 A maximally filled system of click *series* (exemplified with [ǃ]). Dotted circles indicate possible but non-attested items.

	Voiceless			(Pre-)Voiced		
	Plain	Aspirated	Glottalized	Plain	Aspirated	Glottalized
Linguo-Pulmonic Stop	ǃ	ǃʰ	ǃˀ	ᶢǃ	ᶢǃʰ	ᶢǃˀ
Linguo-Pulmonic Sequential Stop	ǃ͡q	ǃ͡qʰ		ᶢǃ͡ɢ	ᶢǃ͡qʰ	
Linguo-Pulmonic Affricate	ǃ͡χ			ᶢǃ͡χ		
Linguo-Pulmonic Nasal	ŋ̊ǃ	ŋ̊ǃʰ	ŋ̊ǃˀ	ŋǃ	ŋǃʰ	ŋǃˀ
Linguo-Glottalic Stop			ǃ’			ᶢǃ’
Linguo-Glottalic Sequential Stop			ǃ͡q’			ᶢǃ͡q’
Linguo-Glottalic Affricate			ǃ͡χ’			ᶢǃ͡χ’
Linguo-Glottalic Sequential Affricate			◌			◌

Note that (linguo-pulmonic and linguo-glottalic) *sequential affricates* are perfectly possible but (to my knowledge) not attested in any language. A *sequential affricate* would be like the corresponding sequential *stop* in having two clearly discernable bursts in sequence, but would differ from it in that the second burst would be affricated. At present I do not know whether there is a *principled* reason why such segments are not attested.

7 Conclusion

In the present work, we have seen that N|uu is a severely endangered language on the verge of extinction, while at the same time, it is a phonetically and phonologically highly complex language.

In Chapter 3, we have for the first time presented an overview of the *entire phoneme inventory*, including not only the click subsystem (here, the terms *simultaneous release* vs. *sequential release* were introduced), but also the non-click consonants and vowels, as well as the as yet unanalyzed *lexical tone* system.

Then, in Chapter 4, the *foot* was for the first time identified as a fundamental organizing unit in the phonological system of the language. Also, the *complexity of phonological words* in N|uu was found to be considerably greater than previously thought.

In the first study presented in Chapter 5, the Back Vowel Constraint (BVC) in N|uu was analyzed in terms of *differences in tongue shape* in different categories of clicks, while in the second study, *stiffness* was identified as an articulatory parameter in N|uu.

Finally, in Chapter 6, a synthesis of the new insights gained in the previous chapters was attempted in the form of a novel scheme for categorizing and transcribing clicks. In this context, the categorical terms *deep concave* vs. *shallow concave* were introduced for the systematic tongue shape differences in clicks, and the terms *tense* vs. *lax* were proposed to capture the categorical effect of articulatory stiffness in clicks.

Fortunately, quite a number of detailed studies on the phonetics and phonology of N|uu have appeared recently. However, many important aspects still await future research. Among these are detailed quantitative studies of the non-clicks and vowels, especially the epiglottalized vowels; an in-depth investigation of the intonation system and its interaction with the lexical tone system (an area that has generally received relatively little attention within the 'Khoisan' descriptive literature so far); and phonetic and phonological phenomena in natural speech and units larger than the phonological word.

References

Baayen, R. H. (2009). languageR: Data sets and functions with *Analyzing linguistic data: A practical introduction to statistics using R* (Version 0.955) [Computer software]. Retrieved from http://CRAN.R-project.org/package=languageR

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes (Version 0.999375-33) [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4

Bleek, D. F. (1927). The distribution of Bushman languages in South Africa. In *Festschrift Meinhof* (pp. 55–64). Glückstadt: Augustin.

Bleek, D. F. (1929). *Comparative vocabularies of Bushman languages* (University of Cape Town, Publications of the School of African Life and Language). Cambridge: Cambridge University Press.

Bleek, D. F. (1937). *|'auni* vocabulary. In J. D. Rheinallt Jones & C. M. Doke (Eds.), *Bushmen of the Southern Kalahari* (pp. 201–220). Johannesburg: University of the Witwatersrand Press.

Bleek, D. F. (1939–1940). A short survey of Bushman languages. *Zeitschrift für Eingeborenen-Sprachen, 30*, 52–72.

Bleek, D. F. (2000). *The ǁŋ !ke or Bushmen of Griqualand West; Notes on the language of the ǁŋ !ke or Bushmen of Griqualand West* (T. Güldemann, Ed.) (Working Paper / Khoisan Forum No. 15). Cologne: University of Cologne, Institut für Afrikanistik.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives* (LOT International Series No. 11). The Hague: Holland Academic Graphics.

Boersma, P. & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.30) [Computer software]. Retrieved from http://www.praat.org/

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Catford, J. C. (1968). The articulatory possibilities of man. In B. Malmberg (Ed.), *Manual of phonetics* (pp. 309–333). Amsterdam: North-Holland.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*, 193–228.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190.

Crawhall, N. (2001). *Written in the sand: Auditing and managing cultural resources with displaced indigenous peoples; A South African case study*. Rondebosch: South African San Institute in co-operation with UNESCO.

Crawhall, N. (2002). Reclaiming language and identity. *Cultural Survival Quarterly, 26*(1), 49–51.

Crawhall, N. (2003). The rediscovery of N|u and the ǂKhomani land claim process in South Africa. In J. Blythe & R. McKenna Brown (Eds.), *Maintaining the links: Language, identity and the land; Proceedings of the Seventh FEL Conference, Broome, Western Australia, 22–24 September 2003* (pp. 13–19). Bath: Foundation for Endangered Languages.

Crawhall, N. (2004). *!Ui-Taa language shift in Gordonia and Postmasburg districts, South Africa* (Unpublished doctoral dissertation). University of Cape Town, Cape Town.

Crawhall, N. (2005a). The story of !Ui: Causality and language shift in Africa. In N. Crawhall & N. Ostler (Eds.), *Creating outsiders: Endangered languages, migration and marginalisation; Foundation for Endangered Languages: Proceedings of the Ninth Conference, Stellenbosch, South Africa, 18–20 November 2005* (pp. 71–81). Bath: Foundation for Endangered Languages.

Crawhall, N. (2005b). Too good to leave behind: The N|u language and the ǂKhomani people of Siyanda district. In R. Finlayson & S. Slabbert (Eds.), *Language and identities in a postcolony: Southern African Perspectives* (Schriften zur Afrikanistik No. 10) (pp. 67–89). Frankfurt am Main: Lang.

Dart, R. A. (1937). The hut distribution, genealogy and homogeneity of the *|'auni-ǂkhomani* Bushmen. In J. D. Rheinallt Jones & C. M. Doke (Eds.), *Bushmen of the Southern Kalahari* (pp. 101–116). Johannesburg: University of the Witwatersrand Press.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1962). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica, 16*, 104–121.

Doke, C. M. (1923–1926). An outline of the phonetics of the language of the ʗhṵ̃: Bushmen of North-West Kalahari. *Bantu Studies, 2*, 129–165.

Doke, C. M. (1937). An outline of *ǂkhomani* Bushman phonetics. In J. D. Rheinallt Jones & C. M. Doke (Eds.), *Bushmen of the Southern Kalahari* (pp. 61–88). Johannesburg: University of the Witwatersrand Press.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology, 23*, 157–191.

Edmondson, J. A., Ziwo, L., Esling, J. H., Harris, J. G., & Li, S. (2001). The aryepiglottic folds and voice quality in the Yi and Bai languages: Laryngoscopic case studies. *Mon-Khmer Studies, 31*, 83–100.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association, 26*, 65–88.

Esling, J. H. (1999). The IPA categories "pharyngeal" and "epiglottal": Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech, 42*, 349–372.

Esling, J. H. (2005). There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics, 50*, 13–44.

Esling, J. H., & Edmondson, J. A. (2002). The laryngeal sphincter as an articulator: Tenseness, tongue root and phonation in Yi and Bai. In A. Braun & H. R. Masthoff (Eds.), *Phonetics and its applications: Festschrift for Jens-Peter Köster on the occasion of his 60th birthday* (Zeitschrift für Dialektologie und Linguistik, Beihefte, No. 121) (pp. 38–51). Stuttgart: Steiner.

Esling, J. H., & Harris, J. G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle & J. Mackenzie Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 347–383). Mahwah, NJ: Lawrence Erlbaum Associates.

Exter, M. (2011). The acoustic modeling of click types. In W.–S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of the Phonetic Sciences: 17–21 August 2011, Hong Kong* [CD-ROM] (pp. 655–658). Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84*, 115–123.

Fulop, S. A., Ladefoged, P., Liu, F., & Vossen, R. (2003). Yeyi clicks: Acoustic description and analysis. *Phonetica, 60*, 231–260.

Greenberg, J. H. (1963). *The languages of Africa* (Publication of the Indiana University Research Center in Anthropology, Folklore, and Linguistics No. 25). Bloomington, IN: Indiana University. (*International Journal of American Linguistics, 29*(1), Pt. 2)

Güldemann, T. (Ed.) (2005). *Studies in Tuu (Southern Khoisan)* (University of Leipzig Papers on Africa, Languages and Literatures Series, No. 23). Leipzig: Institut für Afrikanistik, Universität Leipzig.

Güldemann, T., & Vossen, R. (2000). Khoisan. In B. Heine & D. Nurse (Eds.), *African languages: An introduction* (pp. 99–122). Cambridge: Cambridge University Press.

Hamann, S. (2003). *The phonetics and phonology of retroflexes* (LOT Dissertation Series No. 75). Utrecht: LOT.

Hastings, R. (2001). Evidence for the genetic unity of Southern Khoesan. In A. Bell & P. Washburn (Eds.), *Khoisan: Syntax, phonetics, phonology, and contact* (Cornell Working Papers in Linguistics No. 18) (pp. 225–246). Ithaca, NY: Cornell University.

Herbst, J. F. (1908). *Cape Colony: Report on the Rietfontein area*. Presented to both Houses of Parliament by Command of His Majesty. October, 1908 (Colonial Reports, Miscellaneous, No. 55). London: His Majesty's Stationery Office.

Howell, P., & Rosen, S. (1983). Production and perception of rise time in the voiceless affricate/fricative distinction. *Journal of the Acoustical Society of America, 73*, 976–984.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jessen, M. (2002). An acoustic study of contrasting plosives and click accompaniments in Xhosa. *Phonetica, 59*, 150–179.

Jet Propulsion Laboratory. (2005). *Global SRTM 30-arcsec seamless data: SRTM30* (Version 2) [Data set]. Retrieved from http://www2.jpl.nasa.gov/srtm/

Johnson, K. (1993). Acoustic and auditory analyses of Xhosa clicks and pulmonics. *UCLA Working Papers in Phonetics, 83*, 33–45.

Johnson, K. (2003). *Acoustic & auditory phonetics* (2nd ed.). Malden, MA: Blackwell.

Keating, P. A. (1988). Palatals as complex segments: X-ray evidence. *UCLA Working Papers in Phonetics, 69*, 77–91.

Keating, P. A. (1991). Coronal places of articulation. In C. Paradis & J.-F. Prunet (Eds.), *The special status of coronals: Internal and external evidence* (Phonetics and Phonology No. 2) (pp. 29–48). San Diego, CA: Academic Press.

Keating, P., & Lahiri, A. (1993). Fronted velars, palatalized velars, and palatals. *Phonetica, 50*, 73–101.

Kirchner, R. (2001). *An effort based approach to consonant lenition* (Outstanding Dissertations in Linguistics). New York, NY: Routledge.

Laboissière, R., Lametti, D. R., & Ostry, D. J. (2009). Impedance control and its relation to precision in orofacial movement. *Journal of Neurophysiology, 102*, 523–531.

Ladefoged, P. (1989). Representing phonetic structure. *UCLA Working Papers in Phonetics, 73*, 1–79.

Ladefoged, P. (2003). *Phonetic data analysis*. Malden, MA: Blackwell.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Phonological Theory). Oxford: Blackwell.

Laver, J. (1994). *Principles of phonetics* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Maddieson, I. (1984). *Patterns of sounds*. With a chapter contributed by Sandra Ferrari Disner (Cambridge Studies in Speech Science and Communication). Cambridge: Cambridge University Press.

Maddieson, I. (1997). Phonetic universals. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (Blackwell Handbooks in Linguistics No. 5) (pp. 619–639). Oxford: Blackwell.

Maddieson, I. (2003). The sounds of the Bantu languages. In D. Nurse & G. Philippson (Eds.), *The Bantu languages* (Routledge Language Family Series No. 4) (pp. 15–41). London: Routledge.

Marais, J. S. (1939). *The Cape Coloured people, 1652–1937*. London: Longmans, Green & Co.

McDonough, J., & Ladefoged, P. (1993). Navajo stops. *UCLA Working Papers in Phonetics, 84*, 151–164.

Mielke, J. (2008). *The emergence of distinctive features* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.

Miller, A. L. (2007). Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics, 35*, 56–84.

Miller, A. L. (2010). Tongue body and tongue root shape differences in N|uu clicks correlate with phonotactic patterns. In S. Fuchs, M. Toda, & M. Żygis (Eds.), *Turbulent sounds: An interdisciplinary guide* (Interface Explorations No. 21) (pp. 245–279). Berlin: de Gruyter Mouton.

Miller, A. (2011). The representation of clicks. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Blackwell Companions to Linguistics) (Vol. 1, pp. 416–439). Malden, MA: Wiley-Blackwell.

Miller, A., Brugman, J., & Sands, B. (2007). Acoustic and auditory analysis of N|uu lingual and linguo-pulmonic stop bursts. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences: ICPhS XVI, 6–10 August 2007, Saarbrücken, Germany* (Vol. 1, pp. 769–772). Saarbrücken: Universität des Saarlandes.

Miller, A. L., Brugman, J., Sands, B., Namaseb, L., Exter, M., & Collins, C. (2007). The sounds of N|uu: Place and airstream contrasts. *Working Papers of the Cornell Phonetics Laboratory, 16*, 101–160.

Miller, A. L., Brugman, J., Sands, B., Namaseb, L., Exter, M., & Collins, C. (2009). Differences in airstream and posterior place of articulation among N|uu clicks. *Journal of the International Phonetic Association, 39*, 129–161.

Miller-Ockhuizen, A., & Sands, B. E. (2000). Contrastive lateral clicks and variation in click types. In B. Yuan, T. Huang, & X. Tang (Eds.), *6th International Conference on Spoken Language Processing: ICSLP 2000, Oct. 16 – Oct. 20 2000, Beijing International Convention Center, Beijing, China; The proceedings of the conference* (Vol. 2, pp. 499–502). Beijing: China Military Friendship Publish.

Mitani, S., Kitama, T., & Sato, Y. (2006). Voiceless affricate/fricative distinction by frication duration and amplitude rise slope. *Journal of the Acoustical Society of America, 120*, 1600–1607.

Nakagawa, H. (2006). *Aspects of the phonetic and phonological structure of the G|ui language* (Unpublished doctoral dissertation). University of the Witwatersrand, Johannesburg.

National Geophysical Data Center. (2006). *2-minute gridded global relief data: ETOPO2* (Version 2) [Data set]. Retrieved from http://www.ngdc.noaa.gov/mgg/global/

Naumann, C. (2008). High and low tone in Taa ǂaan (!Xóõ). In S. Ermisch (Ed.), *Khoisan languages and linguistics: Proceedings of the 2nd International Symposium, January 8–12, 2006, Riezlern/Kleinwalsertal* (Quellen zur Khoisan-Forschung No. 22) (pp. 279–302). Cologne: Köppe.

N|u-Sprache bleibt erhalten (2005, May 30). *Nordwest-Zeitung*, p. 9.

Papst [Pabst], H. (1895). Die Kalahariwüste und ihre Bewohner. *Berichte der Rheinischen Missions-Gesellschaft, 52*, 243–247.

Papst [Pabst], H. (1895–1896). Die Kalahariwüste und ihre Bewohner. *Mitteilungen der Geographischen Gesellschaft für Thüringen zu Jena, 14*, 48–54.

Pöch, R. (1909a). [Bericht an die Sitzung am 21. Oktober]. *Anzeiger der Kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse, 46*, 361–365.

Pöch, R. (1909b). [Bericht an die Sitzung am 9. Dezember]. *Anzeiger der Kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse, 46*, 452–453.

Pöch, R. (1910). Reisen im Innern Südafrikas zum Studium der Buschmänner in den Jahren 1907 bis 1909. *Zeitschrift für Ethnologie, 42*, 357–362.

R Development Core Team. (2010). R: A language and environment for statistical computing (Version 2.10.1) [Computer software]. Retrieved from http://www.r-project.org/

Recasens, D. (1990). The articulatory characteristics of palatal consonants. *Journal of Phonetics, 18*, 267–280.

Rheinallt Jones, J. D., & Doke, C. M. (Eds.) (1937). *Bushmen of the Southern Kalahari*. Johannesburg: University of the Witwatersrand Press.

Sands, B., Brugman, J., Exter, M., Namaseb, L., & Miller, A. (2007). Articulatory characteristics of anterior click closures in N|uu. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences: ICPhS XVI, 6–10 August 2007, Saarbrücken, Germany* (Vol. 1, pp. 401–404). Saarbrücken: Universität des Saarlandes.

Sands, B., Miller, A. L., & Brugman, J. (2007). The lexicon in language attrition: The case of N|uu. In D. L. Payne & J. Peña (Eds.), *Selected Proceedings of the 37th Annual Conference on African Linguistics* (pp. 55–65). Somerville, MA: Cascadilla Proceedings Project.

Slis, I. H. (1971). Articulatory effort and its durational and electromyographic correlates. *Phonetica, 23*, 171–188.

Snyman, J. W. (1978). The clicks of Žu|'hõasi. In E. J. M. Baumbach (Ed.), *Tweede Afrikatale-Kongres van UNISA* (Miscellanea Congregalia No. 5) (pp. 144–168). Pretoria: Universiteit van Suid-Afrika.

Stevens, K. N. (1998). *Acoustic phonetics* (Current Studies in Linguistics No. 30). Cambridge, MA: MIT Press.

Stevens, K. N., & Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language, 65*, 81–106.

Stone, M., Epstein, M. A., & Iskarous, K. (2004). Functional segments in tongue movement. *Clinical Linguistics & Phonetics, 18*, 507–521.

Sussman, H. M., Hoemeke, K. A., & Ahmed, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *Journal of the Acoustical Society of America, 94*, 1256–1268.

Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America, 90*, 1309–1325.

Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America, 102*, 562–571.

Traill, A. (1977). The tonal structure of !xõ. In J. W. Snyman (Ed.), *Bushman and Hottentot linguistic studies: Papers of seminar held on 25 October 1975* (Miscellanea Congregalia No. 2) (pp. 11–47). Pretoria: University of South Africa.

Traill, A. (1979). The !Xõo strength hierarchy. In J. W. Snyman (Ed.), *Bushman and Hottentot linguistic studies: Papers of seminar held on 29 October 1977* (Miscellanea Congregalia No. 8) (pp. 1–38). Pretoria: University of South Africa.

Traill, A. (1985). *Phonetic and phonological studies of !Xóõ Bushman* (Quellen zur Khoisan-Forschung No. 1). Hamburg: Buske.

Traill, A. (1986). The laryngeal sphincter as a phonatory mechanism in !Xóõ Bushman. In R. Singer & J. K. Lundy (Eds.), *Variation, culture and evolution in African populations: Papers in honour of Dr Hertha de Villiers*. With a foreword by Professor Emeritus Raymond A. Dart (pp. 123–131). Johannesburg: Witwatersrand University Press.

Traill, A. (1991). Pulmonic control, nasal venting, and aspiration in Khoisan languages. *Journal of the International Phonetic Association, 21*, 13–18.

Traill, A. (1992). A confusion of sounds: The phonetic description of !Xũ clicks. In D. F. Gowlett (Ed.), *African Linguistic contributions: Presented in honour of Ernst Westphal* (pp. 345–362). Pretoria: Via Afrika.

Traill, A. (2009). *A !Xóõ dictionary* (Quellen zur Khoisan-Forschung No. 9) (2nd ed.). Cologne: Köppe.

Traill, A., & Vossen, R. (1997). Sound change in the Khoisan languages: New data on click loss and click replacement. *Journal of African Languages and Linguistics, 18*, 21–56.

Tucker, A. N., Bryan, M. A., & Woodburn, J. (1977). The east African click languages: A phonetic comparison. In W. J. G. Möhlig, F. Rottland, & B. Heine (Eds.), *Zur Sprachgeschichte und Ethnohistorie in Afrika: Neue Beiträge afrikanistischer Forschungen* (pp. 300–323). Berlin: Reimer.

Van Valin, R. D., Jr., & LaPolla, R. J. (1997). *Syntax: Structure, meaning and function* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press

Wessel, P. & Smith, W. H. F. (2010). GMT: The Generic Mapping Tools (Version 4.5.2) [Computer software]. Retrieved from http://gmt.soest.hawaii.edu/

Wodtko, D. S., Irslinger, B., & Schneider, C. (2008). *Nomina im Indogermanischen Lexikon* (Indogermanische Bibliothek, 2. Reihe: Wörterbücher). Heidelberg: Winter.

Wright, R., Maddieson, I., Ladefoged, P., & Sands, B. (1995). A phonetic study of Sandawe clicks. *UCLA Working Papers in Phonetics, 91*, 1–24.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95*, 2240–2253.