

Clicks, Concurrency, and Khoisan

Julian Bradfield

University of Edinburgh

Abstract: I propose that the notions of segment and phoneme be enriched to allow, even in classical theories, some concurrent clustering. My main application is the Khoisan language !Xóǀ, where by treating clicks as phonemes concurrent with phonemic accompaniments, the inventory size is radically reduced, so solving the problems of many unsupported contrasts. I show also how phonological processes of !Xóǀ can be described more elegantly in this setting, and provide support from metalinguistic evidence and experiment evidence of production tasks. I describe a new allophony in !Xóǀ. I go on to discuss other, some rather speculative, applications of the concept of concurrent phoneme.

The article also provides a comprehensive review of the segmental phonetics and phonology of !Xóǀ, together with previous analyses.

1 Opening

1.1 Introduction

Phonology can be said to have emerged as a discipline with the invention, or discovery, of the notion of PHONEME as a contrastive UNIT OF SOUND. Contrast is a much discussed topic, but in this article I concentrate instead on the term UNIT OF SOUND, usually now called a SEGMENT.

When in 2006 the editors of the *Oxford English Dictionary* (OED 2011) revised their entry for PHONEME, to have a short definition in everyday language that would cover all the quotations they have in their files, they wrote ‘a unit of sound in a language that cannot be analysed into smaller linear units and that can distinguish one word from another’. These words, although they still reflect an early 20th century view of the subject, neatly encapsulate both an old problem and the related problem I wish to discuss.

The old problem is what it means to say ‘can be analysed into smaller linear units’. The best known realization of the problem is the question of affricates vs. clusters: the majority view /tʃ/ as a single segment in English, but two in

German, and conversely for /ts/, but seventy years after Trubetzkoy (1939) discussed it, there is still no unanimity among phonologists. Phonologists studying German range from those who admit no affricates at all, to those who admit every phonetic affricate as a phonological affricate – see Wiese (2000) for a brief review.

This article, on the other hand, is concerned with the word LINEAR, which is part of the usual understanding of SEGMENT. I claim that the restriction to linearity is an undue restriction on the definitions of segment (and hence phoneme), and that in some languages, entities traditionally viewed as single segments should be viewed as clusters. The difference is that the clusters are concurrent, rather than sequential. To put the thesis in a sentence, sometimes a co-articulated segment really is better seen as two articulated co-segments.

The notion of concurrent units is already commonplace in certain situations; languages with lexical tone are viewed as placing tones atop segmental units, whether vowels, syllables or words, and sign languages often compose articulations from each hand – though there one can argue about whether the composition belongs in the ‘phonology’. Here I extend it to sounds that are in the segmental layer. My main application is the Khoisan language !Xóǀ, where by treating clicks as phonemes concurrent with phonemic accompaniments, the inventory size is radically reduced, so solving the problems of many unsupported contrasts. I show also how phonological processes of !Xóǀ can be described more elegantly in this setting, and provide support from metalinguistic evidence and experiment evidence of production tasks.

I start with a brief discussion of theoretical assumptions and terms; then I discuss the data and previous analyses for the languages that provide the most compelling example of the thesis; present the new analysis; discuss theoretical and empirical evaluation; and consider some other examples where the thesis might be applied.

1.2 Preliminaries

1.2.1 Theoretical assumptions

My view in this article is representational; adapting a computational process to deal with the new representations is a straightforward task, if it already deals with traditional phonemic representations. Thus, I assume informal notions of segment and phoneme as usually conceived.

Beyond that, I make no commitments in principle to any particular theory. I do not even need to assume the existence of features, though I shall use them descriptively. I do in general assume a mostly linear phonology; the relation to highly non-linear representations such as full-blown autosegmental phonology or gestural phonology is addressed briefly in §4.2.3. For the sake of illustration,

I will exhibit formalizations in the framework of *SPE*; similar illustrations could be done for most currently popular frameworks.

1.2.2 Click basics

I review briefly the phonetics and usage of clicks – for further information, see Ladefoged and Maddieson 1996 and Miller 2011. CLICK is conventionally used to describe a sound which is made by creating a ‘vacuum’ within the oral cavity, part of the cavity being bounded by the back of the tongue against the soft palate, and the rest either by the sides and front part of the tongue against the hard palate, alveolar ridge or teeth, or by the lips. The contact of the tongue back against the soft palate is conventionally called the POSTERIOR CLOSURE, and the other contact is the ANTERIOR CLOSURE. The sound is made by releasing the anterior closure, causing an inrush of air to the cavity. If the anterior closure is released sharply, this causes a distinctive ‘pop’, which is mainly responsible for the very high salience of clicks. If it is released slowly, the ‘pop’ is softer, and overlaid with affricated noise. Usually, the posterior closure is released with or very shortly after the anterior, but it can be maintained.

Traditionally, clicks are described as having VELARIC airstream mechanism, and placed in a separate section of the International Phonetic Alphabet chart (International Phonetic Association 1999). As Miller et al. (2009) point out, the term VELARIC is a little odd, since the velum is purely passive, and I enthusiastically adopt their suggestion of describing clicks as having LINGUAL airstream.

The IPA has notations for five clicks, all of which are widely used across the world paralinguistically:

[ʘ] is a BILABIAL click: the anterior closure is made with the lips, and the cavity is made by closing the tongue body against the front of the soft palate, and then drawing it back. [ʘ] is a kiss sound, though in European cultures the kiss sound is usually made with protruded and rounded lips, whereas linguistically [ʘ] is made with minimal rounding. It is hard to release the closure sharply, and in linguistic use this click always sounds affricated.

[ɰ] is a DENTAL click, in which the anterior closure is made with the blade of the tongue against the top teeth and alveolum. It is the sound used in English cultures as a sign of annoyance: *tut-tut*, or *tsk! tsk!* are conventional representations of [ɰ]. It too is always affricated.

[ɬ] is the LATERAL (ALVEOLAR) click: the cavity is formed by the sides and tip of the tongue against the alveolo-palatal region, and released along one side of the tongue. This usually gives an affricated sound; in Britain, it is conventionally used to urge on a horse. It is possible to make a lateral click with either apical or laminal contact and release; in !Xóǀ, the contact is apical.

[!] is the loudest click: it is ALVEOLAR, with tip and sides of the tongue against the alveolo-palatal area, and then the tongue sharply hollowed and released at the tip to give a ‘full’ pop (with low frequencies, owing to the large cavity created). It has no conventional use in English that I know, but may be used to imitate the sound of a cork drawn from a bottle.

Finally, [ɸ] is the PALATAL click: the closure is made with the blade of the tongue (not the tip) against the alveolo-palatal area, and the cavity is made by hollowing the centre part of the tongue, and then released at the front. This rather smaller cavity gives a ‘sharper’ (higher frequency) pop. It has no conventional uses in English that I know. It is the click taught in Britain to blind people who use clicks for echo location, presumably because the high frequencies and abrupt burst give more precise echoes.

A sixth click, which has not received an IPA symbol but is sometimes notated [!] or [!] is the true retroflex click. This is similar to [!], but the tongue tip is placed a little further back, and the contact may be apical or sublaminar. The impression is slightly softer and higher than [!], and in the Khoisan languages and dialects in which it appears, it corresponds to [ɸ] in the other languages.

A distinctive variation on [!] which is sometimes heard allophonically or idiosyncratically is the ALVEOLAR-SUBLAMINAR PERCUSSIVE click, or a PALATO-ALVEOLAR FLAPPED click. It has the Extended IPA symbol [!̠]. It is made by pronouncing [!], but keeping the front of the tongue relaxed, so that after release the front flaps downward and the underside of the blade strikes the floor of the mouth, which can generate a very audible ‘thud’ after the ‘pop’. In my experience, the ‘cork-drawing’ sound often does this: by opening the jaw and hollowing the tongue to an extreme, so that the tip is drawn back almost to the soft palate before release, a very deep and loud pop is made, and it is hard to prevent the flap from following.

Linguistically, clicks are usually combined with various manners of articulation such as voicing or aspiration applied to the posterior release; this is the topic of this article, and will be discussed in detail in the main body. Traditionally, the term INFLUX is used to refer to the actual click sound created by the release of the anterior closure, and ACCOMPANIMENT or EFFLUX (in older work) to the accompanying pulmonic-initiated sounds from the release of the posterior closure.

1.2.3 Notation

This article deals primarily with Khoisan languages and their click consonants. This topic is particularly bedevilled by notational issues: the ‘correct’ phonological analysis is something on which almost every researcher has their own, different, opinion (this article is not an exception), and therefore their own

notation; but it is even harder than usual to write a neutral ‘phonetic’ transcription, without implicitly subscribing to one or other phonological analysis. In addition, scholars of the languages have used their own practical transcriptions when recording data; for example, Tony Traill, my main source here, used a system that is IPA-like, but not quite IPA. I shall therefore be particularly careful to distinguish notations. In running text, I shall write sounds and words in **bold**, using an IPA-based notation, which tries to give a non-committal but phonological description of the sounds. It uses standard IPA diacritics to indicate modification of the click’s posterior release: for example, **!̤** is a voiced alveolar click, and **ǀ̥** is a voiceless nasal lateral click (the redundant **̥** is added for clarity). An important point is that the writing of a velar or uvular stop next to a click (e.g. **!q**) indicates a phonologically significant prolongation of the posterior closure; it is *not* part of the notation for the click itself, unlike the notation in Ladefoged and Madieson 1996. I use phoneme brackets / / to make explicitly phonemic assertions, phonetic brackets [] when discussing non-phonological detail. Generally, I normalize data to this phonemic notation; when I quote literally from a data source, I shall use *italic sans-serif*.

It is convenient to have a symbol for a generic click – I shall use **Ɂ**. This meta-symbol will be promoted to a phonological symbol during the course of the article.

1.3 Khoisan and clicks

1.3.1 Khoisan languages and language names

KHOISAN, first coined in the form ‘Koïsan’ by Schultze Jena (1928) as an ethnographic term to encompass the Khoekhoe and San ‘races’, is a Greenbergian (Greenberg 1950) classification of those languages of southern Africa which make extensive use of clicks, other than the Bantu languages (which are generally thought to have borrowed the clicks from Khoisan). The relatedness of all the Khoisan languages is no longer accepted, but the term remains as one of convenience in linguistic use, although it is politically sensitive as an ethnographic term.

There are two Tanzanian languages, Hadza (about 800 speakers) and Sandawe (about 40,000 speakers) which are conventionally included under Khoisan. Hadza is not known to be related to other languages; Güldemann and Elderkin (2010) argue that Sandawe is related to Khoe-Kwadi.

The Khoe-Kwadi family includes several living languages of which by far the largest is Khoekhoe, with around 270,000 speakers mainly in Namibia. The Khoekhoe are the groups known as ‘Hottentots’ in colonial times.

The Tuu family has now only one living example: Taa or !Xóǀ, with around 4000 speakers in Namibia and Botswana, which is the main object of my study

here. There are also a few remaining elderly speakers of N|u. It is not generally accepted that Tuu is related to Khoe-Kwadi. Current researchers prefer the name Taa for the dialect cluster which includes !Xóǎ (now spelt !Xoon); however, following my main source, and *Ethnologue*, I shall continue to use !Xóǎ.

Finally, there is the !Kung or Ju family with around 45,000 speakers in Namibia, Botswana and Angola, which includes the well known language Ju|'hoansi, also of high complexity; recently Ju has been related with the previously isolated language †Hoǎ to form a larger Kx'a family (Heine and Honken 2010).

The term SAN is used as an ethnographic term for the (largely hunter-gatherer) Tuu and Ju peoples, as opposed to the (largely pastoralist) Khoe-speaking groups. Some authors use 'San' to include the Khoe speakers also, but this is resisted by some non-Khoe speakers, who also sometimes object to the 'Khoe-San' compound nomenclature. As 'San' is itself a rather derogatory Khoekhoe word, literally 'gatherer, forager', but by extension 'a person who does not own cattle, poor person, outsider' (Haacke and Eiseb 2002), some 'San' prefer to be called by the colonial term 'Bushmen' (Besten 2006).

1.3.2 *Khoisan complexity*

The Khoisan languages are famous for their sometimes huge inventories of consonants. The most complex living language is usually considered to be !Xóǎ. In Ladefoged and Maddieson 1996, the inventory for click consonants alone is given as 85 distinct segments (or rather 83, since two are unattested), and this increases to 115 in Naumann forthcoming. The relatively modest Khoe-Kwadi language Khoekhoe has 20 click consonants, and most of the other languages fall between. (Using the same counting, Zulu has 15, and Xhosa 18.)

The typical Khoisan language has clicks at four places of articulation, of which three are borrowed by Bantu languages such as Zulu. These are alveolar ! (Zulu *q*), dental | (Zulu *c*), lateral || (Zulu *x*) and palatal †. A few surviving languages also have bilabial clicks ǀ. The enormous inventories come from the many ACCOMPANIMENTS with which these four or five basic clicks can be varied. These languages, and !Xóǎ in particular, provide the primary impetus for the thesis of this article.

2 !Xóǎ phonetics and phonology – data

In this part, I review the data that I will use throughout this article. The data is complex, both inherently and because of changes in researchers' understanding, so I aim to provide not just the information necessary for this article, but also a comprehensive overview in a more accessible form than the Khoisanist liter-

ature. The major omission is the tonology, which is complex and not perfectly understood; it is not relevant for the purpose of this article, so I give only a sketch.

2.1 The sounds of !Xóǀ – overview

Until recently, our knowledge of !Xóǀ came mainly from Tony Traill’s thirty-year study of the language, the major publications being the two books Traill 1985 and Traill 1994. Traill chiefly studied an eastern dialect of the language. Recently, a DOBES¹ project team at MPI Leipzig, has, as part of a larger language documentation project, conducted a segment inventory of a western dialect (Naumann forthcoming). There are some differences in the analyses (Naumann finds even more distinctions than Traill), but these differences are not essential for the purposes of this paper. I will adopt the DOBES inventory, but use mainly Traill’s data, supplementing it with DOBES data as appropriate, as the full DOBES data is not yet publicly available.

2.1.1 Morphophonological structure

Although the morphology of !Xóǀ is not fully worked out, analyses by Traill (1994), Naumann (2008) and Kießling (2008) can be somewhat crudely summarized as follows.

!Xóǀ has a very simple word structure. Phonologically, a content word (noun, verb, adjective) has the form C*V{V/CV/C}: that is, there is a first mora, which starts with a possibly complex consonant, and has a vowel (which carries tone and may have several voice qualities); then there is a second mora, which is either a vowel (again with tone and perhaps nasalized), or a consonant (from a small set) and a vowel, or just a consonant (a nasal, which appears to carry tone in some cases). Function words are typically but not invariably monomoraic; and loan words and onomatopoeic words may vary from this structure. With the content words, the first mora is the root, and the second mora carries grammatical information, such as concord class. Most words in a sentence have their second mora determined by that of the ‘head noun’; the concord system is fairly complex. In citing words that inflect concordially, Traill uses the notations **-V**, **-JV**, **-BV**, **-LV** as morphophonological representations of the second mora. For example,

- (1) a. The noun ||χáǀ ‘the point between the shoulder-blades’ is ||χá-ǀ, where ||χá- is the root, and ǀ the suffix (it is a class 2 noun, but the u appears

¹ Volkswagen Foundation programme *Dokumentation Bedrohter Sprachen* (Documentation of Endangered Languages), <http://www.mpi.nl/DOBES/>

to be arbitrary, with the nasality the only observable association with the class 2 forms).

- b. The verb $\text{‡q}^{\text{h}}\text{áJV}$ ‘squash between the nails’ has nominal form $\text{‡q}^{\text{h}}\text{áí}$ and may appear concordially as $\text{‡q}^{\text{h}}\text{áji}$, $\text{‡q}^{\text{h}}\text{ána}$, $\text{‡q}^{\text{h}}\text{áje}$, $\text{‡q}^{\text{h}}\text{áju}$, or $\text{‡q}^{\text{h}}\text{án}$, with surface tones also determined by concord.

Were I to pinch someone at the point between the shoulder blades, the verb would agree with the object and appear (by construction following Traill’s grammatical sketch, not by attestation) as

- (1) c. $\bar{n} \acute{n} \text{ b} \grave{a} \text{ ‡q}^{\text{h}}\text{ána} \parallel \chi \acute{a} \acute{u}$
 I *pres pf* pinch point

whereas with a different noun class it would have a different suffix and tone:

- (1) d. $\bar{n} \acute{n} \text{ b} \grave{a} \text{ ‡q}^{\text{h}}\text{áje} \mid \text{q} \acute{a} \text{m}$
 I squash ant

These words may then be extended with (usually monomoraic) affixes to form longer phonological words; such affixes do not contain clicks. Compound words are also possible, and (at least in the dialect studied by Traill) reduplication of the entire word is a common phenomenon.

2.1.2 Tone

Traill marks four surface tones, which apply to the (bimoraic) word: high (\acute{a}), mid-level (\bar{a}), mid-falling (\hat{a}) and low (\grave{a}). Naumann (2008) analyses this as two monomoraic tones, high and low, so that Traill’s surface tones are HH, LH, HL and LL. This analysis is not completely without problems (Naumann forthcoming), but is mostly successful. There remain some monomoraic words which appear to bear a compound tone. The tones are strongly affected by vowel voice quality, and are extensively modified by the concord system. In this article, I shall use Traill’s markings for surface tone when citing forms.

2.1.3 Consonant overview

Table 1 presents the consonant inventory of !Xóǀ in chart form. The columns are labelled by place of articulation; the rows will be referred to by number. This chart presents the largest inventory: firstly, it includes the DOBES western dialect analysis; secondly it presents, in the lower half, a large number of consonants which are notated as phonetic clusters. I discuss in §3.3 whether these are phonological clusters. In the following sections, I describe the consonants in detail.

	labial	alveolar	alveolar	palatal	labial click	alveolar click	alveolar click	lateral click	palatal click	velar	uvular	glottal
1.	p	t	ts		⊙		!		≠	k	q	ʔ
2.	b	d	dz		⊙		!		≠	g	ɢ	
3.	p ^h	t ^h	ts ^h		⊙ ^h	^h	! ^h	^h	≠ ^h	k ^h	q ^h	
4.	b ^h	d ^h	dz ^h		⊙ ^h	^h	! ^h	^h	≠ ^h	g ^h	ɢ ^h	
5.	p'	t'	ts'		⊙'	'	!'	'	≠'	k'	q'	
6.			dz'		⊙'	'	!'	'	≠'	g'	ɢ'	
7.												q ^{ʔ'}
8.												ɢ ^{ʔ'}
9.					⊙̃	̃	!̃	̃	≠̃			
10.	m	n		ɲ	⊙̃	̃	!̃	̃	≠̃	ŋ		
11.	ʔ _m	ʔ _n			ʔ̃⊙̃	ʔ̃ ̃	ʔ̃!̃	ʔ̃ ̃	ʔ̃≠̃			
12.	f		s								ç	h
13.	w	r,l		j								
14.					⊙q	q	!q	q	≠q			
15.					⊙ɢ	ɢ	!ɢ	ɢ	≠ɢ			
16.					⊙q ^h	q ^h	!q ^h	q ^h	≠q ^h			
17.					⊙q ^h	q ^h	!q ^h	q ^h	≠q ^h			
18.					⊙q'	q'	!q'	q'	≠q'			
19.					⊙q'	q'	!q'	q'	≠q'			
20.	ppq ^{ʔ'}	ttq ^{ʔ'}	tsq ^{ʔ'}		⊙q ^{ʔ'}	q ^{ʔ'}	!q ^{ʔ'}	q ^{ʔ'}	≠q ^{ʔ'}			
21.		dq ^{ʔ'}	dzq ^{ʔ'}		⊙q ^{ʔ'}	q ^{ʔ'}	!q ^{ʔ'}	q ^{ʔ'}	≠q ^{ʔ'}			
22.		tç	tsç		⊙ç	ç	!ç	ç	≠ç			
23.		dç	dzç		⊙ç	ç	!ç	ç	≠ç			
24.					⊙h	h	!h	h	≠h			
25.					⊙h	h	!h	h	≠h			
26.					⊙ʔ	ʔ	!ʔ	ʔ	≠ʔ			
27.					⊙ʔ	ʔ	!ʔ	ʔ	≠ʔ			

Table 1: The consonant inventory of !Xóó

2.2 Non-click consonants

A striking feature of !Xóǀ (and Khoisan more generally) is that all the consonantal complexity occurs word-initially – only a few consonants occur medially or finally. It is therefore natural to consider the positions separately, and I first describe the initial consonants without clicks.

2.2.1 Initial non-clicks

This part of the inventory is already quite rich. In the top left and right of the chart, we have a set of stops with five or six places and five to eight manners, depending on count. Apart from the glottal stop, there are five places of articulation: labial, dental, dental/alveolar affricated, velar, uvular. A typologically unusual feature of !Xóǀ is that oral labial stops are marginal: in Traill 1994, almost all the few words starting with labial stops, and all words starting with **p**, are loanwords.

The manners are more or less as written: the voiceless, voiced and aspirated stops (rows 1–3) are familiar from languages with this distinction: voiceless stops have about zero VOT, whereas voiced stops have voice lead, and aspirated stops voice lag. The voiced aspirated stops (row 4) are, however, not like the familiar breathy-voiced stops of Indic languages: they have voice lead, which persists into the [z] of **dz**, and then at release voicing ceases for the aspiration. Ejectives (row 5) are also familiar; the voiced ejectives (row 6) have voice lead, followed by an ejective release (so **dz'** is rather [ds']).

The uvular ejective affricates **qʰ'**, **gʰ'** (rows 7–8) might be considered another place or another manner; because of their occurrence in clusters, it is convenient to arrange them as manners. They are pronounced as notated, although there is some room for argument about whether they are really velar or uvular – see the discussion in §3.1.

Of the plain nasals (row 10), only **m**, **n** occur initially. The glottalized nasals (row 11) are initials, and are nasal stops with an initial glottal check.

Of the continuants (rows 12–13), **s**, **ʒ** and marginally **h** occur initially in native words; the others may occur in loanwords.

Finally, in the bottom left of the table, there is a group of initials written as phonetic clusters. The pulmonic clusters (rows 22–23) are pronounced as written, with a strong uvular fricative. The ejective clusters (rows 20–21) vary according to dialect and register. Again, the exact place is arguable, and in Eastern careful speech, Traill records pronunciations such as [t'q'], although with no instrumental confirmation of a true double ejective. These clusters are rare in the DOBES data, but reasonably supported by Traill (1994), apart from **pqʰ'**, which occurs only in the superbly onomatopoeic word **pqʰ'ali** ‘the sound of a rapid evacuation of the bowels’.

2.2.2 *Medial consonants*

As remarked in §2.1.1, the bimoraic word may be bisyllabic, with the second syllable starting with one of a very small set of consonants. These are **b**, **m**, **n**, **ɲ**, **ʝ**, **l**, **r**.

ʝ in Traill's data varies from [j] to [ɟ]. In Traill, **r** occurs only in loanwords; in DOBES, **l** occurs only in loanwords, and **r** corresponds to Traill's **l** in native words.

2.2.3 *Final consonants*

The final consonants are **m**, **n**, **ɲ**, **p**, **b**, **r**. All but **m**, **n** are marginal, occurring in loanwords or onomatopoeic words. According to DOBES, final **m**, **n** are more vocalic than consonantal, carrying a mora and a tone. Curiously, Traill does not mention this, although it is very obviously true in his recordings.

2.3 **Click consonants**

All click consonants are initial. I describe the clicks in the order of Table 1.

2.3.1 *Click consonants, simplex*

The clicks in the top half of the chart, in rows 1–11 are notated in a way suggesting a phonetically simplex consonant. The anterior articulation of these clicks matches their non-click counterparts: for example, $\underset{\cdot}{\text{t}}^{\text{h}}$ (row 4) is a palatal click, with voice lead up to the posterior (velar) closure, and aspiration following the posterior release. The voiceless nasal clicks (row 9) such as $\underset{\cdot}{\text{ŋ}}$ have no non-click counterparts. They are pronounced as written: a voiceless $\underset{\cdot}{\text{ŋ}}$ together with velar lowering around the closure period. This accompaniment will be discussed further below, in §5.2.

2.3.2 *Click consonants, complex, long closure*

The clicks given in rows 14–21 are written with a following [q], which as noted at the beginning is intended to indicate a significant prolongation of the posterior closure. Thus in **ɲ**, the click burst is more or less simultaneous with, and so drowns, the posterior release, whereas in **ɲq** the posterior release can be heard after the click burst (and seen on the spectrogram).

The various modifications – aspiration, ejection, ejective affrication – of the posterior release are pronounced as written.

The voiced consonants, in the odd-numbered rows, are pronounced with voice lead into the posterior closure period, and it is not unusual to hear nasalization as well, which is probably simply phonetic enhancement of the pre-voicing. Voicing stops before the posterior release.

2.3.3 Click consonants, complex, other

The final section, in rows 22–27, contains clicks where the click appears to be (phonetically) followed by another sound. It is of course a question to be discussed below whether these are phonological as well as phonetic clusters. Here I just describe the phonetics.

The **Ɂχ** fricative clicks in rows 22–23 are so notated because the fricative is fairly long and prominent, making [Ɂχ] more descriptive than the possible alternative [Ɂʎ] suggesting an affricated posterior release. As I discuss below, there are also systematic reasons for treating them as a click followed by a fricative.

The **Ɂh** clicks in rows 24–25 have received special attention in the phonetic literature. This, or a similar, **Ɂh** accompaniment is found in other languages, including Khoekhoe. It has a distinctive auditory impression, as one hears a long crescendo aspiration (some 200 ms, sometimes even 400 ms) after the click; but the posterior release is not audible. For Khoekhoe (Nama), Ladefoged and Traill (1984) used airflow measurements to establish that the silent start is achieved by nasal venting during the click [Ɂ̃h]; for !Xóǀ, Traill (1991) showed that this is supplemented by breathing in during the click (so [Ɂ̃↓h]), making it the only established example of ingressive pulmonic airflow in normal language.² There is a question about whether the nasalization is phonetic or phonological, which will be touched on below. I treat it as phonetic, and do not write it.

The clicks **Ɂʔ** with glottal stop in rows 26–27 also tend to have nasalization, at least in the voiced version, and this may or may not be phonological – here I have assumed not. They are auditorily distinguished from the ejectives **Ɂ'** in rows 5–6 mainly by the lack of an audible posterior release – similar to the difference between saying [ak'a] and [ak'ʔa].

2.4 Vowels

The vowel system is also rich. Its basis is a simple five-vowel system, **a**, **e**, **i**, **o**, **u**. The front vowels **i**, **e** are fairly well localized around approximately cardinal values; **o**, **u** tend to spread a little more, centralizing in some contexts, sometimes to the extent of neutralizing with each other; **a** is more variable, spreading over most of the lower half of the IPA chart, between [ɑ, a, ɜ]. I shall discuss the behaviour of **a** in some detail later, in §5.1. As most words are bimoraic, long vowels and diphthongs occur; there seems no reason to treat these as anything other than a sequence of two vowels. The following combinations are *not* attested in Traill: **ea**, **eo**, **eu**, **ie**, **io**, **iu**, **uo**, and are also not found in the DOBES data.

² Since Traill was fluent in the language for twenty years before discovering this, it can be assumed that ingressive airflow is a phonetic detail.

Plain	i e a o u	
Nasalized	ĩ ē ã õ ù	
Breathy	ĩ̤ ē̤ ă̤ ɔ̤ ʉ̤	
Creaky	ḭ̃ ḛ̄ ă̰ ɔ̰ ʉ̰	
Pharyngealized	a ^ɣ o ^ɣ u ^ɣ	
Strident	ă ^ɣ ɔ ^ɣ ʉ ^ɣ	
Breathy creaky	ĩ̤̰ ē̤̰ ă̤̰ ɔ̤̰ ʉ̤̰	<i>Traill only</i>
Creaky pharyngealized	ă ^ɣ ̰ ɔ ^ɣ ̰ ʉ ^ɣ ̰	
Creaky strident	ă ^ɣ ̰̰ ɔ ^ɣ ̰̰ ʉ ^ɣ ̰̰	

Table 1': The vowel inventory of !Xóó

The complexity of the vowel system arises from the addition of voice qualities and nasalization to the basic vowels. Phonetically, one hears breathy vowels [V̤], where breathiness may extend over the entire stem; creaky vowels [V̰], where the creak usually occurs in the middle of the first vowel (as with, say, Vietnamese), and may vary from light creaky voice (or even be omitted in fast speech) to a full glottal stop; pharyngealized back vowels [V^ɣ] in the first vowel; and the well-known STRIDENT back vowels [V^ɣ], which have strong epiglottal friction and are often voiceless. Although Ladefoged and Maddieson (1996) treated stridency phonetically as a distinct phonation type (and notated it [V̤̰] to emphasize this), Traill considered (with good reason) that phonologically strident vowels are the realization of breathy pharyngealized vowels /V̤̰^ɣ/. This latter understanding has been continued in the DOBES orthography, and I adopt it here also.

Furthermore, Traill reports breathy creaky vowels [V̤̰], which start breathy and then glottalize; creaky pharyngealized back vowels [V̰^ɣ]; and even strident creaky vowels [V̰^ɣ], which start strident and become glottalized, and are phonemically creaky breathy pharyngealized /V̤̰^ɣ/

Yet further, all of these also occur nasalized, where the nasalization is usually heard over both vowels in the stem. However, there are good reasons to believe that nasalization belongs on the second vowel of a word, whereas the voice qualities belong on the first vowel. Phonemically, therefore, we have the inventory given in Table 1'.

2.5 Phonotactics and phonological processes

There are several phonetic rules given in Traill 1994 which modify the phonetic realization of the inventory given above, and also some phonotactic constraints (from Traill 1985) which limit the number of possible words. Here I will describe

a few which will form part of my argument later.

- (2) Single Aspirate Constraint: A word contains at most one segment that is aspirated, breathy or strident.
- (3) Single Glottal Constraint: A word contains at most one segment that is glottalized or creaky.
- (4) Pharyngeal Constraint: A pharyngealized or strident vowel may not follow an aspirated, ejected, or fricated click. (I.e., it may follow only Ɂ , Ɂ̤ , ʔ̤ , Ɂq and their voiced versions.)

These constraints are strong, but apparently not quite inviolable. Traill 1994 contains four or five lexemes violating (2), and DOBES has two. In every case, non-violating alternatives appear to exist, so they may be instances of phonetic spreading. The appearance of STRIDENT in (2) forms part of the evidence for ‘strident = breathy pharyngealized’.

(3) has two (related) violating lexemes in Traill 1994, and none in DOBES.

(4) applies for the most part with non-click stops as well, but there are a couple of violations there, and in particular, as I shall use later, Traill 1994 gives half a dozen words in **h-** containing pharyngealized vowels.

- (5) Phonetic Back Vowel Constraint (BVC): A BACK consonant may not be followed by a (phonetic) front vowel (**i**, **e**), where the BACK consonants are the velar and uvular non-clicks, and the clicks involving ʘ , **!**, **||**.

The BVC (see, e.g. Miller 2011) applies in some form across the Khoisan languages, with varying notions of BACK (sometimes excluding velars, for example) and different strengths. In the Khoe languages such as Khoekhoe (Nama), the BVC is far from an absolute constraint, but is a clear statistical tendency; in the Tuu and Ju languages the BVC is stronger. The form in (5) addresses surface representations; Traill in fact proposes (Traill 1985, p. 90) the stronger (6).

- (6) Phonological Back Vowel Constraint: A BACK consonant, including any click, may not be followed by a (phonological) front vowel.

He then accounts for (most of) the exceptions by a phonetic rule which creates the exceptional front vowels from underlying **a** in the presence of FRONT clicks. I shall discuss this somewhat counter-intuitive approach at length later in §5.1; for the moment, I just state (7) (Traill 1985, p. 70).

- (7) A-Raising Rule (ARR): First mora plain, breathy or creaky **a** is raised to [ɜ] when
- a. the second mora contains **i**, or is a nasal, and the word starts with a dental non-click or |, ɸ.
 - b. It is further raised to [i] when the second mora is just **i**.

3 Click consonants – questions and analyses

In this part, I review previous work on the phonology of Khoisan click consonants.

3.1 Posterior place distinctions

Before turning to the question of clustering, I discuss one small controversy which interacts with it. In my descriptions, I said that the salient difference between **Ɂ** and **Ɂq** was the prolongation of the posterior closure. However, Ladefoged and Maddieson (1996) describe the difference as one of velar versus uvular place for the posterior closure. This description comes ultimately from Traill, who in his works described **Ɂ** as velar, and **Ɂq** as uvular. He described some of the other complex clicks as having velar articulation; and he also considered the non-click ejective affricates to be phonetically and phonologically **k^x** rather than the DOBES **q^x** that I have adopted. However, in Traill 1994, he was a little more cautious about this, and it is unclear what his final view was.

DOBES, on the other hand, does not need to commit to the exact place of the posterior closure of clicks, and considers the complex prolonged closure clicks to be clusters with members of the uvular non-click series.

The **Ɂ/Ɂq** distinction is widespread in Khoisan, and so has been considered by other researchers. In particular, Miller et al. (2009) raise the question of whether it is even possible to maintain a velar/uvular distinction, and conclude that it is not. They adduce direct articulatory measurements for this – ultrasound imaging (see also Miller, Namaseb, and Iskarous 2007) shows that clicks have a posterior constriction in the uvular to pharyngeal region, depending on the click type.

I have also made some informal experiments deliberately trying to make a velar/uvular posterior contrast (using ultrasound to check the actual articulations), and I cannot convince myself that I can make such a distinction in a plain click, although with a prolonged closure it seems feasible to advance or retract the closure before release.³

³ I am grateful to James M. Scobbie of Queen Margaret University for kindly allowing the use of their ultrasound equipment, and to Stephen Cowen for generous training and support.

I therefore assume here that no velar/uvular posterior place distinction exists in clicks, and refer to Miller 2011 for further discussion.

3.2 Features for clicks

Given their typological rarity, it is not surprising that there is no commonly agreed set of features, or even any several commonly agreed sets of features, for click consonants. Here I briefly review some of the proposals. All authors recognize the separation of click and accompaniment, so all proposals split into a set of features to distinguish the anterior closure/release, and one for the posterior release.

Jakobson (1968) considered how to fit non-pulmonic consonants into his distinctive feature theory. His proposals have had little take-up, so I refer to Traill 1985, ch. 5 for a full description and detailed critique. He has a complex interaction between features for non-pulmonics, but clicks are distinguished by [+checked], and then [tense, lax, strident] etc. can be used to distinguish accompaniments, while [acute] and [compact] can be used for anterior place. However, Traill concludes, in a scathing but solid analysis, that Jakobson's system does not even work for the languages he attempts to describe, let alone for the complexity of !Xóǀ.

Chomsky and Halle (1968) considered clicks in some detail, based mainly on phonetic descriptions of Bantu and Khoekhoe. Clicks carry the distinctive feature [+suction]. The anterior place and release are treated articulatorily in the obvious way by means of [anterior, coronal, lateral, delayed (primary) release]. The accompaniments were described mostly by means of new features introduced for the purpose, such as [delayed release of secondary closure] and [heightened subglottal pressure]. Their system works better than Jakobson's, but again Traill's detailed analysis concludes that it is neither extensive enough to cope with !Xóǀ, nor do the *SPE* features very naturally account for the phonological behaviour of clicks in !Xóǀ. See also §5.1 for a discussion of one aspect of using *SPE* features in clicks.

Snyman (1970) nominally adopts a distinctive feature analysis, but does so, as one might say, pragmatically. He simply invents a feature for each articulatory characteristic: [clear, laryngeal, glottalized ejective] and so on. There is no principled analysis.

Traill (1985), after a long and careful discussion, arrives at a system rather similar to Snyman's, but cleaner and better justified; however, he goes beyond standard feature theory by using contoured values for some features, such as his [friction]. He does not consider this proposal satisfactory. One of the more interesting points is that he several times discusses proposals to give segments internal structure, following e.g. Campbell 1974, so that the cluster phonemes

can be internally split into click and accompaniment while remaining as single phonemes. In Traill 1993 he followed up on this by putting these thoughts into a formal feature geometry setting. However, he was also not fully satisfied with this, and did not adopt it.

Güldemann (2001), as I discuss further in §3.3.3, carries out an extensive and detailed study of sound systems across Khoisan. A notable aspect of his analysis is the emphasis on hierarchical structure: he uses features that are ordered. For example, he has three distinct [stop] features: the first, high in the hierarchy, captures the difference between the nasal clicks and the rest. The second [stop] occurs underneath the scope of an [elaboration] feature, and describes whether the elaboration (meaning any accompaniment except nasality and voice, which are considered more primitive) contains a separate stop in addition to the click. Then there is a second [elaboration] feature, which describes the ejective accompaniments; and below that, the third [stop] feature, which distinguishes \mathfrak{xq}^h from \mathfrak{xq}^{h} (he considers the glottalization in the latter to be phonetic). This is essentially a feature geometry presentation, but as I discuss below, he goes beyond the standard setting.

Miller-Ockhuizen (2003) works mainly at a phonetic rather than formal phonological level; she uses generally articulatory features, but in particular introduces [pharyngeal], characterizing certain clicks, and the acoustic feature [spectral slope] capturing stridency and glottalization.

As I discuss in §3.3.5 below, Miller et al. (2009) go beyond Traill's tentative use of contoured features by introducing contoured airstream features.

In this article, the choice of features for clicks is not a primary concern. Indeed, I am not even committed to the use of features in any particular formal theory; here, it suffices to have some notion of classifying sounds. In the formal development, I will assume *SPE*-like features, and avoid discussion of the details that have vexed previous researchers.

3.3 Clusters or not?

3.3.1 Unitary analyses

Until the 1970s, linguistic descriptions of Khoisan languages recognized the different series of clicks, but did not analyse the accompaniments, which were then called EFFLUXES (Beach 1938). That work itself is a very thorough (and still useful) study of Khoekhoe; but Beach does not classify or analyse the effluxes (of which Khoekhoe has only five: \mathfrak{x} , \mathfrak{x}^h , \mathfrak{xh} , \mathfrak{x}^{h} , \mathfrak{x}^{h}).

Still in 1970, Snyman took the same approach in his study of the Ju language Ju|'hoansi, also called !Xū. This language has the usual four !, |, ||, † click types,

with, according to Snyman (1970), some fourteen accompaniments.⁴ Snyman explicitly presents each such consonant as a phoneme, ascribing *SPE*-style features to each phoneme.

This unitary click analysis (UA for short, following Nakagawa's (2006) analysis of |Gui (*Ethnologue* |Gwi)) has obvious drawbacks, which become more pressing with the increasing number of accompaniments. In the case of !Xóǀ, it leads to the statement that !Xóǀ has 83 (attested) distinct click phonemes per Traill, or 115 per DOBES, as they appear in Table 1. While few things can be said to be impossible, many people find this to be beyond the limits of what human language might be expected to maintain. There are several reasons for this. For one thing, it poses a considerable challenge to the language acquirer. This is especially so when one considers the rarity of many of the 'phonemes'. The size of the !Xóǀ vocabulary is not known, but Traill 1994 lists about 3000 native words (or rather stems), of which about 2000 contain clicks. Though the true native vocabulary may be (or may have been before the enforced sedentarization and migration in the 1980s and 1990s) rather larger, Traill was specifically looking for phonologically illustrative material. Nonetheless, there are three 'phonemes' that occur in only one word each – for example, the sound $\tilde{\Theta}$ is supported only by $\tilde{\Theta}\hat{a}a$ 'sit or stand close together' – and thirty that occur in fewer than ten words each, including every member of the Θ series. Table 2 lists the number of words for each click sound recorded in Traill 1994.

Another indication of the functional load of each phoneme is the incidence of minimal pairs. While there is in general no reason to expect contrasts to be demonstrable between every pair of phonemes, counting the total number of pairwise contrasts gives an indication of the global strength of contrasts. Taking English, for example, with its average sized consonant inventory, more than 95% of the possible pairwise consonant contrasts are illustrated by minimal pairs, even when one only considers monosyllables.⁵

In !Xóǀ, the expected number of minimal pairs is decreased by its very large vowel inventory (as well as the non-click consonants), but increased by the very restricted shape of words: given the basically bimoraic word shape, and the various phonotactic restrictions, there are about 13 000 possible click-initial words

⁴ Miller-Ockhuizen 2003 differs, giving twelve accompaniments. Whether this difference marks a difference in dialect or analysis, I do not know. Generally, Miller-Ockhuizen's analysis is substantially more complex than Snyman's.

⁵ E.g. 'bin/pin/fin/Vin/win/tin/din/thin/sin/nin/rin/Lynne/chin/gin/shin/yin/kin' provide $(17 \times 16)/2 = 136$ of the $(24 \times 23)/2 = 276$ contrasts (assuming 24 consonants in English). Lists of minimal pairs are widely available in speech pathology materials; I used Higgins 2013 to find the 95% with monosyllables. Most of the missing pairs are contrasts involving /ŋ/ and /ʒ/, whose status as phonemes is fairly recent (and dialect dependent).

Ø	9		62		103	!	107	‡	69
Ø̥	6	↓	57	↓↓	74	!↓	70	‡↓	58
Ø ^h	6	^h	28	^h	30	! ^h	21	‡ ^h	39
Ø ^h	2	↓ ^h	5	↓↓ ^h	7	!↓ ^h	7	‡↓ ^h	11
Øχ	2	χ	15	χ	28	!χ	19	‡χ	19
Øχ	1	↓χ	10	↓↓χ	13	!↓χ	9	‡↓χ	10
Øq ^{z'}	1	q ^{z'}	7	q ^{z'}	17	!q ^{z'}	15	‡q ^{z'}	8
Øq ^{z'}	3	↓q ^{z'}	9	↓↓q ^{z'}	13	!↓q ^{z'}	4	‡↓q ^{z'}	10
Øq	5	q	16	q	28	!q	15	‡q	35
Øg	3	↓g	29	↓↓g	29	!g	35	‡g	42
Øq ^h	0	↓q ^h	4	↓↓q ^h	7	!↓q ^h	3	‡↓q ^h	0
Øq'	3	q'	16	q'	25	!q'	12	‡q'	5
Øh	7	h	13	h	18	!h	24	‡h	15
Ø̃	1	↓̃	5	↓↓̃	8	!↓̃	11	‡↓̃	3
Ø̃	9	↓̃	60	↓↓̃	82	!↓̃	104	‡↓̃	48
ʔ̃	2	ʔ̃↓	33	ʔ̃↓↓	41	ʔ̃!↓̃	37	ʔ̃‡↓̃	15
Øʔ	5	ʔ	23	ʔ	32	!ʔ	28	‡ʔ	35

Table 2: Click frequencies in the lexicon (Traill 1994)

in the Traillian analysis, ignoring tone – compare to the 36 000 or so possible English monosyllables. It is perhaps remarkable that !Xóǝ does have a little more than half of the 3403 unitary minimal pairs;⁶ and almost three quarters if one ignores tone.⁷ Nonetheless, combined with the rarity of many unitary phonemes, one must wonder how so many distinctions survive.

If we take a more realistic approach, and only ask for each click to contrast with other clicks of the same anterior place (analogous to looking for contrasts among English /t, d, s, θ, ð, n, l, r/), the picture is somewhat better, but still surprisingly rarefied: almost 30% of such contrasts are not supported by a minimal pair, even if we ignore tone. In English, all contrasts of manner at a given place, except /j/ vs /ɜ/ (if this counts), are supported by multiple minimal pairs, even for such historically recent contrasts as /θ/ vs /ð/.⁸

⁶ The phonology of Traill 1994 has 83 attested click consonants, hence 3403 contrasts.

⁷ The largest minimal set has size 31, with shape /C*aa/; ignoring tone, the set /C*aa/ has size 49.

⁸ The statements in this section about minimal pairs in !Xóǝ were computed by scripts from a manually entered list of the headwords from Traill 1994.

3.3.2 Cluster Analysis

Even a cursory glance at Table 1 must invite the suspicion that at least the more complex accompaniments are really clusters. Consider, for example, the click \mathfrak{xq}^x (row 20). Given that we see also the free-standing consonant q^x (row 7), as well as the non-click pq^x , tq^x , tsq^x combinations also in row 20, the suspicion becomes practically unshakable. Moreover, as I noted above, all these sounds vary similarly with dialect and register – [q^x] (or velar [k^x] according to Traill) itself is a western dialect pronunciation, whereas the eastern dialect pronounces [k^xq] in citation form, with the western form in fast speech (Traill 1993, p. 36).

In Traill’s first book (1985), he assumed unitary analysis, despite its “implausibility”, for most of the book, pleading reluctance to violate tradition. However, at the end of the book, he made the above argument, and proposed what I shall call Cluster Analysis (CA).

As one can see from Table 1, “every one of the simple accompaniments *that forms a phonetic cluster* with a click (except possibly for delayed aspiration) exists as an independent consonant” (Traill 1985, p. 209, original emphasis). Traill therefore proposed a fairly extensive CA, in which the basic clicks are [\mathfrak{x} , \mathfrak{x} , \mathfrak{x} , \mathfrak{x}], and all the others are viewed as clusters. This obviously simplifies the phoneme inventory dramatically: instead of $17 \times 5 = 85$ click phonemes, there are just $4 \times 5 = 20$, and all the others arise from combinations with phonemes already in the non-click inventory. It also (he asserts) has other nice effects on the phonological analysis, mostly by converting complex ‘featural’ rules into natural co-articulatory consequences of the components of the clusters.

This CA is not completely unproblematic. Traill mentioned a couple of “minor details”, such as the awkward absence of free-standing /h/ other than in a couple of interjections; other problems arose later when he (1993) attempted to put !Xóǀ in a feature-geometric framework: the durations of some clusters did not match very well with feature-geometric requirements on timing slots. Despite this, the analysis seems compelling to many.

In recent years, CA has become quite well accepted as the natural way to analyse Khoisan languages. I have already mentioned Güldemann’s (2001) cross-Khoisan analysis, and will discuss it further below.

A recent substantial work discussing cluster analysis at some length is Nakagawa 2006. |Gui is a Khoe language spoken in Botswana of fairly high click complexity, with the usual four clicks, and thirteen accompaniments, which are subset of the !Xóǀ range. Nakagawa adopts a cluster analysis (MCA for Moderate Cluster Analysis) based on Traill’s proposals. Because, unlike Traill (1985), he recognizes plain ejectives (\mathfrak{x}) and aspirates (\mathfrak{x}^h), he includes these as basic

clicks, so ending up with $4 \times 6 = 24$ click phonemes, plus $4 \times 7 = 28$ clusters.

Similarly, Naumann’s (forthcoming) study of western !Xóǒ also adopts a Traillian analysis, largely following and extending MCA – my terms ‘simplex’ and ‘complex’ in §2.3 are chosen to match with the DOBES view that rows 1–13 of Table 1 are phonemes, and rows 14–27 are clusters. As well as the arguments on grounds of parsimony and symmetry of systems, and on the grounds of the phonetic properties that I sketched in §2.3, Naumann also gives some informal observations of speaker behaviour that seem to support CA: for example, his informants sometimes described !q^h- words as starting with !. Under MCA, the phonemes are those in Table 3.

ǀ	ǂ	ǃ ^h	Ǆ ^h	ǅ	ǆ	Ǉ	ǈ
	↓	! ^h	! ^h	!'	!'	!̃	!̃
!	!	! ^h	! ^h	!'	!'	!̃	!̃
		^h	^h	'	'	̃	̃
‡	‡	‡ ^h	‡ ^h	‡'	‡'	‡̃	‡̃

Table 3: !Xóǒ click phonemes under MCA

3.3.3 Güldemann’s analysis

The cross-Khoisan analysis of Güldemann (2001) is quite radical from the point of view of phonological theory. Some of its roots lie in Traill’s discussions of early notions of subsegmental structure, but Güldemann goes further. As I sketched above, he uses a hierarchical structure, so that segments can combine to make bigger segments. One of his main aims is to integrate the click and non-click systems, so there is a top-level featural distinction [suction] (following *SPE*) distinguishing clicks, and then below that a hierarchy of features/subsegments. For him, ‘simple’ stops are the voiced and voiceless stops/clicks. Simple stops can be modulated by aspiration and glottalization (ejectivity is treated as glottalization for phonological reasons, such as the constraint (3)), to produce ‘complex’ stops. Either simple or complex stops can then be sequentially combined with other stops to form ‘cluster’ stops – which are both clusters and single segments with their own featural description.

Güldemann’s discussion brings in a number of aspects of cross-Khoisan phonology, but a detailed review would take more space than is justified for the purposes of this article. Suffice it to make three observations. Firstly, he remains unable to settle firmly on the appropriate set of place features for clicks, owing to some of the issues mentioned above in §3.2. Secondly, for him the !Xóǒ alveolar affricate series (**ts** etc.) is indeed phonologically affricated, whereas Traill treats it (as I implicitly do) as an incidentally affricated series of alveolar stops. Fi-

nally, it is not entirely clear how this approach is to be integrated into formal phonological theories, whether rule- or constraint-based.

3.3.4 *Radical Cluster Analysis*

A section of Nakagawa 2006 that requires special mention for this article is pp. 255–261. Here he considers ‘Radical Cluster Analysis’ (RCA). RCA is ‘radical’ in that it proposes that there is only one click phoneme at each place – which, as will be seen, is precisely the argument this article makes about Khoisan. However, Nakagawa sets up RCA as a straw man to justify his preferred analysis – it is germane, therefore, to explain why he argues that RCA fails. I will go on to argue, as the proposal of this paper, that it is in fact correct to propose such a radical analysis, but a conceptual change in the nature of phoneme and segment is required for it to work as desired.

The difficulty Nakagawa has is choosing which click is basic. Anybody’s first thought would surely be that the plain unvoiced click is the basic click. However, Nakagawa finds this untenable, because although |Gui has the voiced nasal click [ǀ̚] (but not [ǀ̚̚]), it does not have a plain velar nasal [ŋ] in its inventory with which ǀ could cluster. He concludes, therefore, that the only viable choice for the unit click in RCA is the nasal click, with some phonetic rules to explain how it combines with other phonemes to form the other clicks – rules that have to be inelegantly restricted in their application, to avoid destroying the non-click inventory.

As a reviewer observes, it is questionable whether Nakagawa’s reasons are sufficient; ǀ̚ could simply have a defective distribution, or possibly the nasal that combines with clicks is ǀ̚ (which is compatible with my later formulation in which click accompaniments are not specified as velar or back). However, I claim that while radical analysis is correct, a change to parallel clustering brings a number of improvements.

3.3.5 *Arguments against cluster analysis; Miller’s approach*

Although Naumann (forthcoming) adopts CA, he also found some evidence weighing against it. Firstly, it is surprising that the A-Raising Rule (7) still operates following clusters with uvular stops – one would expect a uvular to block any raising effect of the previous click. Secondly, he conducted an informal onset-dropping experiment: two speakers were trained to drop the first sound of words in Afrikaans, and then asked to do the same with !Xóǀ words. Neither speaker simply dropped the click from the cluster; either they dropped the entire cluster, or sometimes produced words starting with **h** or **ʔ**. My proposal will resolve both these issues (see §5.1 and §4.1.1).

Amanda Miller, whose dissertation study (published as Miller-Ockhuizen 2003) of Jul’hoan was mentioned earlier, has recently been working with a number of colleagues on the almost extinct language N|u. Although in 2003 she followed a CA, in Miller et al. 2009 (mentioned briefly above) she and her colleagues argue that cluster analyses are wrong. Instead, they propose to extend the range of features by which clicks are classified, and in particular to add contoured values for the airstream feature. These are to simple airstream values as affricates are to stops and fricatives. N|u has a mid-sized range of accompaniments, which, adapting Miller et al.’s notation to ours, are **ɲ**, **ɲ^h**, **ɲ̤**, **ɲ̤ʔ**, **ɲ̤^h**, **ɲ̤̤**, **ɲq**, **ɲq^h**, **ɲχ**, **ɲχ’**.⁹

The way that Miller et al. classify these clicks by ‘airstream mechanism’ is:

- The simple and nasal clicks **ɲ**, **ɲ^h**, **ɲ̤**, **ɲ̤ʔ**, **ɲ̤^h**, **ɲ̤̤** are said to have simply lingual airstream.
- The clicks **ɲq**, **ɲq^h**, **ɲχ** are said to have ‘linguo-pulmonic’ airstream, reflecting their status (as in the similarly notated !Xóǀ clicks) as moving from a click into a normal pulmonic release, with a clearly audible [q, q^h, χ].
- The click **ɲχ’** is said to have ‘linguo-glottalic’ airstream, similarly.

From the phonetic point of view, this classification allows one to add the click consonants to the standard IPA chart by extending it with new sections for the different values of airstream feature. So we have a block for pulmonic consonants, followed by a block for lingual consonants, followed by a block for linguo-pulmonic, and so on. A concrete motivation for this concerns the difference between **ɲ** and **ɲq**, a distinction shared by !Xóǀ and N|u. As discussed in §3.1, Miller et al. consider (as I agree) that there is no role for velar/uvular place in the contrast; therefore there is only a timing difference, and that this is best seen as a contoured airstream.

From our point of view, this is still a unitary analysis, but with different feature values for the various accompaniments; it does not change the number or identity of phonemes in UA.

Miller’s more phonological arguments for this analysis are laid out in a handbook chapter (Miller 2011). Two of the major arguments are the difficulty of decomposing all clicks into segments that also appear independently (as noted by Nakagawa, see above); and that typologically every language that allows obstruent–obstruent clusters also allows obstruent–sonorant clusters, whereas there are none of the latter in Khoisan languages. My proposal will address both these points (see §6.2).

⁹ The N|u **ɲ̤^h** does not appear to have such a markedly long crescendo aspiration as the !Xóǀ [ɲh]. The **ɲχ’** is probably what I call **ɲqʔ’**.

4 Concurrent phonemes

4.1 Concurrent analysis

Having surveyed the facts and the current analyses, my proposal here may be very simply stated. Namely, *every* click is indeed a cluster. In the case of the basic clicks, the two component segments are the click influx and the accompaniment. Since there is no sequential order between these two components, they are clustered not serially, but concurrently. In IPA notation, this might be written, for example, $\widehat{\mathfrak{x}}$; unfortunately, the tie-bar is widely used to denote a phonetic coarticulation that forms a single phonemic unit, which is exactly not my point. I shall borrow a computer science notation (one of many for the concept) and write $(! \otimes \mathfrak{x})$, where it is stipulated that this is identical to $(\mathfrak{x} \otimes !)$.

Such an analysis brings the advantages of radical cluster analysis, or even of Güldemann's structured cluster analysis, while retaining most of the simplicity of standard segmental and phonemic theories. Formally, it is straightforward enough to be easily incorporated into any theory that works with segments and phonemes.

4.1.1 Concurrent clicks in !Xóǒ

If we apply this idea to the !Xóǒ click inventory (call it CoA 'concurrent analysis'), we obtain a dramatic simplification and reduction. The five clicks become phonemes in their own right; and we can now re-interpret our phonetic meta-notation for accompaniments, such as \mathfrak{xq} , in which the \mathfrak{x} is really a variable ranging over the five click symbols, into a true phonetic and phonemic notation, in which \mathfrak{x} is not a variable, but a novel phonetic symbol to indicate the point at which this sequence of segments synchronizes with any concurrent click segment. The phonetic output now follows from common phonetic rules: $!q$ is phonemically $(! \otimes \mathfrak{xq})$, and an unexceptional phonetic rule unifies the posterior closure required by the click with that required by [q], resulting in a long uvular stop with a click at the beginning.

Thus, even if we retain all 23 unitary accompaniments (call it CoUA), the click inventory size is now $5 + 23$ instead of 5×23 , set out in Table 4. Instead of an exceptionally large array of consonants, we have a modest set, with the formerly apparent complexity being simply clustering. Apart from the fact that the clustering is happening concurrently rather than sequentially, it is no more exceptional than, say, clusters in Russian.

Moreover, all the arguments for a sequential cluster analysis within accompaniments hold just as well in this setting as they do in the traditional setting. MCA, for example, naturally becomes what I might call CoMCA. Now there are five clicks and eight accompaniments, as in Table 5, and all the rest is clustering,

xq	xg	xq ^h	xq ^h	x ^h	x ^h	x'	x'	x̃	x̃	?x̃	xh	xh	x?	x?
				⊙		!		‡						

Table 4: Click phonemes under CoUA

both concurrent and sequential: for example, the click **!q^x** can be analysed as $/(!\otimes xq^x)/$. In this analysis, !Xóó has only 13 click phonemes. For good measure, the arguments against clustering outlined at the start of §3.3.5 no longer obtain: since the ‘onset’ of a word is now a concurrent cluster, it is not surprising that speakers had difficulty deciding how to drop it; and we shall see soon how the failure of uvulars to block A-Raising emerges naturally.

x	x	x ^h	x ^h	x'	x'	x̃	x̃
		⊙		!		‡	

Table 5: Click phonemes under CoMCA

If one adopts Miller’s (§3.3.5) proposal, which is a unitary analysis, one can still adopt CoA: at the phonological level, **xq** will be an accompaniment with linguo-pulmonic airstream, which then combines with a phonological pure click to produce her phonetic ‘linguo-pulmonic’ consonant.

4.1.2 A formal implementation

I intend this proposal as one of basic linguistic theory (Dixon 1997), since it can be understood in any framework, formal or informal, that supports the notions of phoneme and segment. To demonstrate a precise implementation, I give now a version in a variant of *SPE*. I shall use unspecified features in phonemes, rather than go through the formal route of *SPE* markedness theory – it is a routine but unenlightening exercise to re-cast everything in strict *SPE*. Unspecified features are written, e.g., [0voice]. I use *SPE* notation for rules, recalling that X_m means ‘a sequence of at least m X’s’.

For theories such as Optimality Theory (Prince and Smolensky 1993) which also use a feature-based phonemic representation, it is similarly straightforward to add concurrency; and all the rules I exhibit can be routinely translated to ranked constraints.

Recall that in *SPE*, there is a set of binary features, that underlying representations (URs) are strings of feature bundles, which may be unspecified for some features, and that the output of the rewriting rules is a string of fully specified feature bundles. Despite Chomsky and Halle’s express discouragement of such terminology, one can say that PHONEME corresponds to a feature bundle in the UR, and SEGMENT to a bundle in the output, and I will do so henceforth. I assume that features for clicks are as in Table 6, so that clicks share a feature

[+ling(ual)],¹⁰ and all the usual non-click phonemes are specified [–lingual]. The first step is to extend the strings in the URs:

- (8) A phoneme is a one-element CSTRING (‘concurrent string’). There is a commutative and associative binary combinator \otimes on cstrings. Cstrings may be combined with \otimes and concatenation. We let concatenation have higher precedence than \otimes (i.e. $a \otimes bc$ means $a \otimes (bc)$, not $(a \otimes b)c$). The empty cstring ε is the identity for \otimes (i.e. $a \otimes \varepsilon = a$). Every UR is a cstring.

Note that I will use parentheses with the usual mathematical meaning of grouping. This is potentially confusable with the *SPE* use of parentheses to indicate optional elements in rules, but in practice it will always be clear from context which meaning a given parenthesis has.

Definition (8) by itself allows arbitrary combinations; as concurrency is intended to reflect the physical possibility of combining different sounds, I impose (9).

- (9) Weak concurrent airstream constraint: In any UR containing a sub-cstring $a \otimes b$, the phonemes in a may not have contradictory (+/–) values for [lingual]. (And by commutativity, the same holds for b .)

The effect of (9) is to forbid clicks and non-clicks to combine within one half of a concurrent composition. For the moment, I also stipulate (10).

- (10) Strong concurrent airstream constraint: In any UR containing a sub-cstring $a \otimes b$, if a contains a phoneme with a specified value of [lingual], then b may not contain a phoneme with that value.

(10) further restricts \otimes to combining clicks on one side with non-clicks on the other. Next I define the click phonemes.

- (11) The pure click phonemes / θ , |, !, ||, ‡/ are lingual obstruents with features as in Table 6.
- (12) The accompaniment phonemes are specified for laryngeal and manner features (only) as in Table 6. They are notated by / \varkappa / together with diacritics for the positive features.

This is the definition that turns our accompaniment notation \varkappa into a symbol for an actual phoneme. Now / \varkappa , \varkappa^h / etc. are genuine phonemes in the inventory,

¹⁰ *SPE* uses [suction]; I prefer [lingual] as it is now the standard articulatory description of clickness.

albeit with the unusual phonotactic constraint (which can be dispensed with, at least formally – see §6.1) that they occur only in concurrent clusters. This constraint is formulated as (13).

- (13) Click/accompaniment constraint: A UR may contain a [+lingual] phoneme x only if x is in a sub-cstring a of $a \otimes b$ such that b contains a [0lingual] phoneme, and conversely.

This constraint forbids pure clicks and pure accompaniments from appearing by themselves in URs.

Table 6 sets out the featural specifications I assume in the discussion following, both for the click phonemes and for the other phonemes of !Xóǝ. Some choices are of course a little arbitrary; others are justified in the following sections.

Now the ‘simplex’ clicks have underlying representations such as $/(! \otimes \mathfrak{x}^h)/$. The question remains of the ‘complex’ clicks. As I discuss later, there is room for manoeuvre here. For the moment, I assert that ! \mathfrak{q}^h , for example, has the UR $/(! \otimes \mathfrak{x}\mathfrak{q}^h)/$: that is, it is a concurrent cluster, one half being the pure click, and the other being a sequence of \mathfrak{x} and \mathfrak{q}^h .

To complete the formalization, I need to consider whether concurrency survives to the output stage of the *SPE* re-writing process. One may have different views on this, according to where one prefers to draw the phonology/phonetics boundary. My preferred approach is to leave the click concurrency in the output, but to resolve the complex clustering, by adding the following rule late in the *SPE* rule chain:

- (14) Lingual Synchronization Rule:

$$\begin{array}{ccccccc} [+lingual] \otimes [-lingual]_0 & [0lingual] & [-lingual]_0 & & & & \\ 1 & 2 & 3 & 4 & \longrightarrow & 2 & (1 \otimes 3) \quad 4 \end{array}$$

This is *SPE* notation for “a [+lingual] phoneme docks on to a lingually unspecified phoneme in the other concurrent half”. For example, $/(! \otimes \mathfrak{x}\mathfrak{q}^h)/ \rightarrow [(! \otimes \mathfrak{x})\mathfrak{q}^h]$ by this rule, with 1 = $/!/$, 2 empty, 3 = $/\mathfrak{x}/$, 4 = $/\mathfrak{q}^h/$.¹¹

The rule (14) is one of several variations on the technical devices that could be employed to achieve the effect of synchronizing clicks with the pulmonic

¹¹ The rule as formulated allows only one click to dock on a given $/\mathfrak{x}/$. It could be formulated to allow a concatenation of clicks as component 1: it is perfectly possible to make an arbitrarily long sequence of clicks while maintaining the posterior closure. However, no language makes use of this possibility. It could also be formulated to allow a sequence of clicks to dock on to a sequence of accompaniments; but again, I know of no reason to do this.

The pure clicks /ǀ, ǃ, ǂ, ǁ, ǁ̥/ are specified for [+consonantal, –vocalic, –continuant, +lingual] together with the features [ant(erior), cor(onal), high, back, del(ayed) rel(ease)] as follows:

	ant	cor	high	back	del rel
ǀ	+	–	–	+	+
ǃ	+	+	+	–	+
ǂ	+	+	–	+	–
ǁ̥	–	–	+	–	–
ǁ	+	+	–	+	+

The pure accompaniments are specified for the features [voice, nasal, spr(ead) glot(tis), glot(tal) cl(osure)] as follows:

	voice	nasal	spr glot	glot cl
Ɂ	–	–	–	–
Ɂ̤	+	–	–	–
Ɂ ^h	–	–	+	–
Ɂ̤ ^h	+	–	+	–
Ɂ'	–	–	–	+
Ɂ̤'	+	–	–	+
Ɂ̤̃	–	+	–	–
Ɂ̤̤̃	+	+	–	–

Manner features for the pulmonic stops are specified as for the accompaniments using [voice, spr glot glot cl], together with [+del rel] for the alveolar affricated stops and the uvular ejective affricates /qʰ, ɕʰ/. Place features are as in *SPE* with one exception: we distinguish dentals /t, d, .../ from alveolars by [high] (motivated largely by the raising behaviour of dentals described in §5.1). Thus:

	ant	cor	high	back	low
p	+	–	–	–	–
t	+	+	+	–	–
ts	+	+	–	–	–
ɸ	–	–	+	–	–
k	–	–	+	+	–
q	–	–	–	+	–
ʔ	–	–	–	–	+

Continuants, glides, liquids and nasals are as in *SPE*; I tentatively consider the glottalized nasals to be clusters /ʔm, ʔn/.

Vowels are standard, except that we make /a/ unspecified for [back], so

	high	low	back	round
i	+	–	–	–
e	–	–	–	–
a	–	+	0	–
o	–	–	+	+
u	+	–	+	+

Creaky vowels are [+glot cl], breathy vowels are [+spr glot], pharyngealized vowels are [+phar], and strident vowels are [+phar, +spr glot].

Table 6: Feature specifications for CoMCA single phonemes

The representation of the click phonemes as concurrent and sequential combinations of clicks, pure accompaniments and other consonants is, taking alveolar clicks as an example:

Click Reprn	Click Reprn	Click Reprn
! (! ⊗ ɲ)	ʔ̃ (! ⊗ ʔ̃)	!qʔ' (! ⊗ ɲqʔ')
! (! ⊗ ɲ̣)	!q (! ⊗ ɲq)	!qʔ' (! ⊗ ɲ̣qʔ')
!h (! ⊗ ɲ ^h)	!G (! ⊗ ɲG)	!χ (! ⊗ ɲχ)
!h (! ⊗ ɲ ^h)	!q ^h (! ⊗ ɲq ^h)	!χ (! ⊗ ɲχ)
!' (! ⊗ ɲ')	!q ^h (! ⊗ ɲq ^h)	!h (! ⊗ ɲh)
!' (! ⊗ ɲ')	!q' (! ⊗ ɲq')	!h (! ⊗ ɲh)
!̃ (! ⊗ ɲ̃)	!q' (! ⊗ ɲq')	!ʔ (! ⊗ ɲʔ)
!̃ (! ⊗ ɲ̃)		!ʔ (! ⊗ ɲʔ)

Table 7: Representations of clicks in CoMCA

airstream sounds; this one is natural because of the intuition it gives for /ɲ/ being a manner-carrying placeholder waiting to receive a click.

One might wish to eliminate the idea of concurrent segments from the output. This can be done by adding a later rule:

(15) Concurrent Fusion Rule:

$$a \otimes b \longrightarrow a \sqcup b$$

where $a \sqcup b$ is the phoneme whose specified features are the union of those of a and b – it is undefined, and the rule cannot apply, if a and b have inconsistent values for some feature.

The \sqcup operation is not standard *SPE* notation, but has been recently suggested as a useful addition by Bale, Papillon, and Reiss 2013; the rule can of course be written out in standard notation, but is lengthy. The result of applying this rule to $/(! \otimes \mathfrak{N})q^h/$ is the purely sequential cluster $[!q^h]$ where $[!]$ has all its features specified.

4.2 Discussion

4.2.1 Concurrent segments and phonemes – a natural concept

The first question is whether, as I suggested in the introduction, the notion of concurrent segments and phonemes is consistent with the traditional, informal, understanding of segments and phonemes. In basic linguistic theory, the phoneme is still largely defined by structuralist considerations, and the notion of segment is taken as something which we naturally extract from our representations – although, as I remarked, there is not necessarily agreement about what

is or is not a single segment. If we look at clicks, and try to identify segments without preconceptions, I would argue:

- The click influx is *articulatorily* a clearly identifiable gesture, whose only necessary relation with the accompaniment is that it happens during a period of velar closure.
- *Acoustically*, the anterior release is very obvious in its own right, both to any human listener, and on the spectrogram. On the other hand, the accompaniment is easily recognized from the spectrogram, and, I would argue (not least from my own experience) easily heard in its own right by human listeners. The latter claim is supported:
- *Perceptually*, the results of Best et al. (2003) suggest that click place is perceived independently of accompaniment: Zulu speakers discriminate !Xóǀ click places they know, and assimilate !Xóǀ click places they don't know, regardless of a non-Zulu accompaniment. It is also my own experience in learning to discriminate between !Xóǀ clicks, at least once I had learned to hear clicks as speech. In addition, below I cite some evidence from the !Xóǀ lexicon which also suggests perceptual orthogonality.
- Moreover, it appears that *in production* click language speakers can immediately combine newly learned clicks-in-isolation with the accompaniments they already know. To my knowledge this has not been demonstrated before, and so I describe the relevant pilot experiment in the following subsection, and discuss this argument further.

Thus I claim that the notion of concurrent segment is well supported; and if the click and its accompaniment are both segments, they are certainly both phonemes by the usual contrast criterion.

4.2.2 *A click production experiment*

If, as I claim, clicks are separate phonemes from accompaniments, then if one takes a speaker of a click language, and teaches them a new click by itself, it should be the case that if they can use the new click in words at all, they can, without further instruction, use it with all their native accompaniments. If, however, clicks are not so decomposed, then generalizing to all accompaniments involves conscious featural manipulation, which is held by many to be outwith the competence of untrained speakers.¹² There is a considerable debate about

¹² The evidence for features in the mental representation involves both the phonological evidence, and psychological evidence, but as with phonemes (see Dresler 2011 for discussion, and Walsh 2009 for a recent review), the evidence is mixed, and seems to me weaker than for phonemes. For example, Žagar and Locke 1986 found only weak evidence for even subconscious access to features (in association tasks) in 5yo children. With regard to more conscious access, I am not aware of published experimental data. Anecdotally, I

such statements, but it seems plausible that manipulating phonemic segments is at least easier than manipulating features, despite such examples as the difficulty of pronouncing clusters that are not in one's own language.

Here I report a pilot experiment, which aims to test this prediction. Though there are only a couple of participants, the results are interesting and suggestive. I hope to seek support for a full version of this experiment in cooperation with colleagues elsewhere.

The participants¹³ were young adult Nguni speakers, one Zulu and one Xhosa. In my terminology, these languages have three clicks, !, |, ||, written *q*, *c*, *x*. There are five¹⁴ accompaniments, Ɂ, Ɂ̣, Ɂ^h, Ɂ̣̣, Ɂ̣̣̣, written (e.g.) *q*, *gq*, *qh*, *nq*, *ngq*. The two breathy accompaniments have several cues: there is breathy voice during the click, the following vowel is somewhat breathy, and perhaps most importantly, they depress the tone of the following syllable.

The first participant had no linguistic training at all. The second participant had had some exposure to introductory linguistics, mainly in semiology; in debriefing, he appeared to be unaware of standard phonological and phonetic descriptions of Nguni clicks.

The participants were first asked to demonstrate the fifteen UA click phonemes, by reading single words presented in standard orthography (e.g. *ukugcoba*). By chance, one or two of the words were unfamiliar to each participant, and the first speaker had a little difficulty reading out an unknown word, whereas the second read easily from orthography in any case.

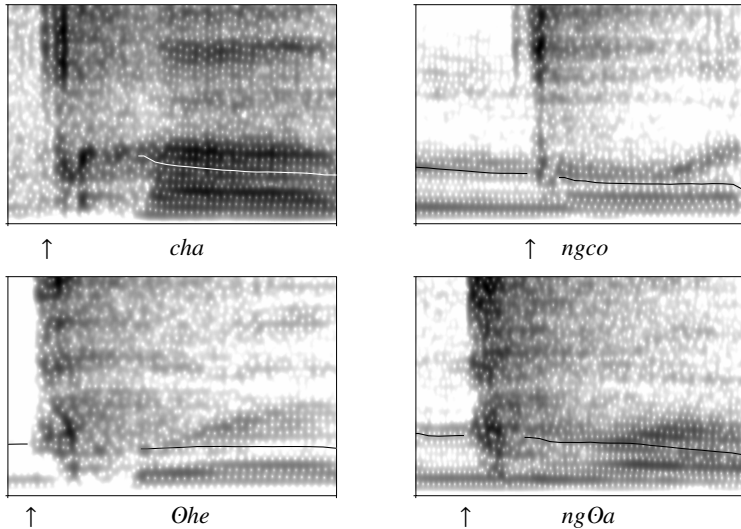
They were then taught, by demonstration, [Ø] and [Ɂ̣̣̣] in isolation, and then asked to read nonce words, presented in orthography with the IPA click symbols (e.g. *ukuOhele*).

The first speaker had a little difficulty incorporating Ɂ̣̣̣ into words, and took several attempts at some, but produced (entirely without prompting) the accompanied versions as expected. For example, her rendition of *ingOabha* shows pre-nasalization, murmur, and lowered tone. With Ɂ̣̣̣, she read fairly smoothly, and apart from intrusive pre-nasalization while hesitating on the first (plain click) word, the results were again as expected. (On subsequent review, I suspect that some of the renditions were the retroflex rather than palatal click; however, the accompaniments are not affected.) Recording quality was not as good as it should

have tried simple tests on several untrained English speakers, and I have yet to find one who can do even such simple analogies (presented in speech) as 'Thinking only about the sounds, /pa:/ is to /ba:/ as /ta:/ is to ___' – the usual answer is /ka:/, but it varies.

¹³ I thank Mabutho Shangase and an anonymous colleague for their kind participation.

¹⁴ Xhosa also has a glottalized nasal Ɂ̣̣̣̣ (*nkq*) but my participant did not recognize my examples for it.



Productions of native (top row) and novel (bottom row) aspirated and breathy nasal clicks, from the orthographically presented words *ukuchaza*, *ukuOhele*, *ukungcola*, *ingOabha*. Spectrogram y-axis from 0 to 5 kHz. Pitch contour marked, with y-axis from 75 to 500 Hz. Samples are 250–400 ms wide; the location of the click burst is marked. The pitch contour interruption in the breathy nasals is probably the analysis being overwhelmed by the click burst. Analysis and rendering by Praat (Boersma and Weenink 2013).

Fig. 1. Production of native and novel clicks

have been, but illustrative spectrograms of some of her native and new clicks are shown in figure 1.

The second speaker found it very difficult to produce **⊙** in words, and after several attempts, this part was abandoned. With **‡**, he read fairly easily, and produced as expected. However, he informed me that **‡** was already known to him, as in his community it is used as a “softer” version of **!** in play language and when talking affectionately to children, so all he had to do was read the nonce words as if talking to a child.

In summary: one speaker successfully produced two previously unfamiliar clicks in all of her native accompaniments; the other speaker did so with one click, but it was already familiar as a (previously unreported, to my knowledge) stylistic variation. However, the very fact that a conscious stylistic variation consistently replaces one click with another across all accompaniments is itself supportive of the hypothesis.

It is also worth remarking that in debriefing, both participants were adamant

that Zulu/Xhosa has three click consonants, and that, e.g. *gq* is *q* combined with *g*. It would be interesting to see whether a speaker uninfluenced by orthography would say the same.

4.2.3 *Concurrent phonemes versus autosegments*

In the original development, particularly as elaborated by Goldsmith (1976), of autosegmental theory, it was conceived as having segments on different tiers, for example the usual phones/phonemes on one tier, and tones on another. Subsequent work looking at the melodic rather than prosodic content of speech moved towards identifying tiers with features (or with elements in the Government Phonology school), so giving a simple and natural account of, say, vowel harmony. Consequently, in such theories both segments and phonemes are emergent concepts, not stipulative concepts, arising from the associations between feature (or element) tiers and the skeletal tier: a (phonological) segment is the bundle of autosegments associated with a particular skeletal point, and the set of phonemes – in so far as the theory admits a notion of phoneme – is simply the set of such segments.

There are several differences between such an approach and my proposal here. In autosegmental theory, the tiers exist throughout, and are specified with binary features (or the presence/absence of an element), and the synchronization between them is effected by association lines. Formally speaking, an autosegmental representation has the form of a parallel composition of a fixed number of sequential tiers, together with synchronization information; multiple such representations may be concatenated sequentially, but then there have to be rules extending the synchronization to the concatenation from its members.

In my approach, however, concurrent and sequential composition act on the same entities, namely phonemes, and can (in principle) be composed with more complex nesting, although in the !Xóó example I imposed constraints to restrict it. Because the entities being composed are phonemes, not features on tiers, they have to be justified as existing with contrastive power in the phoneme inventory of the language.

It is, of course, possible to do some formal encoding: we could analyse Finnish to have abstract phonemes /a, o, u/ and /F/ (for Front), and assert that the Finnish /y/ is really /($u \otimes F$)/, and then express harmony rules. However, to do that, we would have to argue that /F/ is a phoneme in the inventory according to the criteria above. Moreover there is no principled reason for choosing /F/ rather than /B/ (for Back) as the ‘phoneme’. If we choose /F/, then we must argue either that /i/ and /e/ do not contain /F/, despite having all the same acoustic and articulatory signs of it as the other front vowels; or that they do con-

tain it, but there is a very specific phonotactic rule preventing /ɰ/ and /ɣ/ from occurring without it. (Note that above we did claim that click accompaniments do not occur on their own; but firstly they form a natural class, and secondly, it is at least formally possible to avoid this constraint – see below §6.1.) If, instead, we choose /B/, we have to explain why /(i ⊗ B)/ does not appear – again, requiring an ad hoc rule.

In summary, modern autosegmentalism deals with the structure inside segments, whereas the approach here deals with structures built out of segments. However, as I remarked at the beginning of this section, the earliest autosegmental phonology did allow for tiers to contain segments rather than features, and in that sense the proposal here can be seen as similar to it. Ladd (2014) contains a discussion on the historical and current relationships between concurrency, simultaneity and autosegmentalism, and the reader is referred there for a more substantial discussion.

It is possible to modify current autosegmental theories in such a way that my notion of concurrency here is added, above and beyond the built-in notion of tiers. However, a full development of this would occupy some pages in a fairly detailed analysis, which is beyond the intended scope of this article.

4.2.4 *The combinatorial argument*

My claim that clicks and accompaniments are phonemes suggests that they should combine freely, modulo any phonotactic constraints, of which there appear to be none. This raises the question, which requires field investigation, of the gaps in the inventory. Traill heard no occurrence of the clicks ᑖq^h , ᑦq^h over his thirty years of fieldwork. If, as seems to be the case, they do not exist in any word, then from a UA viewpoint it is hard to argue that they exist as sounds in the language. One would therefore expect that if presented with a nonce-word containing them, speakers would fail to recognize the sound correctly, and probably replace it by the nearest extant sound. On the other hand, if the clicks are independent of the accompaniment, one would expect the nonce-word to be perceived and repeated with no difficulty. Naumann (p.c.) concurs that the expected result is the latter, but such an experiment has not yet been carried out. It would be even more compelling in the case of N|u: for !Xóᑖ, the non-concurrent CA would yield the same result, but N|u appears (Miller et al. 2009) to be missing even some basic labial clicks, namely ᑖ^h , ᑦ^h , ᑖ .

Although I have not been able to test this hypothesis in the field, it is supported by the result of the experiment reported in §4.2.2.

Another combinatorial argument relates to the difficulty of learning. As I remarked in §3.3, the huge UA inventory makes it very hard to establish contrasts;

but even the MCA analysis leaves many contrasts without strong evidence. Obviously, a reanalysis like CoA that separates clicks and accompaniments solves these problems – an accompaniment contrast in the context of one click suffices to establish the contrast in the context of any click. For example, there is no support for the contrast between $\tilde{\text{Q}}$ and $\hat{\text{Q}}$; but if these are actually $/(\text{O} \otimes \tilde{\text{x}})/$ and $/(\text{O} \otimes \hat{\text{x}})/$, then the evidenced contrast between $\tilde{\text{!}}$ and $\hat{\text{!}}$ also supports this contrast (and in all the other click places).

It is no surprise that in CoA, even without doing sequential clustering, most of the minimal pairs exist; the exceptions give rise to an interesting observation, discussed in §5.2.

4.2.5 *Metalinguistic evidence*

A small but positive piece of psycholinguistic evidence comes from the !Xóó lexicon. It turns out that not only are clicks very salient for non-speakers, they are also very salient for speakers: so much so that there are words for making the sound of the five basic clicks, and even a word for one variation particularly used in ritual incantations. So important are clicks that some of these words also mean simply ‘to talk about, converse’.

The words follow, in their full pseudo-reduplicated form:

- (16) a. $\tilde{\text{Q}}\acute{\text{u}}\acute{\text{u}}\text{-}\tilde{\text{Q}}\acute{\text{u}}\acute{\text{u}}$ to make the sound of the [O] click
 b. $|\text{h}\grave{\text{e}}\tilde{\text{e}}\text{-}|\text{h}\hat{\text{e}}\tilde{\text{e}}$ or $|\text{ʔ}\grave{\text{e}}\text{e}\text{-}|\text{ʔ}\hat{\text{e}}\tilde{\text{e}}$ make the sound of the [!] click
 c. $!\text{h}\grave{\text{e}}\tilde{\text{e}}\text{-}!\text{h}\hat{\text{e}}\tilde{\text{e}}$ make the sound of the [!] click
 d. $\tilde{\text{!}}\grave{\text{a}}\acute{\text{i}}\text{-}\tilde{\text{!}}\hat{\text{a}}\acute{\text{i}}$ to make the sound of the [!;] click¹⁵; to talk about
 e. $\|\text{h}\grave{\text{e}}\tilde{\text{e}}\text{-}\|\text{h}\hat{\text{e}}\tilde{\text{e}}$ or $\|\grave{\text{!}}\text{a}\text{-}\|\hat{\text{!}}\text{a}$ or $\|\text{ʔ}\grave{\text{a}}\tilde{\text{a}}\text{-}\|\text{ʔ}\hat{\text{a}}\tilde{\text{a}}$ to make the sound of the [!] click
 f. $\# \text{h}\grave{\text{e}}\tilde{\text{e}}\text{-}\# \text{h}\hat{\text{e}}\tilde{\text{e}}$ or $\# \text{ʔ}\grave{\text{e}}\tilde{\text{e}}\text{-}\# \text{ʔ}\hat{\text{e}}\tilde{\text{e}}$ to make the noise of the [ʔ] click; to talk about

It is immediately striking that none of these words for clicks uses the plain unadorned click, at least in UA. Even in the usual CA, the nasal clicks are viewed as primitive, and so some of these words do not contain plain clicks. In CoA, of course, they all do. While this is not a topic on which there is extensive empirical evidence, it seems more plausible for a language to have iconic words for phonemes, than for either a phonetic component of phonemes or for a class of phonemes.¹⁶

¹⁵ Traill actually notes this as “the noisy [!] click”; I am sure that by this he means [!;], the flapped click.

¹⁶ In this connexion, it is interesting that early 20th century researchers such as Beach and Doke used distinct letters for voiced, voiceless and nasal clicks, just as is done in the IPA

5 !Xóõ phonology under concurrent analysis

5.1 A-Raising and the Back Vowel Constraint

The formal development of CoA above defined the representation, and showed some examples of rules involving concurrent clusters. Rules that do not involve concurrent clusters look just as before; but a question arises of whether such rules need to be extended. For example, a rule might refer to properties of the first phoneme of a word – if a word starts with $/(!\otimes\tilde{y})/$, what are those properties? The general form of such a rule in *SPE* is:

$$(17) \quad x \longrightarrow y / \# _$$

where x specifies a class of phonemes and y specifies the modification to the phoneme matched against x . In CoA, this rule will not match a word-initial concurrent cluster – we must explicitly allow for this. For example, (18) is the same rule modified to apply both to initial normal segments, and to initial simplex accompaniments (assuming the Concurrent Airstream Constraints (9) and (10)), but not to initial clicks:

$$(18) \quad x \longrightarrow y / \left\{ \begin{array}{c} \# _ \\ \#[+lingual]\otimes _ \end{array} \right\}$$

or in a more economical notation exploiting $\varepsilon \otimes x = x$ and also allowing complex accompaniments:

$$(19) \quad x \longrightarrow y / \#[+lingual]_0^1 \otimes _ C_0$$

Thus a rule may refer to the initial phoneme, or to the first phoneme of an initial concurrent cluster, as the evidence requires. The Back Vowel Constraint (5) and A-Raising Rule (7) provide good examples of this.

5.1.1 Moderate A-Raising

Recall that the first part of the ARR (7a) raises **a** to [ɜ] if it is before **i**, **Ci** or a nasal, *and* after a dental non-click or a dental or palatal click. This rule applies even in a word like **[qʰán-tá]** [qʰɛn-tá] ‘small *pl*’, showing that the rule targets the click rather than the accompaniment: the apparently intervening uvular, which one would normally expect to block a phonetic raising effect, does not do so. In

for pulmonic sounds. For example, [!] was [ç], and [!] was [ʒ]. The IPA adopted the plain symbols, but refused the others; and then in 1989 it changed to the Africanist symbols (despite the great violence they do to the IPA’s typographic coherence). Possibly the resistance to distinct symbols was subconsciously reinforced by the reluctance to disguise the presence of the click phone itself.

the formal presentations that follow, I shall mostly omit the raising after dental non-clicks; this is merely to simplify the notation.

This rule provides the evidence for how we should distribute concurrent and sequential clustering. *A priori*, it is possible that $|\mathbf{q}'\mathbf{a}\mathbf{n}$ could start with $/(|\otimes \mathbf{x}\mathbf{q}')/$ or with $/(|\otimes \mathbf{x})\mathbf{q}'/$. Indeed, one could even analyse $|\mathbf{q}'\mathbf{a}\mathbf{n}$ as $/(|\otimes \mathbf{x}\mathbf{q}'\mathbf{a}\mathbf{n})/$, and since Khoisan languages allow only one click per stem, this would make some sense from an autosegmental viewpoint. On the other hand, considerations of simplicity and economy suggest that (\otimes) should be applied with the smallest scope, so that all of each half is genuinely concurrent with all of the other half, so favouring $/(|\otimes \mathbf{x})\mathbf{q}'/$. However, the behaviour of the ARR suggests that $/(|\otimes \mathbf{x}\mathbf{q}')/$ is correct.

For the moment, I ignore the question of what it is that the triggering click types have in common, and just list them in rule (20).¹⁷

(20) Formal moderate A-Raising rule:

$$\left[\begin{array}{c} \mathbf{V} \\ +\text{low} \\ -\text{phar} \end{array} \right] \rightarrow [-\text{low}] / \left(\begin{array}{c} \{|\, \#\} \\ \otimes \\ \mathbf{C}_0 \end{array} \right) \left\{ \begin{array}{l} \mathbf{C}_0[\mathbf{V}, +\text{high}, -\text{back}] \\ [+cons, +nasal] \end{array} \right\}$$

Formally, there is little difference between this and the equivalent rule in a standard CA, where the click context would be expressed as the class of dental and palatal simplex clicks followed by \mathbf{C}_0 , instead of a concurrent cluster of the two pure clicks with the accompaniments. Assuming all the constraints and rules in §4.1.2, it can be shown that any set of constraints and rules in this concurrent formalism can be translated into a standard set that will produce the same output; I am adding not expressive power, but naturalness. Here, we avoid the rather peculiar situation in sequential analyses of the raising power of the clicks passing through uvular stops (which one expects to be strongly lowering), because here the target vowel is immediately adjacent to both the click and the accompaniment.

The transparency of the $/\mathbf{C}/$ in $-\mathbf{C}_i/$ requires a little comment – why is it transparent to the raising power of the $-\mathbf{i}$, while (I claim) a sequential uvular should block the licensing from the $[\text{+high}]$ clicks. One could invoke theories that account for VV interactions being long-distance (e.g. Germanic umlaut), while requiring strict adjacency for CV interactions (e.g. English palatalization). However, there is a simpler argument: the permissible $/\mathbf{C}/$ are only $/\mathbf{b}, \mathbf{m}, \mathbf{n}, \mathbf{j}/$,

¹⁷ *SPE* does not have a suitable feature for expressing pharyngealization of vowels, so I use the ad hoc feature $[\text{phar}(\text{yngeal})]$. I assume that $/\mathbf{a}/$ is specified as $[\text{+low}]$, and is unspecified for $[\text{back}]$, so that raising it gives a mid vowel. This is purely for expressibility in the illustrative *SPE*-based framework; I would prefer a formalism with more gradience.

j, l, r/, all of which are either [+high] or do not involve the tongue at all – and the nasals are raising in any case.¹⁸

5.1.2 Full A-Raising

The rules become more interesting when we consider Traill’s account of the Back Vowel Constraint in eastern !Xóõ and the exceptions to it. Recall that his version of the BVC (6) forbids front vowels after any back consonant, including all clicks – arguing that since clicks involve a velar/uvular closure, they are surely at least as back as **k**. He then has to account for the exceptions that he finds. One exception involves just **k**: there is a grammatical particle **kV**, which appears as **ke**, **ki** in some concords. Traill notes that **ke**, **ki** are often pronounced instead as **te**, **ti**, so obeying the constraint phonetically. The other class of exceptions involves the clicks |, ʈ, where phonetic front vowels do appear, for example the words ʈi ‘steenbok’ and |ii ‘to be’. Traill accounts for most of these by asserting that they are underlyingly, e.g., ʈai, and then the full part of the ARR (7b) applies to change **a** to **i**. The evidence for this is partly internal: the plural of ʈii is ʈabatê, with the following morphology:

- (21) a. ʈi- -i
 steenbok class 1 sg
 b. ʈà- -bà -tê
 steenbok class 1 pl pl

where **-tê** is the current productive pluralizer. There is also cross-dialectal evidence: for example, in the DOBES data, ‘steenbok’ is ʈai, pronounced [ʈʔi] with moderate A-Raising. Indeed, although Traill abandons |ii ‘to be’ as an unexplained exception, a reviewer points out that DOBES has what may be the same verb |ai ‘stay, be at a place’, so even that is accounted for.

I have not so far given a precise specification of the pre-context in Full A-Raising. In his descriptions (1985, p. 70 and 1994, p. 40), Traill is not explicit about whether any dental or palatal click triggers it, or just some of them, for example just the plain clicks. However, in the dictionary he marks fully raising words: e.g. ʈii is entered ʈai (>[ʈii]). Thus from the dictionary one can see which posited underlying **-ai** words undergo Full A-Raising – not all of them do. For example, |xai ‘bowstring hemp plant’, which is also a class 1 noun, with plural

¹⁸ The careful reader may recall that final **ŋ** exists in the DOBES inventory, and wonder whether it is included as a raising nasal. The phonemic status of **ŋ** is somewhat shaky – it may just be an allophone of **n** – but in the instances in the DOBES dictionary in which it appears with audio in A-Raising position (e.g. [aŋ]), the vowel is indeed raised.

|χāba-tê, is entered just as /xāi. Indeed, a recording¹⁹ of it is available, and it is pronounced with moderate raising. An examination of all the data shows the following, in my representation:

- (22) Words of the form ($\{|\, \# \} \otimes x$)ai undergo Full A-Raising if x is \mathfrak{x}^h , $\mathfrak{x}\mathfrak{r}$, \mathfrak{x} , $\mathfrak{x}h$, $\mathfrak{x}\mathfrak{r}$; they do not (and therefore undergo only Moderate A-Raising) if x is $\mathfrak{x}\mathfrak{c}$, $\mathfrak{x}\mathfrak{c}^?$, $\mathfrak{x}\mathfrak{q}$, $\mathfrak{x}\mathfrak{q}^?$.

That is, although a uvular segment in the accompaniment does not block moderate A-Raising, it does block full A-Raising. In *SPE*, uvulars are contrastively specified for [+back] and [-high], so there is a choice of which feature to use in the rule. I will accept Traill’s view that A-Raising is indeed raising rather than fronting, and use [high]. So, using the fact that my pure accompaniment phonemes and the two glottal phonemes are unspecified for [high], we can write the Full A-Raising rule as (23):

- (23) Formal Full A-Raising rule:

$$\begin{bmatrix} \text{V} \\ +\text{low} \\ -\text{phar} \end{bmatrix} \rightarrow \begin{bmatrix} +\text{high} \\ -\text{low} \\ -\text{back} \end{bmatrix} / \left(\begin{bmatrix} \{|\, \# \} \\ \otimes \\ [\text{Ohigh}]_1 \end{bmatrix} \right) \text{---} \begin{bmatrix} \text{V} \\ +\text{high} \\ -\text{back} \end{bmatrix}$$

Now consider what distinguishes |, # from the other clicks. There have been several suggestions for features that do so. I tend to prefer Traill’s notion that the difference is that they leave the tongue blade in a high front position, whereas the others pull the tongue lower and backer, which suggests either [back] or [high], or perhaps both. The rules work nicely if both are specified, as I laid out without explanation in Table 6.²⁰ Miller uses [pharyngeal] – see below.

Given this, and a little notation, the following rule suggests itself as a combined description of A-Raising before **i**.

¹⁹ UCLA 2009, Language/NMN/nmn_word-list_0000_01.wav

²⁰ Note that since we have separated the clicks from the accompaniments, there is no interference between specifying the [high] feature for clicks and for the accompaniment; without the concurrent clustering, it is necessary to use a different feature, such as [low] or Miller’s [pharyngeal]. This use of [high] does involve a certain relaxed approach to the intrinsic content of *SPE* features, as does the alternative use of [back]. See Traill 1985, p. 107–108 for an extended discussion, although he was additionally handicapped by the need to include accompaniment features with the clicks.

(24) Formal A-Raising rule:

$$\begin{bmatrix} \text{V} \\ +\text{low} \\ -\text{phar} \end{bmatrix} \rightarrow \begin{bmatrix} (\alpha \wedge \beta)\text{high} \\ -\text{low} \end{bmatrix} // \left(\begin{array}{c} [\alpha\text{high}] \\ \otimes \\ ([0\text{high}][\beta\text{high}]_0)_0^1 \end{array} \right) - \begin{bmatrix} \text{V} \\ +\text{high} \\ -\text{back} \end{bmatrix}$$

where $\alpha \wedge \beta$ is – if either α or β is –, and is + otherwise, and β is 0 if unmatched.

For simplicity, this rule does not explicitly describe the concomitant fronting that results in [i] rather than [i̠] in the full case – as a reviewer suggests, it is probably simplest to assume that a later rule fills in the [–back]. It is also of course possible to incorporate fronting in (24), as we did in (23), at the price of some additional inelegance.

This rule neatly shows the concept that the raising and fronting effect of the following **i** is moderated either by the click or by the accompaniment. Moreover, since I also in Table 6 used [±high] to distinguish dentals from alveolars, this rule also captures the A-Raising with initial dental non-clicks: a non-click initial matches the context by taking the optional lower half to be empty, and then α matches against the initial.

Several similarly complex sets of interactions between different coronal consonants and vowel backness were studied by Flemming (2003), with similar arguments about the different behaviour of the tongue body. The above description can also, as I noted, be cast in terms of fronting rather than raising, and would mostly fit in to Flemming’s (2003) framework.

As examples of the formal application:

(25) a.

$$\#ai = /(\# \otimes \varkappa)ai/ = \left(\begin{array}{c} [+lingual, +high, -back, \dots] \\ \otimes \\ [0lingual, 0high, -voice, \dots] \end{array} \right) /ai/$$

and so $\alpha = +$ and $\beta = 0$ (because unmatched), so $\alpha \wedge \beta = +$, so /a/ changes to [+high, –low, 0back], and then the later rule fills in [–back] from [+high], so. $\#ai \rightarrow [\#ii]$.

b.

$$|\chi ai = /(| \otimes \varkappa \chi)ai/ = \left(\begin{array}{c} [+lingual, +high, -back, \dots] \\ \otimes \\ [0lingual, 0high, \dots][+lingual, -high, \dots] \end{array} \right) /ai/$$

and so $\alpha = +$ and $\beta = -$, so $\alpha \wedge \beta = -$, so /a/ changes to [–high, –low, 0back], i.e. $|\chi ai \rightarrow [|\chi 3i]$.

Note that (24) does *not* agree with Traill’s A-Raising Rule (7), because (24) predicts that there should be moderate raising following a back click without a uvular accompaniment, whereas in (7) only the front clicks trigger any raising. Traill (1994) in fact states that in such contexts **a** undergoes a mild raising to [æ]. However, I have studied his available recordings, and in the readings, all **-ai** words in back clicks appear to show the same degree of raising as other cases of moderate raising. There is not enough data to make any statistically meaningful claim, but both auditory impression and acoustic measurements suggest this. For example, in one recording²¹ **!hai** appears to show considerable assimilation, varying from [əi] to [ɛi] in the same speaker. (On phonetic grounds, one might expect raising to be particularly marked in **xh**, since the long [h] allows plenty of time for the tongue to move away from the position forced by the click. However, there is not enough data available to me to check this.)

It is of course simple to force (25) to match (7), but this requires removing the symmetry between click and accompaniment features, and since the symmetric version appears to be more accurate, there is no call to do so.

5.1.3 *The Back Vowel Constraint*

Though the underlying **a** in most Full A-Raising words is adequately supported by other evidence, part of its motivation is to explain exceptions to Traill’s phonological Back Vowel Constraint (6), which prohibits front vowels after any back consonant. As I noted, there is an alternative formulation (5) of the general Khoisan BVC, which recognizes the distinction between the front and back clicks, and it is perhaps unclear why one should recognize this difference in the ARR but not in the BVC.

A similar situation with regard to the BVC occurs in Ju|’hoan, where also front vowels do in fact occur after the front clicks. Unlike Traill, Miller-Ockhuizen 2003 does not try to explain this away by a phonetic rule operating after the constraint, but rather states the BVC in its (5) form that distinguishes the front |, † clicks from the back !, || clicks. Her technique is to assign the feature [+pharyngeal] to !, ||, and use that in the BVC statement. This use of [pharyngeal] is motivated by the other ‘guttural’ constraints she analyses, but many of these do not appear to apply in !Xóó. The phonetic grounding of this feature is supported by ultrasound; impressionistically, to me it seems to be a consequence of the apical articulation of !, || rather than a primary feature. Miller-Ockhuizen discusses in detail both her own and others’ work on the acoustic and articula-

²¹ UCLA 2009, Language/NMN/nmn_word-list_1983_01.wav. Unfortunately, one of the potentially most useful recordings for this issue is truncated, and the original cannot be traced.

tory properties of the various clicks, and there are a number of ways in which the front clicks can be seen to differ from the back clicks.

In my setting, the choice made above to specify back clicks as [+back, –high] can be exploited to state the BVC in a more refined form:

- (26) Concurrent Back Vowel Constraint: A [–back] vowel must be licensed by an immediately preceding [–back] consonant.

This makes fully A-Raised words licit at the phonological level, and so removes the notion that they are exceptions. It therefore also allows the few remaining unexplained exceptions, such as [i̯] ‘if’, and a dozen or so words in -e- following a dental or palatal click.

It also permits a front vowel to follow a click with an uvular accompaniment, because both the click and the accompaniment immediately precede the vowel; in a non-concurrent formulation the uvular would block the licensing from the front click. According to Traill 1994 there are indeed a couple of such words: [q̣]i-sà, †gēē.

5.2 ‘Delayed aspiration’ and the voiceless nasal

The so-called delayed aspiration accompaniment **xh**, which is widespread in Khoisan, has caused some confusion historically, particularly in terms of its relationship to **x^h** – and as I described in §2.3, it seems that !Xóǝ has **xq^h** in addition, though Traill was unclear about this.

Moreover, as I also noted, **xh** involves nasality, in the form of a (possibly ingressive) voiceless nasal at the beginning. (Beach (1938) had already noted some nasality in Khoekhoe, though he described occasional voiced nasality.) Given that most Khoisan languages have the voiced nasal accompaniment **x̣**, one might wonder whether they are related. However, the arguments are good that the nasality of **xh** is a phonetic detail; for example, both !Xóǝ (per DOBES) and Ju|’hoan have a voiced version **x̣h**, and all voiced accompaniments are pre-voiced and often have phonetic nasality, since nasality is the easiest way to maintain the voicing; similarly, in **xh** the nasality allows for the ‘soft start’ to aspiration – and Naumann (forthcoming) reports that some of his speakers describe **!h** as ‘[!] with a pause’. In any case, !Xóǝ has a distinct voiceless nasal accompaniment **x̣̥**.

However, the !Xóǝ voiceless nasal is somewhat of a puzzle. With the possible exception of †Hoã (Gerlach, p.c.), !Xóǝ is the only extant language to possess this accompaniment, and it is unclear how it emerged.

Güldemann (2001) noted that it appears only before pharyngealized or creaky vowels, and suggested that perhaps it split off from the voiced **x̣** in reaction to “the specific phonetic character of the marked stem vowels”. It is, however, hard

to see how this could have happened, as \tilde{x} still occurs in this environment, and there are even exact minimal pairs, such as $\tilde{t}o^\text{f}li$ ‘*Antizoma angustifolia*’ and $\tilde{t}o^\text{f}li$ ‘wipe or rub the eyes, pick the nose’.

In §4.2.4, I remarked that almost all, but not all, CoA accompaniment contrasts are supported by minimal pairs. It is therefore striking, and not to my knowledge previously observed, that the contrast \tilde{x} vs xh has no support. Not only is there no minimal pair, investigation shows that they are indeed in complementary distribution. As Güldemann observed, \tilde{x} occurs only before creaky or pharyngealized vowels. It follows from the Pharyngeal Constraint (4) that a pharyngealized vowel cannot occur after xh , but checking through Traill 1994 shows the stronger fact that xh occurs only before plain vowels.

Thus \tilde{x} and xh are in complementary distribution, and given the phonetic link between them in terms of voiceless nasality, it is tempting to conjecture (27):

(27) \tilde{x} is an allophone of xh .

If we unify xh and \tilde{x} in Traill’s analysis, and adopt unitary CoA (i.e. sequentially unclustered), then there are 120 minimal pairs of accompaniment phonemes to find, and 115 of these exist, with the remaining 5 also found if we ignore tone.²² For example:

- (28) a. xq^h vs \tilde{x} : no minimal pair
 b. xq^h vs xh : minimal pair $|q^h\acute{a}a$ vs $|h\acute{a}a$ (and many others)

and in fact all the missing minimal pairs in the un-unified system are contrasts with \tilde{x} .

At first sight, phonological arguments cut both ways when considering (28). On the one hand, it is also striking that \tilde{x} does not occur before breathy or strident vowels, whereas x is attested before both. Given the general Single Aspirate Constraint (2), this lends support to the idea that \tilde{x} represents a phonological aspirate. On the other hand, declaring \tilde{x} to be an aspirate then violates the Pharyngeal Constraint (4).

However, as I noted, the Pharyngeal Constraint is violated by several words of the form hV^f -, such as $ho^\text{f}lo$ ‘stand on tiptoe’, so that h itself does not appear to trigger the constraint, and given that I treat xh as a sequential cluster with h , there is no reason to think that xh triggers it. I therefore suggest that indeed the

²² 120, because in the 1994 version of Traill’s phonology, the only source for which there is extensive data, there are 17 accompaniments, so removing one gives $(16 \times 15)/2 = 120$ possible contrasts.

constraint does not apply to \mathfrak{h} , and that its apparent application is due to the formation of \mathfrak{h} as an allophone in the pharyngeal context.

My conjecture as to the emergence of this suggested allophony is that maintaining the long [spread glottis] aspiration characteristic of \mathfrak{h} is awkward when followed by the glottal constriction of creaky vowels, and also when followed by pharyngeal constriction, because then it will tend to lead to stridency, and so the voiceless nasality took over as the main cue. In the context of plain \mathfrak{h} -, however, there was no such alternative cue.

I should note that the dialect recently studied by DOBES slightly muddies the water on this issue. Naumann (forthcoming) reports a word in which \mathfrak{h} occurs before a plain vowel; the same word is reported with a creaky vowel by Traill. Moreover, in the DOBES data, the ‘delayed aspiration’ seems to have considerably stronger aspiration than in the eastern dialect, decreasing the phonetic similarity. The extent of dialectal differences versus differences in analysis requires further investigation, but I might very tentatively conjecture that the distinction is allophonic in eastern !Xóǝ, but in the process of phonologization in western !Xóǝ.

6 Concurrent phonemes – variations and extensions

In this part, I will first discuss some possible alternative choices in the formulation above; and then I will go on to suggest that the notion of concurrent segment and phoneme might be useful beyond the world of clicks. With clicks, the justification of click and accompaniment segments, and hence phonemes, was quite strong. In this section, the justification will become increasingly open to attack, and so I use this part to explore the boundary between concurrent segments and autosegments or features, following on from §4.2.3.

6.1 The nature of \mathfrak{x}

In definition (12) above, the accompaniment phonemes are defined to be specified only for their values of voice, ejectivity, aspiration, and so on, but not for any other values, such as place. The click phonemes are specified for anterior place, height, backness and [+lingual], but for nothing else. Moreover, it is assumed (though not so far explicitly constrained) that neither clicks nor accompaniments occur by themselves in URs, but only in conjunction – in what sense, therefore, are they like other phonemes?

In the case of the pure clicks, I would assert that it is a contingent, rather than necessary, fact about language that clicks do not occur alone. A pure click is a click unconnected to any other airstream – for example, the English *tsk! tsk!* [] consists of pure clicks. A (not very human) language could be constructed out

of pure clicks; but any language that combines clicks with vowels, for example, must synchronize them, and having done so, can take advantage of modifications of the posterior closure.

For the accompaniments, the question is more subtle. I chose to assume that **Ɂ** does not occur on its own in URs; but as I remarked in several places, one can reformulate the theory so that it can. It is debatable whether such reformulations are more or less natural than that of §4.1.2. I shall consider three, the last of which provides an opportunity to discuss the curious nature of !Xõ clusters.

6.1.1 *Accompaniments as pulmonic stops*

One might simply say:

(29) **Ɂ** is just **q** (or **k**)

This is essentially Radical Cluster Analysis, made concurrent instead of sequential: the accompaniments are the existing series of uvular (or velar if preferred) stops. It has the same distributional problem as RCA: there is no **Ɂ**. (There is also no **Ɂ̃**, but that problem goes away if we follow §5.2.) There is a rather marginal **Ɂ̃**, but not in onset position. Either of the solutions suggested for RCA could be applied.

Although this proposal avoids the unusual phonotactic constraint that accompaniments must appear with clicks, it introduces others: why is it that there are no initial clusters **qɁ**, **qh**, **qʔ**? One has to argue that the point of the click clusters **Ɂx**, **Ɂh**, **Ɂʔ** is that the posterior release is inaudible, and that an initial **q** with no release is rather pointless, but then that distinguishes **q** *qua* accompaniment from **q** *qua* independent consonant.

A major drawback to this approach is that now the bare accompaniment has values for height, backness and [lingual], and so all the rules have to be re-cast in a less elegant form. In particular, if **Ɂ** is just **q**, there's nothing to distinguish the two **qs** in **qq**, and so the synchronization rule, which previously could identify **Ɂ**, must instead be written to dock the click onto the first uvular segment in the accompaniment. This happens to work, because there is no **qɁ**, but it is not elegant.

6.1.2 *Accompaniments as clicks*

An alternative suggestion is:

(30) Solitary **Ɂ** is !

In this view, the accompaniment carries with it a 'default' click, which I have somewhat arbitrarily chosen to be !, but this can be changed by concurrent com-

position with a different pure click. In the implementation of §4.1.2, this would be done by leaving the \mathfrak{x} phonemes as they are, and adding a rule that fills in the ! features for an isolated accompaniment.

In such a setting, of course, the chosen default pure click becomes redundant, and can be omitted from the inventory. This solution solves some problems – but there is, to my knowledge, no phonetic or phonological ground for treating one click as more fundamental than another; and more importantly, it makes stating rules such A-Raising and the Back Vowel Constraint complicated, as they apply to the default click too.

6.1.3 *The place of voice*

Another possible, and more substantial, variation has been raised by Daniel Currie Hall (p.c.). I have chosen to put all laryngeal features with the accompaniment. Hall notes that [voice] varies orthogonally to all other features, whereas [spr glot, glot cl, nas] are mutually exclusive (assuming the cluster analysis of $\mathfrak{?}\mathfrak{x}$). Why not, then, place [voice] with the clicks rather than with the accompaniment? This would give basic phonemes /!, !, .../ and accompaniments / \mathfrak{x} , \mathfrak{x}^h , \mathfrak{x}' , \mathfrak{x}^h /.

Such an organisation is also used by Güldemann (2001) in his feature-geometric approach. Hall suggests the following advantages:

- (31) a. The plain and voiced clicks no longer require an accompaniment;
- b. and consequently there is no longer a need for sequential clusters within concurrent clusters (e.g. $!\mathfrak{G}$ is just /(! \otimes q)/), which also explains why
- c. only plain and voiced clicks occur in clusters with other stops.

This suggestion has obvious merits, like those that motivated Güldemann's (2001)'s similar decision. The counter-arguments invoke the conceptual basis of my proposal here. *Ad* (a), plain and voiced clicks require just as much synchronization of separate airstreams as other clicks; and at least in my own experience, voicing clicks is no easier than aspirating them. A click on its own would demonstrate a failure of synchronization. *Ad* (b), if there is no sequential clustering, then one must resort to phonetic rules to explain why the clusters with stops have a prolonged closure after the click rather than before or around it. *Ad* (c), the non-occurrence of ejective, aspirated or nasal clicks in clusters is discussed in the following section.

There is also a more drastic approach to voice, which deserves mention. As is clear from Table 1, the voicing distinction pervades the stop system; and as discussed in §§2.2–2.3 it appears as distinct pre-voicing in most cases, other

than the simple voiced stops. It is therefore tempting to follow the orthographies, and replace the voiced accompaniments \mathfrak{X} , \mathfrak{XG} ... by sequential clusters with a voiced stop: \mathfrak{GX} , \mathfrak{GXq} , ... To the best of my knowledge, there is nothing in !Xóó phonology to argue against this, although it goes against almost all phonological tradition.

6.2 The nature of !Xóó clusters

Although the click clusters seem complex, they are not unreasonably so. The second element of each cluster in rows 14–27 of Table 1 is either uvular or glottal, and so forms either a geminate closure or a simple release when following the posterior closure of the click; and each such second element exists independently.

Formally, in my proposal, the fact that accompaniments do not have the feature [+cons] means that Miller’s (2011) objection (see §3.3.5) to obstruent–obstruent clusters does not obtain: in my $/(!\otimes \mathfrak{X}q)/$, there is a parallel cluster of obstruents, but not a sequential cluster. This reflects the conceptual status of \mathfrak{X} as a synchronization point, which may carry manner features, rather than an obstruent in its own right.

As for the question, raised in (31a), of why there are no click clusters of the form, e.g., $!^h\mathfrak{q} = /(!\otimes \mathfrak{X}^hq)/$, the answer is that realizing the aspiration on $!^h$ would require either releasing the posterior closure and then re-forming it for \mathfrak{q} , so creating a sequential cluster of released obstruents, or transferring the aspiration to the \mathfrak{q} , resulting in something indistinguishable from $!^h\mathfrak{q}$. One may note also that the nasal accompaniment does occur in clusters: I analyse $\mathfrak{?}\mathfrak{X}$ as $/\mathfrak{?}\tilde{\mathfrak{X}}/$, and it may be that $\mathfrak{X}h$ is phonologically $/\tilde{\mathfrak{X}}h/$.

The question remains of the pulmonic clusters in rows 20–23. There is no escaping the phonetic fact that these are sequential obstruent–obstruent clusters, which clearly violate any alleged constraint against such. It is, however, possible to suggest that they are licensed by an analogy with the click clusters, as follows. The click $|\mathfrak{q}^{\mathfrak{X}}|$ is $/(|\otimes \mathfrak{X}q^{\mathfrak{X}})/$. Suppose that the suction is weakened, so that the $/!|/$ switches from [+lingual] to [–lingual]. The result is the illicit parallel cluster $/(\otimes \mathfrak{X}q^{\mathfrak{X}})/$, which can be legitimised by fusing the $/t/$ with the $/\mathfrak{X}/$, resulting in $/tq^{\mathfrak{X}}/$. Thus one can see the \mathfrak{p} , \mathfrak{t} , \mathfrak{ts} clusters as weakened versions of the Θ , $|\mathfrak{}$, $||$ (for example) clusters. However, to quote Traill 1994, p. 161, “[i]t is not the intention of these observations to imply that non-clicks developed from clicks.” Rather, there are many interesting parallelisms between clicks and non-clicks, which, I think, neither Traill, Güldemann nor my proposal has yet fully explained.

6.3 Concurrency in the !Xóõ vowel space

As I described in §2.4, the phonetic vowel space of !Xóõ has five basic vowels, together with (in Traill's view) arbitrary combinations of pharyngealization, creakiness, breathiness and nasalization: so instead of the two dimensional IPA vowel chart, there is a six dimensional chart. The phonological analysis in Table 1' cuts things down somewhat, but even so there are 26 (DOBES) or 37 (Traill) vowel phonemes.

From the point of view of acquisition and stability of the sound system, all the same arguments apply as with clicks. Thirty-seven is a lot of vowels, and as with clicks, some of them are rare, or even unattested. There is, for example, no attested occurrence of **ɔ̥**, but it would be strange indeed if a nonce word including it were not recognized as such.

As with the clicks, there is also morphological evidence that creakiness and nasalization at least behave independently of basic vowel quality. I sketched the principles of the !Xóõ concord system in §2.1.1. For most dependent forms, the vocalic part of the concord is **-ã, -e, -i, -u**, according to the class of the governing noun – the function word described in the lexicon as **kV**, for example, will appear as **kã, ke, ki, ku** according to concord. The demonstrative 'this' is **tYV**, taking the allomorphs **taã, te, ti, tu** – thus the creakiness on the vowel, and indeed the length of the vowel, are part of the lexical specification, while the basic vowel quality and nasalization vary with concord. So the qualities qualify as morphophonemes at least.

I have also noted that strident epiglottal vowels appear to be phonologically breathy pharyngealized; and that there are Single Aspirate and Glottal Constraints (2) and (3).

Then, given the free interplay of voice qualities and nasalization, it is obviously tempting to treat them as phonemes rather than morphophonemes. One could do this by claiming that the first mora of a word may have coda consonants **ʔ, fi, ʔ**, and the second **ŋ**, as is written in the DOBES orthography (with *q, h, ', n*), and that these consonants then spread their quality to the vowels. However, while both creakiness and pharyngealization are (Traill 1985) often realized with a peak that sounds like a light stop, this peak does not appear to occur between moras, but in the first: e.g. **a^ʔi** sounds more like [aʔai] than [aʔi].

Thus, if I wish to admit these qualities as phonemes, the obvious way to do so is to make them concurrent with the vowels, e.g. /(**a** ⊗ **ʔ**)/. In the formal setting, this requires relaxing the Strong Concurrent Airstream Constraint (10) to allow concurrent actions within the pulmonic airstream, and extending the synchronization rules accordingly, but raises no other issues.

Following my discussion in §4.2.3, I also have to justify their existence as

phonemes in the inventory. This requires a rather greater relaxation of the notion of segment than for click accompaniments, and leads into controversial issues.

- *Acoustically*, each of the four basic qualities have measurable correlates.²³
- *Articulatorily*, nasalization and pharyngealization are independent gestures. Breathiness and creakiness are not, as they require opposite laryngeal gestures; but the resolution of the conflict by sequencing permits them to be conceived of as such. Other languages such as Chong (Theraphan 1991) have also been reported to have breathy–creaky vowels implemented by sequencing.
- *Perceptually*, the four basic qualities are independently perceptible without training – even in English they are recognized paralinguistically, either as emotional indicators (breathiness and creakiness) or as stereotypes of other languages: the well known ‘nasal twang’ (Sweet 1877, p. 8, Mayo and Mayo 2011) of some accents of English, or the ‘guttural’ sound of Arabic, arising from the pharyngeal and uvular consonants. It is not always easy to distinguish breathiness and nasalization, as these qualities share a number of acoustic cues (Arai 2006), but other languages (such as Mazatec languages, or Hindi) use both breathiness and nasality.
- In *production*, I predict that, for example, if one teaches a !Xóǀ speaker [y], they will immediately be able to produce [y̥] and [y̜].

6.4 Nasality in other languages

The suggestion of nasality as a phoneme immediately brings to mind other languages. Nasality occurs in many different language families, and its behaviour varies widely, from ‘featural’, through what I am arguing is ‘concurrent segmental’, to something that seems to be supra-segmental, even up to word level, and is naturally seen via autosegmental theory. For example:

In phonetic and purely phonological descriptions of French, the nasal vowels are standardly seen to have phonemic status. The qualities of some of the vowels have drifted far from the oral counterparts – e.g. the historical and orthographic *in* is not [ī] but [ã̃] – and although the connection between nasal and oral is live, in alternations such as masculine *gamin* /-ã̃/ vs feminine *gamine* /-in/, this is usually seen as morphophonological, on a par with the English /ai/ vs /i/ in *divine/divinity*.²⁴

In Portuguese, the nasal vowels have essentially the same quality as their oral

²³ This is not an entirely honest statement: nasality has a rather wide and complex set of acoustic cues (Raphael 2005).

²⁴ Naturally, as with English, there is a movement representing French in full *SPE* style with essentially mediaeval URs, and all the morphophonology included in the rewrite rules. I do not consider this aspect of *SPE* to be within the realm of phonology.

counterparts, and although the morphophonology is similar to French, some analyses of Portuguese phonology propose retaining the historical following nasal, e.g. as an archiphoneme /N/ (Barbosa and Albano 2004), and regarding the nasalization as phonetic. One could argue that the situation is in fact neither of those: rather, nasalization is a concurrent phoneme with the vowel.

Then there is !Xóǀ, where, I have argued, nasalization appears to behave exactly like any other phoneme, save for sitting on top of a vowel rather than after it, and so is a good example of a concurrent phoneme.

Beyond that, in many South American languages, nasality appears as a suprasegmental property, so that, for example, [m] may appear as an allophone of /b/ that occurs in nasal morphemes or syllables. Then there may be spreading rules which may propagate the nasality further in the word, subject to various blocking conditions. (See, e.g., Peng 2000 for illustrations.) This extensive nasal harmony is naturally treated via autosegmental processes; for example, Botma (2004) treats such languages (and others) within the framework of Dependency Phonology. Of course, formally one could claim that Tuyuca (Barnes 1996) [mãĩĩ] and [tĩĩǀ] are underlyingly /($\tilde{\otimes}$ bari)/ and /($\tilde{\otimes}$ tigó)/, but as Barnes' title suggests, there appear to be morphemes marked nasal, marked oral, and unmarked. Asserting nasality as a quasi-segment is one thing, but asserting orality is quite another, and so I would not claim that concurrent phonemes are an appropriate way to analyse nasality in Tuyuca.

6.5 Concurrent phonemes in language change

Returning to the case of French, I would further suggest that the history of French may be understood more easily by the use of concurrent phonemes. A standard philological description of the development of the French nasal vowel in *quand* from Latin *quandō* would be, compressing irrelevant changes:

- (32) a. (/kwando:/ [kwando:] →)
 /kant/ [kant] → /kant/ [kãnt] → /kãt/ [kãt]
 (→ /kã/ [kã])

An equally standard criticism of such accounts is that there is an explanatory lacuna at the phonologization stage: the trigger for the change disappears, and so the nasal vowel is phonologized – but if the trigger disappears, why doesn't the nasalization? The most obvious answer is to invoke generational change: if the children analyse as [kãnt] as /kãt/ (viewing the [n] as excrement) what their parents think of as /kant/ (viewing the [̃] as spreading), then two grammars with the same output can coexist. The phonologization is Ohala's (1981) notion of hypocorrection, but in his account, it is not clear why the children should "fail

to hear” the [n], unless they do hear it and apply his hypercorrection to interpret it as [∅]. The simultaneous hypo-/hyper-correction seems a little contorted.

My preferred answer to this old puzzle is the one that says that phonologization can happen without contrast; or, more generally, that there is a continuum between allophony and phonemic contrast,²⁵ and an allophonic distinction can become gradually internalized in the mental representation, as suggested by, for example, Joan Bybee (Hooper 1981). (See also Peperkamp, Pettinato, and Dupoux 2003 for an experimental study of the allophone/phoneme distinction during acquisition, and Hall 2009 for a model of such systems.) In categorical terms, this amounts to promoting the phonetic intermediate to a non-contrastive but phonological intermediate:

- (32) b. /kant/ [kant] → /kant/ [kãnt] *phonetic spreading*
→ /kãnt/ [kãnt] *hypocorrection*
→ /kât/ [kât] *hypercorrection*

Such an account results in the simultaneous emergence of many unsupported phonemes, one for each oral vowel that gets nasalized, existing without contrastive support for possibly several generations. If we cast the history in terms of concurrency, then the intermediate stage involves only one new phoneme to account for all the vowels that undergo nasalization – and moreover, the use of concurrency avoids interference in existing phonotactics, as the sequential adjacency relation is unchanged. Only when nasalization is completely fused (as perhaps in French but perhaps not in Portuguese) do we really have five new vowel phonemes. Thus we might have:

- (32) c. /kant/ [kant] → /kant/ [kãnt] *spreading*
→ /k(̃ ⊗ a)nt/ [kãnt] *hypocorrection*
→ /k(̃ ⊗ a)t/ [kât] *hypercorrection*
→ /kât/ [kât] *concurrent fusion*

A similar story might be told about palatalization changes. In Gaelic, for example, palatalization emerged from adjacent front vowels in the usual way, but a standard synchronic phonemic analysis simply posits separate palatalized

²⁵ For example, in my fairly conservative RP speech, coda /l/ is dark but fully lateral, and until it was pointed out²⁶ to me at the age of 12 or so, I had never considered coda and onset /l/ to be different. My 10yo son, however, has a fully vocalized coda /l/ [ɣ], and considers this to be clearly a “different sound” from onset /l/ [l], although he has no evidence for a contrast between them, and otherwise shows no particular ability in phonetic discrimination.

²⁶ By Tolkien 1966, p. 392.

and plain (or velarized) versions of most consonants. However, speakers are (at least in the presence of elementary education) well aware of the distinction, and every Gaelic speaker knows that there is broad (*leathan*) /t/ and slender (*caol*) /tʲ/. So one might even say that Gaelic has not yet fused the palatalization, and /tʲ/ ([tʲ ~ tʃ]) is still /t(⊗^j)/ – whereas in English, there is no synchronic relationship at all between /k/ and /tʃ/, although the latter is historically a palatalization of the former.

6.6 Tone

No discussion can be complete without mentioning tone, the concurrent quality *par excellence*. It has always been considered, in both the Western and Chinese linguistic traditions, that Chinese tone is a property of syllables, parallel to the segmental content. Other tone languages also do this, and indeed often tone, despite being contrastive, is not considered worth writing in everyday use, even when the official orthography supports it (e.g. Zulu and Xhosa – and Khoisan languages).

In the case of typical African language families, the tonology is rich and involves sometimes very long-range processes. Such complexity was one of the main motivations for Goldsmith's (1976) elaboration of autosegmental phonology, and for the same reason, it is too rich to be sensibly encompassed within my notion of concurrent phoneme.

With Chinese and similar languages, on the other hand, it seems plain that tone meets every test I have suggested for segmenthood rather than featurehood, and so I would certainly claim that a toneme is a concurrent phoneme. However, unlike the situation with clicks, such a statement is purely a rephrasing of what everybody already agrees, and gives no new insights.

7 Conclusion

In this article, I have proposed a modification of the traditional understanding of the terms SEGMENT and PHONEME to include the notion of parallel as well as sequential clustering. In the case of Khoisan languages, such a modification dramatically reduces the inventory sizes, and thereby makes the languages appear much less exotic – and also much easier to acquire and maintain, if one accepts that maintaining a large number of phonemic contrasts is harder than using contrasts between clusters of phonemes. It also allows a better account of some phonological processes found in the languages. I may note that such a radical reduction in inventory sizes naturally challenges the methodology of some recent proposals (Atkinson 2011) about language dispersion.

In addition, the use of concurrent analyses of clicks exposed hitherto unobserved facts about phonological distributions in !Xóǎ, and thereby suggested an allophonic relationship between two accompaniment phonemes, one of which is a long-standing puzzle for its rarity.

I have also demonstrated a range of other uses for the concept of concurrent phoneme, where an audible character appears to behave more like a segment than a feature; and proposed that this gives a better motivated account of various diachronic processes.

Acknowledgements

In addition to those thanked in the body of the article, I owe debts to many colleagues. I thank especially D. Robert Ladd – our decade-long conversation on our shared interest of concurrency and simultaneity has both encouraged and influenced the work of this article. Also at Edinburgh, I am grateful for encouragement and careful reading of an early draft by Geoffrey K. Pullum and Patrick Honeybone, and for comments from other members of the P-Group, particularly Alice Turk, Bert Remijsen and Satsuki Nakai.

At the Max Planck Institute for Evolutionary Anthropology in Leipzig, I have received much kind assistance from members of the DOBES project. Christfried Naumann provided me with his Taa data, answered many questions, and discussed !Xóǎ phonology at length. Linda Gerlach provided me with her †Hoǎ data, and also commented very carefully on an early draft.

Following talks at the Manchester Phonology Meeting, the University of Toronto and the 2013 Traill Memorial Conference, I received a number of helpful comments, from Dan Silverman, John Goldsmith, Krisztina Polgárdi, Amanda Miller, Tom Güldemann and others. I thank B. Elan Dresher for several discussions on the nature of the phoneme, and Daniel Currie Hall for suggestions on the formalization; and Naomi Nagy for kind hospitality when I visited Toronto.

Finally I thank the reviewers and editors of *Phonology*, whose thorough and insightful comments transformed the article between the submission and the final version.

References

- Arai, Takayuki (2006). Cue parsing between nasality and breathiness in speech perception. *Acoustical Science and Technology* 27(5), 298–301.
- Atkinson, Quentin (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332(6027), 346–349.
- Bale, Alan, Maxime Papillon, and Charles Reiss (2013). Sets of features vs. sets of sets of features. Poster at 21st Manchester Phonology Meeting.

- Barbosa, Plínio A. and Eleonora C. Albano (2004). Brazilian Portuguese. *Journal of the International Phonetic Association* 34(2), 227–232.
- Barnes, Janet (1996). Autosegments with three-way lexical contrasts in Tuyuca. *International Journal of American Linguistics* 62(1), 31–58.
- Beach, Douglas M. (1938). *The Phonetics of the Hottentot Language*. Cambridge: W. Heffer.
- Best, Catherine T., Anthony Traill, Allyson Carter, K. David Harrison, and Alison Faber (2003). !Xóǀ click perception by English, Isizulu, and Sesotho listeners. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9, 2003*, 1292–1295.
- Besten, Michael Paul (2006). *Transformation and Reconstitution of Khoe-San Identities*. Ph. D. thesis, Universiteit Leiden.
- Boersma, Paul and David Weenink (2013). Praat: doing phonetics by computer. Version 5.3.18.
- Botma, Bert (2004). *Phonological Aspects of Nasality: an Element-Based Dependency Approach*. LOT. Ph.D. thesis, University of Amsterdam.
- Campbell, Lyle (1974). Phonological features: problems and proposals. *Language* 50(1), 52–65.
- Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. New York, NY: Harper & Row.
- Dixon, Robert M. W. (1997). *The Rise and Fall of Languages*. Cambridge University Press.
- Dresher, B. Elan (2011). The phoneme. In M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice (Eds.), *The Blackwell Companion to Phonology*, Volume 1, pp. 241–266. Malden, MA & Oxford: Wiley-Blackwell.
- Flemming, Edward (2003). The relationship between coronal place and vowel backness. *Phonology* 20, 335–373.
- Goldsmith, John A. (1976). *Autosegmental Phonology*. Ph. D. thesis, MIT, Cambridge, MA.
- Greenberg, Joseph H. (1950). Studies in African linguistic classification: VI. The click languages. *Southwestern Journal of Anthropology* 6(3), 223–237.
- Güldemann, Tom (2001). *Phonological regularities of consonant systems across Khoisan*. Language and Literature Series 16, University of Leipzig Papers on Africa. Universität Leipzig.
- Güldemann, Tom and Edward D. Elderkin (2010). On external genealogical relationships of the Khoe family. In M. Brenzinger and C. König (Eds.), *Khoisan Languages and Linguistics: Proceedings of the 1st International Symposium January 4–8, 2003, Riezlern/Kleinwalsertal, Köln*, pp. 15–52. Rüdiger Köppe.
- Haacke, Wilfrid H. and Eliphaz Eiseb (2002). *A Khoekhoegowab Dictionary*.

- Windhoek: Gamsberg Macmillan.
- Hall, Kathleen Currie (2009). *A Probabilistic Model of Phonological Relationships from Contrast to Allophony*. Ph. D. thesis, Ohio State University.
- Heine, Bernd and Henry Honken (2010). The Kx'a family: a new Khoisan genealogy. *Journal of Asian and African Studies* 79, 5–36.
- Higgins, John (2013). Minimal pairs for English RP. <http://myweb.tiscali.co.uk/wordscape/wordlist/minpairs.html> accessed 2013–05–09.
- Hooper, Joan Bybee (1981). The empirical determination of phonological representations. In T. Myers, J. Laver, and J. Anderson (Eds.), *The Cognitive Representation of Speech*, pp. 347–357. Elsevier.
- International Phonetic Association (1999). *Handbook of the International Phonetic Association : A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press.
- Jakobson, Roman (1968). Extra-pulmonic consonants (ejectives, implosives, clicks). *Quarterly Progress Report* 90, 221–227.
- Kießling, Roland (2008). Noun classification in !Xoon. In S. Ermisch (Ed.), *Khoisan Languages and Linguistics: Proceedings of the 2nd International Symposium January 8–12, 2006, Riezlern/Kleinwalsertal*, pp. 225–248. Rüdiger Köppe.
- Ladd, D. Robert (2014). *Simultaneous Structure in Phonology*, Chapter Gesture, Feature, Autosegment. Oxford University Press.
- Ladefoged, Peter and Ian Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.
- Ladefoged, Peter and Anthony Traill (1984). Linguistic phonetic description of clicks. *Language* 60, 1–20.
- Mayo, Carolyn M. and Robert Mayo (2011). Normative nasalance values across languages. *ECHO* 6(1), 22–32.
- Miller, Amanda L. (2011). The representation of clicks. In M. van Oostendorp, C. Ewen, E. V. Hume, and K. D. Rice (Eds.), *The Blackwell Companion to Phonology, 5-Volume Set*, Volume 1, pp. 416–439. Malden: Wiley-Blackwell.
- Miller, Amanda L., Johanna Brugmann, Bonny Sands, Levi Namaseh, Mats Exter, and Chris Collins (2009). Differences in airstream and posterior place of articulation among N|uu clicks. *Journal of the International Phonetic Association* 39(2), 129–161.
- Miller, Amanda L., Levi Namaseh, and Khalil Iskarous (2007). Tongue body constriction differences in click types. In J. Cole and J. I. Hualde (Eds.), *Laboratory Phonology* 9, pp. 643–656. Mouton de Gruyter.
- Miller-Ockhuizen, Amanda L. (2003). *The Phonetics and Phonology of Gutturals: a Case Study from Ju|'hoansi*. Routledge.

- Nakagawa, Hiroshi (2006). *Aspects of the Phonetic and Phonological Structure of the G/ui Language*. Ph. D. thesis, University of the Witwatersrand, Johannesburg.
- Naumann, Christfried (2008). High and low tone in Taa Ꞥaa (!Xóó). In S. Ermisch (Ed.), *Khoisan Languages and Linguistics: Proceedings of the 2nd International Symposium January 8–12, 2006, Riezlern/Kleinwalsertal*, pp. 279–302. Rüdiger Köppe Verlag.
- Naumann, Christfried (forthcoming). The phoneme inventory of Taa (West !Xoon dialect). In R. Vossen and W. Haacke (Eds.), *Essays in memory of Anthony Traill*. Rüdiger Köppe Verlag. Manuscript version of 2009.
- OED (2011, March). *OED Online*. Oxford University Press. <http://www.oed.com/>.
- Ohala, John J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, and M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior*, pp. 178–203. Chicago Ling. Soc.
- Peng, Long (2000). Nasal harmony in three South American languages. *International Journal of American Linguistics* 66(1), 76–97.
- Peperkamp, Sharon, Michèle Pettinato, and Emmanuel Dupoux (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, and F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*, Volume 2, pp. 650–661. Cascadilla Press.
- Prince, Alan and Paul Smolensky (1993). Optimality theory: Constraint interaction in generative grammar. Technical Report 2, Rutgers University Center for Cognitive Science.
- Raphael, Lawrence J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni and R. Remez (Eds.), *Handbook of Speech Perception*, pp. 182–206. Blackwell.
- Schultze Jena, Leonhard (1928). *Zur Kenntnis des Körpers der Hottentotten und Buschmänner*. Gustav Fischer.
- Snyman, Jannie Winston (1970). *An Introduction to the !Xū (!Kung) Language*. Johannesburg: Balkema.
- Sweet, Henry (1877). *Handbook of Phonetics*. Oxford: Clarendon Press.
- Theraphan, L-Thongkum (1991). An instrumental study of Chong registers. In J. H. Davidson (Ed.), *Austroasiatic Languages, Essays in honour of H. L. Shorto*, pp. 141–160. London: School of Oriental and African Studies, University of London.
- Tolkien, John Ronald Reuel (1966). *The Return of the King* (2nd ed.). George Allen & Unwin.
- Traill, Anthony (1985). *Phonetic and Phonological Studies of !Xóó Bushman*.

- Hamburg: Buske.
- Truill, Anthony (1991). Pulmonic control in Khoisan. *Journal of the International Phonetic Association* 21(1), 13–18.
- Truill, Anthony (1993). The feature geometry of clicks. In P. v. Staden (Ed.), *Linguistica: Festschrift E. B. van Wyk: 'n huldeblyk*, pp. 134–140. Pretoria: J. L. van Schaik.
- Truill, Anthony (1994). *A !Xóǀ Dictionary*. Köln: Köppe.
- UCLA (2009). *The UCLA Phonetics Lab Archive*. Los Angeles: UCLA Department of Linguistics. <http://archive.phonetics.ucla.edu/>.
- Walsh, Regina (2009, October). Word games: The importance of defining phonemic awareness for professional discourse. *Australian Journal of Language and Literacy* 32(3), 211–225.
- Wiese, Richard (2000). *The Phonology of German*. Oxford University Press.
- Žagar, Linda L. and John L. Locke (1986). The psychological reality of phonetic features in children. *Language, Speech, and Hearing Services in Schools* 17, 56–62.

A Appendix: transcriptions

This Appendix lays out the complex detail and history of notations used for the sounds of !Xóǀ in the primary sources.

A.1 Initial non-click transcriptions

The non-click initials are mostly familiar from other languages, and so there is little confusion in the notations. I give here the transcriptions used by Traill for the Eastern dialect, and by the DOBES project for the Western – the latter transcriptions are being introduced as a practical orthography. As indicated in the tables, not all the sounds found by DOBES were found by Traill.

This article	p	t	ts	k	q	ʔ	b	d	ɖ	g	ɣ
Traill	<i>p</i>	<i>t</i>	<i>ts</i>	<i>k</i>	<i>q</i>	<i>'</i>	<i>b</i>	<i>d</i>	<i>dz</i>	<i>g</i>	<i>ɣ</i>
DOBES	<i>p</i>	<i>t</i>	<i>ts</i>	<i>k</i>	<i>q</i>	<i>'</i>	<i>b</i>	<i>d</i>	<i>dz</i>	<i>g</i>	<i>gq</i>

This article	p^h	t^h	ts^h	k^h	q^h	b^h	d^h	ɖ^h	g^h	ɣ^h
Traill	<i>ph</i>	<i>th</i>	<i>tsh</i>	<i>kh</i>	<i>qh</i>		<i>dth</i>	<i>dtsh</i>	<i>gkh</i>	<i>gqh</i>
DOBES	<i>ph</i>	<i>th</i>	<i>tsh</i>	<i>kh</i>	<i>qh</i>	<i>bh</i>	<i>dh</i>	<i>dzh</i>	<i>gh</i>	<i>gqh</i>

Traill's notation for the voiced aspirates emphasizes the pre-voicing and the voiceless release. As remarked, his notation is phonetically misleading for *dtsh*, as both in the surviving Traill recordings and in DOBES data, the sibilant portion is voiced.

This article	p'	t'	ts'	k'	q'	ɖ'	g'	ɣ'	qʰ'	ɣʰ'	m	n	ʔm	ʔn	s	ɣ	h
Traill		<i>t'</i>	<i>ts'</i>	<i>k'</i>	<i>q'</i>				<i>kx'</i>	<i>gkx'</i>	<i>m</i>	<i>n</i>	<i>'m</i>	<i>'n</i>	<i>s</i>	<i>x</i>	<i>h</i>
DOBES	<i>p'</i>	<i>t'</i>	<i>ts'</i>	<i>k'</i>	<i>q'</i>	<i>ɖ'</i>	<i>g'</i>	<i>gq'</i>	<i>qx'</i>	<i>gqx'</i>	<i>m</i>	<i>n</i>	<i>'m</i>	<i>'n</i>	<i>s</i>	<i>x</i>	<i>h</i>

p' is even more marginal than the other labials – DOBES has one example. Traill did not recognize the simple voiced ejectives, and although he has *gkx'*, for him this belongs in the clusters below.

This article	f	l	r	tqʰ'	tsqʰ'	dqʰ'	ɖqʰ'	tɣ	tsɣ	dɣ	ɖɣ
Traill	<i>f</i>			<i>t'kx'</i>	<i>ts'kx'</i>	<i>dt'kx'</i>	<i>dts'kx'</i>	<i>tɣ</i>	<i>tsɣ</i>	<i>dtɣ</i>	<i>dtsɣ</i>
DOBES	<i>f</i>	<i>l</i>	<i>r</i>	<i>tqx'</i>	<i>tsqx'</i>	<i>dqx'</i>	<i>dzqx'</i>	<i>tɣ</i>	<i>tsɣ</i>	<i>dɣ</i>	<i>dzɣ</i>

Initial **f**, **l**, **r** occur only in loan-words in DOBES, and only **f** in Traill. Traill's *kx'* reflects the question about whether *q^x'* belongs in the velar or uvular series, on which he vacillated; DOBES views it as uvular. The 'double ejective' *t'kx'* is a compromise among the various pronunciations he heard for this series.

A.2 Medial consonants

The transcriptions are straightforward.

This article	b	m	n	ɲ	j	l	r
Traill	<i>b</i>	<i>m</i>	<i>n</i>	<i>ɲ</i>	<i>j</i>	<i>l</i>	<i>r</i>
DOBES	<i>b</i>	<i>m</i>	<i>n</i>	<i>ny</i>	<i>y</i>	<i>l</i>	<i>r</i>

A.3 Final consonants

The transcriptions are similarly straightforward (Traill did not find or recognise **ŋ**, which as noted is marginal in DOBES.)

This article	m	n	ŋ	p	b	r
Traill	<i>m</i>	<i>n</i>		<i>p</i>	<i>b</i>	<i>r</i>
DOBES	<i>m</i>	<i>nn</i>	<i>ng</i>	<i>p</i>	<i>b</i>	<i>r</i>

A.4 Click transcriptions

Owing to the difficulty of distinguishing and identifying the many accompaniments, the transcriptions of clicks present a particularly knotty problem to the reader of the primary sources, and I go in to it in considerable detail, aiming also to elucidate some of the changes in Traill's analysis over the years.

I shall give the notation used by Traill and DOBES, and also the notation used in the clicks chapter of Ladefoged and Maddieson 1996 (henceforth *SoWL*), which is based on Traill's analyses, but makes phonological assumptions that are disputed, as I discuss below. I also give the articulatory descriptions used by Traill (1994) and by Naumann (forthcoming).

There are several confusing aspects of the Traill and *SoWL* notations, so I consider the clicks not in chart order, but grouped by their scope for confusion.

First, there are some fairly straightforward cases:

Row	Traill desc	Traill	here	<i>SoWL</i>	DOBES	DOBES desc
1	basic	ɣ	ɣ	kɣ	ɣ	plain
2	voiced	ɣg	ɣ̣	gɣ	gɣ	voiced
9	voiceless nasal	ɣŋ	ɣ̥̃	ŋɣ	nhɣ	voiceless nasal
10	voiced nasal	ɣn	ɣ̣̃	ŋɣ	nɣ	voiced nasal
11	pre-glottalized nasal	ʔɣ	ʔ̣̃	ʔŋɣ	ʔnɣ	glottalized nasal
14	voiceless uvular stop	ɣq	ɣ̥q	qɣ	ɣq	plain + /q/
15	voiced uvular stop	ɣg	ɣ̣g	gɣ	gɣq	plain + /q/ + voice
22	uvular fricative	ɣx	ɣ̣x	kɣ ^x	ɣx	plain + /x/
23	voiced uvular fricative	gɣx	ɣ̣x	gkɣ ^x	gɣx	plain + /x/ + voice

The main issue here is the *SoWL* notation. Ladefoged and Maddieson chose to notate clicks by combining a click symbol with a preceding velar stop symbol showing the accompaniment. However, in the **ɣq** clicks (rows 14–15), they simply change [k] to [q], suggesting that the difference is purely one of place, and ignoring the prolongation of the closure. As discussed above in §3.1, this is most likely wrong. In the case of the fricative clicks, *SoWL* opts for the affrication symbol, which I rejected on phonetic grounds as well as phonological, and they write it as velar rather than uvular. In order to emphasize the pre-voicing, they write [gkɣ^x] rather than just [gɣ^x].

Next, I consider the clicks that involve aspiration in some way. Traill's notations for these are confusing, as his understanding changed during his studies.

Row	Traill desc	Traill	here	<i>SoWL</i>	DOBES	DOBES desc
3	aspirated	ɣqh	ɣ ^h	kɣ ^h	ɣh	voiceless aspirated
4	vcd. asp. [ʔ94 only]	gɣqh	ɣ̣ ^h	gɣh	gɣh	voiced aspirated
24	delayed aspiration	ɣh	ɣ ^h	ŋɣ ^h	ɣhh	plain + /h/
25	vcd. asp. [ʔ85 only]	gɣh	ɣ̣ ^h	–	nɣhh	plain + /h/ + voice
16	(uvu.) asp. stop	ɣqh	ɣq ^h	–	ɣqh	plain + /qh/
17	vcd. (uvu.) asp. stop	gɣqh	ɣ̣q ^h	gɣh	gɣqh	plain + /qh/ + voice

The DOBES survey finds a set of six clicks involving aspiration: the simple aspirates in rows 3–4; the clicks I write as phonetic clusters with [h] in rows 24–25; and those I write as phonetic clusters with [q^h] in rows 16–17.

At first, Traill (1985) recognized only three of these: two (*ɣh*, *gɣh*) whose descriptions make them clearly rows 24–25, and one that is clearly described as sounding like [ɣq^h] (16), and consequently written *ɣqh*.

Then in Traill 1994, he was less certain about this last click, describing it in ways suggesting that it is actually our **ɣ^h** (row 3). He also added its voiced counterpart, written *gɣqh*; and moreover added a new *gɣqh*, described so as to be our **ɣq^h** (row 17). He also no longer recognized the row 25 clicks, merging their words with the voiceless row 24.

What the true story is, is hard to tell. It is obviously tempting to assume that the DOBES version is correct, and that Traill conflated some of the clicks in different ways at different times. The small number of Traill's recordings available to me do not help.

Note that DOBES has chosen to mark the (possibly phonetic, possibly phonological) nasalization in the voiced delayed aspirate row 25. The *SoWL* notation again marks phonetic detail that blurs apparent phonological patterns.

Finally, I consider the clicks involving ejection or glottalization.

Row	Traill desc	Traill	here	<i>SoWL</i>	DOBES	DOBES desc
5		ɣ'	ɣ'	–	ɣ'	voiceless ejective
6		–	ɣ'	–	gɣ'	voiced ejective
26	glottal stop	ɣ'	ɣʔ	kɣʔ	ɣ''	plain + /' /
27		–	ɣʔ	–	nɣ''	plain + /' / + voice
18	uvular ejective	ɣq'	ɣq'	qɣ'	ɣq'	plain + /q' /
19		–	ɣq'	–	gɣq'	plain + /q' / + voice
20	velar ejective	ɣkɣ'	ɣq ^ɣ '	kɣ ^ɣ '	ɣqɣ'	plain + /qɣ' /
21	voiced velar ejective	gɣkɣ'	ɣq ^ɣ '	gɣk ^ɣ '	gɣqɣ'	plain + /qɣ' / + voice

The story here is similar to the aspirated clicks, though not quite as complex. Traill recognized an accompaniment *ɣq'*, which, it is clear from (1985, p. 143), is our **ɣq'** with delayed posterior release. He did not recognize its voiced counterpart. He also did not distinguish it from a 'plain ejective' **ɣ'**, though he did distinguish it from **ɣʔ**. DOBES, however, finds all three of **ɣq'**, **ɣ'** and **ɣʔ**, together with their voiced counterparts. Again, cross comparison would be inter-

esting – perhaps Traill conflated the two ejectives \mathfrak{x}' and \mathfrak{xq}' . In the DOBES examples for \mathfrak{xq}' , the gap between the click burst and the ejected stop is sometimes quite easy to hear, but sometimes as low as ten milliseconds, even in the formal sentence-speaking context. In the examples for \mathfrak{x}' , the gap is minimal, less than 2 ms – nonetheless, if one cuts away the click burst, one clearly hears the [q']. On the other hand, in Traill's recordings, there are examples of \mathfrak{xq}' (in a word that is also \mathfrak{xq}' according to DOBES) where the only observable difference from \mathfrak{x}' is a slightly lower CoG in the click burst.

In the ejective affricates (20–21), Traill was again a little uncertain about the place of articulation. DOBES considers this to be a cluster with an uvular affricate.

A.5 Vowels

The notations used in the various sources are as follows, taking **a** as an example:

This article	a	ã	ᵛ	ᵛ	a^ʕ	ᵛ^ʕ	ᵛ	a^ʕ	ᵛ^ʕ
Traill	<i>a</i>	<i>ã</i>	<i>ah</i>	<i>a'</i>	<i>ᵛ</i>	<i>ᵛh</i>	<i>ah'</i>	<i>ᵛ'</i>	<i>ᵛh'</i>
DOBES	<i>a</i>	<i>an</i>	<i>ah</i>	<i>a'</i>	<i>aq</i>	<i>aqh</i>			

The notations for strident vowels reflect Traill's view that stridency is phonologically the combination of breathiness and pharyngealization – Traill rather confusingly uses a tilde below to denote pharyngealization, while DOBES uses a fairly natural overloading of *q* (since /q/ does not occur post-vocally).