Journal of Phonetics (2002) **30,** 1–52
doi:10.1006/jpho.2001.0150
Available online at http://www.idealibrary.com on IDEAL®

Voice quality differences associated with stops and clicks in Xhosa

Michael Jessen*

Speaker Identification and Tape Analysis Section, Bundeskriminalamt, Germany and Research Unit for Experimental Phonology, University of Stellenbosch, Republic of South Africa

Justus C. Roux

Department of African Languages and Research Unit for Experimental Phonology, University of Stellenbosch, Republic of South Africa

Received 27th July 1999, and accepted 14th August 2001

Voiced stops/clicks in Xhosa cause f_0 depression in the following vowel. Some authors also claim that these sounds are followed by breathy phonation. An appropriate set of words with different stop and click categories was read by eight speakers of Xhosa. Voice quality was measured as H_1^*–H_2^* and H_1^*–A_3^*. f_0 and F_1 were determined as well. Measurements were made for four periods early in the following vowel and one close to its center. The occurrence of breathy voice after voiced stops/clicks was inferred from levels of H_1^*–H_2 and H_1^*–A_3^* about as high or higher than those generally expected closely after voiceless aspirated stops. The existence of f_0 depression after voiced stops/clicks could be confirmed. F_1 was usually highest after aspirated and lowest after voiced stops/clicks. Indications of breathy voice after voiced stops/clicks were found for some speakers only. It is argued that extensive larynx lowering and vocal fold slackening can explain the specifics of the voicing feature in Xhosa. A similar situation occurs in Shanghai Chinese. Based on that comparison it is suggested that "slack voice" is a more appropriate term for the relevant Xhosa sounds than "breathy voice".

1. Introduction

Xhosa — a Southern-Bantu language of the Nguni group, spoken in South Africa — has a series of stops and click accompaniments that are phonemically transcribed as voiced. However, phonetically, there is usually no closure voicing that exceeds the short and weak voicing into closure pattern also found in the two voiceless cognates of Xhosa. Despite this virtual lack of voicing during closure the term "voiced" is still used here; it will be applied in the sense of "low-frequency property" (see Kingston & Diehl, 1994, 1995 and Section 4.2). The most salient and reliable feature of the voiced stops and clicks

*E-mail: michael.jessen@bka.bund.de

0095-4470/02/010001 + 52 $35.00/0

(meaning "voiced stops and voiced clicks" here and throughout the text) is a lowering of the pitch in the following vowel that is more salient and temporally stable than the pitch perturbation property after voiced obstruents that is found as a universal tendency. In addition to this tonal depression feature, the voiced stops and clicks are claimed by some authors to cause breathy phonation during at least the first part of the following vowel. In this study, measurements are made of correlates of breathy voice including a normalized version of $H1–H2$, f_0 as an index of tonal depression, and F_1 as another correlate of distinctions based on voicing and aspiration in general. Before proceeding with a review of the literature on breathy voice in Xhosa we take a look at the sounds of Xhosa that serve as a basis of comparison in this study and we provide some information about their phonetic realization. Finally, in this introduction, methodological aspects of measuring breathy voice are addressed.

1.1. *Stop types and click accompaniments in Xhosa*

The sounds of Xhosa that are addressed in this study are shown in Table I. These represent only a small subset of the entire Xhosa consonant system. A recent overview of the entire system can be found in Finlayson, Jones, Podile & Snyman (1989). Among the stops, we omitted a palatal series that also exists in Xhosa; since only three places are found among the clicks, it is preferable for statistical reasons to select only three stop places as well. The term "stop" is used in this paper with the same scope as in Ladefoged & Maddieson (1996), hence including plosives, ejectives, and implosives, but excluding nasal consonants.

The different stop types and click accompaniments are arranged in rows and the different places of articulation (following Ladefoged & Traill, 1994 for the characterization

TABLE I. Target sounds of this study arranged according to place and context (columns) and sound category (rows), with separate displays for stops (upper display) and clicks (lower display). Each sound is presented phonetically (in square brackets), followed by the representation in Xhosa orthography

	Bilabial				Alveolar				Velar			
Stops	Postvoc. or initial		Postnasal		Postvoc. or initial		Postnasal		Postvoc. or initial		Postnasal	
Plain/ejective	[pʼ]	p	[pʼ]	p	[tʼ]	t	[tʼ]	t	[kʼ]	k	[kʼ]	k
Aspirated	[pʰ]	ph	—		[tʰ]	th	—		[kʰ]	kh	—	
Voiced	[b̥ʱ]	bh	[bʱ]	b	[d̥ʱ]	d	[dʱ]	d	[g̊ʱ]	g	[gʱ]	g
Implosive	[ɓ]	b	—									

	Dental				Alveolar				Lateral			
Clicks	Postvoc. or initial		Postnasal		Postvoc. or initial		Postnasal		Postvoc. or initial		Postnasal	
Plain	[k|]	c	[ŋk|]	nkc	[k!]	q	[ŋk!]	nkq	[k‖]	x	[ŋk‖]	nkx
Aspirated	[k|ʰ]	ch	—		[k!ʰ]	qh	—		[k‖ʰ]	xh	—	
Voiced	[g̊|ʱ]	gc	[ŋ|ʱ]	ngc	[g̊!ʱ]	gq	[ŋ!ʱ]	ngq	[g̊‖ʱ]	gx	[ŋ‖ʱ]	ngx
Nasal		[ŋ|]	nc			[ŋ!]	nq			[ŋ‖]	nx	

of place in clicks) are arranged in columns, with a further subdivision into "postvocalic or initial" and "postnasal" position. In the present study, the former expression covers word-initial position (initial) as well as morpheme-initial position preceded by a prefix ending in a vowel (postvocalic), while the latter refers to morpheme-initial position preceded by a prefix ending in a nasal (cf. Section 2.1 for further explanation). This distinction is responsible for some allophonic phenomena (most particularly the assumed presence *vs.* absence of closure voicing) and for some phonotactic gaps (indicated by dashes).

Each sound in Table I is presented both in phonetic transcription and in Xhosa orthography. Although Xhosa orthography is phonetically transparent, there is a complication with respect to the bilabial voiced stop and the bilabial implosive, where *b* is used in the double function shown in the table. In keeping with the orthographic convention in Xhosa, the nasals are mentioned together with the postnasal clicks in Table I, whereas for the postnasal stops, the preceding nasals are not mentioned (Pahl, 1989). The postnasal plain clicks (e.g., *nkc*) and the postnasal voiced clicks (e.g., *ngc*) have to be distinguished from the nasal clicks (e.g., *nc*). The nasal clicks are monosegments (and not nasal-plus-click sequences, as *nkc*, *ngc*, etc. are) and have hybrid status with respect to their placement in Table I. On the one hand, they occur in postvocalic context (e.g., morpheme structure *u-nxango* "thirst", with - = prefix boundary), while on the other, they are often investigated in a paradigm with some or all postnasal clicks (e.g., Ladefoged & Maddieson, 1996, p. 260). The most important reported difference between postnasal voiced clicks and nasal clicks in Xhosa lies in the presence *vs.* absence, respectively, of tonal depression and possibly breathy voice (cf. Ladefoged & Traill, 1994; Ladefoged & Maddieson, 1996 on Xhosa click accompaniments).

The transcription of the plain clicks with *k* and the conventions for the voiced, aspirated, and nasal click accompaniments follows from Ladefoged & Traill (1994) and Ladefoged & Maddieson (1996). Other phonetic transcriptions used in Table I reflect certain choices among alternative positions in Xhosa phonetics. We come back to those once we have provided a brief sketch of the issues involved.

One issue involves the stops written as *p*, *t*, *k* in Xhosa. More recent sources, including Pahl (1989) and Finlayson *et al.* (1989), assume that these sounds are produced as ejectives. Lanham (1960, 1969) and Herbert (1987) say that ejection in *p*, *t*, *k* is only produced in slow/formal/careful speech styles or in prominent positions. Stylistic conditioning in the production of ejection is also mentioned for Zulu — another language within the Nguni group — by Doke (1926). In the present study, which is based on rather careful speech, we found clear auditory indications of ejective *p*, *t*, *k* with some, but not all speakers, and further variability within speakers. Upon examination of waveforms and spectrograms, the main acoustic correlates of ejective productions in our material were increased positive VOT and increased burst amplitude relative to the other stop cognates. Most importantly for the present focus, we did not encounter any systematic indications of creaky voice in the vowels after those cases of *p*, *t*, *k* that were auditorily identified as ejectives. Although creaky voice was occasionally found in the following vowel, that possibility occurred for other stops and clicks as well and was not concentrated on *p*, *t*, *k*.

Ejection has also been claimed by some authors to be possible in the plain clicks written *c*, *q*, *x* in Xhosa, again with a suggestion that this is most likely in formal speech styles (Beach, 1938; Lanham, 1960; Louw, 1977*a*; Pahl, 1989). In the present material, clear audible ejection with the plain clicks was only found for one speaker

(M2). These ejective clicks of M2 were produced with long average positive VOT (above 110 ms).

The Xhosa sound categories which attract the greatest controversy are the voiced stops and clicks. The issue of breathy voice associated with these sounds is addressed separately in the following subsection. Beyond that issue, the received view is that the voiced stops and clicks of Xhosa are usually produced without (or with negligible) closure voicing in postvocalic and initial context, and that only after nasals are these sounds literally voiced (Ziervogel, 1967; Pahl, 1989; Finlayson *et al*., 1989). Voicing in postvocalic position remains an option in Xhosa, again partially mediated by stylistic factors (Brand & Roux, 1990), but in the corpus analyzed here only one single token of a voiced stop with voicing during most of the closure was found in this context. Likewise, Sands (1991) found that the voiced clicks of Xhosa are voiceless during closure in (most probably) initial position. In postnasal position, it was found for the present material that voiced stops have very short closure durations ("closure" being used in the sense of occlusion(s) in the oral cavity with raised velum), usually below 30 ms. Postnasal voiced clicks most often have no closure at all in the sense just mentioned. The absence of a *g* in the phonetic transcription in Table I ([ŋ|$^{\text{ɦ}}$], etc.), which follows from Ladefoged & Traill (1994), symbolizes this predominant absence of closure. The absolute closure voicing durations of voiced stops are not systematically longer in postnasal position than in other contexts (nor is closure voicing longer in voiced than in voiceless, i.e., plain/ejective or aspirated, stops and clicks). Thus, voiced stops and clicks in postnasal position are only voiced in the indirect sense that they have a very short or absent closure and that for this reason the percentage of voicing during closure is high, often at 100%. The only sound addressed here with long absolute duration of voicing during closure is the implosive.

Given the absence of closure voicing in most contexts the question arises as to which other property constitutes a more reliable way of distinguishing voiced from other categories, most particularly from the plain stops and clicks if they are produced without ejection. It has been reported that the voiced obstruents of Xhosa induce tonal depression, i.e., the realization of a subsequent vowel with lower pitch than after the other cognates (Rycroft, 1980). In and of itself this is not an unusual phenomenon; it is known as pitch perturbation in many languages (see Kingston & Diehl, 1994, and further references therein). However, the magnitude and temporal scope of this effect and its interaction with other tonal phenomena suggests that tonal depression in Xhosa is more phonologized than the pitch perturbation effect that is found as a cross-linguistic tendency. Tonal depression has also been reported and thoroughly investigated for Zulu (Traill, Khumalo & Fridjon, 1987).

Based on this sketch of controversial areas in Xhosa phonetics, we return again to the phonetic transcriptions chosen for Table I. We have transcribed orthographic *p*, *t*, *k* as ejectives, since that manifestation constitutes a definite possibility. These sounds are classified in the first column as "plain/ejective". Since for the corresponding clicks ejective productions are very rare, we do not transcribe ejective click accompaniments. In the discussion of our data below, we simply talk about plain stops and clicks to refer to orthographic *p*, *t*, *k* and *c*, *q*, *x*. This is the most practical solution, since stops and clicks will usually be analyzed together.

As far as the voiced category is concerned (in postvocalic or initial position), we use voiced *b* (written *bh* in that position), *d*, and *g* together with the diacritic for voicelessness (analogous to the transcription often used for devoiced *b*, *d*, *g* in English). In

that transcription, the voicelessness diacritic refers to the predominant absence of voicing during closure, whereas the voicing symbol refers to the presence of other phonetic indications of phonological voicing, most importantly tonal depression. The phonological significance of this "low-frequency property" (Kingston & Diehl, 1994, 1995) will be addressed in the general discussion. Furthermore for the voiced stops/clicks, we use a transcription in which breathy voice follows the stop/click, rather than occurring simultaneously with it. This deviates from the transcription with simultaneous breathy phonation used for Xhosa by Ladefoged & Traill (1994) and Ladefoged & Maddieson (1996). With our convention, we want to symbolize claims in the literature that breathy voice is primarily manifested in the early part of the following vowel, to be discussed in Section 1.2 (cf. Roux & Dogil, 1997 for further discussion of simultaneous *vs.* sequential diacritics in the transcription of clicks and their accompaniments).

1.2. *Previous studies of breathy voice in Xhosa and Zulu*

Probably, the most controversial issue involving the stop types and click accompaniments in Xhosa is the question of whether the voiced stops and clicks have the effect of inducing breathy phonation in the following vowel. There are some authors who do not mention the occurrence of breathy voice with stops or clicks at all (e.g., Jordan, 1966; Wentzel, Botha & Mzileni, 1972). Others mention it only in association with the voiced clicks in postnasal position (e.g., Lanham, 1960; Ziervogel, 1967; Riordan, Mathiso, Davey, Bentele, Mahlasela & Lanham, 1969). Finally, there is the position that breathy voice occurs with every voiced stop and click in every context (e.g., Pahl, 1989; Finlayson *et al*., 1989). This position is also held by Rycroft (1980) on Nguni languages in general. Rycroft mentions as a likely possibility for Xhosa that roughly the first half of a vowel following the voiced category is breathy voiced, while the latter half of the vowel is produced with modal voice (his Fig. 1.3).

Pahl (1989) explains why some authors assume breathy voice only for the postnasal voiced clicks that are currently written *ngc*, *ngq*, *ngx* and not for voiced stops or voiced clicks in other contexts. According to Pahl, the breathy voiced accompaniment is restricted to "Tshiwo Xhosa (Gcaleka & Rharhabe), i.e., in what was originally regarded as Standard Xhosa" (p. xxxiv). In the "New Orthography" of 1935–1955 (p. xlix) these breathy click types were written as *nch*, *nqh*, *nxh* and phonemically and orthographically distinguished from nonbreathy *ngc*, *ngq*, *ngx*. However, as Pahl explains, this orthographic distinction had to be abandoned in favor of the latter because Xhosa speakers from other dialects, which form the majority of Xhosa speakers, as well as the younger generation of Tshiwo Xhosa speakers, did not make this phonemic distinction and therefore had difficulties with the orthographic distinction as well. It seems that authors including Lanham (1960), Ziervogel (1967), and Riordan *et al.* (1969) assume that this merger between *nch*, *nqh*, *nxh* and *ngc*, *ngq*, *ngx* mentioned by Pahl (1989) created a voice quality in current *ngc*, *ngq*, *ngx* that lies somewhere in-between the presence *vs.* absence of breathy voice in the earlier distinction or that it inherited the breathy voice of earlier *nch*, *nqh*, *nxh*. In fact, Ladefoged & Traill (1994) found indications of breathy voice associated with voiced clicks only with a preceding nasal, not in initial position. This issue will be further investigated in this study (Section 3.5). However, we need to consider the possibility that breathy voice of the kind (and probably magnitude) traditionally associated with postnasal voiced clicks might not be found in the present material, since

only the modern spelling *ngc*, *ngq*, *ngx* was used and no subjects from the older generation of traditional Xhosa speakers were involved.

The only previous experimental or instrumental data on breathy voice associated with the voiced stops or clicks of Xhosa we know of are presented by Louw (1977*a*). Louw, looking for spectrographic evidence of breathy voice in Xhosa, mentions "widely spaced striations indicating murmur" (p. 81; similarly Louw, 1977*b*). What Louw takes as evidence for breathy voice (which is usually assumed to be the same as murmur) can only be meant in an abstract phonological sense. Phonetically, a wide spacing of periods signifies a relatively low fundamental frequency and says nothing about voice quality *per se*.

In contrast to the situation for Xhosa, there are more data on this issue for Zulu, mostly pointing to the absence rather than the presence of breathy voice associated with voiced stops and clicks. Doke (1926) says that "there is no such thing as an aspirated *b* (*bh*) in Zulu, (...)" (p. 51). With the instrumentation available to Doke at the time he might have been unable to detect very subtle indications of breathy voice, but his kymographic waveforms (pp. 51f., 60) at least show no signs of turbulence associated with the voiced stops of Zulu. Kymographic evidence for breathiness/voiced aspiration associated with the voiced category in Zulu has been mentioned by Selmer (1933) and Lanham (1969). However, the kymographic patterns are either poorly reproduced (in the case of Selmer, 1933) or only discussed very briefly (in the case of Lanham, 1969), so the situation remains inconclusive.

Traill *et al.* (1987) is the most informative study on breathy voice in Zulu known to us. Based on electroglottographic recordings of voiced stops and other depressor consonants among the obstruents and sonorants of Zulu, Traill *et al.* found consistent breathy voice only during the constriction phase of voiced fricatives and voiced *h*. Otherwise, no indications were found that the consonants that induce tonal depression are accompanied by breathy phonation, while the remaining consonants are not. Tokens with breathy voice were found, but these were distributed unsystematically among all sound categories. Traill *et al.* also observed that the spread of breathy voice from the consonant into the following vowel often lasted only a few periods. The finding of consistent breathiness during the constriction phase of voiced fricatives and voiced *h* is interesting. However, it might not be easy to make a distinction between the turbulence that derives from breathiness and that from frication, which is expected to be found in a voiced fricative. Fricative production also makes it difficult to distinguish glottal opening patterns that are actively produced from those that result passively from oral constriction (see Bickley & Stevens, 1987; Stevens, 1998). Furthermore, breathy voice during *h* is a common phenomenon that can also be found in English (Klatt & Klatt, 1990).

Outside the realm of the Nguni languages, but still within the Southern Bantu languages, it is interesting to examine the findings on breathy voice in Tsonga nasals presented by Traill & Jackson (1988). Among other things, they report a high degree of speaker specificity in the expression of voice quality. Data that are relevant in this context are also presented by Pongweni (1983) on breathy voice associated with voiced *h* in Shona. Another interesting area is the question of whether breathy voice in Zulu or Xhosa, if it exists, is a feature that was borrowed from Khoisan languages (see Louw, 1977*a*, *b* on this subject).

Beyond stops and clicks, some other sounds are claimed to be associated with breathy voice in Xhosa, but were not included in this study. According to Pahl (1989) and Finlayson *et al.* (1989), breathiness also occurs with some affricates, fricatives, nasals, liquids, and glides (including /ɦ/) in Xhosa. As far as the affricates are concerned, they

present some distributional gaps in Xhosa, since some of them are found only after nasals. This constitutes a phonologized form of stop intrusion between nasal and fricative (cf. Fourakis & Port, 1986). These distributional gaps would have complicated the necessary statistical procedures. For the fricatives we predicted methodological problems, some of which are discussed above in connection with the study of Traill *et al.* (1987). Other methodological problems arise with respect to the normalization procedures we use for the voice quality parameters chosen here. As will become clear later, these procedures are only reliable with sounds that have a relatively high F_1. This condition is not met during the constriction of fricatives or during the hold phase of other consonants. We could have investigated voice quality in the subsequent vowel just as we do for the stops and clicks, but there are two arguments against this. Firstly, voiced fricatives in Xhosa are likely to show voicing during closure, as opposed to voiced stops and clicks. From a functional point of view, that might reduce the need to resort to other means of implementing phonological voicing, such as tonal depression and breathy voice (see "low-frequency property", to be addressed in the general discussion). Secondly, it turned out that for the nonobstruents, our Xhosa informant had difficulties finding suitable examples. It appears — and is also mentioned for some sounds by Lanham (1960) and Finlayson *et al.* (1989) — that some of the claimed plain-breathy distinctions are limited to nonnative vocabulary or have a very low functional load. We did, however, include one near-minimal pair involving the sonorants written *nyh vs. ny*. This distinction is claimed by Lanham (1960) to be of reasonably stable status.

1.3. *Voice quality measurement methodology*

In order to obtain acoustic evidence on the existence of breathy voice associated with some of the stops and clicks of Xhosa we use what might be called H_1-based measurements. The basis of this procedure is to create an FFT spectrum around a relevant vocalic portion and to measure the amplitude of the first harmonic (meaning the fundamental) and the amplitudes of further spectral events, specifically the amplitude of the second harmonic and the amplitudes of the first three formants. Further calculations are applied (among the more recent practitioners of this method), and finally the amplitudes of the higher-frequency events are subtracted from the amplitude of the first harmonic to arrive at the values that are taken as acoustic indices of voice quality.

H_1-based measurements of voice quality were introduced by Fischer-Jørgensen (1967) and have subsequently been used successfully in a number of studies, including Bickley (1982), Ladefoged, Maddieson & Jackson (1988), Traill & Jackson (1988), and Klatt & Klatt (1990), just to mention a few. Of particular interest for us are studies in which voice quality is measured acoustically as an effect caused by the phonological properties of a consonant on the adjacent (mostly the following) vowel, including Chapin Ringo (1988), Cao & Maddieson (1992), Ren (1992), Ní Chasaide & Gobl (1993), and Jessen (1998). A number of methodological guidelines have to be observed with H_1-based measurements of consonant-induced voice quality.

Firstly, for the H_1-based methods to work best methodologically and to have maximum validity, there should be a reasonable spectral distance between the frequency location of the first harmonic and that of the first formant. This condition is met in low vowels and sometimes also in mid vowels. High vowels and consonants however, have F_1 values that are usually too low to maintain a proper distance from the first harmonic. Furthermore, as the degree of constriction increases, there can be a passive vocal fold

abduction due to intra-oral air pressure buildup, whose effects on the acoustics are hard to distinguish from the primary voice quality effect (Bickley & Stevens, 1987; Stevens, 1998). Specifically with nasal consonants, there can be interactions between the acoustic correlates of breathy voice and those of nasality (Klatt & Klatt, 1990). The consequence of these difficulties is that consonant-internal H_1-based measurements and those applied to high vowels are unreliable (Ladefoged *et al*., 1988; Klatt & Klatt, 1990; Stevens and Hanson, 1994; Hanson, 1995, 1997 for a general discussion of these and related issues involving H_1-based in relation to other voice quality measurements).

Secondly, voice quality effects induced by consonants on the adjacent vowel can be short and may not be found with equal strength throughout that vowel. This is analogous to what we know about f_0 perturbation induced by the voiced/voiceless distinction of preceding obstruents (see Hombert, 1978; Kingston & Diehl, 1994), which can also be subject to rapid changes. It is not possible to capture all of these temporal dynamics with H_1-based measurements, especially if there are changes from one period to the next. As is well known, there is an inverse relation between the temporal resolution of FFT analysis and its frequency resolution: with short window lengths, temporal resolution is good and temporal dynamics can be captured optimally, but frequency resolution is poor, so that the harmonics might not be recognized well enough or measured with sufficient precision. The opposite situation occurs with longer window lengths, where harmonic structure is clearly visible, but where there is much temporal smear, so that some of the fine-grained consonant-induced effects might be left uncaptured. (This problem is more serious with lower than with higher voices, since more periods fit into the temporal frame covered by the window in the latter case.) We found a window length of 25.6 ms to be optimal for our purposes, but this is the result of a compromise.

The specific way in which we performed our H_1-based measurements follows the work of Stevens & Hanson (1994), Sluijter (1995), Sluijter, Shattuck-Hufnagel, Stevens & van Heuven (1995), and Hanson (1995, 1997). Their approach extends previous H_1-based approaches by including a set of calculations intended for normalization purposes and for attempting to separate the source characteristics of voice quality variations from the contribution of the filter. These calculations will be addressed in the method section.

It has been found that breathy phonation is characterized by high values of H_1–H_2 or other H_1-based parameters (see Fischer-Jørgensen, 1967; Bickley, 1982; Ladefoged *et al*., 1988; Klatt & Klatt, 1990; Stevens & Hanson, 1994). Among these H_1-based breathy voice correlates, it is H_1–H_2 and H_1–A_1 (A_k = amplitude of formant k) that are mentioned most frequently in the literature. However, H_1–A_3 is also a good breathy voice correlate, as shown by Hanson (1997).

If the voiced stops and clicks of Xhosa were followed by some amount of breathy phonation during at least the early part of the following vowel, high values of H_1–H_2 and related parameters should be observable after voiced stops/clicks compared to other categories. Such an effect can be evaluated against the background of the universal tendency that after voiceless aspirated stops, vowels begin with some amount of breathy phonation as well. This is the result of carry-over coarticulation from the open glottis configuration required for the production of aspiration (see Chapin Ringo, 1988; Löfqvist & McGowan, 1992; Ní Chasaide & Gobl, 1993; Jessen, 1998; Stevens, 1998). In order to be able to infer the presence of breathy voice after voiced stops/clicks, those sounds should be associated with values of H_1–H_2, etc. that are similar to or higher than those found closely after aspirated stops/clicks.

2. Method

2.1. *Stimuli*

Two different word lists were designed, each of them read by four different speakers. In designing the lists, we were assisted by a native Xhosa speaker. The first list (to be referred to as List 1) is presented in Table II.

A number of considerations were taken into account in stimulus selection. Firstly, the target sounds should all occur stem-initially (word-initially in the case of List 1) and in a stressed syllable. We noticed that vowels out of main stress position can get very short and are often breathy or sometimes even voiceless. We were concerned that subtle voice quality effects that might be induced by the target sounds on the following vowel could get neutralized or skewed under low prominence.

Secondly, the vowel after the target sound should be /a/. The normalization procedures to be introduced in Section 2.2 work most reliably if the vowel in which the voice quality measurements are carried out has high F_1. In one case (*gxoba*), this condition could not be met due to the tonal judgment of our informant; here, we selected a mid vowel.

Thirdly, the subsequent consonant should not induce any voice quality of its own on the preceding /a/. For that reason voiced or aspirated obstruents, voiceless fricatives, and /h/ were avoided as much as possible, since they were most likely to carry their own voice quality signatures (also plain stops/clicks were avoided due to their potential for ejective productions and potentially stiff or creaky voice). We preferred /l/, but also allowed for nasals or the implosive in that position. The latter two inclusions are not optimal due to possible creaky or stiff voice in the latter and known nasality/breathiness interactions in the former (Klatt & Klatt, 1990). However, that was the best solution without resorting to nonsense words. We should point out, however, that due to the fact that our crucial voice quality measurements were limited to the first half of the vowel (stressed /a/), which was usually quite long, there was a good temporal safety margin between the measurement sites and the following consonant, so that it is likely that regressive coarticulation had no confounding effects. What we observed, instead, was that the vowel itself often had a voice quality contour, where the first part was modal (except from possible effects

TABLE II. Word List 1. Target sounds are presented in boldface, and separate displays are provided for stops (upper) and clicks (lower). The arrangement in three columns is according to place of articulation. The words are given in Xhosa orthography (notice that the first sound in *bala* is an implosive; *khala* also reoccurs in Table III). The words in Table II were spoken together with the preceding carrier word *uthi* "he says"

Word	Gloss	Word	Gloss	Word	Gloss
pana	the pair	**t**ala	to boast	**k**ama	to comb
phala	to run	**th**ala	to skin	**kh**ala	Cape aloe
bhala	to write	**d**ala	to create	**g**ana	to marry
bala	to count				
cala	the side	**q**ala	to start	**x**aba	to obstruct
chama	to urinate	**qh**ala	to unpack	**xh**ala	to get worried
gcaba	the burst	**gq**ala	to consider	**gx**oba	to contaminate

induced by the target sound) and the latter part was somewhat breathy and reduced in amplitude. This vowel-internal contour seemed to overshadow everything that might have been induced regressively by the following consonant.

Finally, we had to make a decision on the tone carried by the vowel that follows the target sounds. Originally, we planned to have a separate set of stimuli with high- and low-toned vowels, in analogy to the procedures used by Traill *et al.* (1987). It turned out that our informant had difficulties making this distinction with sufficient certainty on his side (cf. Roux, 1998 on the low functional load and perceptual importance of linguistic tone in Xhosa). Neither was there much help from the dictionaries or other Xhosa literature, where tone is transcribed only occasionally. In the case of Pahl (1989) tonal transcription is provided, but only a few letters are covered by this dictionary. We therefore decided to settle on one set of stimuli only. We gave our informant the instruction to look for low-toned examples. In some cases, we found that the tonal judgment made by our informant differed from transcriptions found in the dictionaries; in that case, preference was given to the judgment of the informant. The decision on a low-toned vowel was motivated by the following considerations. Rycroft (1980) claims that consonant-induced breathy voice in Nguni persists throughout the following vowel if that vowel is low-toned. One might in fact expect breathy voice to be more important in low- than in high-toned vowels. It is in the low-toned cases where the tonal depression feature of the voiced consonants will be more difficult to implement, since with the tone of the vowel being low already there will be less of a pitch range left for tonal depression. In fact, some authors claim that tonal depression is only found in high-toned syllables (see Rycroft, 1980; Traill *et al.*, 1987 for literature overview). However, for Zulu, Traill *et al.* (1987) show experimentally that before low tone there is still an f_0 difference between depressors and nondepressors, though one that is reduced relative to high tone context. If the cue value of pitch for the recognition of the voiced category is reduced before low tone, one might expect that voice quality has some compensatory function in the task of distinguishing the voiced from the other categories.

The words in Table II were printed on cards together with the preceding carrier word *uthi* "he says". Most of the words are verbs. However, in some cases examples of a different word class had to be selected to arrive at a word of the desired structure. In that case, an additional prefix is usually required in Xhosa. However, no hesitation or unusual speech behavior resulted from the lack thereof and the resulting speech samples could all be accepted for further analysis.

The list in Table II so far contains no items in postnasal position. In order to elicit the possible existence of breathy voice in that context as well, and in order to be able to compare postnasal examples with those after vowels, a second list (to be referred to as List 2) was created, which is presented in Table III.

The words in Table III were spoken in isolation. They are all nouns, as reflected by the fact that the stem-initial target sounds are preceded by a prefix. For the postvocalic context a prefix ending in or consisting of a vowel was chosen, while for the postnasal context a prefix ending in a nasal was selected. The nasal clicks are also listed with the postnasal clicks, although they are monosegments (and not nasal-cum-click sequences) and in terms of the morpheme structure strictly speaking postvocalic (e.g., *i-ncanda*, *i-nqambi*, *u-nxano*). The nasal clicks are investigated in a paradigm with the postnasal voiced and plain clicks in this study (Section 3.5.1). The items *i-nyele* and *i-nyheke* are mentioned by Lanham (1960) as good examples of words with modal *vs.* breathy palatal nasals, respectively. These are listed in the last division of Table III, together with further

TABLE III. Word List 2. Target sounds (including preceding nasal if present) are presented in boldface and separate displays are provided for stops (upper) and clicks (lower). The words are given in Xhosa spelling (notice that the boldface sound in *ibali* is an implosive, but the one in *imbala* a voiced stop). In each display words with the target sounds in postvocalic context occur first and those in postnasal context second (incl. voiced clicks). Among the display for clicks, a third division contains words with supposedly breathy palatal nasals (written *nyh*) and modal nasals (*ny*), as well as more words with some of the sounds in the second division

Word	Gloss	Word	Gloss	Word	Gloss
i**p**ali	pole	i**t**ayi	tie	isi**k**ali	scales
i**ph**ala	wanderer	i**th**ala	library	i**kh**ala	Cape aloe
i**bh**ali	bale of wool	i**d**ama	dam	i**g**ala	squirrel
i**b**ali	story				
i**mp**ala	impala	i**nt**amo	neck	i**nk**alo	neck of land
i**mb**ala	mark of a burn	i**nd**ano	disappointment	i**ng**alo	arm
u**c**ango	door	isi**q**alo	beginning	i**x**amba	pack of sugar
isi**ch**aso	opposition	i**qh**alo	proverb	i**xh**ala	fear
isi**gc**awu	spider	i**gq**ala	expert	isi**gx**ala	mark from sting
i**nkc**aso	opposition	i**nkq**ali	starter (person)	i**nkx**aso	support
ii**ngc**ango	doors	i**ngq**ambu	tongue ligament	u**ngx**alo	act of stuffing
i**nc**anda	Cape porcupine	i**nq**ambi	tabooed item	u**nx**ano	thirst
i**nkc**aza	comb	i**ny**ele	edge, border		
i**ngx**awa	gun charge	i**nyh**eke	upper hare lip		

examples with supposedly breathy as opposed to nonbreathy clicks mentioned in the literature. The conditions for stimulus selection mentioned for Table II (low-toned following /a/, etc.), also apply to the words in Table III, as far as possible.

In relation to the terminology used in Table I, the target sounds in Table II occur in initial position, while those in Table III occur either in postvocalic position or in postnasal position. Since the target sounds in Table II are preceded by a vowel as well (from the carrier word *uthi*), for the remainder of this paper we will use the term "postvocalic" to refer both to "initial" and to "postvocalic" in its narrower sense of word-internal vocalic precedence. As demonstrated in Table I, the two contexts form a natural class in terms of phonotactic and allophonic behavior.

2.2. *Subjects, recording, processing*

The two lists of stimuli were presented on cards in random order, and were read by native Xhosa speakers. Four speakers who will be referred to as F1, F2, M1, M2 read List 1, and four speakers who will be referred to as F3, F4, M3, M4 read List 2. M and F indicate the sex of each speaker. Speakers F1, F2, F3, M1, and M3 are students of a high school/college in Stellenbosch and close to 19 years of age. They were all raised by Xhosa-speaking parents. Our informant, M2, grew up in Stellenbosch as well but is frequently in contact with traditional Transkei Xhosa speakers. Speakers F4 and M4 were born and raised in the Transkei. M4 still lives there today, while F4 has lived in

Stellenbosch for 2 years. M2, M4, and F4 are between 35- and 50-years old. All subjects were paid, except for F4, who is a colleague.

Recordings were made in recording studios with sound treatment. The cards with the stimuli were presented to each subject one by one. Instructions were given to read the short phrases (List 1) and words (List 2) in a comfortable and natural tempo. In the case of reading errors the affected item was asked to be repeated and the error was ignored for further analysis. The words of List 1 were read 3 times, those of List 2 usually twice. Subjects spoke into an ECM-MS5 Electret Condenser Microphone (Sony) from a distance of about 35 cm. The microphone was positioned in such a way as to avoid direct impacts from the breath stream in the production of stop and click bursts. Recordings were made on DAT tape with a TCD-D10 Pro II Digital Audio Tape-Corder (Sony). The tape recordings were later transferred to CSL (Kay) for further acoustic analysis. During that process, the data on the DAT tape were downsampled to 20 kHz, and low-passed filtered accordingly, to allow spectral analysis for frequencies up to 10 kHz.

For six of the speakers, measurements were made for one of the repetitions only, with the remaining one(s) serving as a backup in case of errors. To investigate whether repetition had a statistical effect, velar stops and alveolar clicks were measured for all three repetitions in the speech of F1 and F2. One-way ANOVAs with Repetition as the factor and the four acoustic parameters investigated in this study as the dependent variables were calculated. For that purpose, data were pooled across speaker F1 and F2 and all periods, on the basis of the material that was read more than once. Repetition was non-significant for all four acoustic parameters. After it was determined that Repetition had no systematic effects on the results, these data were pooled with the remaining data of these speakers.

Since our interest is in those voice quality effects that are induced by the different stops and clicks in Xhosa, we established a closer net of measurements at the beginning of the following vowel than later in that vowel. Even if it is possible that consonant-induced effects last throughout the vowel, it is more likely that the strongest and most interesting effects are found in its earlier portions. We selected a set of five labels associated with the waveform for each token, serving as the center around which spectra were to be created. The first label was located in the first period after the target consonant. The second label was located two periods into the vowel counting from the position of the first. The next two labels were located further into the vowel, again taking every second period. The fifth and final label was located close to the midpoint of the vowel. These five periods will be referred to as p(eriod)1–5 in this paper. Naturally, the absolute time span covered by the first four labels differs depending on the average fundamental frequency of the speaker, and is shortest for high-pitched voices. Since the amplitudes of the periods usually increase after the stops and clicks, it was not always easy to decide on the location of the first label; we took the first clearly identifiable period, even if still small in amplitude.

For each token, an FFT spectrum with superimposed LPC spectrum was created around each of the five temporal locations just described. A Hamming window of 25.6 ms duration was used for this purpose. For LPC analysis, an order of 20 was used and pre-emphasis was applied. A number of amplitudes and frequencies were measured in each spectrum. The amplitude values that we measured are that of the first harmonic (abbreviated as H_1) and of the second harmonic (H_2), as well as the amplitudes of the first formant (A_1), the second (A_2), and the third (A_3) formant. On the frequency dimension, we measured the frequency of the first harmonic (i.e., the fundamental frequency; f_0) and the frequencies of the first three formants (F_1, F_2, F_3).

Identification of the first harmonic was straightforward in almost all cases.[1] H_1 was taken as the maximum amplitude value of that harmonic in the FFT spectrum. Taking the frequency of the first harmonic as f_0 turned out to be too inaccurate. Instead, we determined f_0 by measuring the duration of the period around which the spectrum was created and taking the reciprocal of period duration in seconds. The f_0 value thus obtained also served to indicate where to expect the first harmonic in case of doubt. The formant amplitudes A_1, A_2, A_3 were measured as the peak amplitudes of those harmonics that occur closest in frequency to the respective formants. If the harmonic that was literally closest in frequency to the formant was of unrepresentatively low amplitude, the harmonic of highest peak amplitude in close vicinity of the frequency of the LPC formant was selected. Where a formant could not be identified in LPC spectra, we looked for global maxima in the distribution of harmonics visible in the FFT spectra that would roughly correspond to the frequency where the formant was to be expected. This procedure was supported by further verification from the overall dynamics of formant structure visible in spectrograms and from related speech portions produced by the same speaker. In those cases, the relevant FFT maxima were used not only for their amplitude values but also for their frequency values, which served as an estimate of formant frequency. In very few cases, no measurements with sufficient certainty could be made at all; these were excluded from the analysis.

The measured values of H_1 and H_2 were normalized for the influence of F_1 and f_0 in order to achieve better comparability across speakers (and vowels, when considering about the cases with /o/ instead of /a/), and to achieve a better estimate of pure source characteristics. Following the procedure introduced by Hanson (1995, 1997) and Sluijter (1995) the result of (1) was subtracted from H_1 and the result of (2) from H_2, in order to arrive at what is referred to as H_1^* and H_2^*, respectively. The latter was subtracted from the former and the resulting parameter H_1^*–H_2^* was used as a dependent variable in this study.

$$20 \log_{10}(F_1^2/((F_1 + f_0)\ (F_1 - f_0))) \tag{1}$$

$$20 \log_{10}(F_1^2/((F_1 + 2f_0)\ (F_1 - 2f_0))) \tag{2}$$

With the same intent, A_3 was corrected for the influence of F_1 and F_2. Following Hanson (1995, 1997), the result of the calculation in Equation (3) was added to A_3, to arrive at A_3^*. This value was subtracted from H_1^*, resulting in the dependent variable H_1^*–A_3^*.

$$20 \log_{10}(((1 - (F_3/F_1)^2)(1 - (F_3/F_2)^2))/((1 - (F_3/F_{1N})^2)(1 - (F_3/F_{2N})^2))) \tag{3}$$

[1] In very few tokens (not at all limited to the plain/ejective and implosive category) there was a series of periods at vowel onset that were diplophonic (as characterized primarily by alternating period amplitudes) and auditorily creaky. Those tokens were not included for the following reason. From some test spectra created in these creaky portions we could determine a low-frequency peak of very low amplitude. The frequency of this peak corresponds to the reciprocal of the period durations that are determined if the intermediate low-amplitude peak in the diplophonic waveform is ignored. Had we taken these low-frequency low-amplitude peaks in the FFT spectrum as the first harmonic, H_1–H_2 and the other H_1-based parameters would have become extremely low. Low values of these parameters in creaky voice are consistent with reports in the literature (Ladefoged *et al*., 1988; Klatt & Klatt, 1990). However, it seems that when voice quality on the creaky voice end of the voice quality continuum turns diplophonic and results in the existence of first harmonics that are unusually low in frequency and amplitude for the speaker, application of the H_1-based methods is not advisable. Instead, it seems more sensible in those cases to treat creakiness as a qualitative phenomenon rather than one that is located within the space of values that are found with H_1-based measurements.

F_{1N} and F_{2N} in formula (3) stand for the "neutral formants". These are average frequency values of the first two formants across all tokens produced by all speakers of the relevant vowel. In almost all cases, this was the vowel /a/, while for the word *gxoba* "to contaminate" in List 1 it was /o/. For /a/, $F_{1N} = 802$ Hz, $F_{2N} = 1586$ Hz; for /o/, $F_{1N} = 542$ Hz, $F_{2N} = 988$ Hz. (Normalization in the case of /e/ in the words *inyele* and *inyheke* was not always successful, and therefore only the unnormalized H_1–H_2 and H_1–A_3 are reported for these words; see Section 3.5.) In addition to H_1^*–H_2^* and H_1^*–A_3^*, we also determined H_1^*–A_1 and H_1^*–A_2^* (the correction of A_2 into A_2^* was performed according to a calculation proposed by Sluijter, 1995). The results from these latter parameters did not differ in any substantial or systematic way from those from the two former voice quality parameters; in the interest of space, we do not report the results for H_1^*–A_1 and H_1^*–A_2^* here.

Measurement of f_0 and F_1 was necessary since this information is required in the calculations. However, these parameters were also used as dependent variables in their own right. Measurement of f_0 allows us to directly investigate claims about tonal depression in Xhosa. Including F_1 is also of interest, since this parameter is known to be sensitive to voicing and aspiration. F_1 is higher after aspirated than unaspirated stops. Low F_1 can result from larynx lowering, which is likely to be prominent with the voiced stops and clicks of Xhosa. Those and other influences are made more explicit later in this paper.

2.3. *Statistical treatment*

Two types of ANOVAs were calculated. For the first type, all stops and clicks were included that occur after a vowel — either separated by a word boundary (Table II) or by a word-internal morpheme boundary (Table III). In Section 2.1, we decided to unify these two contexts with the term "postvocalic". In postvocalic position, all different stop types and click accompaniments are possible, including the aspirated one. For this material, three-way ANOVAs were determined with stop type and click accompaniment (to be referred to as Category) as one independent variable (plain, voiced, aspirated), Speaker as the second (M1–4, F1–4), and the stop/click distinction (Manner) as the third.

The second type of ANOVA includes the stimuli in postnasal position. Since these were only spoken by half of the speakers, only the speech from these speakers was included (hence, only material from Table III was used). In order to most directly compare postnasal sounds with postvocalic ones, the aspirated category, which only occurs in the latter case, was not taken into account. On this material, four-way ANOVAs were calculated, with Category (plain, voiced), Speaker (M3–M4, F3–F4), Manner (stop, click), and Context (postnasal, postvocalic) as the factors.

Some aspects are the same for both types of ANOVAs. Firstly, different places of articulation in stops and clicks were not differentiated in statistical analysis, since the labial/alveolar/velar distinction in the stops does not carry over to the clicks.[2] Secondly, data for different periods (1–5) were treated separately and were not pooled. This was advisable since dynamic changes into the vowel are expected. Finally, the dependent variables used with all ANOVAs were f_0, F_1, H_1^*–H_2^*, and H_1^*–A_3^*.

[2] In 2-way ANOVAs with Category and Place, run separately for stops and clicks due to the different sets of place of articulations involved, Place and its interaction with Category remained mostly non-significant, in contrast to the expected (mostly) significant effects for Category. The only exception occurred for clicks and H_1^*–H_2^*. Here Place was significant for periods 3 and 4. It turned out that the alveolar click induced the highest and dental clicks the lowest H_1^*–H_2^* values. (In these statistical tests, the postnasal sounds were not included.)

Since Category, Speaker, and their interaction turned out to be significant in many cases, it was advisable to carry out additional one-way ANOVAs separately for each Speaker, with Category as the factor. *Post hoc* tests from these statistics were used to differentiate the effects of the three different categories plain *vs.* voiced *vs.* aspirated. This was done separately for Period, Speaker, and Context, but pooled across stops/clicks and across different places of articulation. Since a full statistical report is space consuming, it is only provided for H_1^*–H_2^*, which turns out to be the parameter most informative to the main topic of this paper. For the other three dependent variables, a shorter report is presented in which only the presence of significance at the level $p < 0.05$ in *post hoc* comparisons is mentioned. Unless postnasal context is mentioned explicitly the results from these one-way ANOVAs refer to postvocalic position.

The results of the statistics mentioned so far are reported in Sections 3.1–3.4, separately for each of the four dependent variables. Further issues are addressed on the basis of descriptive statistics only (Section 3.5), since they require work on more narrowly defined subsets of the data, resulting in numbers too low for reasonable inferential statistics. Some results for the implosive are mentioned briefly in Section 4.2 but are not reported numerically.

3. Results and discussion

3.1. *Tonal depression* (f_0)

3.1.1. *Results*

In the three- and four-way ANOVAs mentioned in Section 2.3, significant results were obtained for Category, Speaker, and the interaction between the two in all five periods, and most often at the level $p < 0.001$. Significant effects involving Manner and Context were limited to individual periods, did not show a systematic pattern, and had p values usually higher than 0.001. A full report is presented in Table AI in Appendix A.

Means and standard deviations of f_0 for the first four measured periods of the following vowel plus a fifth period at vowel midpoint are presented in Fig. 1. Note that, as mentioned in Section 2.2, periods 1–4 in the figure represent every second period, and that period 5 represents the center of the vowel. For Fig. 1, data were pooled across places of articulation and across stops and clicks. Different speakers were also pooled (F1–4, M1–4), except that female and male speakers were differentiated due to well-known gender differences in average fundamental frequency. Fig. 1 contains the data for postvocalic position, where all three stop and click categories can occur and not those in postnasal position, where aspirated stops and clicks are ruled out. It turned out that for speakers F3–4 and M3–4, who produced both postvocalic and postnasal tokens, the f_0 values of plain and voiced stops/clicks and their dynamic patterns according to period number are very similar in postvocalic and postnasal position. Due to these similarities between the two contexts, the results for postnasal position are presented not as figures here but as a table in Appendix B (Table BI). Analogous similarities between the two contexts occur with the remaining parameters, F_1, H_1^*–H_2^*, and H_1^*–A_3^*, and need not be mentioned further; those postnasal data are all listed in Table BI in Appendix B.

Fig. 1 shows that the vowels after voiced stops and clicks begin with substantially lower f_0 than vowels after the aspirated or plain (ejective) category. Although the

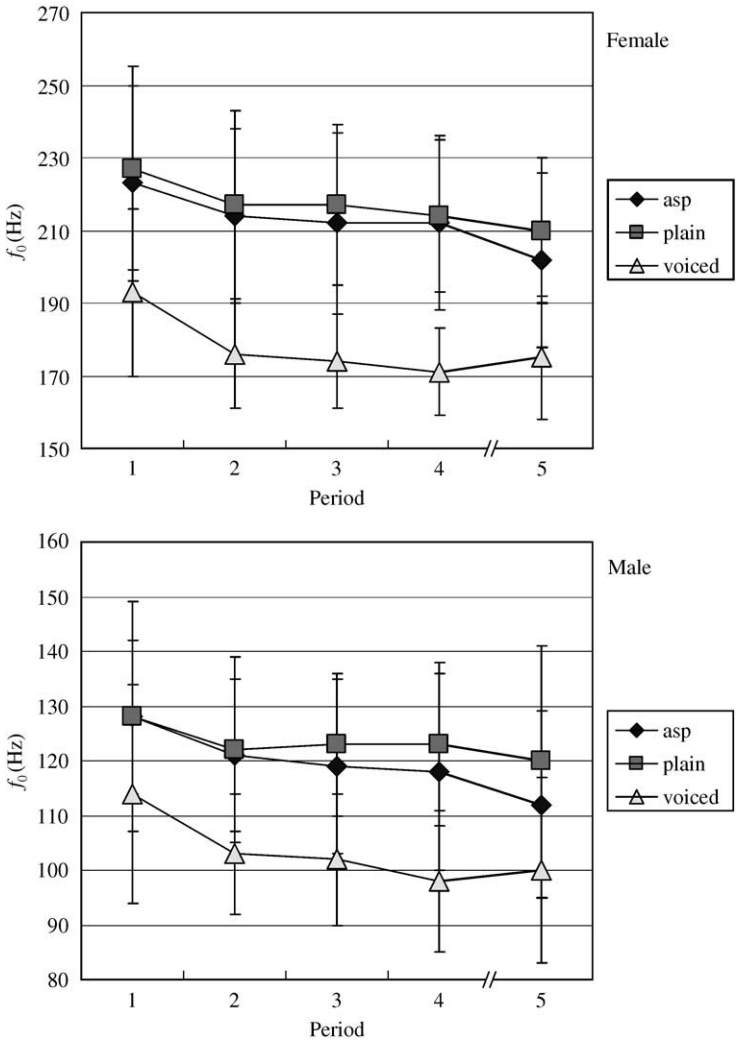

Figure 1. Results for f_0. Means and standard deviations (in Hz on the *y*-axis) for each period (*x*-axis) and each of the three stop and click cognates (aspirated, plain, voiced). Separate graphs for female speakers (upper) and male speakers (lower).

difference is largest for the first four measured periods, especially p2–4, there is still a clear difference at the center of the vowel (p5). The aspirated and plain category do not differ much with respect to each other. The direction of the differences between aspirated and plain is subject to speaker differences. For speakers F2 and M2, there is lower f_0 after aspirated than after plain stops and clicks, whereas in the speech of M1 f_0 is lower after the plain than after the aspirated category; for the remaining speakers f_0 is essentially identical after both categories, often with changes in direction for different periods. Fig. 1 also shows that after plain and aspirated stops and clicks, there is an f_0 lowering movement from the first period to vowel center. For the voiced stops and clicks, we observe f_0 lowering as well, but it only lasts until the fourth measurement point; from that point until vowel center an f_0 rise can be seen. (Given the lack of measurements

between the fourth pair of periods and vowel center it cannot be determined where exactly between vowel onset and vowel center the lowest f_0 value occurs.)

All the f_0 patterns mentioned so far are quite uniform across speakers. However, it needs to be mentioned that for one speaker (M4), the distance between the low f_0 values in the voiced category and the higher f_0 values of the other two categories (especially the aspirated) is less striking than in the average patterns shown in Fig. 1.

Post hoc tests based upon one-way ANOVAs separately for each speaker with Category as the factor showed for most speakers and periods that f_0 was significantly lower after voiced stops/clicks than after their aspirated on plain counterparts, with no significant difference between the aspirated and the plain category. It was only in the speech of M2 (p2 and p3) that significant differences were also found between the aspirated and plain category. In a few cases (p5 by F2, p1 by M1, p1 and p5 by M2, p4 by M4), only the comparison between the voiced category and the one of the remaining categories with the highest f_0 was significant. No significant effects occurred for the first period in the speech of F1, F2, F3, M3, and M4, probably due to reduced mean differences compared to periods 2–4 and to the relatively large standard deviations for that period. Consistent with the observations on M4, mentioned above, there were no significant effects for his stops/clicks in postvocalic position, except for period 4, where significantly smaller values for voiced than plain were found. However, this speaker showed a significant effect for periods 2–5 after nasals. Significantly lower values after voiced than plain stops/clicks was the general pattern in postnasal position, except for p1 by F3, M3 (and M4) as well as p2 by M3, where no significant effects were obtained.

3.1.2. *Discussion*

These results show that f_0 after voiced stops and clicks in Xhosa is substantially lower than after aspirated and plain cognates. The f_0 differences induced by the different stop and click types are essentially maintained for at least the first half of the following vowel. This is different from the f_0 perturbation patterns occurring in many languages, where the difference in f_0 between voiced and voiceless stops is often substantially reduced with increasing distance from vowel onset (Hombert, 1978). Lowered pitch after voiced stops and clicks probably has more prominence in the phonological system of Xhosa speakers than the redundant voicing-induced pitch differences found in many other languages. If that is the case, it is reasonable to assume that Xhosa speakers use direct means, such as larynx lowering with activation of extrinsic laryngeal muscles, to ensure implementation of the low f_0 values observed here.

The f_0 lowering movement from the first period to vowel center observed after plain and aspirated stops/clicks could reflect the intended low-tone target of the vowel (which might not be reached until very late in the vowel; cf. Traill *et al.*, 1987, p. 261) and/or it could be the implementation of a declarative intonation contour. The complex lowering–rising f_0 contour after voiced stops and clicks suggests that the f_0 pattern after voiced stops and clicks is more than carry-over coarticulation from a low consonant-inherent f_0 value. If that were the case, we would expect to find a (nearly) straight interpolation between low pitch at vowel onset and the low intonational or tonal target at vowel center (or later), but not a complex pattern where the f_0 value at vowel onset and vowel midpoint is higher than during the intervening period. It seems that the speaker actively plans the first part of the vowel after voiced stops and clicks to be implemented with a low f_0 value and that this implementation works better a few periods into the

vowel than immediately at vowel onset. If this is correct, the f_0 values that are implemented after the voiced stops/clicks in Xhosa are on average below the lexical and/or intonational low tone target of the vowel.

3.2. F_1

3.2.1. *Results*

The three- and four-way ANOVAs present a picture similar to that for f_0 in that by far, most significant effects were found for Category, Speaker, and the interaction between the two. However, while Speaker was significant for all periods at $p < 0.001$, Category and the interaction were sometimes not significant for periods 2–5 and the p values were often lower. As before, significant effects involving Manner and Context were not particularly systematic and statistically robust. The details are presented in Table AII in Appendix A.

The descriptive results for all stops and clicks in postvocalic position across speakers are presented in Fig. 2. Its format is analogous to Fig. 1, except that this time no differentiation according to gender was necessary.

Fig. 2 shows that after aspirated stops/clicks, F_1 starts relatively high and undergoes little change up to the vowel center. After the voiced and plain category, on the other hand, there is a rising F_1 transition from p1 to p5. The difference between the influence of the voiced and the plain stops/clicks is relatively small in the average data presented in Fig. 2, but across periods 1–4, F_1 is consistently lower after the former than the latter. Contrary to the situation for f_0, most of the difference between the categories disappears towards the center of the vowel.

With F_1 there is more speaker variability than with f_0. Those speaker differences are addressed now, but the average patterns shown in Fig. 2 provide a good point of

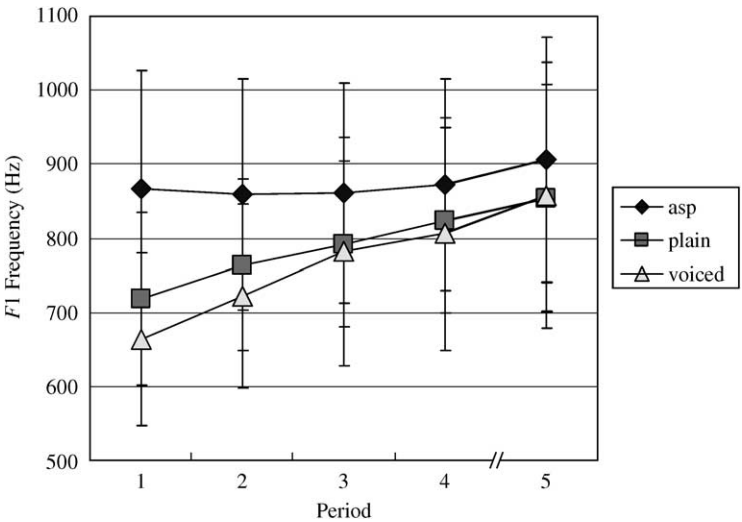

Figure 2. Results for F_1 pooled across all speakers. Means and standard deviations in Hz.

departure for the following discussion. For all speakers except M1 and M4 F_1 after aspirated stops/clicks is higher in all or nearly all periods than F_1 after the voiced or plain category. The voiced category induces the lowest F_1 values in comparison to plain and aspirated stops/clicks for speakers F1, M1, M2 and F4 (with few exceptions only), as well as for F3 but only in postnasal position. One-way ANOVAs and *post hoc* tests separately for each speaker give an impression of the magnitude of the observed patterns. The higher F_1 position after aspirated stops/clicks compared to their voiced and plain counterparts is significant for the first two periods of speakers F1 and F2 and for the first period of speakers F3 and M3. For the third period of M2, the first period of F4, and the second period of M3, F_1 is significantly higher only for the aspirated set when compared to the voiced category. The lower F_1 position after voiced stops/clicks compared to both other categories is significant for the first period of M1 and the first two periods of M2. In postnasal position, significant effects (voiced lower than plain) were obtained for all periods in the speech of F4 but for none in the productions of the other three subjects.

3.2.2. *Discussion*

That F_1 should start higher in a vowel that follows an aspirated stop than after a corresponding unaspirated stop (or click) is predicted from a number of interrelated factors. Firstly, voiceless aspiration causes masking of most or all of the rising F_1 transition from the low F_1 target of the stop (due to oral constriction) to the higher F_1 target of the following vowel (most clearly in /a/). By the time aspiration ends, F_1 is therefore relatively high. Secondly, F_1 is raised due to the effect of the trachea as a resonator. Tracheal coupling results from the glottal opening gesture in aspirated stops and is still detectable at vowel onset. Tracheal influence is also the reason for the broadening of F_1 bandwidth, which results in a reduction of the amplitude of the first formant. This is a factor that enhances the overall masking effect of aspiration on formant transitions (see Stevens, 1998, Section 8.3). The absence of a rising F_1 transition after aspirated stops and clicks shown in Fig. 2 is entirely consistent with these expectations about the influence of aspiration on F_1 in the following vowel.

Relatively low F_1 onset after voiced stops (and clicks) can be interpreted as the result of progressive coarticulation of the cavity-enlarging maneuvers that are necessary for maintaining voicing during closure (Fischer-Jørgensen, 1968, p. 92; cf. Stevens, Blumstein, Glicksman, Burton & Kurowski, 1992 for fricatives). Tongue root advancement and larynx lowering, which are among the most effective means of oral cavity enlargement (Westbury, 1983), both have a lowering effect on F_1. As mentioned in Section 1.1, closure voicing in Xhosa voiced stops and clicks is usually not maintained. However, as will be addressed in the general discussion, there is evidence that larynx lowering has achieved a high degree of importance in Xhosa.

Compared to the f_0 pattern after voiced stops and clicks, the F_1 pattern after these sounds shown in Fig. 2 really does seem like a case of coarticulation. Here, the lowest values are found at vowel onset and there is a gradual increase towards the center of the vowel. This pattern — together with the fact that there is only little distance to the F_1 values after plain stops/clicks and that the difference between all three categories is nearly equalized towards the center of the vowel — suggests that F_1 after voiced stops and clicks is not actively controlled in the implementation of the voicing feature in Xhosa.

3.3. H_1^*–H_2^*

3.3.1. *Results*

As discussed in Section 1.3, H_1^*–H_2^* is a parameter that allows us to investigate voice quality differences induced by the different stop and click categories in Xhosa, which is the main research interest of this paper. The results from this parameter are now fully described, including speaker-by-speaker illustration of means and standard deviations and a full report of statistical analysis. This degree of explicitness is justified not only by the main topic of this study but also by the striking differences between speakers in

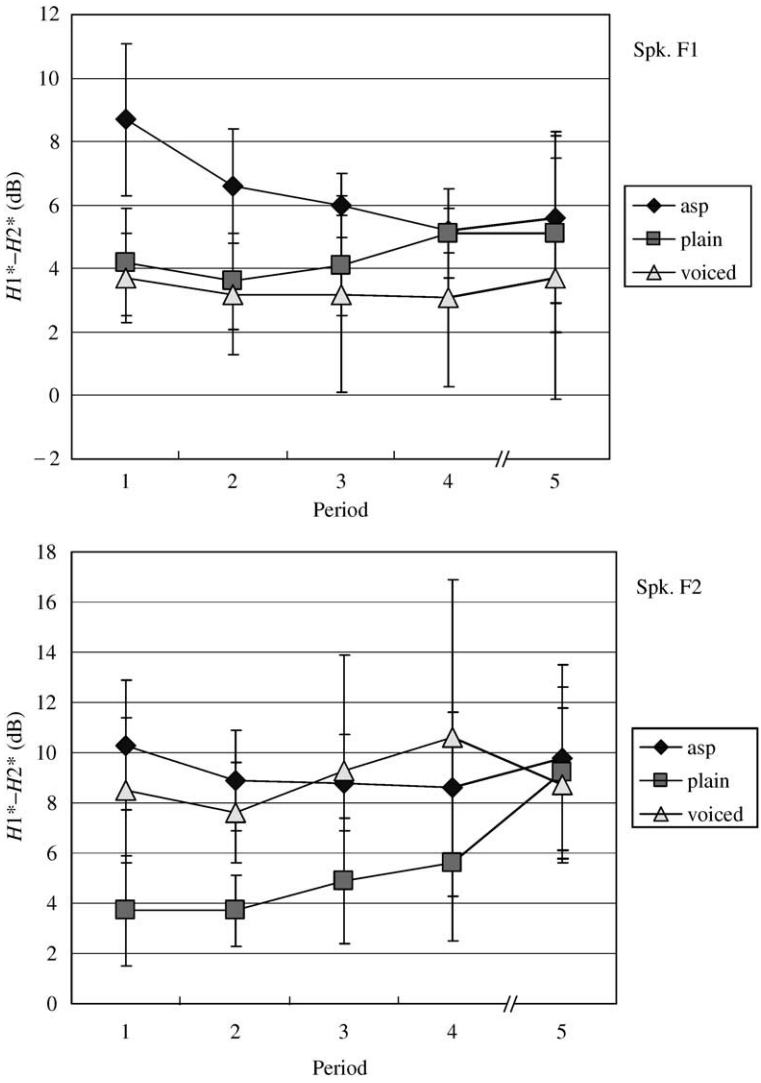

Figure 3. Results for H_1*–H_2*. Means and standard deviations in dB separately for each speaker. Individual graphs for speakers (in this order) F1, F2, M1, M2, F3, F4, M3, M4 (the latter four in postvocalic position). Continued on pp 21–23.

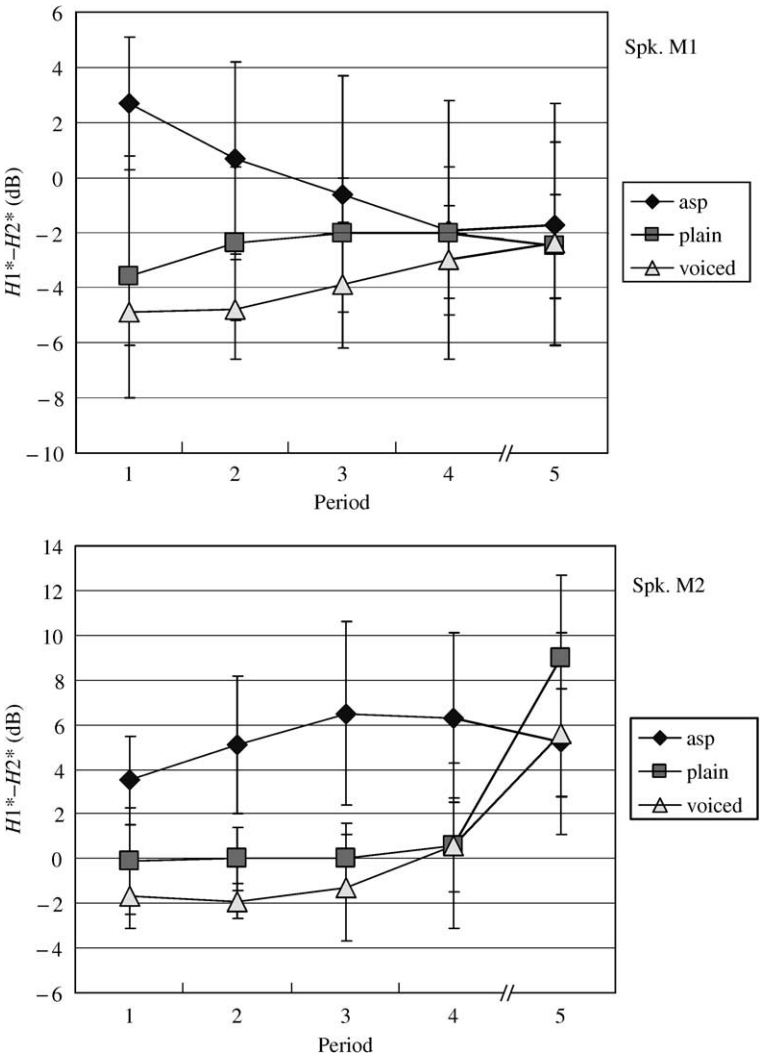

Figure 3. Continued.

the way aspirated, voiced, and plain stops and clicks turn out to affect the following vowel.

Three- and four-way ANOVAs of the type mentioned in Section 2.3 show a pattern quite similar to the one found for the parameter f_0. With very few exceptions only, all of them occurring at p5, Category, Speaker, and the interaction between these factors are significant at the level $p < 0.001$ (or in one case at $p = 0.001$). The remaining significant effects, that involve other factors, are few in number and have p values that are only slightly below 0.05 (Table AIII).

Means and standard deviations for H_1^*–H_2^* in postvocalic context are illustrated in Fig. 3. As before, data are pooled across the stop/click distinction as well as different places of articulation, but this time a separate display is provided for each speaker. Other conventions remain as in Fig. 1.

The patterns shown in Fig. 3 can be divided roughly into two different groups. In the first group, manifested by speakers F1, M1, and M2, H_1^*–H_2^* is lower after the voiced category compared to the aspirated and plain category. Hence, there are no indications in the speech of these subjects that voiced stops and clicks induce breathy voice, if the voice quality effects of the other categories are taken as the basis of comparison. This is different in the second group of speakers, manifested by the remaining five speakers, where H_1^*–H_2^* after voiced stops and clicks is comparatively high. The levels of H_1^*–H_2^* can either be similar to those found after aspirated stops/clicks (speaker F2) or H_1^*–H_2^* can be predominantly the highest among all phonological categories (F3, F4, M3, M4). In postnasal position (listed in Table BI, Appendix B), these latter speakers all show higher values after voiced than plain stops/clicks (with the exception of p5 by M4). The results for these five speakers indicate that there is in fact some amount of breathy

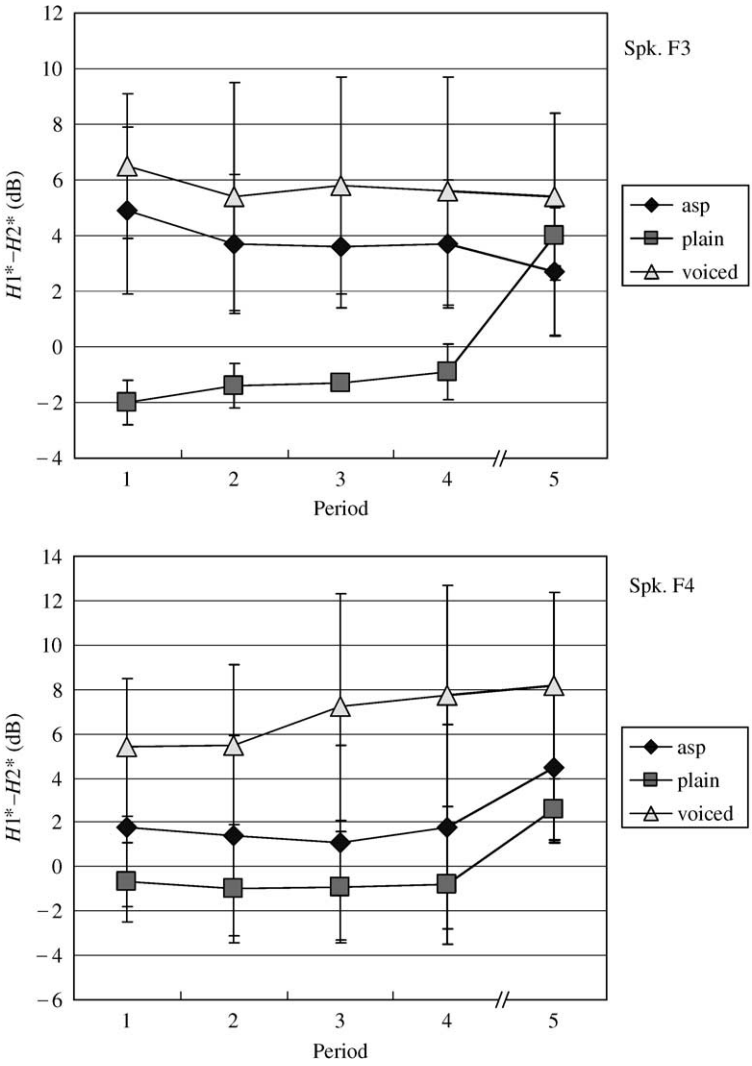

Figure 3. Continued.

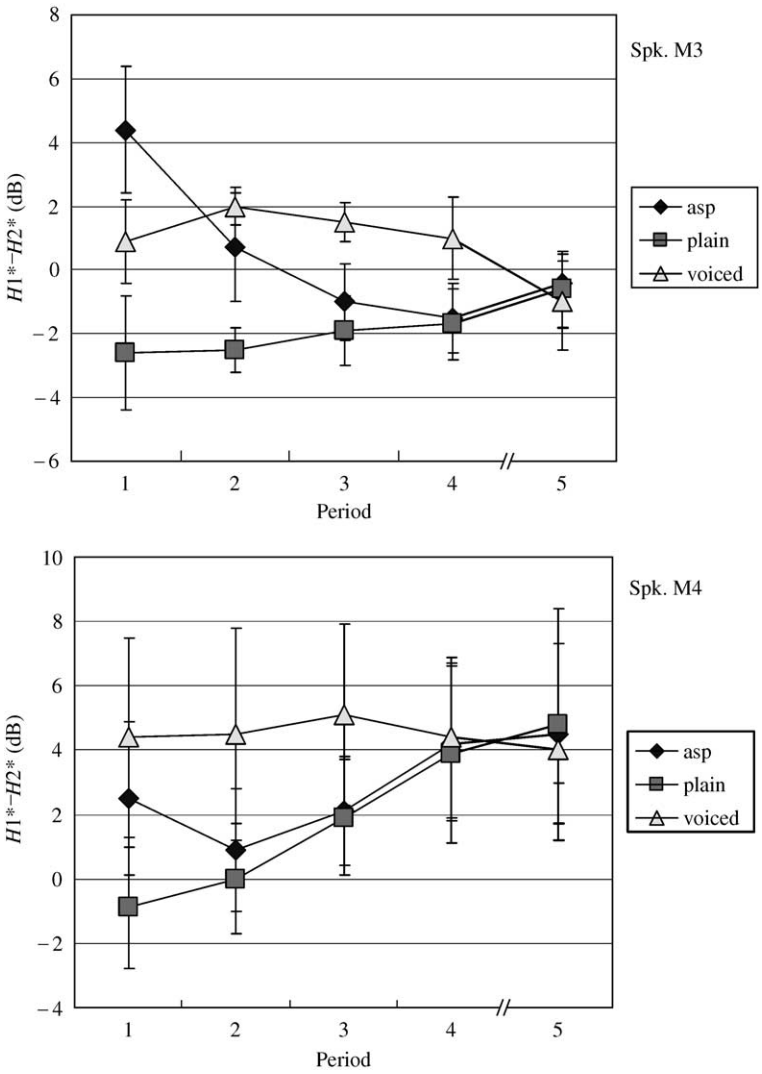

Figure 3. Continued.

phonation after the voiced stops and clicks of Xhosa. A different way of visualizing speaker differences is presented in Fig. 4, based on a subset of the data shown in Fig. 3 (thanks are due to one of the reviewers for proposing this layout).

Focusing on the third period, Fig. 4 shows that all speakers have higher H_1^*–H_2^* after aspirated than plain stops and clicks. The differences occur with the shift from the aspirated to the voiced category. For three speakers, there is a downward shift from aspirated to voiced (the first group of speakers, mentioned above), while for the other five there is an upward shift (the second group). A similar pattern would occur with a focus on periods 1, 2, or 4. Fig. 4 is also an informative way of visualizing statistical interactions between Category and Speaker, that were reported at the beginning of this section.

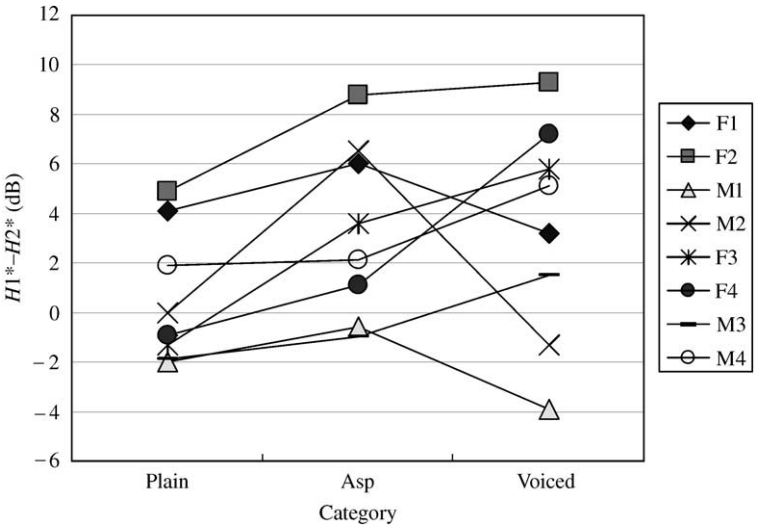

Figure 4. Means of H_1*–H_2* (in dB on the *y*-axis) for period 3, differentiated according to speaker. Stop and click category (aspirated, plain, voiced) on the *x*-axis for each of the eight speakers.

TABLE IV. Statistical results for the parameter H_1*–H_2*. Subjects and contexts (where more than one) are presented in the top row, and the number of the period in the first column. The first display presents the results for List 1, the second one for the postvocalic stimuli of List 2, and the third one for the postnasal stimuli of List 2. For the first two displays, where three phonological categories are compared, the *F* values and *p* values of one-way ANOVAs with Category as the factor are presented first in each cell, followed by the results of *post hoc* tests, where significant at $p < 0.05$ (Tukey HSD multiple comparisons). If significant at $p < 0.01$ the relevant *post hoc* comparison is presented in boldface. (All significant effects are discussed in the text, no matter whether at $p < 0.05$ or at $p < 0.01$.) The directions of the effects are presented for convenience (they can also be seen in Fig. 3), with the abbreviation "a" for "aspirated", "p" for "plain", and "v" for "voiced". For the third display, where only a two-way comparison is made, only the *F* and *p* values are presented (with the exception of period 5 by M4 the direction is always $p <$ v in postnasal position)

	F1	F2	M1	M2
1	$F(2, 22) = 17.924$; $p < 0.001$ **a > p; a > v**	$F(2, 26) = 17.617$; $p < 0.001$ **a > p; p < v**	$F(2, 15) = 11.184$; $p = 0.001$ **a > p; a > v**	$F(2, 14) = 10.841$; $p = 0.001$ a > p; **a > v**
2	$F(2, 23) = 8.975$; $p = 0.001$ a > p; **a > v**	$F(2, 25) = 21.847$; $p < 0.001$ **a > p; p < v**	$F(2, 15) = 5.804$; $p = 0.014$ a > v	$F(2, 15) = 19.157$; $p < 0.001$ **a > p; a > v**
3	$F(2, 23) = 3.875$; $p = 0.035$ a > v	$F(2, 26) = 5.655$; $p = 0.009$ a > p; p < v	$F(2, 15) = 1.662$; $p = 0.223$	$F(2, 15) = 12.220$; $p = 0.001$ **a > p; a > v**
4	$F(2, 23) = 3.243$; $p = 0.057$	$F(2, 26) = 3.148$; $p = 0.060$	$F(2, 15) = 0.226$; $p = 0.801$	$F(2, 15) = 5.835$; $p = 0.013$ a > p; a > v
5	$F(2, 23) = 0.855$; $p = 0.438$	$F(2, 26) = 0.222$; $p = 0.802$	$F(2, 15) = 0.088$; $p = 0.916$	$F(2, 15) = 1.996$; $p = 0.170$

TABLE IV. Continued.

	F3 postvocalic	F4 postvocalic	M3 postvocalic	M4 postvocalic
1	$F(2, 14) = 20.677$; $p < 0.001$ **a > p; p < v**	$F(2, 15) = 6.324$; $p = 0.010$ **p < v**	$F(2, 15) = 24.959$; $p < 0.001$ **a > p; a > v; p < v**	$F(2, 15) = 7.007$; $p = 0.007$ **p < v**
2	$F(2, 14) = 9.442$; $p = 0.003$ a > p; **p < v**	$F(2, 15) = 5.110$; $p = 0.020$ p < v	$F(2, 15) = 25.284$; $p < 0.001$ **a > p; p < v**	$F(2, 15) = 5.772$; $p = 0.014$ p < v
3	$F(2, 14) = 11.638$; $p = 0.001$ a > p; **p < v**	$F(2, 15) = 6.066$; $p = 0.020$ p < v	$F(2, 15) = 16.859$; $p < 0.001$ **a < v; p < v**	$F(2, 15) = 4.061$; $p = 0.039$
4	$F(2, 14) = 9.118$; $p = 0.003$ a > p; **p < v**	$F(2, 15) = 6.305$; $p = 0.010$ **p < v**	$F(2, 15) = 9.892$; $p = 0.002$ **a < v; p < v**	$F(2, 14) = 0.064$; $p = 0.938$
5	$F(2, 14) = 1.917$; $p = 0.184$	$F(2, 15) = 4.633$; $p = 0.027$ p < v	$F(2, 15) = 0.390$; $p = 0.684$	$F(2, 15) = 0.128$; $p = 0.881$
	F3 postnasal	**F4 postnasal**	**M3 postnasal**	**M4 postnasal**
1	$F(1, 11) = 20.889$; $p = 0.001$	$F(1, 11) = 29.186$; $p < 0.001$	$F(1, 12) = 21.172$; $p = 0.001$	$F(1, 12) = 12.381$; $p = 0.007$
2	$F(1, 11) = 18.442$; $p = 0.001$	$F(1, 12) = 54.562$; $p < 0.001$	$F(1, 12) = 15.097$; $p = 0.002$	$F(1, 12) = 14.068$; $p = 0.003$
3	$F(1, 11) = 19.889$; $p = 0.001$	$F(1, 12) = 56.759$; $p < 0.001$	$F(1, 12) = 1.549$; $p = 0.237$	$F(1, 12) = 18.076$; $p = 0.001$
4	$F(1, 11) = 18.942$; $p = 0.001$	$F(1, 12) = 48.797$; $p < 0.001$	$F(1, 12) = 5.724$; $p = 0.034$	$F(1, 12) = 0.885$; $p = 0.365$
5	$F(1, 11) = 3.269$; $p = 0.098$	$F(1, 11) = 31.857$; $p < 0.001$	$F(1, 12) = 0.485$; $p = 0.499$	$F(1, 12) = 0.483$; $p = 0.500$

We proceed with the results of one-way ANOVAs and *post hocs* for H_1^*–H_2^*. A full statistical report is presented in Table IV.

For the first group of speakers, H_1^*–H_2^* is significantly higher after the aspirated set compared to both other phonological categories for p1 and p2 in the speech of F1, for p1 in the speech of M1, and for p1–p4 by M2. For p3 by F1 and p2 by M1 only the difference between the aspirated and voiced categories was significant. As far as the second group is concerned the following statistical patterns were obtained. For speaker F2 H_1^*–H_2^* after plain stops/clicks was significantly lower than after both aspirated and voiced ones for p1–p3, while no significant differences were found among the latter categories. The same statistical pattern occurred for p1–p4 by F3. For F4 only the comparison plain (lower) *vs.* voiced (higher) was significant, but for all five periods. For M4 this was the case only for p1–p2, with no significant effects for the other periods. For M3 all comparisons were significant in the first period; at p2 the aspirated and voiced categories were significantly higher than the plain set, while at p3 and p4 the voiced set was significantly higher than both the plain and aspirated counterparts. As far as the postnasal cases are concerned all

comparisons were significant (H_1^*–H_2^* after voiced higher than after plain), except for p5 by F3, p3 and p5 by M3, and p4 and p5 by M4.

So far in this section, the focus has been on the different levels of H_1^*–H_2^* after the three categories. As far as the period-by-period dynamics are concerned, some observations can be made. Firstly, a decline in H_1^*–H_2^* up to p3/p4 can be seen after the aspirated stops/clicks of most speakers, most strikingly with F1, M1, and M3. Secondly, for most speakers, H_1^*–H_2^* converges to relatively high values towards the center of the vowel (p5); for the plain category there is a rise from p1 to p5 with all speakers. Thirdly, by the time the center of the vowel is reached (p5) most of the H_1^*–H_2^* differences between different stop and click categories are strongly reduced or eliminated. It is only for F4 where elevated H_1^*–H_2^* after voiced stops/clicks remains striking even at the vowel center.

3.3.2. *Discussion*

It has been demonstrated, most transparently in Fig. 4, that the relation between plain and aspirated stops/clicks is uniform across speakers, whereas on the other hand, there are striking speaker differences with respect to the H_1^*–H_2^* values associated with the voiced category.

Turning to the former observation first, the results are consistent with the expectation that H_1^*–H_2^* begins higher in a vowel that follows an aspirated than a plain stop (and click), for the following reason. Voiceless aspirated stops are produced with a large glottal opening gesture, whose maximum is roughly coordinated with stop release (Löfqvist & Yoshioka, 1980). Usually, the adduction phase of this glottal opening gesture is not complete at the voicing onset of the following vowel. For that reason, it is expected that there is an interval during which voicing associated with the vowel occurs together with transglottal turbulent airflow associated with the yet incompletely closed glottis. In airflow studies, this effect is manifested as a gradual decrease in open quotient from voicing onset after an aspirated stop into the following vowel (Löfqvist & McGowan, 1992). Since the most direct acoustic correlate of open quotient is (normalized or raw) H_1–H_2 (Klatt & Klatt, 1990; Stevens *et al.*, 1995), this acoustic parameter should start relatively high after aspirated stops as well and then reduce later in the vowel. Relatively high H_1–H_2 at vowel onset after aspirated stops is confirmed by Chapin Ringo (1988) for English as well as by Ní Chasaide & Gobl (1993) and Jessen (1998) for German. The increased H_1–H_2 values after aspirated stops are indicative of a type of breathy voice quality that also occurs in voiced intervocalic /h/ in English (Stevens, 1998).

These considerations are entirely consistent with the high start and the decline in H_1^*–H_2^* after aspirated stops/clicks that could be observed in Fig. 3, most clearly for speakers F1, M1, and M3. However, for speaker M2, H_1^*–H_2^* actually increases from vowel onset. So it seems that another influence is at work as well. It was mentioned in Section 2.1 that long vowels in Xhosa often become more breathy during their second half. Notice that in the speech of M2 H_1^*–H_2^* at p5, which is located around the midpoint of the vowel, is relatively high independent of the influence of the preceding stop or click. It is therefore conceivable that the increase of H_1^*–H_2^* during the early part of the vowel is due to the influence of this voice quality contour often observed in Xhosa. A similar situation occurs in the speech of M4. As with M2, there is an increase of H_1^*–H_2^* after aspirated stops/clicks early in the vowel and a high value at vowel midpoint, which is independent of phonological category. Interestingly, H_1^*–H_2^* first decreases from period 1 to 2 and then starts to increase in the speech of M4. This initial decrease is probably due

to the coarticulatory influence of aspiration mentioned above. It seems that coarticulatory breathy voice induced by a preceding aspirated stop or click competes with the prosodic voice quality contour of the entire vowel, which is independent of the phonological status of the preceding stop and click. Note again that for the plain stops/clicks, H_1^*–H_2^* increases from p1 to p5 for all speakers, which seems to reflect the influence of the overall consonant-independent voice quality contour of the vowel. The fact that for some speakers (in particular F2, M2, F3, M4) H_1^*–H_2^* becomes comparatively high towards the center of the vowel after all stop and click categories does not detract from our findings about the consonant-induced voiced quality effects. The fact remains that the vowels after all three stop and click categories were investigated under identical conditions and that according to our results, there are five speakers who show levels of H_1^*–H_2^* with the first four pairs of vowel periods after voiced stops and clicks that are relatively high when compared to the H_1^*–H_2^* levels measured during the same temporal interval after the other categories, in particular the plain.

Turning now to the voice quality effects of the voiced stops and clicks, we have found that for some speakers H_1^*–H_2^* is relatively low, while a second set of speakers (F2, F3, F4, M3, M4; cf. Fig. 4) shows values as high as and often higher than those found after aspirated stops/clicks. If it is assumed — based on the preceding discussion — that aspirated stops are followed by breathy voice and that this breathy voice quality is manifested with relatively high H_1^*–H_2^*, the (predominantly) equally high or even higher H_1^*–H_2^* values after the voiced stops/clicks of the second set of speakers should result from breathy phonation as well. It is likely that these speakers produced (at least part of) the vowels after voiced stops/clicks with some amount of glottal leakage, which would cause high values of open quotient and ultimately H_1^*–H_2^*. If this reasoning is correct, it is still possible that the physiological cause of glottal leakage is different after the voiceless aspirated stops of Xhosa than after the voiced set. It can safely be assumed that breathy voice after aspirated stops/clicks results from the final stages of an active glottal opening gesture, caused by activation of the posterior cricoarytenoid muscle (Löfqvist & Yoshioka, 1980). The interpretation of glottal leakage after voiced stops, found in the second group of Xhosa speakers mentioned earlier, is more difficult. Probably, it results from overall vocal fold slackness, which is related to larynx lowering and tonal depression. This important issue is addressed further in the general discussion.

The fact that any existing voice quality differences between the different stop and click categories are usually strongly reduced or eliminated by the time the center of the vowel is reached supports the general claim of Rycroft (1980) that breathy voice associated with voiced stops and clicks is only found during the first and not during the second half of the following vowel. (More precisely, we need to consider the relatively high breathy voice levels after voiced compared to other stops and clicks during the early part of the vowel; in absolute terms, the same elevated levels of H_1^*–H_2^* are found for some speakers after all three categories at vowel center, and therefore it would be incorrect to say that there is no breathy voice during the second half of the vowel.) However, his more specific claim that with lexical low tone the entire vowel is affected cannot be supported. It also needs to be pointed out again that there is speaker specificity with respect to the occurrence of breathy voice after voiced stops and clicks. This challenges Rycroft's claims, as he seems to regard breathy voice as an obligatory component of voiced stops and clicks in Xhosa.

One final point needs to be addressed in this discussion. We mentioned in Section 1.1 that according to the literature and also according to our own observations on the present material the plain clicks and especially the plain stops are sometimes produced

with ejection. However, we did not find any systematic indications (auditorily or in waveform/spectrogram) of creaky voice after stops or clicks produced with ejection. Creaky voice is expected to show relatively low H_1^*–H_2^* (or H_1–A_1; Ladefoged *et al.*, 1988). The possibility remains that there are weaker kinds of creakiness, of the type referred to by Ladefoged and Maddieson (1996) as "stiff voice", which might occur after ejected stops/clicks but which we had not been able to detect. Given that possibility, it could be the case that for some speakers, the H_1^*–H_2^* values found after the plain category are slightly lower than the values that are expected from cases that are unequivocally produced with modal voice quality. We doubt, however, that such a possible effect would be strong or systematic enough to invalidate any of the conclusions that we make about the crucial issue of voice quality associated with the voiced stops and clicks in Xhosa.[3] The more general issue that lies behind this point is that with the H_1-based methods used here voice quality can only be evaluated relatively. For that reason, evaluation of voice quality after voiced stops/clicks needs a baseline from the other categories, which should be as predictable and well-understood as possible. We took the values after aspirated stops and clicks as that baseline.

3.4. H_1^*–A_3^*

3.4.1. *Results*

H_1^*–A_3^* is another voice quality parameter that can reflect the occurrence of breathy voice (Hanson, 1997). Many of the results for the parameter H_1^*–A_3^* are similar to those for H_1^*–H_2^*, though there are differences worth mentioning. Statistical evaluation will be briefer than for H_1^*–H_2^*. Specifically, we leave out the table with the results of speaker-by-speaker one-way ANOVAs and *post hocs*. As far as the results for three- and four-way ANOVAs are concerned (Table AIV), they present a picture similar to H_1^*–H_2^*, but with often lower levels of significance for Category and the Category × Speaker interaction (plus two more non-significant effects at p5) and a higher number of significant effects involving other factors. Speaker-by-speaker descriptive results are presented in Fig. 5 in the format similar to the documentation of H_1^*–H_2^*.

Recall from the results of H_1^*–H_2^* that in what was referred to as the first group of speakers values after aspirated stops/clicks were substantially higher than after either voiced or plain stops/clicks, at least for some of the early periods. From Fig. 5, we can see that those differences are reduced with H_1^*–A_3^* (especially in the speech of F1) and that the decline after aspirated stops/clicks is less striking (especially with F1 and M1). We also observe that for speaker M1, who was part of this group, the level for the voiced category is elevated relative to the situation for the parameter H_1^*–H_2^*, such that with respect to H_1^*–A_3^*, he now belongs more to the second group, where the levels for the aspirated and voiced category are both relatively high. Consistent with the measurements for H_1^*–H_2^*, H_1^*–A_3^* usually remains highest after the voiced category for speakers F3 and F4, though the differences from the other categories are reduced (most clearly with F4). This also holds for p3 and p4 by M3. However, inconsistently, the values for the

[3]Ejective productions are most salient for speakers M2 (stops and clicks) and F4 (only stops). But according to Fig. 3 H_1^*–H_2^* is not relatively lower in plain compared to other stops/clicks for these speakers than for other speakers. For M2, who had the strongest signs of ejection, the plain stops/clicks are not even the category with the lowest values of H_1^*–H_2^*. These observations further substantiate our impression that ejective productions have no systematic effect on the following vowel in the form of stiff or creaky voice quality in Xhosa. Recall from footnote 1 that the few tokens with strong creaky voice (whether from ejectives or other sounds) were excluded from analysis.

voiced category of M4 fall slightly behind those for the aspirated category. In postnasal position H_1^*–A_3^* is mostly higher after voiced than plain stops/clicks, though again with a reduced distance between the two relative to the results for H_1^*–H_2^*.

Turning to the statistical significance of the differences we find that for speaker F1 only the difference between aspirated (higher) and plain (lower) stops/clicks is significant, and only for the first period. For F2 all three comparisons are significant for p1. For p2 H_1^*–A_3^* after plain stops/clicks is significantly lower than after either the aspirated or voiced set. For p3 and p5 the aspirated set is significantly higher than the plain set, and for p4 the voiced set is significantly higher than the plain. For M1 the plain stops/clicks have significantly lower values than either the aspirated or voiced ones at p2, while no significant effects are found for the other periods. For p1–p4 of speaker F3 we again find

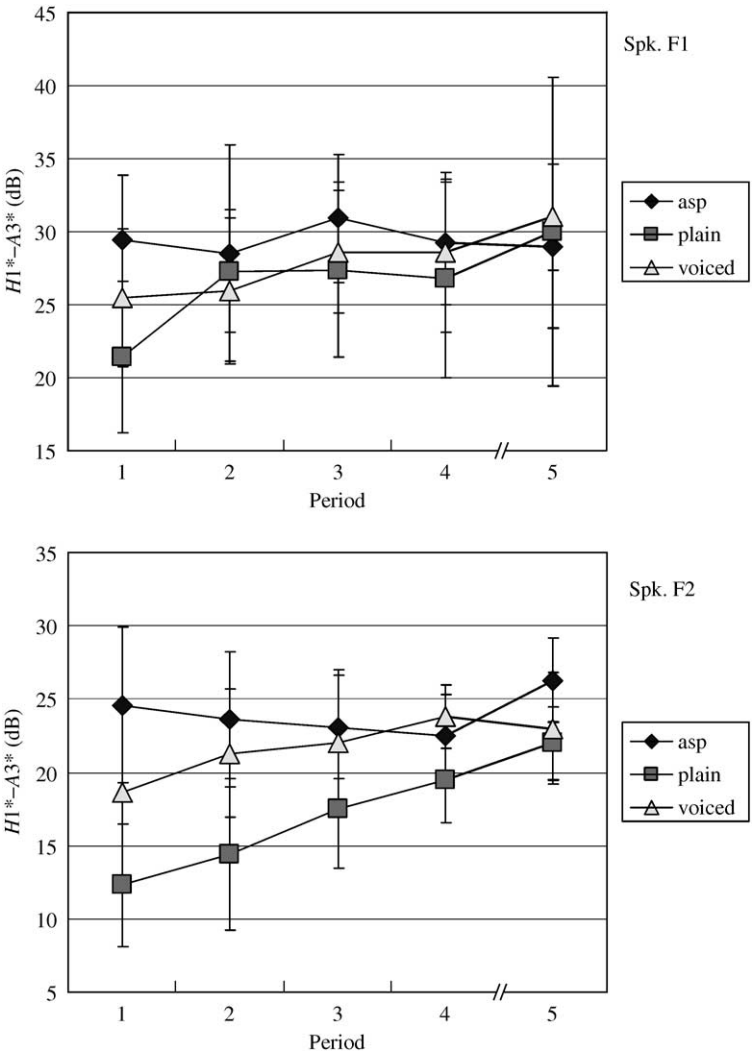

Figure 5. Results for H_1*–A_3*. Means and standard deviations in dB, with individual graphs for each speaker. Continued on pages 30–32.

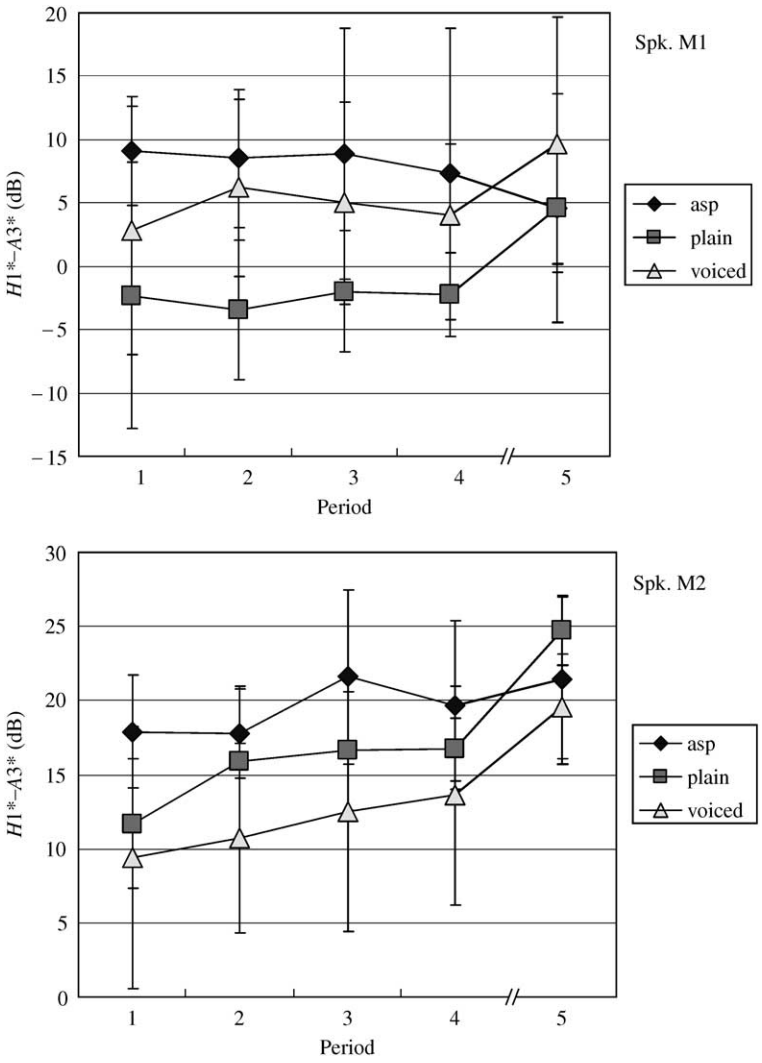

Figure 5. Continued.

the pattern that plain stops and clicks show significantly lower values than either voiced or aspirated ones. For p1 and p2 in the speech of M3 the aspirated stop/clicks have significantly higher values than the voiced set. In postnasal position, significantly higher values were obtained for the voiced than for the plain set for p1–p4 in the speech of F3, but for the other speakers no significant effects were found in that position.

3.4.2. *Discussion*

Results for the parameter H_1^*–A_3^* have shown a predominance of higher values after aspirated and voiced stops/clicks than after plain ones, although there were fewer significant effects than for H_1^*–H_2^*. The division into those speakers with relatively low values after voiced stops/clicks on the one hand, and those where the voiced category

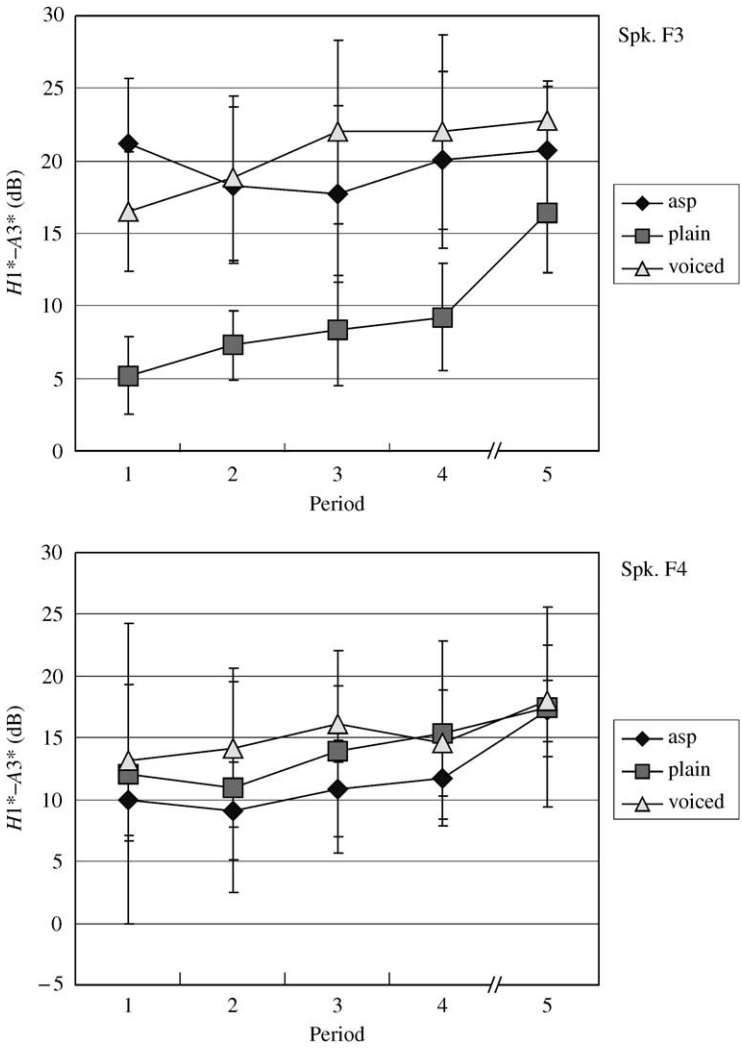

Figure 5. Continued.

induces levels comparable to aspirated stops/clicks on the other, has shifted from three *vs.* five respectively with H_1^*–H_2^* to two *vs.* six, or perhaps even one *vs.* seven, with H_1^*–A_3^* (only M2 remains as a clear case of a group-one speaker, though at a non-significant level). That voiced stops/clicks actually induce higher values than aspirated ones in some speakers is much less common and statistically robust for H_1^*–A_3^* than for H_1^*–H_2^*.

The results for H_1^*–A_3^* are consistent with the claim in (some of) the literature that voiced stops and clicks are followed by some amount of breathy voice quality. This follows from the observation that levels of H_1^*–A_3^* after voiced stops/clicks are often similar to those after the aspirated set and higher than after the plain set, combined with the expectation that aspirated stops/clicks are also followed by breathy voice. However, values after aspirated stops/clicks are often slightly and in a few cases even significantly

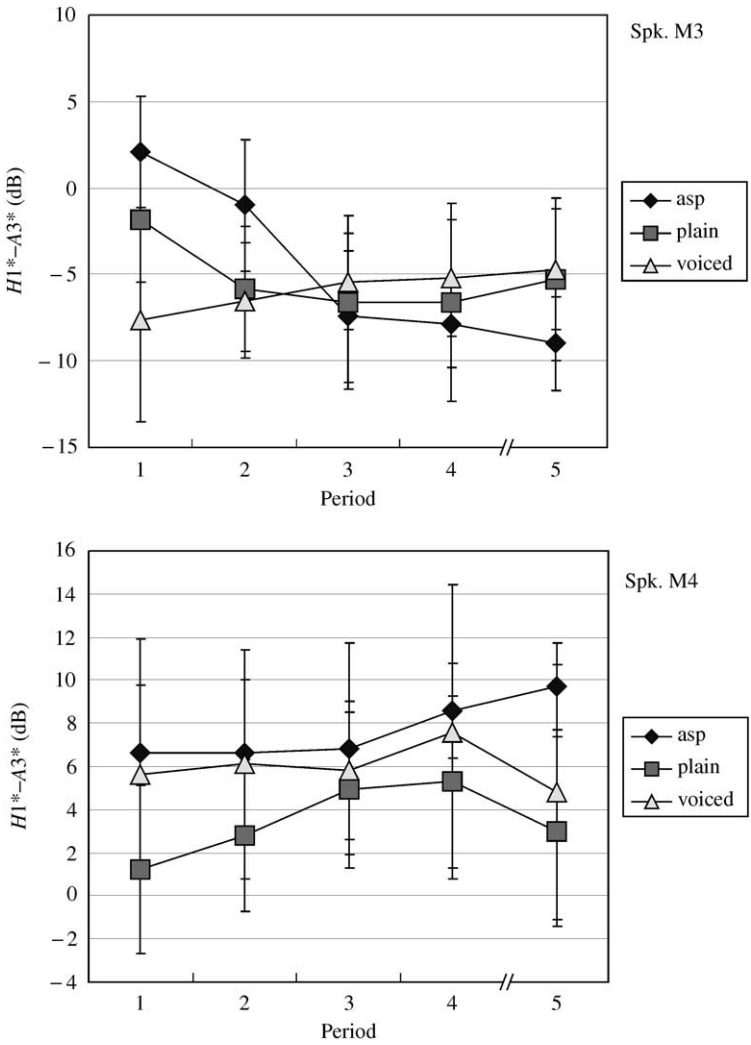

Figure 5. Continued.

higher than after the voiced set. So the level of breathiness after the latter must be relatively weak, when interpreted on the basis of the H_1^*–A_3^* results. In comparison to the results for H_1^*–H_2^*, the evidence from the parameter H_1^*–A_3^* shows on the one hand, more uniformity across speakers, but on the other, less statistical robustness.

3.5. *Data on (post)nasal clicks and sonorants*

3.5.1. *Nasal clicks and clicks in postnasal position*

Fig. 6 provides the results for nasal clicks in comparison to plain and voiced clicks that occur in postnasal position. This distinction occurs in the speech of F3, F4, M3, and M4. The means and standard deviations of the data pooled across periods 2 and 3 and across the three different places of articulation for clicks (dental, alveolar, lateral) are represented.

Postnasal voiced clicks induce lower f_0 in the following vowel than nasal clicks and postnasal plain clicks. f_0 depression is therefore available as a cue to the distinction between nasal clicks and postnasal voiced clicks, the latter of which exhibit tonal depression while the former do not (cf. Ladefoged & Traill, 1994; Ladefoged & Maddieson, 1996 on Xhosa click accompaniments). As mentioned in Section 1.1, it is important to have such distinguishing cues since like the nasal clicks, the postnasal voiced clicks most often have no closure (in the sense of occlusion(s) in the oral cavity with raised velum). With respect to F_1, nasal clicks pattern closely with voiced clicks, so no stable cue to this distinction is available from this parameter. The behavior of nasal

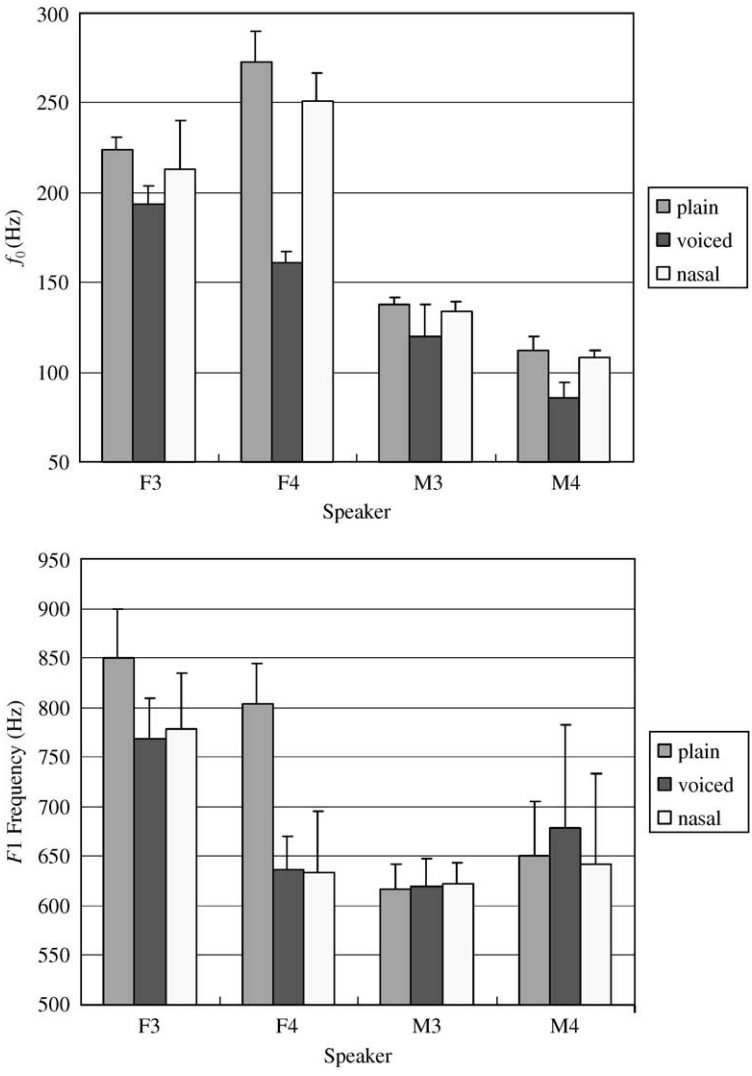

Figure 6. Means and standard deviations for (post)nasal clicks spoken by F3, F4, M3, M4 (*x*-axis), pooled across periods 2 and 3. For each speaker, the three columns represent plain postnasal clicks, voiced postnasal clicks, and nasal clicks. Separate graphs for (in this order) f_0 in Hz, F_1 in Hz, H_1*–H_2* in dB, and H_1*–A_3* in dB.

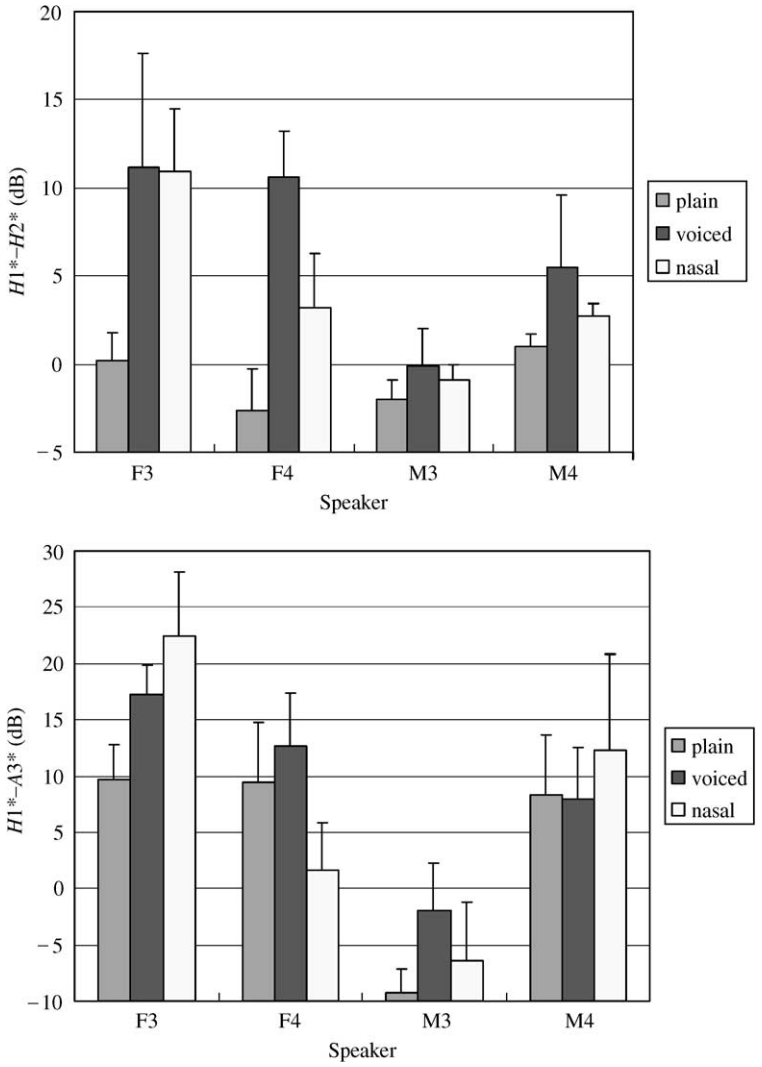

Figure 6. Continued.

clicks compared to the plain and voiced postnasal clicks is rather heterogeneous with respect to H_1^*–H_2^*. For speaker F3, nasal clicks have values that are almost as high as those after voiced clicks, which suggests the presence of breathy voice in both cases. For F4 and M4, H_1^*–H_2^* is clearly lower after nasal than voiced clicks, though still higher than after plain clicks (for M3 differences are overall small). That particular pattern would be consistent with the view that postnasal voiced clicks have breathy voice while nasal clicks do not. However, overall voice quality seems of only limited value for the distinction between nasal clicks and voiced postnasal clicks, and the tonal patterns seem to carry most of the distinctive load. A generally similar situation is found for H_1^*–A_3^*, except that for F3 and M4 nasal clicks emerge with the highest values of all. That would go against the expectation that only postnasal voiced clicks but not nasal clicks are followed by breathy voice and underlines the need for tonal depression to differentiate the former from the latter.

These rather indifferent results for H_1^*–H_2^* and H_1^*–A_3^* also raise doubts about the claims made by Lanham (1960), Ziervogel (1967), and Riordan *et al.* (1969) that breathy voice is only found with the postnasal voiced clicks of Xhosa. If that was true for the speech of our subjects, we would have found a more consistent and stronger effect showing higher values of H_1^*–H_2^* and H_1^*–A_3^* after postnasal voiced clicks than after nasal clicks. We also tested this claim by comparing voiced clicks in postvocalic position with those in postnasal position in the speech of F3, F4, M3, and M4. One-way ANOVAs (pooled across p2 and p3) showed that there was usually no significant difference in terms of H_1^*–H_2^* or H_1^*–A_3^* between voiced clicks after vowels compared to those after nasals. Where a significant difference occurred (in terms of H_1^*–A_3^* for F3 and in terms of H_1^*–H_2^* for M3) it was after the vowel, not after the nasal, where higher values of the two voice quality parameters were obtained. It remains entirely possible, however, that breathy voice, possibly including aspiration turbulence, exists in the postnasal voiced clicks written *nch*, etc., as produced by traditional Xhosa speakers of the older generation. It would be an interesting task for future research to carry out phonetic fieldwork in the rural areas of the Transkei to settle this issue.

3.5.2. *Sonorants*

Table V shows the results for the words *inyheke* "upper hare lip" and *inyele* "edge". According to Lanham (1960), these words contain palatal nasals produced with and without breathy voice, respectively. These words were produced once by the four

TABLE V. Results for *inyele* (left in each cell) *vs. inyheke* (right in each cell). Separate displays for different speakers, in each of them acoustic parameter in rows and results for the five different periods in columns

	p1	p2	p3	p4	p5
Speaker F3					
f_0 (Hz)	240 *vs.* 222	240 *vs.* 222	246 *vs.* 215	246 *vs.* 206	235 *vs.* 186
F_1 (Hz)	472 *vs.* 288	472 *vs.* 340	472 *vs.* 334	472 *vs.* 281	583 *vs.* 308
H_1–H_2 (dB)	−1.1 *vs.* 7.1	−2.6 *vs.* 5.9	−3.6 *vs.* 6.9	−4.0 *vs.* 8.7	−1.8 *vs.* 12.7
H_1–A_3 (dB)	18.3 *vs.* 34.5	17.8 *vs.* 28.9	15.5 *vs.* 28.1	13.5 *vs.* 28.5	10.4 *vs.* 23.6
Speaker F4					
f_0 (Hz)	281 *vs.* 176	281 *vs.* 169	281 *vs.* 168	277 *vs.* 168	232 *vs.* 162
F_1 (Hz)	531 *vs.* 511	550 *vs.* 511	550 *vs.* 511	550 *vs.* 511	681 *vs.* 531
H_1–H_2 (dB)	−0.2 *vs.* 10.0	−2.2 *vs.* 11.5	−4.4 *vs.* 9.0	−7.7 *vs.* 11.9	−1.2 *vs.* 9.0
H_1–A_3 (dB)	16.4 *vs.* 18.2	15.8 *vs.* 18.8	15.2 *vs.* 24.3	15.7 *vs.* 25.2	12.7 *vs.* 21.8
Speaker M3					
f_0 (Hz)	138 *vs.* 152	145 *vs.* 140	138 *vs.* 134	138 *vs.* 130	135 *vs.* 118
F_1 (Hz)	373 *vs.* 462	445 *vs.* 413	439 *vs.* 400	432 *vs.* 393	504 *vs.* 459
H_1–H_2 (dB)	−9.5 *vs.* −10.0	−8.5 *vs.* −7.3	−7.6 *vs.* −5.6	−7.5 *vs.* −2.5	−4.5 *vs.* 0.3
H_1–A_3 (dB)	12.3 *vs.* 9.5	6.6 *vs.* 7.1	4.5 *vs.* 4.6	5.6 *vs.* 4.1	3.1 *vs.* −0.4
Speaker M4					
f_0 (Hz)	112 *vs.* 92	106 *vs.* 88	104 *vs.* 83	94 *vs.* 83	86 *vs.* 80
F_1 (Hz)	249 *vs.* 275	249 *vs.* 550	321 *vs.* 583	511 *vs.* 537	531 *vs.* 577
H_1–H_2 (dB)	0.4 *vs.* 5.3	0.3 *vs.* 4.0	0.2 *vs.* 5.1	1.1 *vs.* 3.1	−0.2 *vs.* 1.5
H_1–A_3 (dB)	21.6 *vs.* 14.5	18.2 *vs.* 11.8	11.9 *vs.* 8.0	15.9 *vs.* 7.3	10.7 *vs.* 3.6

speakers that read List 2. The normalization procedures were not always successful, due to the low F_1 position of /e/. We therefore present the results as unnormalized H_1–H_2 and H_1–A_3. (Usually, in the data analyzed in this paper we found a high correlation between the normalized and the unnormalized values, especially between H_1–H_2 and H_1^*–H_2^*.)

In terms of f_0, the difference between *nyh* and *ny* is similar to the difference between voiced and plain stops and clicks, respectively (not very important) in the speech of F3, F4, and M4 (considering both the postvocalic data analyzed so far and the postnasal data listed in Table BI in Appendix B). However, the difference in f_0 produced by M3 is usually smaller than for his stops and clicks, and for p1, it is in the opposite direction. Differences in F_1 are not entirely consistent, although there is a clear tendency that F_1 is lower after *nyh* than after *ny*. H_1–H_2 after *nyh* and *ny* in the speech of F3, F4, M4 is comparable to H_1^*–H_2^* after the voiced and plain stops/clicks, respectively, matching most closely to the values for stops/clicks in postnasal position, listed in Appendix B. For speaker M3, on the other hand, the difference in H_1–H_2 after *nyh vs. ny* is usually smaller in comparison to his voiced *vs.* plain stops/clicks, and opposite in direction for the first period. (Part of the difference between sonorants and postnasal stops/clicks is due to presence *vs.* absence of normalization and the different target vowels.) For the parameter H_1–A_3, there are higher values after *nyh* than *ny* only for F3 and F4. For M4 the opposite pattern occurs and for M3 no consistent pattern emerges.

Altogether, there is evidence that the distinction between *nyh* and *ny* has an influence on voice quality similar to the distinction between voiced and plain stop/clicks in postvocalic, and most particularly in postnasal position. However, the interspeaker variability is somewhat greater for the occurrence of breathy voice after sonorants than after obstruents.

4. General discussion

4.1. *Main results*

Among the acoustic properties investigated here it was f_0 in the following vowel that turned out to be the statistically most robust, the most speaker-invariant, and the temporally most stable means of distinguishing the voiced stops/clicks from the plain (sometimes ejective) and aspirated cognates. With few exceptions only, most noticeably speaker M4, voiced stops and clicks showed significantly lower f_0 in the following vowel than either the aspirated or plain cognates. With only a slight reduction of the differences the effect was still found at the center of the vowel. Upon speaker-to-speaker evaluation of f_0 plots (not shown in this paper), it can be seen that some speakers have no reduction of f_0 differences towards the center of the vowel at all.

Lower f_0 after voiced than voiceless stops is known as a widespread phenomenon (see among others, Hombert, 1978; Ohde, 1984; Löfqvist, Baer, McGarr & Seider Story, 1989; Kingston & Diehl, 1994; Jessen, 1998).[4] However, in Xhosa, the f_0 difference between the

[4]The references cited here include both languages with an essentially unaspirated/aspirated difference (English, German) and those with a voiced/voiceless difference in a more literal sense (French, Dutch). We believe that those two constitute different phonological types and that it is important to carefully differentiate the influence of voicing *vs.* its absence from the one of aspiration *vs.* its absence on f_0 (Jessen, 1998 for discussion). In the present discussion, we focus on explanations regarding the influence of voicing.

voiced and the other categories seems more phonologized than this crosslinguistically known f_0 perturbation. Perhaps even more than from the absolute differences in Hertz, this can be inferred from the temporal stability of f_0 depression in Xhosa, where reduction of the differences towards the center of the vowel is only slight (or, as mentioned above, completely absent for some speakers). This is different from f_0 perturbation patterns in languages like English or French (references above), where a clear period-by-period reduction of the difference can be observed. As far as the absolute f_0 differences are concerned, it needs to be emphasized again that in our stimuli, the target vowel is specified for a lexical low tone. With a lexical high tone, the magnitude of the f_0 differences after voiced *vs.* other sounds would probably have been even stronger (cf. Traill *et al.*, 1987 on Zulu).

Greater phonological importance of f_0 depression in Xhosa compared to f_0 perturbation in languages like English and French can also be inferred from the fact that in those languages, other phonetic differences are available to cue the relevant distinction (primarily an aspiration difference in English and a closure voicing difference in French). In Xhosa, on the other hand, there are situations where the distinction rests exclusively or primarily on f_0 depression (and possibly breathy voice, to be discussed below). This is the case when voiced stops and clicks are compared to their plain cognates and when the latter are produced without strong perceptual signs of ejection. As mentioned in Section 1.1, the voiced stops and clicks of Xhosa are not produced with longer closure voicing than any of the other stops and clicks (except the implosive). Given the importance of f_0 depression in Xhosa, it is likely that speakers of this language actively plan and execute speech gestures that ensure low f_0 after voiced stops and clicks.

Another phonetic property associated with voiced stops and clicks that was found in this study is breathy voice in the following vowel. However, in contrast to f_0 depression, occurrence of breathy voice was more speaker dependent, less temporally stable, and less statistically robust. We inferred breathy phonation from high values of H_1^*–H_2^* and H_1^*–A_3^*. More specifically, we used the values found after the aspirated stops and clicks as the baseline, since some breathy voice after them is generally expected and empirically supported from other languages. With equally high or higher H_1^*–H_2^* and H_1^*–A_3^* after voiced than aspirated stops/clicks, breathy phonation can be assumed after the former. Indications of breathy voice after voiced stops and clicks, as measured and inferred in this manner, were found for about five of our eight speakers (depending on the acoustic parameter and the period number).

Except for the influence of gender, our experiment was not designed for the systematic study of dialectal and sociolectal variables. Therefore, we are not able to relate the voice quality differences between speakers to such factors. There seems to be a tendency in our data that breathy voice after voiced stops/clicks is found most often with speakers from or in close contact with the traditional Transkei Xhosa community and less frequently among those with an urban background in the Cape area. For example, the two speakers with the highest prominence of H_1^*–H_2^* after voiced compared to other categories (F4 and M4) are also the ones with the closest affiliation to the traditional Transkei community. However, more research is necessary to differentiate those and other influences, including also age. As for now, we have to treat the variability in consonant-induced voice quality differences as speaker specific (cf. Traill & Jackson, 1988). In fact, that at least some pure speaker specificity is at work is suggested by the fact that within the dialectally, sociolectally, and otherwise homogeneous group of speakers F1, F2, F3,

M1, and M3 there are some speakers with clear signs of breathy voice after voiced stops and clicks, whereas for others this is not the case.

Given this variability of breathy phonation as a means of differentiating the voiced stops and clicks from the other stop and click categories (especially the plain set), it is more likely that breathy phonation in Xhosa occurs concomitantly with gestures that have a different primary goal (in particular f_0 depression) than that breathy phonation is something that is targeted directly and independently from other phonetic manifestations of the voicing feature. (This is not meant to imply that breathy phonation cannot be targeted independently from tonal depression at all; see Section 4.2.) Ideally, we want to identify a unitary gesture and associated principles (aerodynamic, physiological, etc.) that can explain all the three main properties of voiced stops and clicks in Xhosa together: lack of closure voicing, tonal depression, and (optional) breathy phonation. The tendency for low F_1 found with the voiced stops/clicks should also be accounted for in such an explanation. An attempt along those lines will be presented in the following section.

4.2. *Explaining the specifics of the voicing feature in Xhosa*

In summary, it seems that in the voiced stops and clicks of Xhosa, extensive vocal fold slackening creates a strong f_0 cue together with optional breathy voice and low F_1 but is fatal to closure voicing itself. We proceed to spell out this hypothesis in more detail.

Normally in the production of voiced stops, slack vocal folds are beneficial for the maintenance of closure voicing since they permit vocal fold vibration at the low transglottal pressures found in obstruent production. Vocal fold slackening is created in part by larynx lowering. Larynx lowering in turn increases the size of the oral cavity. This allows for longer maintenance of a pressure drop across the glottis and therefore constitutes another advantage for the production of obstruent voicing (Stevens, 1977, 1998, pp. 466–470). The vocal fold slackening brought about by larynx lowering also leads to a reduction of f_0 — a principle that has been formulated as the "vertical tension hypothesis" (Hombert, Ohala & Ewan, 1979). Due to progressive coarticulation this f_0 reduction also affects the following vowel, i.e., it takes a certain time until f_0 in the following vowel has reached its target value. Ohala (1978) mentions activity of the sternohyoid and sternothyroid muscles as major factors in f_0 reduction through larynx lowering. According to Ohala, this principle of f_0 reduction is crucial for f_0 values which are too low to be achievable by relaxation of the cricothyroid alone. Denning (1989, p. 83) (based on Painter, 1976), also discusses the possibility that larynx lowering is further enhanced by activity of the muscles that are primarily responsible for tongue root advancement, which is another oral-cavity enlarging maneuver. Larynx lowering and tongue root advancement are also the cause for the lowering of F_1 associated with voiced stops.

The principles mentioned so far capture the situation in a language like French or Russian, where the voiced stops are produced with clear (phonologically primary) closure voicing and some additional (phonologically redundant) f_0 lowering. However, the voiced stops (and clicks) of Xhosa are predominantly voiceless during closure (in postvocalic position), so something more needs to be said in order to understand the specifics of the voicing feature in Xhosa. Kingston (1985, p. 7) and Denning (1989, p. 89) point out that beyond a certain level of vocal fold slackening transglottal air flow becomes so high that it actually works against the maintenance of closure voicing, since it leads to a fast buildup of intraoral air pressure. If such extensive levels of vocal fold slackening are achieved by additional larynx lowering, it is expected that f_0 lowering

becomes more extensive as well and achieves the status of a strong perceptual cue (additional F_1 lowering is predicted as well). It is also likely that the high levels of transglottal air flow in extensive vocal fold slackening lead to levels of breathy voice that are equally high as or even higher than those found after aspirated stops. Some percentage of breathy voice might also be attributed to passive glottal opening that results from an increased buildup of intraoral air pressure (Bickley & Stevens, 1987). If this scenario is correct, Xhosa speakers by opting for strong vocal fold slackening in the implementation of the voicing feature have replaced closure voicing by a strong f_0 cue, usually accompanied by low F_1 and optionally by breathy voice.

What we have hypothesized here as an explanation for the specifics of the voiced stops and clicks in Xhosa could be a stage in the phonologization of pitch from a former voiced/voiceless distinction. (Although breathy voice is supported by the actions that result in pitch-lowering, it is unlikely that this voice quality is subject to phonologization by itself in Xhosa; otherwise, more stability across speakers in the occurrence of breathy voice after the voiced stops and clicks should occur.) However, here we do not intend to make any specific claims about the historical linguistics of Xhosa. It needs to be kept in mind that there are cases where real closure voicing (still) plays a role in Xhosa, e.g., the implosive and presumably the voiced fricatives and affricates. On the other hand, there is further evidence for a certain phonological autonomy of tonal depression, together with optional breathy voice and F_1 lowering, from the fact that there is a distinction among the postnasal palatal nasals written *nyh vs. ny* that is essentially phonetically parallel to the one between voiced and plain stops and clicks (see Section 3.5 and Appendix B). Note that for voiced sonorants there is no natural motivation for properties like low f_0, F_1, or breathy voicing, contrary to the situation for voiced obstruents. Therefore, one can assume that in sonorants this distinction is created actively in congruity with the distinction among stops and clicks (and probably fricatives and affricates) in Xhosa.

In this scenario hypothesized for Xhosa extensive larynx lowering is the primary gesture of the voiced stops and clicks, leading to vocal fold slackening with glottal leakage and potentially breathy voice, to devoicing, strong f_0 lowering, and to some F_1 lowering. With the exception of devoicing, these phonetic properties do not only form a natural association on the production level but there are also perceptual/auditory associations. Kingston & Diehl (1994, 1995) use the concept of "low-frequency property" as the common denominator for voicing during closure, low F_1 and low f_0. All of these acoustic properties express a predominance of energy in the low-frequency domain of the spectrum. Breathy voice in the form of an increased prominence of the first harmonic relative to higher harmonics can be added straightforwardly to the list of low-frequency properties. Kingston, Macmillian, Dickey, Thorburn & Bartels (1997) show that breathy voice integrates with low F_1 values in the perception of English vowels. According to Kingston & Diehl (1994, 1995), the various manifestations of the low-frequency property are different ways of implementing the phonological feature [voice]. What is unusual about Xhosa are the particular weights assigned to the individual low-frequency properties in this language as compared to many other languages. Whereas voicing during closure might be the most stable and perceptually salient property of the set /b, d, g/ in most languages for which an analysis with [voice] is motivated (for discussion see Kohler, 1984; Kingston & Diehl, 1994; Iverson & Salmons, 1995; Jessen, 1998), Xhosa assigns little weight to this property. Instead, tonal depression and potentially breathy voice and low F_1 play a more prominent role in Xhosa. This perceptual interpretation leads to the hypothesis that one low-frequency property can compensate for the lack of

or low salience of another. Since closure voicing is not reliably present in the voiced stops/clicks of Xhosa, it is conceivable that low F_1, breathy voice, and in particular, low f_0 can perceptually compensate for this predominant lack of closure voicing. Compensation among low-frequency properties might also explain the patterns of speaker M4. We saw that f_0 depression after voiced stops and clicks is not very prominent with this speaker, but that he shows clearly elevated levels of H_1^*–H_2^* after voiced compared to other cognates (though less striking H_1^*–A_3^* patterns). This perceptual perspective suggests that on top of the "natural" physiological associations discussed here, the speaker is also able to manipulate the low-frequency properties individually, knowing about their acoustic similarities (cf. Perkell, Matthies, Svirsky & Jordan, 1995).

One further remark is necessary about the status of implosives. Implosives are generally assumed to be produced with extensive larynx lowering. Notice first that while implosives can be largely voiceless, they are often fully or almost fully voiced (Lindau, 1984). This is indeed the case for the majority of speakers and tokens in the present Xhosa material (cf. Roux, 1991). Secondly, in applying measurements of H_1^*–H_2^* and H_1^*–H_3^* on productions of the Xhosa implosive by speakers F1, F2, M1, and M2, we found values that were clearly lower than the H_1^*–H_2^* and H_1^*–A_3^* values found after aspirated stops/clicks. The frequent presence of voicing and relatively low values of H_1^*–H_2^* and H_1^*–A_3^* show that implosives are not usually voiceless nor breathy voiced (and if voiceless it is due to strong adduction, not abduction of the vocal folds). This is in contrast to the voiced stops and clicks of Xhosa which presumably also have strong larynx lowering, but which are often voiceless and often breathy voiced. We did, however, find f_0 depression similar to that found after voiced stops with those speakers that produced fully voiced implosives. For the present discussion, these facts are important, since they demonstrate the flexibility of the interactions between laryngeal gestures and states discussed here. While the scenario proposed here, according to which larynx lowering enhances low f_0, low F_1, glottal leakage, and devoicing, is a "natural" possibility (cf. Denning, 1989), there are no strict causal relationships in this area of laryngeal physiology (cf. also Kingston, 1985). Lack of pure physiological conditioning can also be inferred from the fact that there is no strict temporal synchronization between the different low-frequency properties (see, for example, the asynchronous movements of f_0 and F_1 after the voiced stops/clicks in Figs 1 and 2). These insights strengthen the hypothesis, expressed in the previous paragraph, that part of the motivation for the observed cluster of properties in the voiced stops and clicks of Xhosa is perceptual.

4.3. *Parallels and differences in other languages*

The question arises whether the particular way in which the voicing feature is implemented in Xhosa is found elsewhere among the languages of the world. In this section, we argue that a close parallel to the Xhosa system is found in Shanghai Chinese, and that on the basis of this comparison it is reasonable to characterize the breathy type of voice quality found optionally with the voiced stops and clicks of Xhosa as "slack voice". We also argue that the phonetic implementation and phonological status of the voiced stops in Xhosa is different from that of the stops in Hindi and other languages that have been characterized as breathy voiced or voiced aspirated, and we finally make a comparison between breathy voice in Hindi and the universal tendency for breathy voice after voiceless aspirated stops.

4.3.1. *Slack voice in Shanghai Chinese*

An interesting parallel to Xhosa emerges from the phonetic study of Shanghai Chinese by Ren (1992). (Relevant data for Wu dialects more generally are also reported in Cao & Maddieson, 1992, though with no data on the voiceless aspirated stops.) Shanghai Chinese has three stop categories: voiceless unaspirated (possibly associated with some stiff voice similar to Korean reinforced stops), voiceless aspirated, and a third category that might be phonologically characterized as voiced. But as in Xhosa, the "voiced" stops in Shanghai are not produced with any significant amount of voicing during closure in word-initial postvocalic (or utterance-initial) position. (However, in word-medial poststress position "voiced" stops are fully voiced in Shanghai, which might be different from Xhosa; in our study, we have not investigated that particular context, since all target sounds were in a stressed syllable.)

Phonetic descriptions of Shanghai mention complementary distribution in initial position with respect to the possible tones after aspirated and plain stops on the one hand, and voiced stops on the other: tones beginning with low pitch (low rising, short low rising) only occur after the voiced stops, those beginning with mid or high pitch only after the remaining stop cognates (Ren, p. 7f.). f_0 measurements (p. 137) show a pattern quite similar to the one found here. According to Ren's Fig. 4.26 (p. 137), f_0 is substantially lower after voiced stops than after the other two stop categories (which differ only little), and by period 11–13 (which is probably around the center of the vowel or later) the differences are still maintained. Both the tonal transcription and the f_0 measurements suggest that Shanghai has a type of phonologized f_0 depression similar to that found in Xhosa. It is also interesting that Iwata, Hirose, Niimi & Horiguchi (1991) report sternohyoid activity (indicating active larynx lowering) associated with f_0 lowering after "voiced" stops in Suzhou — another Wu dialect.

Furthermore, the voiced stops of Shanghai are traditionally described as being produced with breathy voice or voiced aspiration. Ren (1992) makes acoustic as well as aerodynamic and fiberscopic measurements to investigate those claims. He finds confirmation from increased values of H_1–H_2 and H_1–A_1 after voiced compared to the plain stops (see also Cao & Maddieson, 1992), which is also shown by Ren to have perceptual relevance. In terms of significant *post hocs* reported for his data pooled across speakers (p. 22), H_1–H_2 and H_1–A_1 after voiced stops are either on the same level as after aspirated stops, or significantly lower, but never significantly higher. Hence, if we apply the same baseline method to Shanghai as we used for Xhosa (breathy voice after voiced only if H_1-based values are at least as high as after aspirated), a similar result is found. From these results, there is no more reason to talk about breathy voice in Shanghai than in Xhosa. As with Xhosa, there is also speaker variability in Shanghai. For some speakers, H_1–H_2 and H_1–A_1 after voiced stops can fall clearly behind the values after aspirated stops — most clearly at vowel onset, but also at vowel center (see pp. 30–47),[5]

[5]While the comparison of voice qualities after the voiced stops with those after the voiceless aspirated stops presents a similar picture in Shanghai and Xhosa, the comparison between voiced and plain stops leads to a somewhat different result. Whereas for three Xhosa speakers (F1, M1, M2) normalized H_1–H_2 is lower after voiced stops (and clicks) than after the plain cognates, H_1–H_2 is almost always higher after voiced than plain stops in Shanghai (though Ren, p. 46, shows exceptions in the speech of his speaker H). Due to the relationality of voice quality (cf. Traill & Jackson, 1988) and the difficulty of identifying different voice qualities from absolute H_1-based values, it is possible that low H_1–H_2 after plain stops in Shanghai reflect the presence of stiff voice, whereas voiced stops are followed by modal voice. Since Ren (1992) frequently points out similarities between the plain stops in Shanghai and the reinforced stops of Korean, where the existence of stiff voice is established (Ladefoged & Maddieson, 1996), this is a possible interpretation (cf. also footnote 6 on Javanese).

and, as in Xhosa, significant voice quality differences early in the vowel usually become non-significant at vowel midpoint in Shanghai (see also Table VII in Cao & Maddieson, 1992). Overall, though the relevant phonetic patterns of Xhosa and Shanghai Chinese are not exactly the same, there is enough evidence to observe a parallel in the voice quality effects in the two languages.

Given this parallel between Shanghai Chinese and Xhosa in the voice quality after voiced stops, together with other similarities in the implementation of the voicing features in these languages, it should be considered whether certain terminology that has been proposed for the voiced stops of Shanghai should be applied to Xhosa as well. Ladefoged & Maddieson (1996, pp. 63–66) use the term "slack voice" for the type of voice quality that occurs after the voiced stops of Shanghai and other Wu dialects of Chinese, suggesting that slack voice is a voice quality that is similar to breathy voice but less extreme in terms of the degree of glottal opening and the amount of airflow (p. 63). As a difference they suggest that real breathy voice, as found in Hindi, involves active separation of the arytenoids, while slack voice results from a configuration where the vocal folds are vibrating "loosely", without separation at the arytenoids (p. 66) (the Hindi case will be discussed separately below).[6]

The spectrogram by Ladefoged & Maddieson (1996, p. 65) of a voiced stop in Shanghai shows no turbulence early in the following vowel, much in contrast to their spectrogram of a breathy voiced stop in Hindi (p. 59). This (predominant) lack of turbulence seems to be a good way of capturing the weaker type of breathiness in slack voice compared to the stronger type in genuine breathy voice. Likewise, we found no particular signs of turbulence in the spectrograms of vowels after voiced stops and clicks in our material. Or, more precisely, turbulence, where it was visible in spectrograms at all, occurred after all three stop and click categories and hence was a speaker-specific feature. Likewise, no turbulence noise can be seen in the spectrograms of vowels preceded by voiced stops in Xhosa that are provided by Louw (1977*a*). This again suggests that the amount of breathy voice after voiced stops and clicks in Xhosa is only small at best. As far as the voiced clicks of Xhosa are concerned, the same conclusion is drawn by Ladefoged & Traill (1994) and Ladefoged & Maddieson (1996). They also point out that any existing breathy voice in Xhosa is not as strong as the breathy voice that is found in Hindi or Marathi. Perhaps then it is more appropriate to use "slack voiced" instead of "breathy voiced" (or "murmur") for the relevant click accompaniments in Xhosa (see the tables for the possible click accompaniments in Ladefoged and Traill, 1994 and in Ladefoged & Maddieson, 1996, p. 278).

4.3.2. *Hindi and other languages with breathy voice/voiced aspiration*

We agree with Ladefoged & Maddieson (1996) that the stronger type of breathy voice found in the (commonly called) voiced aspirated stops of Hindi should be distinguished

[6]As another language with slack voiced stops, Ladefoged & Maddieson (1996) mention Javanese. However, Javanese has no voiceless aspirated stops, which makes it impossible to investigate this language with the methodology used here, according to which the phonetic correlates of breathy voice after voiceless aspirated stops serve as a baseline for the identification of slack voice. This makes the matter complicated, since there is evidence that the cognates in opposition to the slack voiced stops are produced with stiff voice. Given the relationality of H_1-based measures, the contrast in Javanese could theoretically be any one of the following three: stiff voiced *vs.* slack voiced, stiff voiced *vs.* plain, or slack voiced *vs.* plain (cf. Hayward, 1995 on this issue). However, the fact that Fagan (1988) found aperiodic energy around F_3 after at least some tokens of /b, d, g/ suggests that Javanese does in fact have a phonation type that projects into the slack/breathy end of the voice quality continuum.

from the weaker kind found in Shanghai Chinese (based on our findings we would like to add Xhosa here as well). In keeping with this, henceforth we refer to the stronger type as breathy voice in a narrow sense and to the weaker type as slack voice (up to this point in the present paper we have used the term breathy voice in a broad sense, to encompass slack voice). We already mentioned the Ladefoged & Maddieson (1996) position that only in breathy voice of the type in Hindi is there active spreading of the arytenoids, while slack voice results from vocal fold looseness. According to this view, to which we subscribe, glottal leakage in real breathy voice is the direct result of active glottal opening with activation of the posterior cricoarytenoid muscle (PCA), while glottal leakage in slack voice is a more indirect result of strong vocal fold slackening, induced in large part by larynx lowering, but with no active glottal abduction. Although we know of no relevant fiberoptic or electromyographic studies of Xhosa, we strongly doubt that there will be any indications of active glottal opening with the voiced stops and clicks of Xhosa.

However, there is one more accessible piece of evidence in support of our position that Xhosa uses slack rather than breathy voice. In Hindi, breathy phonation can vary with voiceless aspiration in the realization of the voiced aspirated stops (Schiefer, 1992; Davis, 1994). Similarly, Traill (1992) reports variation between breathy voice and voiceless aspiration in the realization of the nasalized aspirated clicks of !Xũ). In Xhosa, on the other hand, we never observed any significant amount of voiceless aspiration after the voiced stops and clicks.

Phonologically, breathy voice in Hindi and !Xũ) seems to be part of (whatever feature represents) aspiration, whereas in Xhosa, slack voice seems to be part of the voicing feature (see our discussion of "low-frequency property" and the feature [voice] in Section 4.2). This is in contrast to the phonological feature analysis of Lanham (1969), where the voiced stops and clicks in Xhosa are specified like the voiced aspirates of Hindi.

There is also the possibility that breathy voice in Hindi is actually a combination of both breathy voice and slack voice. The former would be the result of an active glottal opening gesture (coordinated late with respect to oral release), as the implementation of the aspiration feature, while the latter would be part of the implementation of the voicing feature. This analysis would be consistent with the fact that breathy voiced stops in Hindi are followed by lower f_0 than any of the other three stop cognates in Hindi (Kagaya & Hirose, 1975).

4.3.3. *Breathy voice after voiceless aspirated stops*

The kind of breathy voice that is found after voiceless aspirated stops in Xhosa (and which is expected generally in languages with voiceless aspirated stops) should be qualitatively similar to the one found in the Hindi voiced aspirated stops, since in both cases there is a mixture of transglottal turbulent airflow, resulting from active glottal opening caused by PCA activity, with voicing (abstracting here from the further possibility, suggested above, that Hindi has a combination of both breathy and slack voice). But quantitatively, breathy voice after voiceless aspirated stops is more comparable to slack voice, since most of the glottal opening gesture is completed at vowel onset and the breathiness level will therefore be relatively low after voiceless aspirated stops. (This consideration is the basis of our method, explained earlier, of using the spectral patterns after aspirated stops as the baseline for the interpretation of voice quality after voiced stop and clicks.) In our data, both breathy voice after aspirated stops/clicks and slack

voice after voiced stops/clicks are manifested by increased H_1^*–H_2^* and H_1^*–A_3^*, but they can usually be distinguished by the shorter temporal spread of the former than the latter. This is particularly evident in the speech of M3 (Fig. 3), where for the first period H_1^*–H_2^* is highest after the aspirated category, but where H_1^*–H_2^* is on its rapid decline and soon surpassed by the increasing and longer-lasting values after voiced stops/clicks. The same kind of pattern occurs in Shanghai Chinese according to Ren (1992, pp. 32–47): at vowel onset H_1–H_2 and H_1–A_1 is usually highest after voiceless aspirated stops, but at vowel center it is often highest after the voiced stops. Relatively long temporal scope of slack voice is probably the result of a long persistence of the larynx lowering gesture and its consequences for low f_0 and vocal fold slackness.

4.4. *Conclusion*

The results from the measurements of the voice quality parameters H_1^*–H_2^* and H_1^*–A_3^* lend support to the view expressed by Rycroft (1980), Pahl (1989), and Finlayson *et al.* (1989) that the voiced stops and clicks of Xhosa are associated with some amount of breathy voice in the following vowel. But if the intention of these authors is to claim that breathy voice is as much a stable and speaker-independent feature of voiced stops and clicks in Xhosa as tonal depression, the results of this study tell otherwise. We found that whereas tonal depression occurs in the speech of all of our subjects, breathy phonation with the voiced stops and clicks was only found for some of our subjects. It is this between-subject variability that was meant when talking about "optional" breathy voice in this paper. (However, strictly speaking, breathy voice associated with the voiced stops and clicks is not optional for those speakers who use it; thanks are due to one of the reviewers for pointing this out.) Since regional, social, and age-related factors were not varied systematically in this study, we can only speculate about the nature of this variability in voice quality. It seems, however, that some proportion of this variation between presence and absence (or degree) of breathy voice is purely speaker-specific, since some of the variation was found to occur between speakers of practically identical age, location, and social status. It should also be kept in mind that the occurrence of breathy voice associated with voiced stops and clicks might be limited by prosodic and contextual factors. Possibly, it is more salient in low- than high-tone vowels. Tonal depression, on the other hand, is probably more invariant. Clearly, more research is necessary on these issues.

In the speech of those speakers who show nonmodal voice quality with the voiced stops and clicks, this voice quality did not include any consistent signs of turbulence noise. This suggests that the term "breathy voice" is too strong, and that instead the characterization of the relevant voice quality as "slack voice" is in better agreement with the results. To the extent that this interpretation is convincing, this study also encourages a revision of the set of possible click accompaniments, changing "breathy voiced" (or "murmur") to "slack voiced" in the naming of the relevant Xhosa click accompaniments by Ladefoged & Traill (1994) and Ladefoged & Maddieson (1996).

Our study has shown that f_0 depression is the statistically most salient, temporally most stable, and the most speaker-invariant phonetic property to distinguish the voiced stops and clicks from their plain (optionally ejective) and aspirated cognates. We hypothesize that f_0 depression is created by a strong larynx lowering gesture which also enhances vocal fold slackening (and some F_1 lowering). The glottal leakage created by this vocal fold slackening is high enough to cause devoicing and optional slack (weak

breathy) phonation. The particular implementation of the voicing feature in Xhosa is rare among the languages of the world. Based on our knowledge, we think that its closest equivalent is found in Shanghai Chinese. We agree with Ladefoged and Maddieson (1996) that stops produced with slack voice should be distinguished from the breathy voiced stops found in languages like Hindi.

This paper benefits from many constructive comments (only a few of them explicitly acknowledged in the text) by Peter Ladefoged and a second reviewer. Financial support from the Alexander von Humboldt Foundation and from the Research Unit for Experimental Phonology at University of Stellenbosch to the first author is gratefully acknowledged.

References

Beach, D. M. (1938) *The phonetics of the Hottentot language*. Cambridge: Heffer.

Bickley, C. (1982) Acoustic analysis and perception of breathy vowels, *MIT Research Laboratory of Electronics Speech Communication Group Working Papers*, **1,** 71–81.

Bickley, C. A. & Stevens, K. N. (1987) Effects of a vocal tract constriction on the glottal source: data from voiced consonants. In *Laryngeal function in phonation and respiration* (T. Baer, C. Sasaki & K. S. Harris, editors), pp. 239–253. Boston: College-Hill Press.

Brand, H. S. P. & Roux, J. C. (1990) Devokalisasie in Xhosa: 'n herinterpretasie, *South African Journal of African Languages*, **10,** 109–117.

Cao, J. & Maddieson, I. (1992) An exploration of phonation types in Wu dialects of Chinese, *Journal of Phonetics*, **20,** 77–92.

Chapin Ringo, C. (1988) Enhanced amplitude of the first harmonic as a correlate of voicelessness in aspirated consonants, *Journal of the Acoustical Society of America*, **83** (Suppl. 1), S70 [Abstract].

Davis, K. (1994) Stop voicing in Hindi, *Journal of Phonetics*, **22,** 177–193.

Denning, K. (1989) *The diachronic development of phonological voice quality, with special reference to Dinka and the other Nilotic languages*. PhD dissertation, Stanford University.

Doke, C. M. (1926) *The phonetics of the Zulu language*. Johannesburg: The University of the Witwatersrand Press.

Fagan, J. L. (1988) Javanese intervocalic stop phonemes: the light/heavy distinction. In *Studies in Austronesian linguistics* (R. McGinn, editor), pp. 173–200. Athens, OH: Center for International Studies Ohio University.

Finlayson, R., Jones, J., Podile, K. & Snyman, J. W. (1989) *An introduction to Xhosa phonetics*. Constantia, Cape: Marius Lubbe.

Fischer-Jørgensen, E. (1967) Phonetic analysis of breathy (murmured) vowels in Gujarati, *Indian Linguistics*, **28,** 71–139.

Fischer-Jørgensen, E. (1968) Voicing, tenseness and aspiration in stop consonants, with special reference to French and Danish, *Annual Report of the Institute of Phonetics of the University of Copenhagen*, **3,** 63–114.

Fourakis, M. & Port, R. (1986) Stop epenthesis in English, *Journal of Phonetics*, **14,** 197–221.

Hanson, H. M. (1995) *Glottal characteristics of female speakers*. PhD dissertation, Harvard University.

Hanson, H. M. (1997) Glottal characteristics of female speakers: acoustic correlates, *Journal of the Acoustical Society of America*, **101**(1), 466–481.

Hayward, K. (1995) /p/ vs. /b/ in Javanese: the role of the vocal folds, *SOAS Working Papers in Linguistics and Phonetics*, **5,** 1–11.

Herbert, R. K. (1987) Articulatory modes and typological universals: the puzzle of Bantu ejectives and implosives. In *In honor of Ilse Lehiste* (R. Channon & L. Shockey, editors), pp. 401–413. Dordrecht: Foris.

Hombert. J.-M. (1978) Consonant type, vowel quality, and tone. In *Tone: a linguistic survey* (V. A. Fromkin, editor), pp. 77–112. New York: Academic Press.

Hombert, J.-M., Ohala, J. J. & Ewan, W. G. (1979) Phonetic explanations for the development of tones, *Language*, **55,** 37–58.

Iverson, G. K. & Salmons, J. C. (1995) Aspiration and laryngeal representation in Germanic, *Phonology*, **12,** 369–396.

Iwata, R., Hirose, H., Niimi, S. and Horiguchi, S. (1991) Physiological properties of "breathy" phonation in a Chinese dialect — a fiberoptic and electromyographic study on Suzhou dialect, *Proceedings of the XIIth International Congress of Phonetic Sciences*, Vol. 3, 162–165.

Jessen, M. (1998) *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: Benjamins.

Jordan, A. C. (1966) *A practical course in Xhosa*. Cape Town: Longmans.

Kagaya, R. & Hirose, H. (1975) Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants, *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, **9,** 27–46.

Kingston, J. C. (1985) *The phonetics and phonology of the timing of oral and glottal events*. PhD dissertation, University of California, Berkeley.

Kingston, J. & Diehl, R. L. (1994) Phonetic knowledge, *Language*, **70,** 419–454.
Kingston, J. & Diehl, R. L. (1995) Intermediate properties in the perception of distinctive feature values. In *Phonology and phonetic evidence. Papers in laboratory phonology IV* (B. Connell & A. Arvaniti, editors), pp. 7–27. Cambridge: Cambridge University Press.
Kingston, J., Macmillian, N. A., Dickey, L. W., Thorburn, R. & Bartels, C. (1997) Integrality in the perception of tongue root position and voice quality in vowels, *Journal of the Acoustical Society of America*, **101**(3), 1696–1709.
Klatt, D. H. & Klatt, L. C. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers, *Journal of the Acoustical Society of America*, **87**(2), 820–857.
Kohler, K. J. (1984) Phonetic explanation in phonology: the feature fortis/lenis, *Phonetica*, **41,** 150–174.
Ladefoged, P. & Maddieson, I. (1996) *The sounds of the world's languages*. Oxford: Blackwell.
Ladefoged, P., Maddieson, I. & Jackson, M. (1988) Investigating phonation types in different languages. In *Vocal physiology*: *Voice production, mechanisms and functions* (O. Fujimura, editor), pp. 297–316. New York: Raven Press.
Ladefoged, P. & Traill, A. (1994) Clicks and their accompaniments, *Journal of Phonetics*, **22,** 33–64.
Lanham, L. W. (1960) *The comparative phonology of Nguni*. PhD dissertation, University of the Witwatersrand, Johannesburg.
Lanham, L. W. (1969) Generative phonology and the analysis of Nguni consonants, *Lingua*, **24,** 155–162.
Lindau, M. (1984) Phonetic differences in glottalic consonants, *Journal of Phonetics*, **12,** 147–155.
Löfqvist, A., Baer, T., McGarr, N. S. & Seider Story, R. (1989) The cricothyroid muscle in voicing control, *Journal of the Acoustical Society of America*, **85**(3), 1314–1321.
Löfqvist, A. & McGowan, R. S. (1992) Influence of consonantal environment on voice source aerodynamics, *Journal of Phonetics*, **20,** 93–110.
Löfqvist, A. & Yoshioka, H. (1980) Laryngeal activity in Swedish obstruent clusters, *Journal of the Acoustical Society of America*, **68**(3), 792–801.
Louw, J. A. (1977*a*) The adaptation on non-click Khoi consonants in Xhosa. In *Khoisan linguistic studies* 3 (A. Traill, editor), pp. 74–92. University of the Witwatersrand, Johannesburg.
Louw, J. A. (1977*b*) Clicks as loans in Xhosa. In *Bushman and Hottentot linguistic studies* (J. W. Snyman, editor), pp. 82–100. University of South Africa, Pretoria.
Ní Chasaide, A. & Gobl, C. (1993) Contextual variation of the vowel voice source as a function of adjacent consonants, *Language and Speech*, **36,** 303–330.
Ohala, J. J. (1978) The production of tone. In *Tone*: *a linguistic survey* (V. A. Fromkin, editor), pp. 5–39. New York: Academic Press.
Ohde, R. N. (1984) Fundamental frequency as an acoustic correlate of stop consonant voicing, *Journal of the Acoustical Society of America*, **75**(1), 224–230.
Pahl, H. W. (1989—editor in chief). *The greater dictionary of Xhosa*. University of Fort Hare, Alice, South Africa.
Painter, C. (1976) Pitch control and pharynx width in Twi: an electromyographic study, *Phonetica*, **33,** 334–352.
Perkell, J. S., Matthies, M. L., Svirsky, M. A. & Jordan, M. I. (1995) Goal-based speech motor control: a theoretical framework and some preliminary data, *Journal of Phonetics*, **23,** 23–35.
Pongweni, A. J. C. (1983) An acoustic study of the qualitative and pitch effect of breathy-voice on Shona vowels, *Journal of Phonetics*, **11,** 129–138.
Ren, N. (1992) *Phonation types and stop consonant distinctions*: *Shanghai Chinese*. PhD dissertation, University of Connecticut.
Riordan, J., Mathiso, M., Davey, A. S., Bentele, S. V., Mahlasela, B. & Lanham, L. W. (1969) *Lumko Xhosa self-instruction course*. Institute of Social and Economical Research, Rhodes University, Grahamstown.
Roux, J. C. (1991) On ingressive glottalic and velaric articulation in Xhosa, *Proceedings of the XIIth International Congress of Phonetic Sciences*, Vol. 3, 158–161.
Roux, J. C. (1998) Xhosa: a tone or pitch-accent language?, *South African Journal of Linguistics*, Suppl. **36,** 33–50.
Roux, J. C. & Dogil, G. (1997) On the phonetic representation of clicks: some experimental phonetic considerations. *Proceedings of the XVIth international congress of linguists.*
Rycroft, D. K. (1980) The depression feature in Nguni languages and its interaction with tone, *Communication No.* 8, Department of African Languages, Rhodes University, Grahamstown.
Sands, B. (1991) Evidence for click features: acoustic characteristics of Xhosa clicks, *UCLA Working Papers in Phonetics*, **80,** 6–37.
Schiefer, L. (1992) Trading relations in the perception of stops and their implications for a phonological theory. In *Papers in laboratory phonology II. Gesture, segment, prosody* (G. J. Docherty & D. R. Ladd, editors), pp. 296–313. Cambridge: Cambridge University Press.
Selmer, E. W. (1933) *Experimentelle Beiträge zur Zulu Phonetik*. Norwegian Academy of Sciences, Oslo: Broggers.

Sluijter, A. M. C. (1995) *Phonetic correlates of stress and accent*. The Hague: Holland Academic Graphics.
Sluijter, A. M. C., Shattuck-Hufnagel, S., Stevens, K. N. & van Heuven, V. J. (1995) Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English, *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 2, 630–633.
Stevens, K. N. (1977) Physics of laryngeal behavior and larynx modes, *Phonetica*, **34,** 264–279.
Stevens, K. N. (1998) *Acoustic phonetics*. Cambridge, MA: The MIT Press.
Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M. & Kurowski, K. (1992) Acoustic and perceptual characteristics of voicing in fricatives & fricative clusters, *Journal of the Acoustical Society of America*, **91**(5), 2979–3000.
Stevens, K. N. & Hanson, H. M. (1994) Classification of glottal vibration from acoustic measurements. In *Vocal fold physiology*: *voice quality control* (O. Fujimura & M. Hirano, editors), pp. 147–170. San Diego: Singular.
Traill, A. (1992) A confusion of sounds: the phonetic description of !Xũ clicks. In *African linguistic contributions presented in honour of Ernst Westphal* (D. F. Gowlett, editor), pp. 345–362. Pretoria: Via Africa.
Traill, A. & Jackson, M. (1988) Speaker variation and phonation type in Tsonga nasals, *Journal of Phonetics*, **16,** 385–400.
Traill, A., Khumalo, J. S. M. & Fridjon, P. (1987) Depressing facts about Zulu, *African Studies*, **46,** 255–274.
Wentzel, P. J., Botha, J. J. & Mzileni, P. M. (1972) *Xhosa-Taalboek*. Perskor: Johannesburg.
Westbury, J. R. (1983) Enlargement of the supraglottal cavity and its relation to stop consonant voicing, *Journal of the Acoustical Society of America*, **73**(4), 1322–1336.
Ziervogel, D. (1967—editor). *Handbook of the speech sounds and sound changes of the Bantu languages of South Africa*. University of South Africa, Pretoria.

Appendix A

TABLE AI. Significant results from first type of ANOVA (upper) and second type of ANOVA (lower) for parameter f_0, as explained in Section 2.3. First column: period; second: main effect for Category; third: main effect for Speaker; fourth: Category/Speaker interaction; fifth: other significant effects

	Category (plain, aspirated, voiced)	Speaker (F1–4, M1–4)	Category × Speaker	Other
1	$F(2, 112) = 27.278$, $p < 0.001$	$F(7, 112) = 142.069$, $p < 0.001$	$F(14, 112) = 2.511$, $p = 0.003$	
2	$F(2, 113) = 106.076$, $p < 0.001$	$F(7, 113) = 345.057$, $p < 0.001$	$F(14, 113) = 5.679$, $p < 0.001$	
3	$F(2, 114) = 115.476$, $p < 0.001$	$F(7, 114) = 331.445$, $p < 0.001$	$F(14, 114) = 4.646$, $p < 0.001$	
4	$F(2, 113) = 154.792$, $p < 0.001$	$F(7, 113) = 387.211$, $p < 0.001$	$F(14, 113) = 4.928$, $p < 0.001$	Category × Manner: $F(2, 113) = 4.111$, $p = 0.018$
5	$F(2, 114) = 48.015$, $p < 0.001$	$F(7, 114) = 217.380$, $p < 0.001$	$F(14, 114) = 2.317$, $p = 0.007$	
	Category (plain, voiced)	Speaker (F3–4, M3–4)	Category × Speaker	Other
1	$F(1, 69) = 30.956$, $p < 0.001$	$F(3, 69) = 97.579$, $p < 0.001$	$F(3, 69) = 9.208$, $p < 0.001$	Manner × Context: $F(1, 69) = 4.261$, $p = 0.042$
2	$F(1, 70) = 73.590$, $p < 0.001$	$F(3, 70) = 194.488$, $p < 0.001$	$F(3, 70) = 12.406$, $p < 0.001$	
3	$F(1, 70) = 98.197$, $p < 0.001$	$F(3, 70) = 263.468$, $p < 0.001$	$F(3, 70) = 15.319$, $p < 0.001$	Category × Manner: $F(1, 70) = 4.421$, $p = 0.039$
4	$F(1, 69) = 301.557$, $p < 0.001$	$F(3, 69) = 543.838$, $p < 0.001$	$F(3, 69) = 35.274$, $p < 0.001$	Context: $F(1, 69) = 12.799$, $p < 0.001$ Category × Context: $F(1, 69) = 6.051$, $p = 0.016$
5	$F(1, 70) = 175.283$, $p < 0.001$	$F(3, 70) = 465.845$, $p < 0.001$	$F(3, 70) = 15.166$, $p < 0.001$	Category × Context: $F(1, 70) = 7.489$, $p = 0.007$ Speaker × Manner × Context: $F(3, 70) = 3.420$, $p = 0.021$

TABLE AII. Significant results from first type of ANOVA (upper) and second type of ANOVA (lower) for parameter F_1

	Category (plain, aspirated, voiced)	Speaker (F1–4, M1–4)	Category × Speaker	Other
1	$F(2, 113) = 57.895$, $p < 0.001$	$F(7, 113) = 22.237$, $p < 0.001$	$F(14, 113) = 3.788$, $p < 0.001$	
2	$F(2, 114) = 23.193$, $p < 0.001$	$F(7, 114) = 20.208$, $p < 0.001$	$F(14, 114) = 2.691$, $p = 0.001$	
3	$F(2, 114) = 7.717$, $p < 0.001$	$F(7, 114) = 20.784$, $p < 0.001$		
4	$F(2, 114) = 5.287$, $p = 0.006$	$F(7, 114) = 21.134$, $p < 0.001$		
5		$F(7, 114) = 30.785$, $p < 0.001$		Manner: $F(1, 114) = 4.478$, $p = 0.036$
	Category (plain, voiced)	Speaker (F3–4, M3–4)	Category × Speaker	Other
1	$F(1, 70) = 4.351$, $p = 0.040$	$F(3, 70) = 25.981$, $p < 0.001$	$F(3, 70) = 4.335$, $p = 0.007$	
2	$F(1, 70) = 5.090$, $p = 0.027$	$F(3, 70) = 26.934$, $p < 0.001$		Manner × Context: $F(1, 70) = 4.137$, $p = 0.045$
3		$F(3, 70) = 23.674$, $p < 0.001$	$F(3, 70) = 3.425$, $p = 0.021$	Context: $F(1, 70) = 4.989$, $p = 0.028$
4		$F(3, 70) = 26.748$, $p < 0.001$	$F(3, 70) = 3.287$, $p = 0.025$	Context: $F(1, 70) = 5.320$, $p = 0.024$ Manner × Context: $F(1, 70) = 7.379$, $p = 0.008$
5		$F(3, 70) = 38.371$, $p < 0.001$		Manner: $F(1, 70) = 7.062$, $p = 0.009$ Speaker × Category × Manner × Context: $F(3, 70) = 3.243$, $p = 0.027$

TABLE AIII. Significant results from first type of ANOVA (upper) and second type of ANOVA (lower) for parameter H_1^*–H_2^*

	Category (plain, aspirated, voiced)	Speaker (F1–4, M1–4)	Category × Speaker	Other
1	$F(2, 112) = 58.979$, $p < 0.001$	$F(7, 112) = 33.987$, $p < 0.001$	$F(14, 112) = 5.765$, $p < 0.001$	
2	$F(2, 113) = 30.155$, $p < 0.001$	$F(7, 113) = 28.020$, $p < 0.001$	$F(14, 113) = 6.092$, $p < 0.001$	
3	$F(2, 114) = 15.270$, $p < 0.001$	$F(7, 114) = 23.890$, $p < 0.001$	$F(14, 114) = 4.686$, $p < 0.001$	
4	$F(2, 113) = 8.092$, $p < 0.001$	$F(7, 113) = 21.285$, $p < 0.001$	$F(14, 113) = 3.058$, $p < 0.001$	
5		$F(7, 114) = 32.129$, $p < 0.001$		
	Category (plain, voiced)	Speaker (F3–4, M3–4)	Category × Speaker	Other
1	$F(1, 69) = 149.607$, $p < 0.001$	$F(3, 69) = 13.093$, $p < 0.001$	$F(3, 69) = 7.881$, $p < 0.001$	Category × Manner: $F(1, 69) = 4.472$, $p = 0.038$
2	$F(1, 70) = 137.660$, $p < 0.001$	$F(3, 70) = 9.658$, $p < 0.001$	$F(3, 70) = 5.637$, $p = 0.001$	Manner: $F(1, 70) = 5.137$, $p = 0.026$
3	$F(1, 70) = 112.898$, $p < 0.001$	$F(3, 70) = 13.059$, $p < 0.001$	$F(3, 70) = 9.511$, $p < 0.001$	
4	$F(1, 69) = 72.013$, $p < 0.001$	$F(3, 69) = 13.577$, $p < 0.001$	$F(3, 69) = 12.233$, $p < 0.001$	
5	$F(1, 70) = 17.533$, $p < 0.001$	$F(3, 70) = 39.166$, $p < 0.001$	$F(3, 70) = 14.892$, $p < 0.001$	

TABLE AIV. Significant results from first type of ANOVA (upper) and second type of ANOVA (lower) for parameter H_1^*–A_3^*

	Category (plain, aspirated, voiced)	Speaker (F1–4, M1–4)	Category × Speaker	Other
1	$F(2, 112) = 21.517$, $p < 0.001$	$F(7, 112) = 49.107$, $p < 0.001$	$F(14, 112) = 2.272$, $p = 0.009$	Manner: $F(1, 112) = 4.095$, $p = 0.045$
2	$F(2, 113) = 12.647$, $p < 0.001$	$F(7, 113) = 72.788$, $p < 0.001$	$F(14, 113) = 2.646$, $p = 0.002$	
3	$F(2, 114) = 8.045$, $p < 0.001$	$F(7, 114) = 89.216$, $p < 0.001$	$F(14, 114) = 2.447$, $p = 0.004$	
4	$F(2, 113) = 6.152$, $p = 0.002$	$F(7, 113) = 93.999$, $p < 0.001$	$F(14, 113) = 2.055$, $p = 0.019$	
5		$F(7, 114) = 116.259$, $p < 0.001$		Manner: $F(1, 114) = 5.846$, $p = 0.017$
	Category (plain, voiced)	Speaker (F3–4, M3–4)	Category × Speaker	Other
1	$F(1, 69) = 6.607$, $p = 0.012$	$F(3, 69) = 40.890$, $p < 0.001$	$F(3, 69) = 6.290$, $p < 0.001$	Speaker × Manner: $F(3, 69) = 2.839$, $p = 0.044$
2	$F(1, 70) = 11.674$, $p = 0.001$	$F(3, 70) = 77.229$, $p < 0.001$	$F(3, 70) = 3.520$, $p = 0.019$	
3	$F(1, 70) = 16.712$, $p < 0.001$	$F(3, 70) = 123.029$, $p < 0.001$	$F(3, 70) = 4.969$, $p = 0.003$	Speaker × Category × Manner × Context: $F(3, 70) = 3.236$, $p = 0.027$
4	$F(1, 69) = 12.292$, $p < 0.001$	$F(3, 69) = 100.802$, $p < 0.001$	$F(3, 69) = 3.902$, $p = 0.012$	Speaker × Manner: $F(3, 69) = 2.907$, $p = 0.040$
5		$F(3, 70) = 124.212$, $p < 0.001$		

Appendix B

TABLE BI. Means and standard deviations (in parentheses) for the postnasal data. First column: acoustic parameter; second: data pool (female = F3, F4, male = M3, M4, all = F3, F4, M3, M4); remaining columns: values for plain and voiced postnasal stops/clicks (pooled) for each of the five periods (indicated in top row). This table contains the information for postnasal position corresponding to the postvocalic data shown in Figs 1, 2, 3, 5

		1		2		3		4		5	
Parameter	Period Data from	Plain	Voiced	Plain	Voiced	Plain	Voiced	Plain	Voiced	Plain	Voiced
f_0	Female	244 (52)	194 (37)	236 (42)	183 (22)	244 (40)	186 (21)	254 (25)	182 (17)	242 (22)	179 (15)
f_0	Male	141 (20)	120 (29)	129 (14)	110 (23)	125 (15)	107 (20)	131 (18)	103 (19)	120 (20)	96 (17)
F_1	All	690 (117)	636 (100)	731 (115)	669 (80)	744 (102)	703 (86)	757 (107)	716 (111)	812 (132)	784 (150)
H_1^*–H_2^*	Spk. F3	−0.5 (2.7)	11.3 (6.2)	−0.5 (1.9)	9.0 (5.5)	0.5 (1.6)	9.5 (5.0)	1.1 (1.6)	9.3 (4.7)	3.7 (1.5)	6.1 (3.2)
H_1^*–H_2^*	Spk. F4	−2.9 (3.6)	8.5 (4.0)	−3.6 (2.4)	8.6 (3.6)	−2.3 (2.2)	10 (3.7)	−0.3 (1.5)	12.1 (4.4)	1.2 (2.4)	9.2 (2.7)
H_1^*–H_2^*	Spk. M3	−2.7 (1.2)	0.7 (1.5)	−2.6 (0.6)	0.0(1.6)	−1.8 (1.0)	−0.6 (2.1)	−1.6 (1.0)	0.8 (2.4)	−1.3 (1.0)	−0.8 (1.5)
H_1^*–H_2^*	Spk. M4	−0.6 (1.5)	3.9 (3.4)	−0.3 (1.3)	5.5 (3.9)	0.3 (1.6)	5.6 (2.8)	3.0 (2.5)	4.7 (3.8)	4.0 (2.7)	2.8 (3.6)
H_1^*–A_3^*	Spk. F3	7.8(4.1)	15.6 (1.7)	11.1 (4.7)	17.8 (2.0)	11.3 (4.1)	16.8 (2.9)	13.2 (4.0)	19.0 (3.9)	17.8 (7.5)	21.0 (1.9)
H_1^*–A_3^*	Spk. F4	8.3 (7.1)	11.6 (4.5)	10.3 (7.3)	13.2 (3.8)	12.2 (5.5)	14.0 (3.8)	12.7 (3.4)	13.9 (4.6)	17.6 (2.0)	13.3 (5.9)
H_1^*–A_3^*	Spk. M3	−2.0 (6.4)	−5.0 (3.0)	−6.6 (4.7)	−4.0 (4.3)	−6.7 (3.7)	−4.5 (5.4)	−5.8 (3.2)	−2.6 (4.3)	−5.4 (3.4)	−3.5 (5.8)
H_1^*–A_3^*	Spk. M4	3.8 (6.5)	9.6 (9.1)	8.0 (6.3)	7.0 (3.6)	6.9 (3.3)	8.9 (4.2)	4.2 (4.3)	6.7 (5.5)	6.5 (6.9)	8.2 (4.9)