CHAPTER 3

THE SOUNDS OF THE BANTU LANGUAGES

Ian Maddieson and Bonny Sands

1 INTRODUCTION

Among phoneticians, the Bantu languages have a reputation as not having many interesting features, with the exception of the clicks introduced in some languages of the southern area. Although it’s true that many languages within the Bantu group are phonetically quite similar to each other, there is considerably more diversity in their phonetic patterns than is often believed. Some of this diversity may be disguised by the widespread use of simplifying transcriptions and orthographies which normalise away variation within and between languages or underrepresent distinctions. Part of the aim of the present chapter is therefore to draw greater attention to this diversity. Since the Bantu languages have received very extensive historical analysis, this group of languages also provides a fertile field for examining inferences about the nature of phonetic sound change. The phonetic differences which exist between closely related languages provide opportunities for testing theories about phonological organisation.

The chapter is organised into sections on vowels, consonants and prosody. There are many important interactions between these three aspects of phonetic structure and some of these will be taken up at the point where it seems appropriate to do so. For example, the Bantu languages provide very striking examples of vowels affecting consonant realisations, particularly considered diachronically, and the nature of particular segments also has significant impacts on prosodic quantity and on tonal patterns. Special attention is paid to consonants with complex articulations, including clicks and the so-called “whistling fricatives.” It is hoped that the brief discussions of selected issues here will encourage more attention to be paid to phonetic aspects of these languages.

2 VOWELS

2.1 Vowel spacing

The majority of Bantu languages – with some notable exceptions, particularly in the North-West – have simple-looking systems of five or seven vowels in which the expected relationships between the features of vowel height, backness and rounding hold. That is, the back non-low vowels are rounded, and the low and front vowels are unrounded. The vowels of the five-vowel systems are therefore usually transcribed as /i e a o u/ and the seven-vowel systems are most often transcribed as /i e e a o o u/ (Hyman 1999). However, these standardised transcriptions may disguise significant differences between languages, especially with respect to the nature of the vowels written /e/ and /o/.
In the five-vowel system of Xhosa S41, for example, /e o/ are genuinely mid in character. The positions of vowels in an acoustic space are often shown by plotting values of the first two formants. Readers unfamiliar with acoustic analysis might see Ladefoged (2000) for an introduction to the concept of a formant. The mean formant values for Xhosa S41 vowels given by Roux and Holtzhausen (1989) are plotted in this way in Figure 3.1. In this and following figures of the same type, the origin of the axes is in the upper right, with first formant (F1) values increasing down from the origin, and second formant (F2) values increasing to the left. The distances along the axes are scaled to reflect auditory/perceptual intervals; F2 is plotted using a logarithmic scale. This kind of display closely parallels the traditional auditorily based vowel space based on perceived “height” and “backness” values used, for example, in the IPA Handbook (1999), but has the advantage of being based on verifiable measurement. In Figure 3.1, it can be seen that in Xhosa S41 /e o/ are located almost equidistant from the high vowels /i u/ and the low vowel /a/. There is a raising process in Xhosa S41, which results in higher variants of /e o/ when /i u/ occur in the next syllable. The means for /e o/ plotted here do not include tokens of these raised variants.

Compare the spacing of Xhosa vowels with those of Kalanga S16, shown in Figure 3.2. The maxima in Figure 3.2 are higher compared to Figure 3.1 due to male/female differences in formant range. Note that there are different ways to normalise vowels across...
speakers; eight different methods are evaluated by Wissing and Pienaar (2014) for a corpus of Southern Sotho S33 vowels, for instance. In the case of Kalanga S16, the “mid” vowels /e o/ are relatively close to the high vowels /i u/ and far from /a/. As a rough rule of thumb, vowels with a first formant lower than 400 Hz may be considered high vowels in a female voice. On this basis these particular vowels would not quite justify being considered high, but they are clearly markedly higher than those of Xhosa S41.

There have been relatively few acoustic studies of other Bantu five-vowel systems, but Swahili G42 (Nchimbi 1997) has a pattern similar to Xhosa S41, while Bemba M42 (Hamann & Kula 2015), Ndebele S44 and the Zezuru variety of Shona S12 (Manuel 1990) have a pattern similar to Kalanga. The distribution seen in Xhosa S41 or Swahili G42 is similar to that most typically found cross-linguistically in five-vowel systems transcribed /i e a o u/, such as Spanish, Hadza or Hawaiian. This is also the pattern predicted by computational models of vowel system structure from Liljencrants and Lindblom (1972) to Schwartz et al. (1997) based on the principle that vowels should be expected to be roughly equally dispersed in a space defined by the major formant resonances. In Kalanga S16, on the other hand, the vowels are crowded into the upper part of the vowel space, with the front pair in particular being very close together.

A Bantu five-vowel system consisting of /i e a o u/ has been described for Soga JE16 (Nabirye et al. 2016) and Fwe K402 (Gunnink 2016). Means of Fwe vowel formants are shown in Figure 3.3. The high vowels /i/ and /u/ are lower and more centralised than those in Xhosa S41 and Kalanga S16 and could be transcribed [ɨ] and [ʊ], respectively. Both Soga JE16 and Fwe K402 have a vowel length contrast. Impressionistically, there appear to be no differences in vowel quality between pairs that differ in length in the two languages.

Variations in the structure of seven-vowel systems occur which are similar to those of the five-vowel systems. A plot of vowel distribution in Nyamwezi F22 is shown in Figure 3.4. In this language, the vowels are to a large degree placed where they might be expected, given a respect for dispersion principles. This is particularly apparent for the front vowels, which are equally spaced from each other. The word list available for measurement included a more balanced sample of front than of back vowels, and the back vowels are probably in reality more separated than this plot indicates. These data suggest that transcription of this vowel set as [i e a ə o u], as in Figure 3.4, is appropriate rather than the [i i e a o o u] preferred by Maganga and Schadeberg (1992). For Sukuma F21, Batibo (1985) also provides acoustic evidence for a relatively wide separation of the seven vowels, with /e ə o/ all being clearly mid vowels.

The relationship between the seven vowels of Vove B305 is notably different, as demonstrated in Figure 3.5. Here a pair of vowels in the front and a pair of vowel in the back have such low values of F1 that they are all appropriately considered to be high vowels. The means are 248 Hz for /i/, 313 Hz for /i/, 277 Hz for /u/, and 334 Hz for /ʊ/. The next lower vowels are markedly lower. In addition we may note that the front pair /i/ and /i/ and the back pair /u/ and /u/ have F2 values which are identical or nearly so, whereas Nyamwezi F22 /e o/ have F2 values intermediate between the higher and lower vowels in the system.

The acoustic phonetic characteristics of the eight- and nine-vowel systems of some Mbam languages (A40+A60) are detailed in Boyd (2015). Mande A46, Nen A44 and Gunu A622 all have an eight-vowel system with [-ATR] /i a ə o/ and [+ATR] /i a o u/. Nine-vowel languages in the Mbam group, such as Mmala A62B and Baca A621, have a contrast between /e/ and /e/ not found in the eight-vowel systems. The F1 averages of /i/ and /i/ in Mbam languages is typically higher than that of /e/ and /e/. The [-ATR] high
**FIGURE 3.3** FWE Vowel Formant Means According to Measurements by the Second Author on Recordings Made Available by Hilde Gunnink. Each point represents the mean of between seven and 27 tokens of unreduced stem-initial vowels spoken by a male speaker.

**FIGURE 3.4** Nyamwezi F22 Vowel Formant Means According to Measurements by the First Author. Each point represents the mean of between nine and 23 tokens of unreduced final or penultimate vowels in a word list spoken by a male speaker.

**FIGURE 3.5** Vove B305 Vowel Formant Means According to Measurements by the First Author on a Recording Made by Jean-Marie Hombert, Made Available by Lolke van der Veen. Each point represents the mean of 20 or 30 measurements on minimal sets of words differing only in the penultimate vowel, spoken by a male speaker.
vowels may thus be misinterpreted as being lower than the [+ATR] mid vowels, but the high F1 values may be instead attributed to a retracted tongue root position. Another nine-vowel Bantu language is Liko D201 (De Wit 2015: 45).

### 2.2 Tongue root position in harmony systems

Co-occurrence restrictions of a harmonic nature between vowels, very typical of sub-Saharan African languages, are quite commonly found in Bantu languages, though often limited in extent, e.g., only applying in certain morphological contexts, such as between verb roots and extensions. Bantu vowel harmony constraints do not seem to be a survival of an older Benue-Congo or even Niger-Congo harmony (Stewart 2000), but to be mostly more or less local innovations with diverse patterns of implementation (Hyman 1999). Vowel height, backness and rounding can all be factors in control of Bantu harmony. Vowel harmony in Africa often involves the independent use of pharyngeal cavity size, that is, adjustments of pharynx volume which cannot be accounted for as a function of the height and frontness of the tongue body (see Ladefoged & Maddieson 1996 for discussion). This is usually discussed as a contrast between advanced and retracted (or neutral) tongue root position, i.e., ± ATR. An interesting issue is therefore whether the Bantu languages, particularly those with seven or more vowels, make use of the ATR feature in this phonetic sense. Phonologists often use [ATR] as a diacritic feature, even to distinguish pairs of vowels such as i/ɪ in English ‘beat’/‘bit’ where tongue root position is not the phonetic mechanism involved. The question of the role of ATR interacts with the question of the nature of the high vowels, as the *super-high/*high contrast might have been an expression of an ATR contrast or transformed into one in daughter languages.

It is difficult to be certain that ATR contrasts exist in a language unless direct articulatory data on the vocal tract configuration during vowel production is available. There are very few studies of this type available so far for Bantu languages, but one data set is shown in Figure 3.6. These pictures are magnetic resonance images of sustained vowels produced by Pither Medjo Mvé, a speaker of the Bitam variety of Fang A75 (Demolin et al. 1992). Figure 3.6 shows very clearly that independent tongue root adjustment does not contribute to the distinctions between any members of the front vowel set /i ɛ e/ nor the back vowel set /u o ɔ/. The pharynx width, measured as the distance from the tongue root surface to the back wall of the pharynx at the height of the top of the epiglottis, in /e/ is intermediate between that in /i/ and /ɛ/, and that in /o/ is intermediate between /u/ and /ɔ/. It can be predicted from tongue body position: front vowels have wider pharynx than back vowels, lower vowels have narrower pharynx than higher vowels. The three front vowels and the three back vowels can therefore be distinguished one from another solely by height.

The Bitam variety of Fang A75 has eight vowels and seven peripheral vowels, plus mid central /ə/ (Medjo Mvé 1997). An acoustic plot of these vowels is given in Figure 3.7. Note particularly the slope of a line connecting the back vowels which points roughly to the position of the central vowel /a/, similar to that seen in Figure 3.1 and Figure 3.2, and attributable to the fact that F1 and F2 frequencies co-vary in these vowels. This pattern is typical of that found in vowel systems where the back series is distinguished by degrees of height with no other factors being significantly involved. In this variety, lexical stems are marked by a strong tendency for V1 and V2 to be identical except if V2 is /a/, when /i a a o u/ are all relatively common as V1, but /e e ɔ/ are not. Note that as many PB
FIGURE 3.6 ARTICULATORY POSITIONS OF SIX OF THE VOWELS OF FANG A75 (VARIETY OF BITAM). TOP ROW, FRONT VOWELS /i/, /e/, /ɛ/; BOTTOM ROW, BACK VOWELS /u/, /o/, /ɔ/.

Source: Mid-sagittal MRI scans of isolated vowels, made available by Didier Demolin.

FIGURE 3.7 FANG A75 VOWEL FORMANT MEANS. EACH POINT REPRESENTS THE MEAN OF SIX MEASUREMENTS, THREE OF ISOLATED VOWEL TOKENS, PLUS THREE TOKENS IN FINAL VOWELS IN /alV/ NONSENSE WORDS.

Source: Recording by Pither Medjo Mvé made available by D. Demolin; measurements by the first author.
*CVCV items have become monosyllabic in Fang, the V2 in these cases is often not the *V2 of the reconstructed form. This pattern of co-occurrences is not one which suggests a phonological role for ATR.

Nande JD42 contrasts with Bitam Fang A75 in that it uses ATR for phonetic distinctions. Ultrasound images clearly show differences in tongue root position across vowel pairs (Gick 2002, Gick et al. 2006), as seen for the ATR /e/ and RTR /e/ vowels in parts a) and b) of Figure 3.8. Note that the back of the mouth is found on the left side of an ultrasound image, but on the right side of an MRI image. The ATR vowel /e/ and the RTR vowel /e/ differ both in the shape of the tongue body and in the amount of tongue root retraction, which can be estimated by the volume of tongue mass which occurs to the left of the white dotted line. The RTR vowel shows more tongue root retraction than the ATR vowel while the ATR vowel (on top) shows more of a bunched tongue shape.

The ATR/RTR contrast in Nande JD42 is also suggested by the harmonic behaviour and acoustic characteristics of vowels. Mean formant values of the 10 surface vowels for one speaker are plotted in Figure 3.9. Harmonically related pairs are noted by the use of the same symbol with and without a -ATR diacritic. In each case the putatively [-ATR] vowel has a substantially higher first formant (hence a lower position on the chart) than its harmonic counterpart. Most strikingly, the “high” vowels /i u/ are placed lower than the “mid” vowels /e o/. Narrowing the pharynx raises the first formant, other things being equal. The pair /u u/ where F2 is the same are thus quite likely (almost) solely different in pharynx width. The other back vowel pair /o o/ shows a smaller than expected F2 difference given the size of the difference between their first formants; a substantial pharynx width difference coupled with a degree of opening of the oral constriction may be inferred. Note that a sloping line can be fit to the vowel set /u o a/ and a second roughly parallel lower one to the set /u o a/, but a straight line cannot be fit to the set /u o a/ as is possible for Fang A75 /u o a/. The pattern for the front vowels suggests a greater interaction of the major features of vowel height and backness with pharynx width. This is not surprising, as retracting the tongue root is more likely to pull the tongue back and down when the tongue body position is front.

We may now revisit the Kalanga S16 and Vove B305 high vowels in Figure 3.2 and Figure 3.5. The members of the high vowel pairs /i i/ and /u u/ in Vove B305 have virtually the same second formant values as each other and differ only in F1. The Kalanga S16 vowel pairs transcribed /i i/ and /u u/, which are acoustically equally as high as the Vove B305 pairs, differ in both F1 and F2. Plausibly, the Vove B305 vowel pairs differ phonetically in pharynx width, which is consistent with the auditory impression they create, while the Kalanga S16 pairs differ in height and to a lesser degree in backness, which is consistent with the auditory impression they create. Although these acoustic measurements are suggestive, it should be borne in mind that inferences from simple formant measures concerning vowel articulation must be made with caution.

Acoustic evidence for tongue root retraction of vowels in several Bantu languages has been provided by Starwalt (2008). She found that [-ATR] vowels with a constricted voice quality tend to have higher center of gravity values, while [+ATR] vowels with a “hollow” quality have lower center of gravity values (Starwalt 2008: 441). In her study, F1, B1 (F1 bandwidth), center of gravity and A1-A2 (relative amplitudes of F1 and F2) help distinguish vowel pairs that differ in [ATR] value to varying degrees depending on the vowel pair and speaker. This suggests that speakers of the same language may differ in the degree to which they use tongue root position to contrast vowels that are described as differing in the phonological feature [ATR].
FIGURE 3.8 ULTRASOUND IMAGES OF NANDE JD42 VOWELS A) ATR /e/ B) RTR /e/, TAKEN ALONG THE MID-SAGITTAL PLANE. NOTE THAT THE TONGUE TIP IS ON THE RIGHT AND THE TONGUE ROOT ON THE LEFT, THE REVERSE OF THE IMAGES IN FIGURE 3.6. THE TONGUE SURFACE APPEARS AS A CURVED WHITE LINE. A VERTICAL WHITE DOTTED LINE HAS BEEN ADDED TO FACILITATE COMPARISON BETWEEN THE TWO IMAGES.

2.3 Pharyngealised vowels

Pharyngealised vowels /iˤ uˤ ɛˤ oˤ/ have been reported for Kwasio A81 (Duke & Martin 2012). These vowels are produced with a retracted tongue root, causing a constriction in the upper pharynx. The upper pharyngealised vowels of Kwasio A81 do not have the harsh voice quality associated with lower pharyngealised vowels, i.e., epiglottalised or aryepiglottalised vowels, as found in Tuu, Kx’a and Khoe languages of southern Africa (cf. Miller 2007, Miller et al. 2009a). Pharyngealised vowels occur in a few other Bantu languages including Gyele A801 (Blench 2011) and Jarawan Bantu (Rueck et al. 2009, cited in Blench 2015).

Kwasio A81 pharyngealised vowels differ significantly in vowel quality compared to their non-pharyngealised counterparts. For instance, /uˤ/ and /oˤ/ are produced as the lower and more centralised vowels [ʊˤ] and [ɞˤ], respectively (Duke & Martin 2012: 220). Figure 3.10 shows a typical example of /o/ in the word /ko/ ‘to go’; /o/ has a low F2 (below 1000 Hz). The pharyngealised /oˤ/ in Figure 3.11 in the word /koˤ/ ‘avarice’ has a higher F2 (above 1000 Hz), and the higher formants are much more prominent than those of /o/. The pharyngealised vowel is longer than the plain vowel, which reflects the origin of the pharyngealisation from a reduced velar stops in C1 position in roots of the shape C VC V (Duke & Martin 2012: 220). Pharyngeals have developed from velars in other Niger-Congo languages. In the Bantoid language Mundabli (Völl 2012: 535), pharyngealised vowels correspond to final /k/ and /ʔ/ in cognates in its close relative Mufu. In the Gur language Minyanka, the pharyngeal fricative [ʕ] is a variant of /ɡ/ (Dombrowsky-Hahn 1999: 52).

Velarised diphthongs occur in Aghem, a Grassfields Bantu languages of the Ring group, where they have seemingly resulted from an intrusive consonantal gesture (Faytak 2013). This differs from Kwasio A81 pharyngealisation which likely results from the reduction of a consonant.
2.4 Nasalised vowels

Nasal vowels are not particularly common in the Bantu languages, but are found in certain mostly western areas, for example in Ngungwel B72a of the Teke group (Paulian 1994), in Umbundu R11 (Schadeberg 1982), in Gyele A801 (Renaud 1976) and in a few words in the Bitam variety of Fang A75 (Medjo Mvé 1997). As is generally the case cross-linguistically, there are fewer nasal vowels than oral ones, at least in lexical stems. In Ngungwel B72a, there are three oral and three nasal vowels in prefixes [e o ẽ ã õ]. Lexical stems have a system of seven oral vowels but only five nasal vowels. Nasal vowels in the stem are reported to have the qualities [ĩ ë ã õ ũ] and to be invariably long. The article of Paulian (1994) does include a few words with short nasalised vowels in stems, but these may be misprints. There are thus seven phonetic qualities among the nasalised vowels, but no contrast between all seven in any environment. Examples are given in Table 3.1.

A role for vowel nasalisation in the transmission of nasal consonant harmony across intervening vowels seems likely in the history of Bantu (Greenberg 1951, Hyman 1995).
2.5 Vowel length

Vowel length contrasts occur in some Bantu languages, which may or may not be accompanied by changes in vowel quality and/or various processes of vowel lengthening (cf. Sections 6-7). Bemba M42 short vowels /i e a o u/ tend to be lax compared to their long vowel counterparts /iː eː aː oː uː/ (Hamann & Kula 2015): short high and mid vowels tend to be lower and more centralised than long ones, while /a/ is higher than /aː/. Rwanda JD61 contrasts long and short vowels yet also has vowel lengthening before NC as well as after a consonant-glide sequence (Myers 2005). Another language with a vowel length contrast, Vili H12, lengthens vowels before liquids (in the environment /C_L/) and before nasals /C_N/ (Roux & Ndinga-Koumba-Binza 2011), but not before NC (Ndinga-Koumba-Binza 2011).

2.6 Super-close vowels

Fricated vowels occur in Kom and Oku, two Grassfields Bantu languages of the central Ring group (Faytak 2014, Faytak & Merrill 2014), as well as in several Bantoid languages of the northern Cameroon Grassfields (Faytak 2015). These vowels bring to mind the super-high or super-close vowels /i̧ u̧/ used by Meeussen (1967, 1969) and Guthrie (1967, 1970a, 1970b, 1971) and notated as /î û/ by Meinhof (1899), in addition to “normal” high /i u/. Super-close vowels were reconstructed in order to account for the set of sound changes known as “Bantu Spirantization,” but recent reconstructions have abandoned this explanation (Schadeberg 1995, Bastin et al. 2002, Bostoen 2008). Faytak (2014) reconstructs back vowels *u, *o, *o for the Central Ring group of Grassfields Bantu languages. He argues that, in two of these languages, Kom and Oku, *o raised to /u/ and *u became fricated, sometimes occurring with a schwa [ə] offglide (Faytak & Merrill 2014).

Matumbi P13 has been claimed to have super-close vowels /i̧ u̧/ (Odden 1996: 5), but the description of the contrast between /i̧ u̧/ and /i u/ as being “roughly equivalent to the contrast between [ɛ], [o] and [i], [u]” suggests that the vowels likely contrast tongue root position (ATR) rather than tongue height.

A particularly rare phenomenon reported in Hendo C82 involves the class 5 prefix, which is actually the reflex of the Proto-Bantu augment *di- followed by the noun prefix *i- (cf. De Blois 1970: 155). Demolin et al. (2002) describe it as an unreleased voiced palatal implosive [ʄ] before a voiceless stop or affricate, e.g., in [ʄpaka] ‘moth.’ MRI scans indicate that this segment is appropriately viewed as a hyperarticulation of the vowel /i/.

3 CONSONANTS

3.1 Consonant overview

Most Bantu languages are reported as having two series of plosives, voiced and voiceless, and this follows the Proto-Bantu reconstruction of Meeussen (1967). Except in post-nasal environments and sometimes before his reconstructed super-high vowels, the reconstructed voiced plosives most commonly correspond to voiced continuants of one type or another or to implosives in the modern languages. An alternation of some kind is probably to be reconstructed to an early stage, possible even pre-Bantu. Bantu orthographies usually do not indicate these alternations, unless subsequent developments have created
a contrast between, say, /b/ and /ɓ/, or /b/ and /β/. This illustrates one instance where the occurrence of cross-linguistically less common phonetic segments may be disguised by notational practices. Aspiration is a contrastive property of voiceless stops (and affricates) in some languages where it is often a reflex of an earlier voiceless prenasalised stop (cf. Kerremans 1980).

Most Bantu languages have a full set of nasals at each place of articulation where a stop or affricate appears, but often intricate (morpho)phonological processes govern nasal/oral alternations and syllabification and other prosodic processes concerning nasals. Most of the languages have relatively limited sets of fricatives of the cross-linguistically common types, although lateral fricatives (and affricates) have developed in or been borrowed into a number of the southern languages, such as Sotho-Tswana S30, Xhosa S41 and Zulu S42. Particularly striking in this connection is the velar ejective lateral affricate \[kʟ̥\] of Zulu S42 (cf. Naidoo 2007), which is auditorily reminiscent of a lateral click. There is often only one contrastive liquid, i.e., /l/, /ɾ/ or /r/, though Chaga E60 is among those with more (Davey et al. 1982, Philippson & Montlahuc 2003). High front vowels condition tap allophones of /l/ in Ganda JE15 (Myers 2015) and Tsonga S53 (Bennett & Lee 2015), and of /ɾ/ in the Washili variety of Ngazidja G44a (Patin 2013). The two vocoid approximants /j/ and /w/ occur in many languages, often alternating with high vowels /i u/.

Phonetic studies of labial consonants include the study of plain and prenasalised bilabial trills /b mɓ/ in Medumba, a Narrow Grassfields language, by Olson and Meynadier (2015). Differences in lip posture appear to enhance the contrast between labio-dental /f v/ and labial fricatives /ɸ β/ in Kwangali K33 and in Manyo K332 (Ladefoged 1990). Labial flaps reportedly occur in various Southern Bantu languages, such as Nyanja N31a, Korekore S11, Manyika S13, Ndau S15 and Kalanga S16, and they may contrast with the labio-dental approximant /ʋ/ in the Zezuru S12 variety of Shona (Olson & Hajek 1999).

Although most Bantu languages use only one coronal (typically alveolar) and one dorsal (velar) place of articulation, contrasts between dental and alveolar places are found in several languages, and contrasts between velars and uvulars are found in Kgalagari S311 (Dickens 1987, Monaka 2001, 2005). Tswana S31 has a voiceless uvular affricate and voiceless uvular fricative (Bennett et al. 2016). A voiced pharyngeal fricative /ʕ/ is found in Nyokon A45 (Lovestrand 2011). Consonant gemination has developed through internal processes in languages such as Ganda JE15 (Clements 1986) and by contact with Cushitic languages in Ilwana E701 (Nurse 1994).

In the rest of this section, three of the particular issues of phonetic interest are discussed: the dental/alveolar place contrast, the possible occurrence of articulatorily complex consonants, and the nature of the so-called ‘whistling fricatives.’ Longer sections of the chapter will be devoted to aspects of laryngeal action in consonants, to the description of clicks and their distribution in Bantu, and to some of the interesting aspects of nasality which occur in these languages.

### 3.2 Dental vs. alveolar place of articulation

The phonetic realisation of dental and alveolar consonants is dependent on the airstream mechanism. Dental and alveolar implosives and clicks may display constriction patterns that differ from those of corresponding pulmonic stops. Electropalatography (EPG) of Mvita Swahili G42b shows that implosive /ɗ/ has a more retracted occlusion than
pulmonic /t/ and /ɗ̪/ has a shorter occlusion than /g/ (Hayward et al. 1989: 54). The (post: 303) alveolar // clicks in Zulu S42 (Thomas-Vilakati 2010) and Xhosa S41 (Doke 1926: 303) are retracted in comparison to pulmonic alveolar consonants such as /tʰ/ and /s/.

The typical pattern for dental/alveolar contrasts is that the dentals are laminal while the alveolars are apical. Soga JE16 follows this pattern, as shown in the palatograms in Figure 3.12 and Figure 3.13 (Nabirye et al. 2016). The width of the constriction for the apical alveolar nasal in Figure 3.12 is narrower than the width of the constriction of the laminal dental in Figure 3.13, as indicated by the positioning of the horizontal white lines superimposed on each photograph.

### 3.3 Complex or simple consonants?

Doubly articulated labial-velar stops (and nasals) are found almost exclusively in the languages of Africa, but they occur in only relatively few of the Bantu languages, including Londo A11 (Kuperus 1985), “Sawabantu” languages of Guthrie’s groups A10–20–30 (Mutaka & Ebobissé 1996–1997), Fang A75 (Medjo Mvé 1997), and Mijikenda E70 (Nurse & Hinnebusch 1993, Kutsch Lojenga 2001) among others. However, from the phonetic point of view, the Bantu languages have fewer articulatorily complex consonants than is sometimes suggested. An interesting process of intensification of secondary articulations into obstruents occurs, inter alia, in Rwanda JD61 (Jouannet 1983) and Shona S10 (Doke 1931a). This process does not result in double articulations that are almost totally overlapped, as in labial-velars, but sequential articulations which are overlapped either not at all or no more than is typical of sequences such as /tk/ or /pk/ in

---

**FIGURE 3.12** PALATOGRAM OF [ana] SPOKEN BY A SOGA JE16 SPEAKER. THE WHITE HORIZONTAL LINES INDICATE THE WIDTH OF THE MAXIMUM CONSTRICION.


**FIGURE 3.13** PALATOGRAM OF [ana] SPOKEN BY A SOGA JE16 SPEAKER. THE WHITE HORIZONTAL LINES INDICATE THE WIDTH OF THE MAXIMUM CONSTRICION.

English words like ‘fruitcake’ or ‘hopkiln.’ On the other hand, it does produce rather unusual consonant sequences in onset positions.

Examples of the Rwanda JD61 strengthening of an underlying /u/ or /w/ into a velar stop after a non-homorganic nasal or stop are illustrated by the spectrograms in Figures 3.14–3.16. As these show, the first segment is released before the closure for the second is formed. When the sequence is voiced, as in /mg bg/, a quite marked central vocoid separates the two segments. When the sequence is voiceless, as in /tk/, there is a strong oral release of the first closure. There is no overlap in the closures for the two segments, except optionally in the case of the nasal sequence /mp/. Somewhat similar facts have been shown for the Zezuru S12 variety of Shona (Maddieson 1990). However, as was observed long ago by Doke (1931b, 1931a), the phonetic patterns vary quite considerably across the different varieties of Shona S10.

A particularly interesting claim is made by Mathangwane (1999) concerning her pronunciation of parallel forms in Kalanga S16. She suggests that elements like the /pk/ which evolves from earlier or underlying /pw/ are pronounced with almost fully overlapped closures and their duration is similar to that of simple /k/ and /p/ segments, i.e., they are [p̊k, b̊g]. She reports that the labial closure is formed first. This would therefore be an important counter-example to the more common pattern found in labial-velar doubly articulated segments in other languages in which the labial closure is formed very slightly later (10–15 ms) than the velar one. The one spectrogram of a word containing /pk/ published in this study actually shows that the duration of the element is considerably longer than a simple stop, suggesting it contains a sequence of articulations, although no burst is visible for the /p/. Recordings made by the first author of two other female speakers of Kalanga S16, one from Francistown in Botswana and one from Zimbabwe, did not replicate the pattern suggested by Mathangwane. For example, the word meaning...
‘ampit,’ transcribed by Mathangwane as [hapʰkʰa], could receive three pronunciations – [hakʰwa] with no labial closure, [hapxa] with a labial stop followed by a fairly long velar fricative, or [hapkʰa] with a sequence of stops with clearly separate releases, as illustrated in Figure 3.17. This third pronunciation was characterised by one of the speakers as being more typical of speakers of 50 or more years of age. Evidently more study of the phonetic and sociolinguistic variation in this area would be of great interest.

3.4 ‘Whistling’ fricatives

Shona S10 and Kalanga S16 are also marked by the occurrence of a type of labialisation co-produced with alveolar fricatives which have led to these segments being named “whistled,” or “whistling fricatives” (Doke 1931a, Bladon et al. 1987). Unlike “ordinary” labialisation, which involves rounding and protrusion of the lips accompanied by a raising of the tongue back, i.e., a [w]-like articulation, this labialisation involves primarily a vertical narrowing of the lips with little or no protrusion and no accompanying tongue back raising. The gesture is also timed differently from “ordinary” labialisation in that it covers the fricative duration rather than being primarily realised as an offglide; hence “whistling fricatives” can themselves be labialised in their release phase. Similar segments are very rare in the world’s languages, but do occur in the Dagestani language Tabasaran (Kodzasov & Muravjeva 1982).

A detailed study of a weakly “whistled” fricative in Tsonga S53 shows that the narrowed lip posture is accompanied by a retroflex lingual gesture and thus may be transcribed with a retroflex fricative symbol [ʂ], e.g., [s̥ɪr̥á] ‘disasters’ (Lee-Kim et al. 2014). The whistled fricative has more peaked and compact spectra than its non-whistled counterpart, and the fricatives also differ in other acoustic measures. Although lip positions have not been reported for Tshwa S51, the acoustic findings are similar to those in Tsonga.
S53 in that the “whistling” fricatives have narrower spectral peak bandwidths and lower spectral peak frequencies when compared to their non-whistled fricative counterparts (Shosted 2006).

In Changana S53, “whistling” fricatives occur with a rounded lip posture (Shosted 2011) rather than the narrowed lip posture seen in Shona S10, Kalanga S16 and Tsonga S53. Since a rounded lip posture can also be seen in non-whistled fricatives, such as in the sequence [usu], the labial constriction alone cannot account for the whistle-like concentration of the frication noise, but it must be due to a particular linguopalatal configuration that is yet undescribed. This study of Changana S53 “whistling” fricatives underscores the fact that the phonetic realisation of a cross-linguistically rare sound may differ from one language to the next. “Whistling” fricatives are very rare cross-linguistically, but they do occur in Mozambican Portuguese (Ashby & Barbosa 2011), clearly due to the influence of Bantu languages.

4 LARYNGEAL ACTION IN CONSONANTS

4.1 Voice onset time

Though most Bantu languages are reported as having voiced and voiceless series of plosives, three-way contrasts in plosives based on Voice Onset Time (VOT) do occur. Engstrand and Lodhi (1985) study one such contrast in Swahili G42 and Monaka (2001, 2005) examines a three-way contrast in Kgalagari S311. Monaka’s detailed study combines acoustic data with data about larynx height and vocal fold vibrations obtained using a laryngograph. VOT differs, as expected, between voiced, voiceless unaspirated and aspirated stop categories in Kgalagari S311, and it also varies by place of articulation within each category. She shows that voiceless palatal and velar stops tend to have longer VOT measurements than bilabial, dental or uvular stops (Monaka 2005). Voiced stops tend to be made with a downward movement of the larynx, presumably to help sustain voicing (Monaka 2001). She also uses electropalatography (EPG) to show the susceptibility of stops to coarticulation varies not only by place of articulation, but also according to voice category; aspirated stops are the least susceptible to coarticulation and voiced stops are the most (Monaka 2001). Other studies of coarticulation in Bantu languages have not looked at voicing contrasts (Manuel 1987, Beddor et al. 2002, Malambe 2015), but Dogil and Roux (1996) argue that ejectives and clicks in Xhosa S41 are more resistant to coarticulation than other consonants.

An unusual VOT contrast between partially voiced plosives and fully voiced stops, possibly implosives, has been described in Bekwel A85b (Cheucle 2014: 287) and the Kanincin variety of Ruwund L53 (Demolin 2015: 495).

4.2 Implosives and ejectives

Languages of the North-West, the Eastern coastal area and the South-East often have at least one implosive, most frequently a bilabial, but implosives are generally absent in the languages of the Congo basin and the South-West. Ejective stops and affricates are more rarely found in the Bantu languages, although they occur as variants of the unaspirated voiceless stops in languages of the South, especially in post-nasal contexts. The ejection is generally weak compared to that found in languages of the Afro-Asiatic family, except for Ilwana E701 where the ejectives are in borrowed Cushitic vocabulary, and the ejective
lateral affricate of Zulu S42 mentioned earlier. In the Ngwato S31c variety of Tswana S31, ejectives are weak and are sometimes lenited, with loss of ejection: /tˈkʰ/ ~ [ðt̚ ʃʰ] (Gouskova et al. 2011: 2127).

The segments labelled as implosives are sometimes described as if a glottal constriction is characteristic of their production. In Bantu, this is typically not the case; the vocal folds are in the normal position for voicing. Rather, what is critical is that the larynx is lowering during their production, so that the size of the supralaryngeal cavity is being enlarged while the oral closure is maintained. This may have two principal effects. Firstly, it allows the amplitude of vocal fold vibration to increase during the closure, giving a particularly strong percept of voicing at the time of the release. Secondly, it may mean that the intra-oral pressure is relatively low at the time when the closure is released so that at the moment of release the initial airflow is ingressive (Hardcastle & Brasington 1978). The waveform of an intervocalic bilabial implosive in Tonga S62 is shown in Figure 3.18. Dashed vertical lines mark the onset and offset of the bilabial closure. Voicing is continuous through the closure; upper and lower lines have been constructed on the figure linking respectively the positive and negative peaks in the waveform in order to dramatise the growing amplitude of the voicing during the closure.

In Zulu S42, implosive [ɓ] tends to have a shorter closure duration and lower amplitude burst than plosive [b] (Naidoo 2010). In Mpiemo A86c, implosives have a slight rise in F₀ before the onset of a following vowel while voiced plosives have a sharp dip in F₀ (Nagano-Madsen & Thornell 2012).

### 4.3 Depressor consonants

Another special laryngeal action occurs in the “depressor” consonants which are characteristic of certain Bantu languages of the Eastern and Southern regions. This term was originally applied to consonants which have a particularly salient lowering effect on the pitch of the voice in their neighbourhood (Lanham 1958). It has since sometimes come

---

to be used for any consonant which has any local lowering effect on pitch or, more accurately, on the fundamental frequency of vocal fold vibration, abbreviated $F_0$, such as an ordinary voiced plosive. It has even been used for those which may simply block a raising or high-tone spreading process. However, the original notion of a depressor consonant is quite different from this expanded use. The most detailed study remains that of Traill et al. (1987) on depressor consonants in Zulu S42. This study shows that the $F_0$ associated with depressors is lower than a low tone, and the lowest pitch is centered on the depressor consonants themselves. At vowel onset, the $F_0$ difference between High and Low tones after a set of non-depressor consonants is 22 Hz, but a High tone onset after depressor consonants is 44 Hz lower than after the non-depressors and a Low after depressors is 23 Hz lower than after non-depressors. This is the mean across three speakers, two male and one female. Thus a High after a depressor begins considerably lower than a Low elsewhere. Figure 3.19 compares the pitch contours of the Swati S43 words /lihála/ ‘aloe’ and /lihála/ ‘harrow,’ where /̤/ is a diacritic to mark the fact that the consonant is a depressor in the second word. Despite the fact that the lexical tone after the depressor is high (Rycroft 1981), the onset $F_0$ is about 30 Hz lower than the low tone onset after the non-depressor, and a rapid pitch fall begins during the vowel which precedes the depressor.

Voiceless, voiced, prenasalised and even aspirated stops may all pattern as depressor consonants (Chen & Downing 2011, Cibelli 2015, Lee 2015). Figure 3.19 also illustrates the fact that depression is not necessarily associated with voicing as both /h/ and /h̤/ are voiceless (Downing & Gick 2001, Downing 2009). Equally, voiced segments such as nasals and approximants may contrast in depression (Traill & Jackson 1988, Wright & Shryock 1993, Mathangwane 1998). Since these segments make for easy
tracking of $F_0$ through the consonant, the centring of the depression on the consonant can be most easily visualised with them. Two examples from Giryama E72a are illustrated in Figure 3.20. In these cases there is a substantial fall in $F_0$ from the onset to the middle of the nasal, and pitch begins to rise before the consonant is released; the pitch peak on the vowel is 40Hz (left panel) or 50Hz (right panel) higher than the lowest pitch in the nasal. In these words there is noticeably breathy phonation during part of the consonant and at the vowel onset which is transcribed as [ŋ]. However, breathiness is not an invariable accompaniment of depression as had been proposed by Rycroft (1980). Following Traill et al. (1987), we understand true “depression” to consist of a special laryngeal posture consistent with very low pitch co-produced with the consonant it is associated with. This gesture may become associated with any class of consonants and thus is capable of becoming itself an independent phonological entity deployed for grammatical effect as in the “depression without depressors” described by Traill (1990).

5 CLICKS

Though cross-linguistically rare, clicks are used by millions of people speaking various Bantu languages. They occur in two separate geographical clusters, the South-East (SEB) and the South-West (SWB), as shown in Figure 3.21. Clicks in the South-East cluster were borrowed from Khoe and possibly also from Taa and Kx’a languages into Nguni S40 (Louw 2013, Pakendorf et al. 2017); from Nguni (primarily Zulu S42), they subsequently spread into other SEB languages (Letele 1945, Bailey 1995). Clicks in the South-West cluster were borrowed independently from those in the South-East. In Fwe K402, they were borrowed from Khoe and Ju languages (Bostoen & Sands 2012, Gunnink et al. 2015). In languages of both the South-East and the South-West clusters, clicks can be found in Bantu roots as well as in loanwords. The functional load of clicks varies across languages, as detailed in Pakendorf et al. (2017) and Sands & Gunnink (forthcoming), both in terms of the number of contrastive click consonants, and in terms of the percentage of lexical items which contain clicks.
In the South-East, the core is formed by the languages of the Nguni group (S40), especially Xhosa S41, Zulu S42, Phuthi S404 and Zimbabwean Ndebele S44, which have between 12 and 15 click consonants; Swati S43 has fewer clicks (Doke 1954, Pakendorf et al. 2017). Clicks are found in many words in Southern Sotho S33 (Guma 1971), but only occur in a few sound symbolic words and interjections in Northern Sotho S32 (Poulos & Louwrens 1994). Clicks do not occur in Venda S21 (Ziervogel et al. 1981, Poulos 1990). In the Shona S10 group, clicks have only been reported to occur in midlands varieties of Kalanga S16 and in the Ndau S15 variety in Mozambique (Borland 1970, Mkanganwi 1972, Afido et al. 1989, Pongweni 1990). In the Tswa-Ronga S50 group, clicks have been reported to occur in Tswa S51, Tsonga S53, Konde S54, Nkuna S53D and Ronga S54 (Passy 1914, Persson 1932, Doke 1954, Baumbach 1974, Afido et al. 1989, Sito 1996), but their functional load in these lects is not well known. Clicks have also been reported to occur in Chopi S61 (Bailey 1995) and in the Mzimba variety of Tumbuka N21 (Moyo 1995).

In the South-West, the area near where the borders of Namibia, Angola, Botswana and Zambia meet, the largest number of clicks is found in Yeyi R41. Namibian Yeyi is described as having 19 click consonants (Gowlett 1997: 257), while Botswana Yeyi speakers vary, having as few as 12 or as many as 22 distinct click consonants (Fulop et al. 2003). Clicks are also found in Manyo (Geiriku) K332, Sambyu K331, Kwangali K33, Mbukushu K333 and Fwe K402 (Baumbach 1997, Möhlig 1997, Gunnink et al. 2015). Mbalangwe K401 has clicks, but whether it is a sociolect of Subiya K42 (Maho 1998:

51) or of Yeyi (Baumbach 1997: 307) is unclear. Clicks have been reported to occur in Ikuhane, or Botswanan Subiya (Ndana et al. 2017), but they have not been documented in Namibian or Zambian varieties (Baumbach 1997: 311, Jacottet 1896). Click consonants do not occur Herero R31, Umbundu R11, Totela K41 or Lozi K21, nor are they found in languages of the Wambo R20 cluster, such as Kwayama R21, Mbalanhu R214 and Ndonga R22. Clicks are marginal in Tswana S31 and Kgalagari S311, with the possible exception of the Shetjauba variety of Kgalagari (Tlale 2005, Lukusa and Monaka 2008).

The separate South-East and South-West groups of Bantu languages with clicks can be seen in the map of Southern Africa in Figure 3.21. Areas in black on the map represent the geographical distribution of languages with large click inventories, and areas in grey represent smaller click inventories. Areas north of Swati S43 and east of Ndebele S44 with grey patterns show the S10, S50 and S61 zones where clicks have been sporadically attested. Like most linguistic maps, this map represents a somewhat fictitious ethnographic idealisation not corresponding precisely with any exact time or population distribution. It is also not possible to definitively state the number of Bantu languages with clicks; clicks may occur in some varieties and not others, as in the case of Fwe K402 (Pakendorf et al. 2017: 20, Gunnink forthcoming), and may have even been lost where they were once attested. Tswana S51 may be one such case, as the last attestation was by Persson (1932). It is noteworthy that none of the Bantu languages of East Africa appear to have acquired clicks from the surviving or former languages of this area with clicks (Maddieson et al. 1999).

Languages which lost clicks entirely include “Northern” Ndebele of South Africa S408 and Lozi K21 (Ziervogel 1959, Gowlett 1989, Skhosana 2009), though it seems some “Northern” Ndebele S408 speakers are borrowing clicks back from Zulu S42. Some speakers of Southern Ndebele S407 have a reduced click inventory (Schulz & Laine 2016). Click loss is an on-going process in Chopi (Bailey 1995) and in Imusho Fwe (Gunnink forthcoming). Sowetan Zulu S42, too, has a reduced number of click consonants, likely due to contact with Southern Sotho S33 (Gunnink 2014). A rapid reduction in the number of click contrasts occurred more than 100 years ago in the far-flung varieties of Nguni known as Ngoni N12 (Elmslie 1891, Spiss 1904, Doke 1954); Ngoni speakers subsequently shifted from Nguni to languages of the Manda N10 group (Maho 2003). Note that languages of Malawi and Tanzania are not shown on the map in Figure 3.21. Clicks have not been reported for Manda group languages and are unlikely to occur unless efforts to revitalise Malawian Ngoni on a Zulu model prove effective (Kishindo 2002).

The mechanism of producing clicks is now fairly well understood and is illustrated by the sequence of midsagittal real-time MRI in Figure 3.22. A closure in the vocal tract is formed by the back of the tongue contacting the roof of the mouth in the velar or uvular area and a second closure is formed in front of the location of this closure by the tip or blade of the tongue or the lips, as shown at timestep 1. A small quantity of air is entrapped inside the sealed oral cavity. Although not seen in a mid-sagittal diagram, the sides of the tongue are also raised to complete the seal between anterior and dorsal closures. The center portion of the tongue is then lowered while the two main closures are maintained (timesteps 2–3), enlarging the volume of the space between them. In addition, there may be retraction of the tongue tip, dorsum or tongue root for some clicks (Miller 2008, Miller & Finch 2011). Expansion of the closed cavity causes the pressure in the air inside the space to be reduced well below that of the air outside the mouth. Next, the closure at the front and/or side of the mouth is released (timestep 4) and the abrupt equalisation of air pressures inside and outside the mouth results in a sharp acoustic transient. Finally, the
FIGURE 3.22 TIME-ALIGNED AUDIO AND VIDEO DATA OF A SWATI S43 DENTAL CLICK AND FOLLOWING VOWEL IN THE SYLLABLE ngca /ŋɡaː/. THE TOP AND MIDDLE ROWS SHOW A WAVEFORM AND SPECTROGRAM, RESPECTIVELY. BROKEN VERTICAL LINES INDICATE THE FIVE POINTS IN TIME CORRESPONDING TO THE rtMRI IMAGES SHOWN IN THE BOTTOM ROW. AN ACOUSTIC ARTEFACT OF RECORDING IN THE CYLINDRICAL METALLIC MRI SCANNER BORE IS A SERIES OF ECHOES SPACED AT 53 MS INTERVALS. THE WHITE BOW-SHAPED LINE CROSSING EACH MIDSAGITTAL IMAGE IS ALSO AN ARTIFACT. SEE PROCTOR ET AL. (2013), PROCTOR ET AL. (FORTHCOMING) FOR A DESCRIPTION OF THE METHODOLOGY USED TO OBTAIN THE IMAGES.

Source: Recording and images made available by Michael Proctor.
back closure is released, and this release may be separately audible. The posture of the vowel following the click is seen in timestep 5.

Clicks in Bantu languages are often made with a back closure that is velar, however uvular constrictions also occur, particularly for post-alveolar [!] and lateral clicks [l] (Miller 2008). Kx’a, Tuu and Khoe (“Khoisan”) languages tend to favour uvular rather than velar constrictions (cf. Miller et al. 2007, Miller et al. 2009b, Miller 2010, 2016). The velar release of a Xhosa S41 dental click is shown in Figure 3.23, which has a wave-form and spectrogram of the word caca /ǀàǀà/ ‘be clear.’ The first unaspirated dental click has a velar burst 17 ms after the anterior click burst. The click in the second syllable has a dorsal release that is closer in time to the release of the anterior click closure. The second click also has a velar closure. This can be seen by the converging F2 and F3 transitions at the end of the first vowel, (as indicated by the arrow), which indicate a velar constriction.

The basic click mechanism does not determine what the larynx is doing while these movements are taking place in the oral cavity, nor whether the velum itself is raised or lowered to block or permit air from the lungs to flow out through the nose. Thus, a click can be accompanied by simple glottal closure, by modal or breathy voicing, by open vocal folds, or by use of the ejective mechanism. Changes in larynx activity can be variously timed in relation to the action in the oral cavity, and to the timing of movements raising and lowering the velum. The possible variations are thus very numerous, and many different categories of individual clicks are found when all the languages which use them are considered (Ladefoged & Maddieson 1996).

In describing clicks, it is customary to talk of the click type and the click accompaniment. The click type refers to the location of the front closure and the manner in which it is released, which may be abrupt or affricated, central or lateral. The accompaniment refers to all the other aspects of the click: laryngeal action and timing, nasal coupling, and the location (uvular or velar) and manner of release (abrupt or affricated) of the back closure. In South-East Bantu languages, three contrastive click types are found, and probably no more than seven accompaniments are used. Zulu S42 and Xhosa S41 have dental /ǀ/, alveolar lateral /ǁ/ and apical post-alveolar /ǃ/ click types. The last of these was often described as “palatal” in older literature. In South-West Bantu languages, Yeyi has these three click types as well as a contrastive laminal post-alveolar type /ǂ/, variously called “alveolar” or “palatal” in different sources. The palatal click type † may be found as a variant of /ǃ/ used in child-directed speech in Zulu and Xhosa (Bradfield 2014: 27). The bilabial click /ʘ/ is not found in Bantu except in paralinguistic utterances, and as a variant pronunciation of a sequence of labial and velar stops, as in Rwanda JD61 (Demolin 2015: 483). Dental and lateral clicks are sometimes called “noisy,” “affricated,” or “pre-affricated” (Roux 2007), while the (post-)alveolar is described as “abrupt” or “unaffricated.” Palatal clicks in Yeyi R41 are somewhat fricated (Fulop et al. 2003), though they are typically produced with an “abrupt” or “unaffricated” release in Khoisan languages. Yeyi R41 has eight different accompaniments (Fulop et al. 2003), including several contrasts which are not found in other Bantu languages. Yeyi R41 contrasts clicks with a velar fricated and ejective velar fricated release (/ǀǂ/, /ǀǂ’/) as well as glottalised and ejected clicks (e.g., /ǀʔ/, /ǀ’/) (Fulop et al. 2003). Fwe has four accompaniments including a voiceless nasal accompaniment (Gunnink forthcoming) not known to occur in any other Bantu language. It is possible that phonetic studies of other South-West Bantu click languages will reveal additional click accompaniments.

Thomas-Vilakati’s analysis of Zulu click types (Thomas 2000, Thomas-Vilakati 2010), combining insights from acoustic, aerodynamic and electropalatographic techniques, is
anterior release
dorsal release

FIGURE 3.23 SPECTROGRAM OF THE XHOSA S41 WORD CACA /lä∂a/ ‘BE CLEAR’ SPOKEN BY SIZWE SATYO, A MALE SPEAKER OF XHOSA. THE SMALL ARROWS ON THE WAVEFORM SHOW A DISTINCT ANTERIOR AND DORSAL BURST ON THE FIRST CLICK. THE ARROW IN THE SPECTROGRAM POINTS TO A CONVERGENCE OF F2 AND F3 CHARACTERISTIC OF VELARS.

Source: Recording made by Peter Ladefoged in 1979 and archived at the UCLA Phonetics Lab Archive (Ladefoged 2007).
the most detailed study of click production in a Bantu language to date. Thomas-Vilakati confirms that the velar closure always precedes the front closure; this accounts for the fact that nasals preceding clicks assimilate in place to velar position, and corrects a misobservation by Doke (1926), who believed the front closure was formed first: the velar closure must be released after the front closure for the click mechanism to work, but it could in principle be formed later. Dorsal closures for all three click types in Thomas-Vilakati’s data are held for about 175 milliseconds, but the front closures show some significant timing differences. The front closure for dental clicks is formed earlier and held longer (about 105 ms) than that for post-alveolar or lateral clicks (about 80 ms). The relative timing and durations of velar and front closures deduced from acoustic and aerodynamic data are graphed in Figure 3.24.

More details on the articulations of clicks are given by electropalatography (EPG). Speakers wear a thin custom-made acrylic insert moulded to the shape of their upper teeth and hard palate in which a number of electrodes are embedded which sense contact between the tongue and the roof of the mouth. In Thomas-Vilakati’s study, inserts with 96 electrodes were used, together with software allowing a sweep of the contact patterns to be made every 10 ms. The articulatory contacts can then be examined using stylised displays such as those in Figure 3.29, Figure 3.30 and Figure 3.31, which represent the arc of the teeth and the vault of the palate. Contacted electrodes are shown as black squares and uncontacted ones as grey dots. Figure 3.25 shows the production of a dental click. The first frame, numbered 0, is close to the time that velar closure is first made, as detected from the accompanying acoustic record. Because the insert does not cover the soft palate, this closure cannot be observed on the EPG record at this time. The seal around the inside of the teeth is made by 40 ms later, and as the contact area of the back of the tongue enlarges, the front edge of the velar contact is now visible as a line of contacted electrodes at the bottom of the arc. The closures overlap for 100 ms, until frame 140. During this time, rarefaction is occurring. This figure makes clear that the expansion of the cavity is not solely due to moving the location of the back closure further back. That Zulu dental clicks are produced with a controlled fricated release is also clear from the way the front release initially involves formation of a narrow channel, clearly visible in frame 150.

![Figure 3.24](#)

**FIGURE 3.24** CLOSURE DURATIONS AND TIMING RELATIONS IN THE THREE CLICK TYPES OF ZULU $S^42$; MEANS FOR VOICELESS CLICKS IN THREE VOWEL ENVIRONMENTS SPOKEN BY THREE SPEAKERS, ADAPTED FROM THOMAS-VILAKATI (2010). FRONT CLOSURE DURATIONS ARE SHOWN AS HEAVILY STIPPLED BARS.
Production of a lateral click is illustrated in Figure 3.26. In this particular token there is a long lag between the time the velar closure is made and when the front closure is sealed, about 80 ms later. The contact of the front of the tongue is asymmetrical, as the side of the tongue opposite to where the release will be made is braced contra-laterally against the palate. The release of a lateral click is also affricated, occurring initially through a narrow channel quite far back, as shown in frame 170 and continuing in frame 180. In contrast to these two affricated click types, a post-alveolar click is released without affrication. As Figure 3.27 shows, the shift from sealed to open occurs rapidly and completely, here between the two frames numbered 170 and 180. These frames also illustrate the retraction of the tongue tip which occurs just before release of this click type. From frame 150 through to frame 170 the contacted area moves back, so that the configuration at the moment of release is clearly post-alveolar.

During the time period in which the two closures of a click overlap, lowering of the center of the tongue creates a partial vacuum in the cavity between them. Thomas-Vilakati’s work provides the first direct measures of how powerful the energy generated by
this gesture is. Air pressure in the oral cavity is measured in relation to the ambient atmospheric pressure in hectoPascals (hPa, equivalent to the pressure required to support 1 cm of water). For an ordinary pulmonic stop, peak pressure behind the closure ranges between about 5 and 20 hPa, depending on the loudness of the voice. The peak negative
Pressures reached in clicks are typically -100 hPa or more and may reach over -200, as shown in Figure 3.28. Post-alveolar clicks have the greatest rarefaction, lateral clicks the least, perhaps because the contra-lateral bracing of the tongue in the lateral clicks may constrain the amount of tongue-center lowering that is possible. Thomas-Vilakati’s aerodynamic data also reflect the different dynamics of the affricated and abrupt clicks. The equalisation of internal and external pressure at release occurs much more quickly in post-alveolar clicks than for dental and lateral clicks. This can be shown by calculating the average rate of pressure change over this phase of the click, which is 14.4 hPa/ms for post-alveolars, 7.9 for dentals, and 4.2 hPa/ms for laterals. Only a small part of this difference can be accounted for by the difference in peak pressure between the click types.

Zulu S42 has four different accompaniments to its three click types: “plain” (voiceless unaspirated), voiceless aspirated, voiced and voiced nasalised. Because the place of the dorsal closure is not contrastive, it is not necessary to indicate the (velar in this case) place before the click type symbol. Our recommended IPA transcription and corresponding Zulu S42 orthographic symbols is given in Table 3.2. There are several ways of indicating the same click following IPA principles, e.g., /ɡǃ/, ɡǃ/, ǃ/ are equivalent ways of representing a voiced (post-)alveolar click. Similarly, /ŋǃ/, ₀!, ₀/ all represent a voiced nasal (post-)alveolar click. Xhosa S41 has five accompaniments, three of which are the same as in Zulu S42. The Xhosa S41 “voiced” clicks are breathy or slack voiced (Jessen & Roux 2002) and may even be devoiced (Maphalala et al. 2014, Braver 2017). There is a distinct breathy/slack voiced nasalised accompaniment; these two series are “depressor consonants.” Some speakers of Xhosa S41 produce “plain” clicks with ejection (Jessen 2002).

One of the most striking things about clicks in Bantu is the lack of respect for place distinctions when few categorical contrasts exist. In Nkuna S53D, Baumbach (1974) indicates that clicks are indifferently pronounced as dental or post-alveolar. In Mbugusu...
K333, the one series of clicks is reported to be pronounced “either as dental, palatal or [post-]alveolar sounds” (Fisch 1998). Southern Sotho S33 only has a single click type which may vary in place. Older accounts of Southern Sotho S33 describe both post-alveolar or sub-laminal retroflex articulations (Doke 1923: 713, 1926: 301). In Manyo K332, clicks are “mostly dental, however, with a broad individual variation” (Möhlig 1997). There are four click accompaniments in Fwe K402: voiceless unaspirated, voiced oral, voiced nasal and voiceless nasal, but the language has no contrast for click type or place (Gunnink forthcoming). In the central (Imusho) variety of Fwe, the word ‘papyrus’ may be pronounced with an unaspirated dental click (\[ruǀom\]), as in Figure 3.29, or as an unaspirated alveolar click (\[ruǃom\]), as in Figure 3.30, with no difference in meaning. The dorsal constriction of clicks in Fwe is typically velar. In Figure 3.29 and Figure 3.30, the anterior click burst has a higher amplitude than the velar release burst, as is typical for clicks cross-linguistically. In Figure 3.31, however, the dorsal burst has a higher amplitude than the anterior click burst. The click in the word [ruǀoma] ‘papyrus’ in Figure 3.31 is a very weak click, as indicated by the extended IPA (extIPA) diacritic for a weak articulation, e.g., \[^c\]’, which is similar to the diacritic for an unreleased stop e.g., \[^c\]’, but placed under the consonant rather than after it. The word ‘papyrus’ may also be articulated with a velar stop in place of the click [rukoma], as seen in Figure 3.32. Because the velar stop burst in the weak click [\[^c\]’] is louder than the anterior click burst, it is perhaps not surprising that \[^k\]’ has come to replace [\[^c\]’] for some speakers. The current variation between clicks and velars in Imusho Fwe may eventually lead to the loss of clicks in the variety altogether, as clicks are replaced by velars. Anecdotally, it seems that clicks in other Bantu languages may also vary in amplitude, depending on the individual speaker, stylistic or scophonetic variables, and prosodic environment. The context-free liberty to vary place of articulation of clicks in some Bantu languages is rarely encountered with other classes of consonants.

### 6 NASALS AND NASALITY

The special phonetic interest of consonantal nasality in the Bantu languages involves principally the prenasalised segments and the realisation of “voiceless” nasals. In both cases aspects of timing are particularly relevant. Another feature of interest is the presence of a cross-linguistically rare contrast between nasalised and oral glottal approximants (/\(h\)/ and /\(h\)/) found in Kwangali K33 (Ladefoged & Maddieson 1996: 132).
Detailed studies of timing in prenasalised stops are included in Maddieson (1993), Maddieson and Ladefoged (1993) and Hubbard (1994, 1995). Using data from these sources, Figure 3.33 compares the durations of nasals and voiced prenasalised stops as well as of the vowels that precede them in two languages, Ganda JE15 and Sukuma F21. In both languages the oral stop duration in voiced prenasalised stops is very short, so the total segment...
duration is not so very different from that of a simple nasal. Both languages have contrasts of vowel quantity and compensatory lengthening of vowels before prenasalised stops, but there are interesting differences between the two. Lengthened vowels are much closer in duration to underlying long vowels in Ganda JE15 than they are in Sukuma F21. Sukuma F21 lengthened vowels are almost exactly intermediate between underlying short and long vowels and the nasal portion is quite long. Nyambo JE21 is similar to Sukuma in its pattern.

**FIGURE 3.30** WAVEFORM AND SPECTROGRAM OF THE MIDDLE SYLLABLE OF THE FWE K402 WORD [ruǃoma] ‘PAPYRUS,’ SPOKEN BY A MALE SPEAKER.

Source: Recording made available by Koen Bostoen.
Hubbard (1994, 1995) suggests that the difference from Ganda is related to the fact that lengthened vowels count in a different way in tone assignment rules in these languages.

Hubbard (1994, 1995) also compared the durations of vowels in three further languages with different patterns. The mean results are given in Table 3.3. Ndendeule N101
has no long vowels and no lengthening. Yao P21 has a long/short contrast and significant compensatory lengthening so that vowels before prenasalised stops are as long as underlying long vowels and have more than double the duration of short vowels. Tonga M64 has long vowels but does not show any compensatory lengthening before NC. This difference seems to be related to the different origin of long vowels; Yao P21 maintains
Proto-Bantu vowel length distinctions and adds to them. Tonga M64 does not preserve Proto-Bantu vowel length, but has developed long vowels from intervocalic consonant loss.

Several recent detailed studies have looked at the timing and laryngeal characteristics of stops after nasals in Tswana S31 and Kgalagari S311. These closely related languages have been argued to violate a constraint against voiceless stops after nasals. Post-nasal stops are devoiced in Kgalagari S311 (Solé et al. 2010), and in Tswana S31 only for some speakers (Coetzee & Pretorius 2010). There is evidence for post-nasal fortition rather than devoicing in the Ngwato S31c variety (Gouskova et al. 2011, Boyer & Zsiga 2013). In this variety, some speakers fail to devoice, and others devoice intervocically as well as after nasals (Zsiga et al. 2006).

In several areas earlier voiceless prenasalised stops have developed into voiceless nasals or related types of segments, including in Sukuma F21 (Maddieson 1991), Pokomo E71, Bondei G24 (Huffman & Hinnebusch 1998), Kalanga S16 (Mathangwane 1998) and Rwanda JD61 (Demolin & Delvaux 2001). Aspects of the original sequencing of nasal + oral and voiced + voiceless portions found in prenasalised stops are sometimes retained and small variations in the timing and magnitude of the different component gestures create quite large variability in the acoustic pattern of these segments as critical alignments are made or missed.

\[\text{FIGURE 3.33 COMPARISON OF SELECTED VOWEL AND CONSONANTS LENGTHS IN GANDA JE15 AND SUKUMA F21 (SEE TEXT FOR EXPLANATION).}\]

<table>
<thead>
<tr>
<th>Vowel duration in . . .</th>
<th>Ndendeule N101</th>
<th>Yao P21</th>
<th>Tonga M64</th>
</tr>
</thead>
<tbody>
<tr>
<td>. . . CVC</td>
<td>148</td>
<td>61</td>
<td>100</td>
</tr>
<tr>
<td>. . . CVNC</td>
<td>146</td>
<td>130</td>
<td>101</td>
</tr>
<tr>
<td>. . . CV:C</td>
<td>–</td>
<td>132</td>
<td>241</td>
</tr>
</tbody>
</table>

Source: After Hubbard 1995
Variation in the realisation of “voiceless nasals” is at least in part correlated with position in a word. Figure 3.34 shows a spectrogram of the Nyamwezi F22 word /ŋəpo/ ‘basket’ spoken in isolation. Dotted vertical lines separate the major phonetic components of the first syllable. The portion marked A, between the first two lines, is phonetically a voiceless velar nasal [ŋ]. The second line marks the time-point at which the velar closure is released. The fragment marked B has voiceless oral airflow, with resonances similar to those of the following /a/ vowel. Fragment C is the voiced portion of the vowel /a/. This type of segment might well be described as an “aspirated voiceless nasal.” Figure 3.35 shows a realisation of a medial instance of the same segment in the word /kɔŋᾶ/ ‘to suck.’ In this case there is no consonantal nasality. The nasal feature is realised as nasalisation of the latter part of the vowel /ɔ/ in Fragment C, following an oral portion, B, and the aspiration of the initial stop, A. Fragment D, which is the consonantal part of the /ŋ/ is voiceless but oral, and as often in an [h]-sound, the transition of the formants of the flanking vowels can be traced through its duration.

In Sukuma F21, the nasal portion of the “voiceless nasals” is often at least partly voiced or breathy voiced, as described in Maddieson (1991), whereas the parallel segments in Rwanda JD61 are fully voiced (except after voiceless fricatives), but produced with a modified kind of voicing described by Demolin and Delvaux (2001) as whispery-voice.

7 PROSODIC CHARACTERISTICS

A discussion of Bantu phonetics would not be complete without reference to some of the studies of the major prosodic characteristics of the languages. Stress in Bantu often falls
on the penult, typically with vowel lengthening, but stem-initial prominence also occurs (Downing 2010). A majority of Bantu languages have a tonal distinction of High and Low tones, which often may combine into contour tones. High tone is generally the phonologically marked tone, with Low tone being unmarked (Stevick 1969, Downing 2011) (see also Chapter 5). Languages without tone do occur, e.g., Swahili G42, Mwiini G412, Nyakyusa M31, as do ones with more than two level tones, e.g., Kamba E55 and Oku (Grassfields Bantu) (Downing 2010, Hyman 2014). Aspects of prosody to be covered here include: patterning of tones, phonetic implementation of tones, positional restrictions, intonation, focus prosody and question prosody. Typically, studies of intonation in Bantu languages tend to look at \( F_0 \) and duration; measures of intensity and spectral tilt are less often used to identify prosodic cues (Zerbian & Barnard 2008).

The patterning of tones in many Bantu languages resembles that of pitch-accent systems. For instance, the number of High tones which may surface in a word or a stem may be limited to one and prominent peaks tend to occur in a predictable position, often the penult (Downing 2010). In Northern Sotho S32, however, there is speaker variation in the position of the \( F_0 \) peak, which may occur “somewhere between the second and the third syllable, counting from the high-tone-bearing, verbum initial syllable” (Zerbian 2009).

The phonetic shapes of tone sequences can usually be modeled on the basis of the position and height of local H targets, with the Low tones treated as automatically filled valleys between these points. Certain more complex patterns, such as those noted by Hombert (1990) in Fang A75, and by Roux (1995) in Xhosa S41, may require a more elaborate model. High tones tend to fall on the antepenult in Nguni S40 languages such as Xhosa S41, though the penult is stressed/lengthened (Downing 2010). Provisions
have to be made for the special effects of depressor consonants on tone in Nguni languages. Depressors also occur in Digo E73 and other Mijikenda E70 group languages and in Kalanga S16 and other Shona S10 group languages (Downing 2010). In Zulu S42, the phonetic effects of depressor consonants on pitch differ from pitch lowering effects caused by implosive consonants (Chen & Downing 2011). In Rwanda JD61, there is anticipatory coarticulation of tone, with the $F_0$ of a syllable being affected by a High tone in a later syllable (Myers 2003).

The most extensive body of work on the phonetics of tone in a single Bantu language concerns Chewa N31b (Carleton 1996, Myers 1996, Myers & Carleton 1996, Myers 1999a, b). Detailed studies of this type not only illuminate the individual language studied but may provide insights into diachronic issues. For example, in Chewa N31b, as is common cross-linguistically, the High pitch peak is realised at the end of the syllable to which it is associated (Kim 1998, Myers 1999a). This pattern may form the basis for the frequent shifting of a High tone to a later syllable. These studies also address several issues in the relation between intonation and tone. For example, Myers (1999b) shows that syntactically unmarked yes/no questions are characterised by a slower rate of pitch declination than statements. Carleton (1996) demonstrated that units of paragraph length are organised by long-range patterns of tonal declination and resetting.

Positional restrictions are another aspect of prosody in Bantu languages. Tonal contrasts and vowel length contrasts are often restricted to stem-initial syllables (Downing 2010). In Jita JE25, for instance, only the initial syllable of verb roots may contrast in tone (Downing 2011). In Mwiini G412, however, long vowels may surface on the penult or antepenult and only occur word-initially in loanwords (Kisseberth & Abasheikh 2004: xvii). Contour tones may be restricted to heavy syllables. For instance, a contrast between HL and LH contours is restricted to long vowels in Rwanda JD61 (Myers 2003). Stem-initial syllables typically have a greater number of segmental contrasts than found elsewhere (Downing 2010).

Surveys of intonation in Bantu languages include Zerbian and Barnard (2008) and the volume edited by Downing & Rialland (2016a). Both surveys reveal a great deal of variety across Bantu languages. For instance, there are languages with and without down-drift, though the former are more common (Downing & Rialland 2016b). Mbochi C25, which does not have downdrift, still has final lowering due to a L% boundary tone (Rialland & Aborobongui 2016). There are different types of downstep attested in some Bantu languages. Downstep due to a floating Low tone is attested in Basaa A43a (Makasso et al. 2016). Downstep affects the second of two adjacent High tones in Tswana S31 (Zerbian & Kügler 2015) and Bemba M42 (Kula & Hamann 2016). The interaction of final lowering and downstep in Pare G22 is detailed in Herman (1996). Final lowering is fairly common across Bantu, but is not attested in Basaa A43a (Downing & Rialland 2016b). Final lowering associated with a L% boundary tone at the end of a sentence in Ngazidja G44a is often associated with a devoiced final syllable (Patin 2016). In Tswana S31, declarative sentences are primarily marked by penultimate lengthening and a reduced or devoiced final vowel (Zerbian 2016).

Focus in Bantu is often marked using morphosyntactic means rather than through the use of prosody (Downing & Hyman 2016, Downing & Rialland 2016b). In Bemba M42, however, new information focus is indicated on a subject by its placement in post-verbal position and by pitch raising of the pre-focus constituent (Kula & Hamann 2016). Some North-Western Bantu languages which have stem-initial accent, such as Eton A71, have a
focus prosody that causes the lengthening of stem-initial consonants and vowels (Van de Velde & Idiatov 2016). Focus and emphasis are associated with pitch raising in Mwiini G412 (Kisseberth 2016), but this seems to be the exception rather than the rule in Bantu. Chewa N31b and Tumbuka N21, for instance, do not have focus prosody (Downing 2016).

A wide range of means of marking question prosody have been noted for Bantu languages. Rialland’s (2007) survey includes seven different prosodic types found in Bantu languages, the most common being the use of register expansion along with the reduction of downdrift. Both falling and rising intonation patterns are found in question prosodies. Final High or rising intonations are found in Ganda JE15, Chewa N31b and Saghala E74b, while final High-Low or High-falling intonations are found in Jita JE25. Mongo-Nkundu C61 has reduction of final lowering, while Zulu S42 and Southern Sotho S33 cancel penultimate lengthening in question prosody. Polar or mid tones are found in Holoholo D28 and Nyanga D43. Zamba C322 and Ganda JE15 raise a final High tone in question prosody. These seven prosodic types do not account for all of the details of the individual languages. For instance, the final High in yes-no questions in Zamba is preceded by a sharp fall (Bokamba 1976: 19). In Bemba M42, polar questions are marked by a final boundary L% on the final syllable, but pitch range expansion is also used (Kula & Hamann 2016).

8 CONCLUSION

Bantu languages provide an opportunity to compare phonetic differences between fairly large numbers of related languages. Phonological theories, phonetic theories, and hypotheses about patterns of sound change can be tested in this real-world laboratory, ensuring the popularity of Bantu languages as subjects of research for years to come. The increasing availability of ultrasound and MRI technologies should lead to future studies examining the effect of prosodic environment on articulation. There is much work that remains to be done on cross-linguistic, intra- and inter-speaker variation of typologically unusual sounds such as clicks and “whistling fricatives.” Undoubtedly, studies of intonation and prosody in Bantu languages will continue to increase in number. Corpus studies of Bantu languages are currently few in number (Prinsloo & de Schryver 2001, Niesler et al. 2005, Allwood et al. 2010), but the increasing availability of such corpora may encourage phonetic studies of natural (unelicited) speech.

REFERENCES


Bokamba, E. G. (1976) *Question Formation in Some Bantu Languages*. Bloomington: Indiana University, PhD.


Gunnink, H. (forthcoming) Click Loss and Click Insertion in Fwe. In Sands, B. (ed.), Handbook of Click Languages. Leiden: Brill.


M October 21, 2021; For: 9781315755946, chapter3, 10.4324/9781315755946-3


