

The Effects of Speech Rate on VOT for Initial Plosives and Click Accompaniments in Zulu

Patrick J. Midtlyng
University of Chicago

1. Introduction

1.1. Overview

Cross-linguistic studies of acoustic correlates have shown that for phonological voicing distinctions (e.g., voiced/voiceless, aspirated/unaspirated, some ejective/non-ejective) there are ranges of values for voice onset time (henceforth VOT) that are associated with those distinctions (Stevens 1989, VanDam 2007, Lisker & Abramson 1964, Miller et al. 1983, Hodgson & Miller 1996, Volaitis & Miller 1992, Cho & Ladefoged 1999). VOT is the interval from the release of a plosive (burst) to the beginning of glottal excitation. There are three VOT categories: pre-voiced (negative VOT), short-lag (short positive VOT), and long-lag (long positive VOT). VOT varies with place of articulation (henceforth POA): the farther back the closure the longer the VOT (Fischer-Jørgenson 1954, Peterson & Lehiste 1960), but other factors including prosodic context, lexical frequency, speaker variability, and speech rate can affect VOT ranges (Miller et al. 1983, Kessinger & Blumstein 1997, Miller 1977, Miller & Baer 1983, VanDam & Port 2003; 2005, Theodore et al. 2007).

1.2. Phonetics of Zulu

Table 1 shows the distribution of phonological voice types across the places of articulation for Zulu plosives and click accompaniments. Notice that not all POAs are complete.

Table 1. Plosives and click accompaniments by phonological manner and place of articulation. The phonetic transcription of the target sound is on the left in brackets and the orthographic representation of that sound is on the right in parentheses.

Plosives	Bilabial		Alveolar		Velar	
Plain/Ejective	[pʰ]	(p)	[tʰ]	(t)	[k/kʰ]	(k)
Aspirated	[pʰ]	(ph)	[tʰ]	(th)	[kʰ]	(kh)
‘Voiced’	[b/b̥]	(bh)	[d/d̥]	(d)	[g/g̥]	(g)
Implosive	[ɓ]	(b)				
Click	Dental		Alveolo-Palatal		Alveolar-Lateral	
Plain	[k]	(c)	[kʰ]	(q)	[kʰ]	(x)
Aspirated	[kʰ]	(ch)	[kʰ]	(qh)	[kʰ]	(xh)
‘Voiced’	[g/g̥]	(gc)	[gʰ/g̥ʰ]	(gq)	[gʰ/g̥ʰ]	(gx)

* The author wishes to thank the audience at ACAL-40, especially Jennifer Cole and Mary Paster for their helpful questions and comments. Thanks also to Marc Ettlinger, Jason Riggle and the two anonymous reviewers for their assistance in the completion of this project. All errors are solely my own.

In his *Textbook of the Zulu Language* (1950), Doke lays out manner for plosives in two categories: explosive and implosive. Doke divides the explosive manner into four categories: radical, ejective, aspirated and voiced. The radical explosive is “pronounced without any accompanying vibration of the vocal cord, or closure of the glottis, or aspiration” (Doke, 1950). The only radical explosive is <k>. Orthographic <k> represents either a phonetically plain or ejective plosive. Doke describes the radical <k> as devoid of aspiration although it has slight voicing. Taljaard & Synman (1991) and Poulos & Msimang (1998) describe Doke’s radical <k> as “non-ejective partially voiced plosive” and use the [k̤] symbol. Doke calls the /b d g/ series “voiced” as does Poulos (although he uses the symbol for voiceless, e.g., [b̥]), but Taljaard & Synman use “delayed breathy voice” for this series. At the bilabial and alveolar places of articulation there are only ejective plosives and not plain voiceless plosives. The only implosive occurs at the bilabial place of articulation. The same distinctions among these sources hold for the click accompaniments as well, although there is debate about the “ejective” nature of clicks (especially of noisy clicks such as the dental and alveolo-lateral), but that will not be addressed here. Maddieson (1984) summarizes Zulu’s use of the voicing contrasts for plosives as follows: implosive, plain voiced, plain voiceless, plain aspirated, and voiceless ejective. Although, it is important to note that not all of the series are complete. The plain voiceless is Doke’s velar radical *k* and along with the bilabial implosive [ɓ], these are the only segments of their type across all places of articulation. For clicks the voicing contrasts are “voiced” (Doke uses “voiced”, Taljaard & Synman use “breathy-voiced”), voiceless, and voiceless aspirated. We acknowledge that many studies have analyzed the Zulu plosive series and have concluded that what is here called a ‘voiced’ plosive or click accompaniment is actually a short-lag voiceless stop. Our reason for adopting the “phonological voice type category” is in keeping with the discussion of Kessinger & Blumstein’s (1997) notion that phonological category reflects phonetic inventory. Phonetically, the Zulu implosives are pre-voiced: the ‘voiced’ consonants are short-lag and aspirated consonants are long-lag. This leaves the plain/ejective consonants as somewhere in between. This is where our discussion turns to the depressor/non-depressor contrast and the role of tonal depression as an acoustic correlate to voicing categorization.

Depressors are the phonologically ‘voiced’ plosives and click accompaniments. They have two important phonetic effects: 1) they lower the tone of the following vowel and 2) they lengthen surrounding vowels to accommodate the tonal lowering (Traill et al. 1987, Russell 2000). The tonal lowering has been argued to be the crucial acoustic correlate that differentiates depressor consonants from plain/ejective consonants. The data from Traill et al. data bear this out: VOTs between depressors and ejectives do not vary significantly, so the only cue to the distinction is most likely the accompanying pitch perturbations.

If we control for speech rate, will we find that plain/ejective consonants are regularly indistinguishable from their ‘voiced’ counterparts (and are thus also short-lag) and confirm the findings of Traill et al.? Or will we find that when speech rate is controlled for that VOT serves as an acoustic correlate to voicing distinctions and that plain/ejective plosives and click accompaniments are actually somewhere in between the short-lag ‘voiced’ consonants and the long-lag aspirated consonants?

Altering speech rate affects both the production and the perception of voicing contrasts (Summerfield 1981, Miller et al. 1986, Pind 1995, Kessinger & Blumstein 1997, Volaitis & Miller 1992, and Miller et al. 1997). On the production side, the effect of speech rate on VOT is not equal for all voicing contrasts. This asymmetric effect of speech rate on VOT is not a universal feature, but a language specific one. Miller et al. (1986) found that speech rate had a disproportionate effect on the production of /bi/ versus that of /pi/ (while both showed decreases in VOT as speech rate increased, the effect on the “voiced” plosive was smaller). Pind (1995) showed similar effects in Icelandic with long-lag plosives being significantly affected while short-lag plosives were not. The phonological differences between two or more languages for the same phonetic category reflect differing phonetic inventories (Kessinger & Blumstein 1997). A phonetically short-lag plosive can therefore be

phonologically voiced or voiceless based on the contrasting phonetic category; phonetic categories behave differently as a function of different phonological inventories.¹

The purpose of this study is to find the answer to the question above by examining the Zulu plosives series, which has not been examined within the cross-linguistic frame of speech rate/VOT interactions. Moreover, as Thomas-Vilakti (1999) points out, no study has examined the effect of speech rate on click accompaniments (henceforth CAs) in Zulu.

2. Method

2.1. Subject

One female Zulu native speaker took part in this study and was paid for her participation. The speaker was in her mid-40s and is a native from the kwaZulu-Natal region of South Africa. She was in the United States studying at the time of the experiment.

2.2. Stimuli

A wordlist was compiled using C.M Doke's *English-Zulu Dictionary* (Doke 1958:572) and finalized with the help of the Zulu consultant. The same segmental and prosodic conditions apply for the stimuli: all words are disyllabic with the target segment in the onset of the stressed syllable and the tonic vowel (low-toned /a/) was held constant. By holding tone level constant we control for the reported effects of tonal depression. The second tone was also low. The token list is in Table 2.

Table 2. Wordlist with targeted segments bolded (orthographic) with phonetic transcription. The bolded segments are the target segments. Phonetic transcriptions are in brackets. The stimuli were put together using an English-Zulu dictionary and aided by the native speaker consultant.

Word	Gloss	Word	Gloss	Word	Gloss
Plosives					
pana	hobble	taka	place in a	kama	to comb
[p'a:na]		[t'a:ka]	heap	[k/k'a:ma]	
phala	to scrape	thala	to look	khala	to cry
[p ^h a:la]		[t ^h a:la]	about	[k ^h a:ma]	
bhala	to write	dala	to create	gana	to marry
[b ^h a:la]		[d ^h a:la]		[g ^h a:ma]	
bala	to count				
[b ^h a:la]					
Click Accompaniments					
-cala	The side	-qala	to start	xaba	to block
[k a:la]		[k a:la]		[k a:ba]	
chama	to pass	qhana	button up,	xhama	women's girdle
[k ^h a:ma]	water,	[k ^h a:ma]	fasten	[k ^h a:ma]	
	urinate				
gcaba	to cut small	gqala	to look at	gxaba	stride out, walk
[g a:ba]	incisions	[g a:la]	intently	[g a:ba]	with long steps

¹ A reviewer points out that many researchers (Traill et al. 1987, Giannini et al. 1988) have argued that /bh d g gc gx/ are phonetically [p t k | ! |]. We acknowledge this and our categorization by phonological voicing category does not negate the fact that the "voiced" stops and click accompaniments are much like English or German "voiced" plosives: short-lag.

2.3. Procedure

The speaker was presented with two randomized lists of tokens with five repetitions of each token in each list for a total of 10 repetitions of each token in each condition. Tokens were spoken in isolation as well as set in the carrier phrase ‘Ngithi _____’ (I say _____) for the slow/fast conditions. Before recording, the speaker was prompted by a metronome and a contextual description of each condition (i.e., speak slowly and clearly or speak quickly). Recording was made using a unidirectional microphone on a Marantz PMD670 Solid State Recorder in PCM monophonic wav format at 48.0 kHz sampling rate 24 bit depth in a sound proof booth in the Phonology Lab at the University of Chicago.

2.4. Acoustic analysis

The sound files were segmented and word duration, tonic vowel duration, and voice onset time was recorded. Total word duration was measured to ensure that the speaker changed speaking rates. Tonic vowel duration was measured for two reasons: 1) to ensure that shorter word durations were not a result of final vowel deletion² and 2) to confirm the effect of depressor consonants on vowel length. VOT was measured from the burst of the plosive consonant/CA to the beginning of voicing. Tokens were labeled by POA, phonological voicing category (implosive, voiced, ejective, voiceless aspirated), and speech condition. A click is comprised of two closures: an anterior closure (influx) which gives the click its name and a posterior closure (efflux) where the velum is raised creating a pocket of air in the oral cavity (Ladefoged & Traill 1994). CAs are the posterior release and are either voiced, plain voiceless, or voiceless aspirated. Differentiating the release of the efflux from the influx can be a difficult task; and in many cases (but not all) the release of the accompaniment is fused closely to the click release itself. Measurements were taken in accordance with the methodology of Ladefoged & Traill (1994), Ladefoged & Maddieson (1995:425), and Roux (2007). The data were analyzed in SPSS and tested for main effects of POA, speech rate, and voice type and any interaction effects. The bilabial implosive (which is the only pre-voiced plosive or CA) was omitted so as not to bias the VOT results with respect to phonological voice type or POA.

3. Results

3.1. Mean word duration

Mean word durations for plosives were 642 ms (SD = 78) for the isolated condition, 573 ms (SD = 57) in the slow condition and 419 ms (SD = 41) in the fast condition. Mean word durations for CAs were 601 ms (SD = 62) for the isolated condition, 557 ms (SD = 57) in the slow condition and 421 ms (SD = 46) in the fast condition. Speech rate does have an effect on total word duration for both plosives and click accompaniments.

3.2. Tonic vowel duration

Mean tonic vowel durations for plosives were 320 ms (SD = 51) in the isolated condition, 244 ms (SD = 30) in the slow condition and 190 ms (SD = 22) for the fast condition. Mean tonic vowel durations for CAs was 290 ms (SD = 38), 233 ms (SD = 32) in the slow condition and 180 ms (SD = 28) for the fast condition. As with mean word duration, tonic vowel duration also decreased as speech rate increased. Figures 1a (CA) and 1b (plosives) show the effect of voice type on mean tonic vowel duration.

² In discussions with the Zulu consultant and pilot recordings, increased speech rate was often accompanied by final vowel deletion (e.g., /k^hana/ → [k^han]). By measuring both word duration and tonic vowel duration we ensure that the speaker actually abided by the speech conditions.

Figure 1. Mean tonic vowel duration (ms) by phonological voice type and speech condition

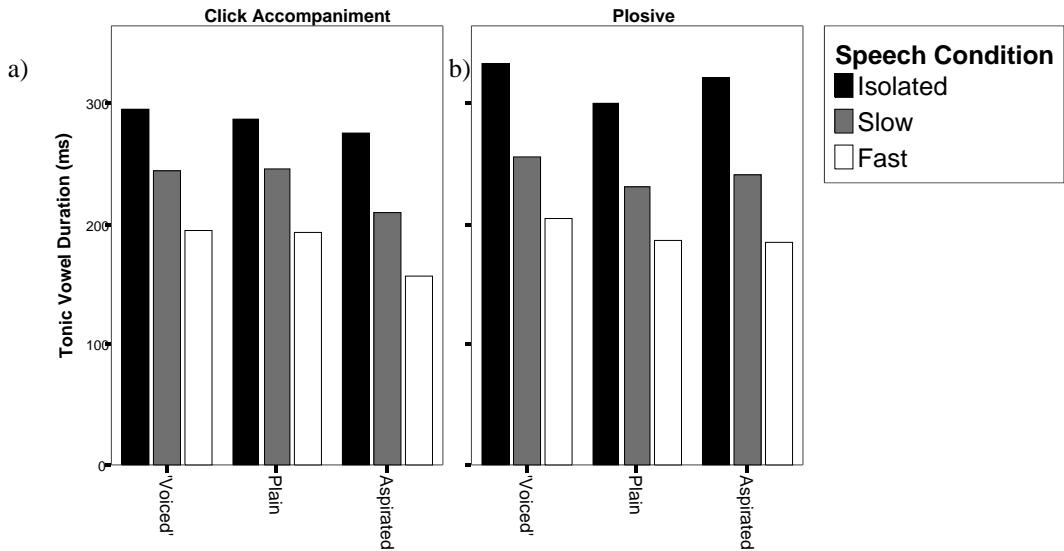


Figure 1. Click (a) and Plosives (b) initial tokens both show mean tonic vowel reduction. The mean tonic vowel lengths for the phonologically ‘voiced’ segments are longer than their respective non-depressor counterparts. The difference is not as clear for click accompaniments.

For both lists the effect of speech rate was significant on mean tonic vowel duration (plosives ($F(2, 8) = 352.15, p < .05$) and CAs ($F(2, 8) = 220.07, p < .05$)). There was also a significant effect of voice type (plosives ($F(2, 8) = 37.11, p < .05$) and CAs ($F(2, 8) = 23.80, p < .05$)). For neither list were there interaction effects. Post hoc Student-Newman-Kuels test ($p < .05$) showed a significant difference for speech rate for both lists. Post hoc Student-Newman-Kuels test ($p < .05$) showed a significant difference for phonological voice type in the plosives list between phonologically ‘voiced’ plosives on one hand and both ejective and aspirated plosives. There is no significant difference between ejective plosives and aspirated plosives. For CAs, the Student-Newman-Kuels ($p < .05$) tests showed significant differences between the aspirated CAs and the ‘voiced’ and plain CAs. There is no significant difference between the ‘voiced’ CAs and the plain CAs for voice type.

A Mann-Whitney U test was performed on mean tonic vowel duration by depressor/non-depressor. The results showed that over all speech conditions the non-depressors had significantly lower means, $z = -3.801, p < .001$. The resulting means for depressors versus non-depressors support the findings of Russell (2000) and are in Table 3.

Table 3. Depressors (“voiced”) versus Non-depressor Tonic Vowel Length (ms)

	“voiced”	ejective/plain	aspirated
Isolated	322.07	294.01	288.73
Slow	254.40	238.51	218.15
Fast	200.24	189.10	166.51

3.3. Voice onset time

The VOT results for initial plosives are given in Figure 2. Table 4 has the means and ranges for all plosives by POA, phonological voice type, and speech condition.

Figure 2. Mean voice onset time (ms) for target segments by speech condition (plosives)

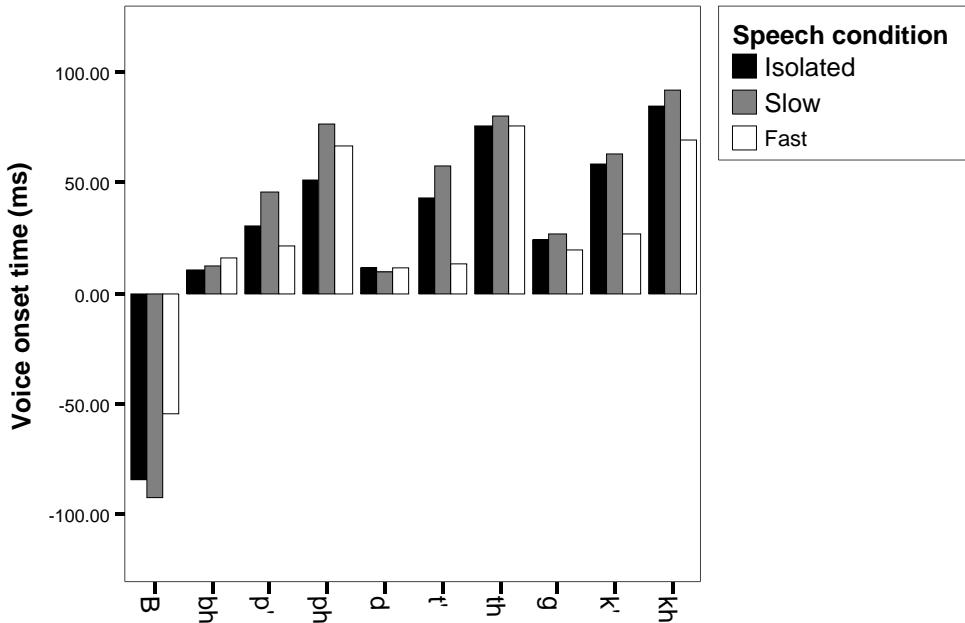


Figure 2. Above we have the mean VOTs for initial plosives. Target segments are arranged by place of articulation moving backward and by voice type moving from pre-voiced to voiceless aspirated. ‘B’ stands for the bilabial implosive.

Table 4. Mean VOT and ranges by target sound and condition for initial plosives (milliseconds).³ The Table shows the mean VOT ranges for each target sound in the plosive list. Means are **bolded**. The ranges for each sound are given underneath the means

	[b]	[bʰ]	[p]	[pʰ]	[d]	[dʰ]	[t]	[tʰ]	[g]	[k]	[kʰ]
Isolated	-84.32 -104 - -49	10.27 8 - 13	30.60 16 - 41	50.85 40 - 67	11.59 8 - 14	42.75 32 - 53	75.89 61 - 94	23.87 18 - 29	58.61 38 - 76	85.13 32 - 123	
Slow	-92.59 -121 - -74	12.06 9 - 15	46.20 14 - 66	76.81 51-123	9.69 6 - 13	57.79 30 - 91	80.20 65 - 101	36.96 19 - 40	63.19 32 - 83	92.15 72 - 128	
Fast	-55.12 -76 - -35	16.27 5 - 29	21.79 12 - 43	66.56 37-174	11.08 4 - 20	13.08 8 - 22	75.49 54 - 96	19.66 11 - 39	27.06 14 - 82	69.50 28 - 132	

³ The bilabial implosive is listed in the figure and the table to illustrate the effect of speech condition but was omitted from the ANOVA statistics because it would bias the statistic for voice type since only the bilabial series has the implosive. This is in keeping with other similar studies (Jessen 2002).

Looking at the data in Figure 2 and Table 4 we see that voiced segments have shorter VOTs than ejectives, which are in turn shorter than aspirated segments. From Table 4 we see that the implosive is pre-voiced in all conditions. In addition to voice type generalization, we also see other cross-linguistic tendencies for VOT in Zulu: as POA moves farther back in the mouth VOT tends to increase and as speech rate increases VOT tends to decrease. The differences in VOT among the voiced segments are less than the differences for other voicing types across all places of articulation and speech rates. Moreover, the voiced segments are relatively resistant to the effects of speech rate. Figure 2 shows this trend while the means in Table 4 corroborate this finding. While the ejective plosives' VOTs are longer than their voiced counterparts in the isolated and slow conditions, the differences between voiced consonants and ejectives in the fast condition are smaller and there is a marked increase in the overlap in their respective ranges for all places of articulation. All plosives' means and ranges converge toward the voiced values.

Kessinger & Blumstein (1997) showed that for some slow context cases VOTs were longer than the VOTs for the same segments in an isolated context. Our data confirm this, but it is not consistent over POA or voice type. A three-way univariate analysis was performed on the plosive list (speech rate x voicing x POA). There was a significant main effects of speech rate ($F(2, 26) = 51.15$; $p < .05$), voicing ($F(2, 26) = 512.06$; $p < .05$), and for POA ($F(2, 26) = 37.48$; $p < .05$). There were interaction effects for speech rate x voicing ($F(4, 26) = 18.70$; $p < .05$), speech rate x POA ($F(4, 26) = 6.52$, $p < .05$) and POA x voicing ($F(4, 26) = 5.39$; $p < .05$) and for speech rate x POA x voicing ($F(8, 26) = 2.67$; $p < .05$). Post hoc Student-Newman-Keuls tests ($p < .05$) showed a significant difference for voicing, POA, and between slow and fast rates of speech. All factors had significant main effects. Significant interaction effects were also found across the board.

The VOT results for initial CAs are given in Figure 3. Table 5 has the means and ranges for all plosives by POA, voicing type, and speech rate condition.

Figure 3. Mean voice onset time (ms) for target segments by speech condition (click accompaniments)

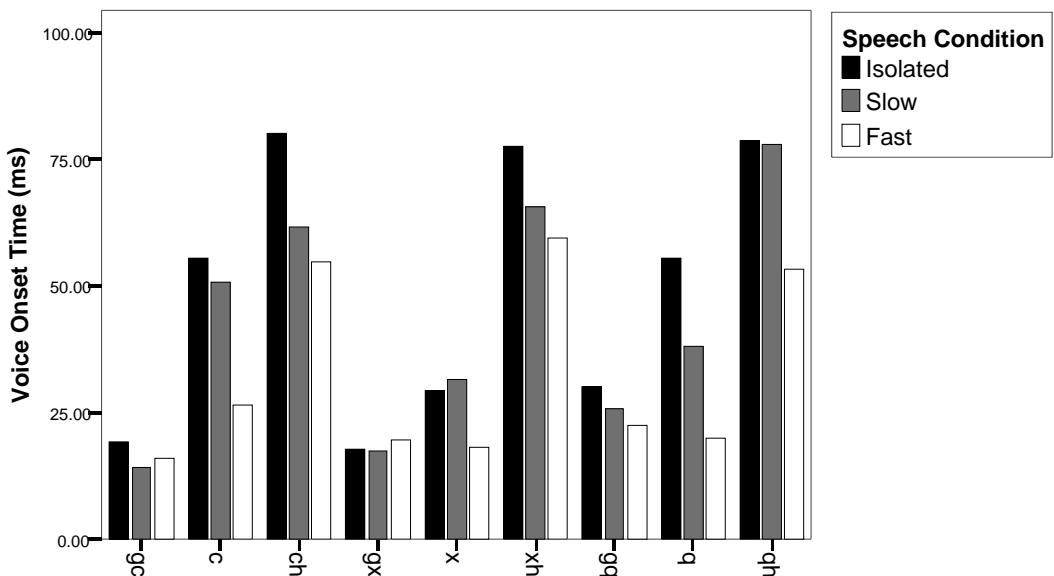


Figure 3. Above we see mean VOTs for initial click accompaniments (From left to right: Dental, alveolor-lateral and alveolo-palatal grouped by voiced, plain, and aspirated). Target segments are arranged by place of articulation moving backward and by voice type moving from pre-voiced to voiceless aspirated.

Table 5. Mean VOT and ranges by target sound and condition for initial click accompaniments. The mean VOT ranges for each target sound in the click accompaniments. Means are bolded. The ranges for each sound are given underneath the means

	[g̥]	[k]	[kʰ]	[g̥l]	[kl]	[kʰ]	[g̥!]	[k!]	[kʰ!]
Isolated	17.97	55.49	73.98	17.66	29.49	87.07	27.58	55.56	78.87
	12-24	35-79	55-98	9-23	11-48	72-108	15-42	45-68	56-114
Slow	14.09	53.82	61.81	14.98	27.47	60.19	25.92	34.83	77.89
	10-18	41-65	21-102	10-19	17-33	35-78	7-43	23-50	55-89
Fast	13.74	26.34	50.26	19.58	18.13	63.99	14.63	17.22	53.22
	10-18	13-36	31-75	7-40	12-26	53-74	8-24	9-29	20-109

The CA list (Figure 3) resembles the plosive list (Figure 2). Both lists show similar results with respect to the effect of voice type and speech rate on VOT. As speech rate increases VOT decreases. The voiced and aspirated alveolo-lateral CAs' VOT lengthen from the slow to fast context, but decrease for both from the isolated to the slow contexts. All other targets, including the voiceless alveolo-lateral CAs, show decreases. The means of the voiceless CAs and aspirated CAs VOTs converge towards the means for the voiced CAs. The tendency is also for the alveolo-lateral CAs to be shorter than their dental and alveolo-palatal counterparts. The convergence is toward the voiced accompaniments, similar to the result we found for the plosive list.

A three way univariate analysis was performed to on the CAs (speech rate x voicing x POA). POA was used to test the effect if any of the influx on the efflux. For CAs there was a significant main effect of voicing ($F(2,26) = 340.44$; $p < .05$), speech rate ($F(2, 26) = 48.62$; $p < .05$), and POA ($F(2, 26) = 4.98$; $p < .05$). There were significant interaction effects for voice x place ($F(4, 26) = 10.31$; $p < .05$), voice x rate ($F(4, 26) = 7.05$; $p < .05$), place x rate ($F(4, 26) = 3.34$; $p < .05$), and place x rate x voice ($F(8, 26) = 2.43$; $p < .05$). Post-hoc Student-Newman-Keuls tests ($p < .05$) showed significant differences for voicing and rate for all CAs. For place, post hoc analysis shows that the difference between alveolo-lateral and {dental, alveolo-palatal} was significant.

3.4. Velar plosive versus click accompaniment VOT

The VOTs for velar plosives were compared to the VOTs for velaric CAs for the effect of speech rate and voice type. Our working hypothesis was that we would find similar ranges with respect to voicing contrasts. However when tested the differences in VOT between the two groups we found a significant difference. The explanation for this is not immediately clear, but there are two possible explanations: 1) either CA releases are not uniformly velaric, or 2) they are uniformly velaric and the influx release has a shortening effect on the efflux release. The two lists were tested for differences in articulation type (plosive versus CA). Figures 4a-4c show the differences in articulation type by both speech rate and voice type.

Figure 4. Mean voice onset time (ms) for target segments by speech condition, voice type and articulation type

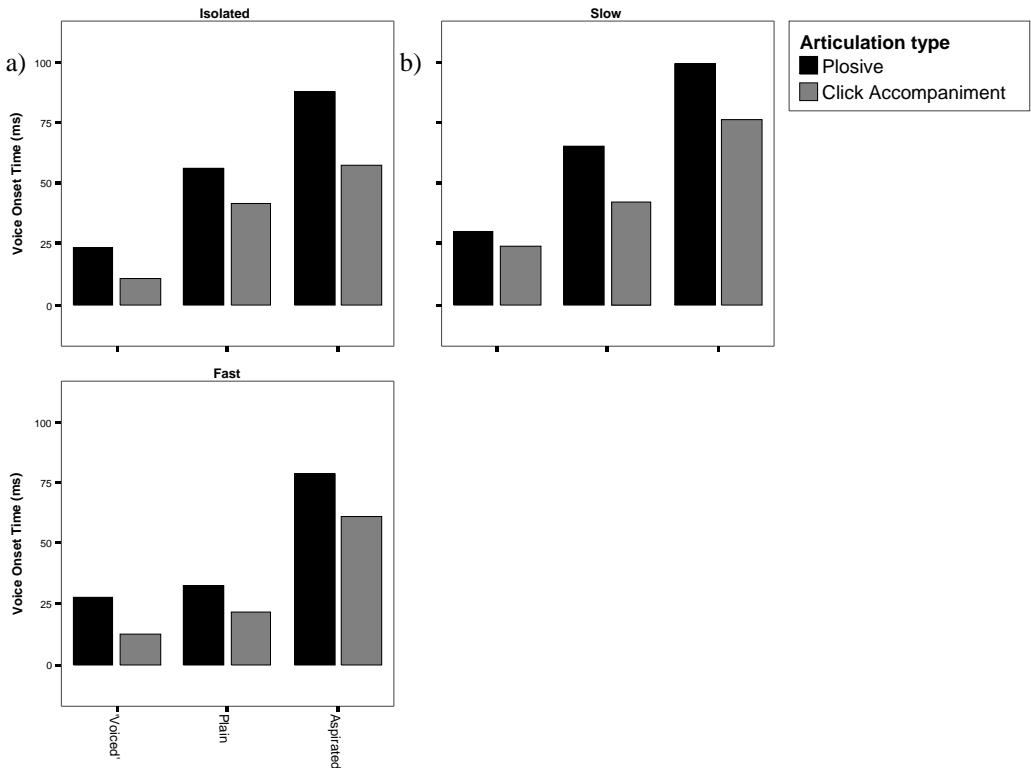


Figure 4. 4a-4c show the mean VOTs for velar plosives and click accompaniments by articulation type, voice type and speech rate. The bars represent articulation type. The horizontal axes represent phonological voice type. The panels represent each speech rate condition.

Table 6. Mean VOT and ranges by target sound and condition for velar plosives and initial click accompaniments (milliseconds). The tokens are arranged by voice type. The means are **bolded** and the ranges are underneath.

	[<u>g</u>]	[<u>g</u>]	[<u>g</u>]	[<u>g</u> !]	[kʼ]	[k]	[k]	[k!]	[kʰ]	[kʰ]	[k ʰ]	[k!ʰ]
Isolated	23.87 18-29	17.97 12-24	17.66 9-23	27.58 15-42	58.60 38-73	55.49 35-79	29.49 11-48	55.56 45-68	85.13 33-123	73.98 55-98	87.07 72-108	78.87 56-114
Slow	26.96 19-39	14.09 10-18	14.98 10-19	25.92 7-43	63.18 32-83	53.82 41-65	27.47 17-33	34.83 23-50	92.15 72-129	61.81 21-102	60.19 35-78	77.89 55-89
Fast	19.66 11-29	13.74 10-18	19.58 7-40	14.63 8-24	27.06 14-46	26.34 13-36	18.13 12-26	17.22 9-29	69.5 29-132	50.26 31-75	63.99 53-74	53.22 20-109

Figures 4a-4c show and Table 6 show that the mean VOTs for CAs and velar plosives are similar, but the mean VOT is for plosives is slightly longer than the mean VOTs for CAs. This tendency is clear in panel 4b, the slow condition. The voiced velars and CAs are similar with respect to their VOT means over all speech conditions, with the ejective plosives and voiceless CAs and aspirated plosives and CAs showing more variation (in that order). The mean values tend to be shorter for CAs and the ranges in nearly all cases show considerable overlap.

A three way univariate ANOVA analysis (speech rate x voice type x articulation type) showed that there are significant differences between the velar plosives and the CAs for voice type ($F(2, 17) = 266.96$; $p < .05$), speech rate ($F(2, 17) = 34.80$; $p < .05$), and articulation type ($F(1, 17) = 29.81$; $p < .05$). There were significant interaction effects for speech rate x voice type ($F(4, 17) = 6.24$; $p < .05$), articulation type x voice type ($F(2, 17) = 3.49$, $p < .05$), and articulation type x speech rate ($F(2, 17) = 4.29$; $p < .05$). There was no significant interaction effect of articulation type x voice type x speech rate. A post-hoc S-N-K test showed significant differences ($p < .05$) for speech rate and voice type only (there are only two articulation types).

A Mann-Whitney *U* test was conducted to evaluate the hypothesis that mean VOTs for CAs are similar to the mean VOTs for velar plosives. The results of the test showed that over all speech rates and voice types, the CAs have significantly lower VOT means, $z = -3.32$, $p < .001$. CAs had an average VOT of 170.91 while velar plosives had a mean VOT of 213.03.

4. Discussion

The purpose of this paper is to answer a few questions about voicing distinctions and the role that VOT plays as an acoustic correlate to these distinctions in Zulu.

The phonological voicing categories for Zulu are as follows: implosive, voiced, ejective, and aspirated (plosives) and voiced, plain, and aspirated (clicks). The phonetic implementation of these phonological categories shows that only the bilabial implosive is pre-voiced. The aspirated segments are all long-lag. Their VOTs are much longer than either the ejective plosives/plain clicks or voiced plosives/clicks. The phonologically voiced segments are short-lag. The ejective plosives and plain CAs are statistically shorter than the aspirated plosives under all speech rate conditions and are longer than the short-lag voiced plosives in both the isolated and slow conditions. In the fast context their VOTs shorten significantly. It seems they are neither short-lag nor long-lag, but somewhere in between (dependent on speaking rate). They are similar to the results that VanDam & Port (2003) found for English voiceless plosives in different prosodic positions. The variable nature of ejective production cross-linguistically is key to understanding the variation we see in Zulu.

These findings are interesting for several reasons. First, we review the results from Traill et al. (1987) about acoustic correlates to voicing distinctions in Zulu (as laid out by Rycroft), “[it is] where VOT values do not differ [between /bh d g/ and /p t k/], that the pitch clues are crucial” (272).

This means that there is some disagreement about the role that VOT plays as an acoustic correlate to voicing distinctions. Traill et al. showed that tonal depression is an acoustic correlate to differentiating depressor/voiced plosives from their ejective counterparts by experimental splicing of /p’ t’ k’/ to depressed high tones. These cases were perceived as /bh d g/ and not ejective. On the other hand, no artificially added pitch perturbations could cause the misperception of /ph th kh/. The same tonal depression does not account for the difference between depressor/voiced plosives and aspirated plosives. These observations concerning depressors’ pitch perturbations and VOT as cues to voicing distinctions are ‘based only on preliminary tests and require to be pursued in a systematic investigation’ (272).

In order to do this we tested VOT as an acoustic correlate to voicing by controlling for speech rate. If VOT is not a crucial ‘clue for distinguishing’ depressor/voiced plosives from ejective plosives in all speech conditions then we would expect that VOT ranges and means would be close or overlap in all speech conditions. We did not find this. VOT is a significant acoustic correlate for both lists to all voicing distinctions save voiced and ejective/plain segments in the fast context only.

4.1. Plosives

The results for the plosives in Zulu confirm what others have found with research on VOT and the effect of speech rate. Theodore et al. (2007) showed that both the effects of POA and speech rate are talker-specific (more on this and ejective production below). POA is talker-specific because while VOT for velars is always significantly longer than for labials, the magnitude by which it is longer varies significantly by speaker. They also found that the magnitude by which it decreases is stable across POA for individual speakers. This was shown by the discussion and comparison to other speakers in other studies.

Other studies on the effect of speech rate have found similar patterns. Perception studies (Summerfield 1981) show that variable speech rates are significant in the perception of phonetic categories and that timing is an intrinsic part of the measure. Altering speaking rate affects the onset of voicing and was shown to affect listeners' perceptions of what plosive they were hearing (in English). Kessinger & Blumstein (1997) showed that while speech rate affects timing, the effect that it has on VOT is not symmetric; given a basic contiguous three-way distinction, it has the greatest affect on the pre-voiced and long-lag plosives. Short-lag plosives are more 'stable' in the face of increased speech rate. That is to say that given three phonemic categories, one category acts as an anchor towards which the other categories' ranges may approach as speech rate increases. VOT work done on other Niger Congo and Nilo Saharan languages show similar results to our Zulu data for both initial plosives and initial click accompaniments (in addition to his own research on Xhosa, Jessen (2002) cites work by JC Roux (1990) and PW Lewis (1998) see also Sands 1991, Jessen & Roux 2002, Jessen 2002, Roux 2007).

We have shown that VOT is a significant acoustic correlate to voicing category distinctions for all voice types in isolated, slow, and fast speech rate conditions. The phonological categories are voiced, voiceless ejective/plain, and voiceless aspirated. Phonetically none of the categories save the implosive are pre-voiced. The implosive's average VOT values are consistent with cross-linguistic evidence presented for bilabial implosives in four-category languages.

The short lag stops /b̥ d̥ g/ have mean VOTs that hover in the high single digits to the mid-thirties in milliseconds. The aspirated segments have a very long VOT ranging from the fifty to close to 100 milliseconds. The ejectives are much more variable, with ranges that vary from as long as the aspirated in the isolated and slow contexts to as short as the voiced plosives in the fast context. This is especially apparent in [d̥a:la] (11.08 milliseconds) vs. [t̥'aka] (13.08 milliseconds). Lisker & Abramson (1967) report for some languages the VOT intervals for [p] and [g] overlap in the +20 - +30 ms range, which is the case in Zulu (where [p] is ejectivized).

It is our position that the phonologically 'voiced'/depressor plosives provide a phonetic anchor in Zulu. This is clear from the data in § 3.3. Any increase in speech rate is not accompanied by a large shift in VOT, but all other categories' VOTs converge toward these values. There is still a significant difference between the VOTs for the ejectives and the aspirated plosives, but both VOTs are reduced dramatically by increased speaking rate. The phonologically voiced category is phonetically voiceless, indicating that the contrasting phonetic category is long-lag in Zulu (for plosives). We will examine the nature of CAs below.

This study introduces data showing that the ejectives and the voiced plosives do not have similar VOT values and ranges in all contexts when speech rate is controlled. This raises an interesting question about ejection production. Jessen (2002) presented data on ejectives in Xhosa that questioned the nature of ejective production in that language. Jessen remarks that ejection is variable in Xhosa. He kept track of auditorily ejective consonants and looked at the VOT and burst amplitude to determine which of those two correlates was predictive of the ejective nature of the production. Of the possible combinations of VOT and burst amplitude only low burst amplitude combined with short VOT was not a possible combination producing auditory ejection. The other combinations seemed to be consistent for individual speakers. That is to say individual speakers tend to produce ejectives in the same way. Cho & Ladefoged (1999) present data that shows that long VOT is not necessarily a universal feature of ejection. Moreover, the manner of ejection production is likely language specific. And as Jessen further shows, the production trade off is likely to be speaker specific. Wright et al. (2002) has a similar discussion about the typology of ejective production and finds that in Witsuwit'en the average VOT /t'/ is somewhere between the voiceless unaspirated and the voiceless aspirated, but that the speaker variation makes the notion "average /t'/" problematic (62-63). This does not negate the findings here with respect to ejective VOT in Zulu, but only serves to highlight that for some Zulu speakers (and in controlled conditions), VOT is indeed an acoustic correlate to voicing distinctions (cf. Traill, et al. 1987).

Traill et al. (1987) found in their data that ejectives and voiced plosives/depressors had very similar mean VOTs (which we may logically assume means that their consultant's ejectives had low VOTs and high burst amplitude in the speech condition of their tests). We confirmed this in the fast

speech condition, but we also found that in isolation and in slow speech conditions, ejective VOTs varied considerably. It is possible that the speaker in this study normally produces ejectives with long VOTs (we did not test burst amplitude) and is able to maintain the audible ejection with shorter VOTs in the fast speech contexts by compensating with burst amplitude (or perhaps the ejection is the product of high burst amplitude and the VOT is irrelevant to the perception of ejection). Regardless, the ejectives in this study have VOTs that are very susceptible to speech rate effects. In the fast context, the VOTs of the ejectives are virtually identical to the VOTs of the voiced consonants/depressors. It is in this context that we might revisit the findings of Traill et al. (1987) and argue that when the VOTs are not significantly different, the differentiating acoustic correlate between these two classes might very well be the pitch perturbations present from the depressors. This last point is conjecture, but it is logical, given the cross-linguistic variability in ejection production. Isolated and slow speech is more likely to produce longer VOTs with respect to ejection.

4.2. *Click accompaniments*

The effects of speech rate on CAs were shown to be similar to speech rate effects on plosives. Voice type was also shown to be similar, but POA is another issue. It is tradition to say that clicks are made with a velaric ingressive airstream. The release of the first closure creates a rushing of air into the oral cavity which is followed by the release of the back closure. Since it is assumed that the secondary closure is the same for all clicks we would expect two things: 1) VOT values would be positive and 2) the mean differences between values for POA and voice type would be relatively small and VOT intervals would exhibit overlap.

On these two points we find that all VOTs are positive and there is considerable overlap in their ranges. Comparisons to work on Xhosa by Sands (1991) show that while the Zulu VOTs are slightly shorter for nearly all places of articulation and voice type, they exhibit the same degree of overlap. The overlap comes from the fact that the posterior release for all is (presumably) in approximately the same place. We may then remark that POA of the anterior release seems to have an effect on the posterior release but it is not predictable. The alveolo-lateral click showed significant differences (shorter!) from the dental and alveolo-palatal clicks, but the latter two were not significantly different from each other. The front closure in the dental click is farther forward than either the alveolo-lateral or alveolo-palatal, but yet has a longer VOT than the alveolo-lateral. The difference between the alveolo-palatal and the dental was not significant. With this in mind we cannot make a prediction based on the distance between the anterior and posterior closures, nor can we make any claim that the influx has a predictable effect on the release of the efflux.

One tendency that clicks and their accompaniments follow is the voice type distinction within series. The VOTs are significantly different from voiced clicks to plain voiceless and from plain voiceless to voiceless aspirated. The effect of speech rate on VOT categories is similar to what we found for the plosives. The phonetic anchor of the click and their accompaniments is the voiced CAs, which like the plosives, are phonologically voiced.

The CA's VOTs are shorter than the velar plosive VOTs. We see that the relative difference between velar plosive VOTs and CA VOTs are stable. Beyond this difference, the Figures in 4 show that the efflux releases produce VOTs that behave similarly to the VOTs of velar plosives with respect to speech rate. Which leads to a question for further research: why are VOTs for the velar releases of CAs shorter than the velar releases for velar plosives? Aerodynamically, what allows for the shorter onset? Is it a result of some articulatory effect the release of the influx has on the release of the efflux? Based on the results reported in § 3.4, CA releases are only 80% of velar plosive releases. There are two possible explanations: 1) the posterior closures are not velaric or 2) the posterior release is affected idiosyncratically by the anterior release (e.g., a degree of articulatory slip caused by air pressure differentials). Both seem plausible, but instrumental research on click production using electropalatography and ultrasound technologies have shown that the posterior release for CAs varies with the anterior release (Thomas-Vilakati 1999, Miller et al. 2007). Thomas-Vilakati showed that the alveolo-palatal click actually has an uvular release. Ultrasound work on N|uu by Miller et al. (2009). showed that “different click types differ not only in their anterior constriction, but also the locations of their posterior releases” (45).

Miller et al. (2009) go further showing that in N|uu central alveolar clicks [!] and palatal clicks [‡] both have post velar posterior places of articulation while they get the impression that dental clicks [!] have an upper-pharyngeal posterior constriction and the lateral alveolar click [||] has again a post-uvular posterior constriction (45). Miller et al. (2007) conclude that the “notion of velaric airstream mechanism is too simplistic...as the posterior releases of both clicks involve retraction into the post-velar region.” They propose that the term “lingual airstream mechanism should be used instead” and call for more ultrasound imaging (9).

This confirms that the first hypothesis is the correct one. With this in mind it does not surprise us that the timing of an acoustic correlate such as VOT would show overlap, but also that there would be no predictable pattern with respect to the influx release. In essence, they argue that the posterior release point is dependent on the anterior POA, which taking into account the physical apparatus of click production, is logical.

4.3. Depressors/Non-depressors

One of the impetuses of this study was the data showing that it was tonal depression and not VOT that is the major acoustic cue to voicing distinctions between depressor plosives and ejective plosives and depressor clicks and plain clicks. Along the way we confirmed the findings of Russell (2000) on the vowel lengthening effects of tonal depression. However, post hoc tests did not show a significant difference between the depressor clicks and the plain clicks for VOT. These findings call for a more detailed study of depressor consonants by a more natural class distinction: plosives and clicks, for example. Both depressor clicks and plosives have similar lengthening affects on neighboring vowels but they do not behave similarly with respect to VOT.

We also partially confirmed the findings of Traill et al. with respect to VOT and tonal depression as acoustic correlates to voicing contrasts in Zulu. We found that ejectives are distinct from depressor plosives (and plain clicks from depressor clicks) in isolated and slow contexts but not in the fast speaking context. These findings seem contrary to their data in both isolated and normal speech contexts. According to the phonetic literature (Ladefoged & Traill 1994, Russell 2000) voiced clicks are also depressor consonants and Traill et al. did not elicit them. Moreover, the CAs show the same behavior as plosives—voiced and plain voiceless CAs are distinct with respect to mean VOTs in the isolated and slow contexts but not the fast context. We countered this possible complication with the plosives list by including a discussion of the ejectives produced by this speaker. All of them had very high VOTs. Work by Cho & Ladefoged (1999), Wright et al. (2002) and Jessen (2002) present the variable, language specific, and in some languages, talker specific nature of ejective production.

References

- Cho, Taehong, and Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27. 207-29.
- Doke, Clement Martyn. 1923. A dissertation on the phonetics of the Zulu language. *Bulletin of the School of Oriental Studies, University of London* 2. 685-729.
- Doke, Clement Martyn. 1926. *The phonetics of the Zulu language*. Johannesburg: University of the Witwatersrand Press.
- Doke, Clement Martyn. 1950. *Text-book of Zulu grammar*. Johannesburg; London: University of the Witwatersrand Press; Longmans, Green.
- Doke, Clement Martyn. 1958. *English-Zulu dictionary*. Johannesburg: Witwatersrand University Press.
- Fischer-Jørgenson, Eli. 1954. Acoustic analysis of stop consonants. *Miscellanea Phonetica* 2. 42-59.
- Giannini, Antonella, Massimo Pettorino and Maddalena Toscano. 1988. Some remarks on Zulu stops. *Afrikanistische Arbeitspapiere* 13. 8-25.
- Hodgson, Phillip, and Joanne L. Miller. 1996. Internal structure of phonetic categories: Evidence for within category trading relations. *Journal of the Acoustical Society of America* 100. 565-76.
- Jessen, Michael. 2002. An acoustic study of contrasting plosives and click accompaniments in Xhosa. *Phonetica* 59. 150-79.
- Jessen, Michael, and Justus C. Roux. 2002. Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics* 30. 1-52.

- Kessinger, R. H., and S. E. Blumstein. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25. 143-68.
- Ladefoged, Peter, and Ian Maddieson. 1995. *The sounds of the world's languages*. Oxford, UK; Cambridge, Mass.: Blackwell.
- Ladefoged, Peter, and Anthony Traill. 1994. Clicks and their accompaniments. *Journal of Phonetics* 22. 33-64.
- Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384-422.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge Cambridgeshire ; New York: Cambridge University Press.
- Miller, Amanda L., Johanna Brugman, Bonny Sands, Levi Namaseb, Max Exter, and Chris Collins. 2009. *Differences in Airstream and Posterior Place among N/uu Lingual Stops*. *Journal of the International Phonetic Association*, 39.2, 129-161.
- Miller, Amanda, L. Namaseb, and K. Iskarous. 2007. Tongue body constriction differences in click type. In *Laboratory Phonology 9*, ed. by Jennifer Cole and Jose Ignacio Hualde. New York: Mouton de Gruyter. 643-656.
- Miller, Joanne L. 1977. Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America* 20. 641-8.
- Miller, Joanne L., and Thomas Baer. 1983. Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America* 73. 1751-5.
- Miller, Joanne L., Cynthia M. Connine, Trude M. Schermer, and Keith R. Kluender. 1983. A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America* 73. 2124-33.
- Miller, Joanne L., Kerry P. Green, and Adam Reeves. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43. 106-15.
- Miller, Joanne L., Timothy B. O'Rourke, and Lydia E. Volaitis. 1997. Internal structure of phonetic categories: Effect of speaking rate. *Phonetica* 54. 121-37.
- Peterson, Gordon E., and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32. 693-703.
- Pind, Jörgen. 1995. Speaking rate, VOT, and quantity: The search for higher order variants in two Icelandic speech cues. *Perception & Psychophysics* 57. 291-304.
- Poulos, G., & Msimang, C. T. (1998). *A linguistic analysis of Zulu*. Cape Town: Via Afrika.
- Roux, Justus C. 2007. Unresolved issues in the representation and phonetic click articulation in Xhosa and Zulu. *Language Matters* 38. 8-25.
- Russell, Margaret A. 2000. Phonetic Aspects of Tone Displacement in Zulu. In *Proceedings from the 36 Meeting of the Chicago Linguistic Society: The Main Session*, ed. Akira Okrent and John Boyle. The Chicago Linguistic Society: Chicago, IL.
- Sands, Bonnie. 1991. Evidence for click-features: Acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics*. 6-37.
- Stevens, Kenneth, N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17. 3-45.
- Summerfield, A. Quentin. 1981. On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7. 1074-95.
- Taljaard, P. C., & J.W. Snyman 1991. *An introduction to Zulu phonetics* (Second impression ed.). Constantia: Marius Lubbe Publishers.
- Theodore, Rachel M., Joanne L. Miller, and David DeSteno. 2007. The Effect of Speaking Rate on Voice-Onset-Time is Talker-Specific. *Proceedings of the XVIth International Congress of Phonetic Sciences*. Saabruken, Germany.
- Thomas-Vilakati, Kimberly Diane. 1999. *Coproduction and coarticulation in IsiZulu clicks*. Los Angeles, CA: University of California, Los Angeles: PhD Dissertation.
- Traill, Anthony, James S. Khumalo and Paul Fridjohn. 1987. Some Depressing Facts about Zulu. *African Studies* 46. 255-274.
- VanDam, Mark. 2007. *Plasticity of Phonological Categories*. Bloomington, IN: Indiana University Dissertation.
- VanDam, Mark, and Robert Port. 2003. VOT of American English stops with prosodic correlates. Presented at the 145th Meeting of the Acoustical Society of America. Nashville, TN.
- VanDam, Mark, and Robert Port. 2005. VOT is shorter in high frequency words. Poster presented at the 145th Meeting of the Acoustical Society of America. Vancouver, BC, Canada.
- Volaitis, Lydia E., and Joanne L. Miller. 1992. Phonetic prototypes: Influences of place of articulation and speaking rate on the internal structure of voicing contrasts. *Journal of the Acoustical Society of America* 92. 735.
- Wright, Richard, Sharon Hargus and Katharine Davis. (2002). On the categorization of ejectives: data from Witsuwit'en. *Journal of the International Phonetic Association* 32. 43-77. doi:10.1017/S0025100302000142

Selected Proceedings of the 40th Annual Conference on African Linguistics: African Languages and Linguistics Today

edited by Eyamba G. Bokamba,
Ryan K. Shosted, and Bezza Tesfaw Ayalew

Cascadilla Proceedings Project Somerville, MA 2011

Copyright information

Selected Proceedings of the 40th Annual Conference on African Linguistics:
African Languages and Linguistics Today

© 2011 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-446-1 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Midtlyng, Patrick J. 2011. The Effects of Speech Rate on VOT for Initial Plosives and Click Accompaniments in Zulu. In *Selected Proceedings of the 40th Annual Conference on African Linguistics*, ed. Eyamba G. Bokamba et al., 105-118. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2569.