NOTES ON UNENCODED SPEECH: CLICKS AND THEIR ACCOMPANIMENTS IN XHOSA

G. Dogil* and J. Roux**

*Chair of Experimental Phonetics
University of Stuttgart

**Research Unit for Experimental Phonology
University of Stellenbosch

ABSTRACT - In this paper we argue that clicks are a prototypical case of unencoded speech. Unlike other speech sounds clicks do not coarticulate with their phonetic environment, they block vowel coarticulation across them and may not be recovered from the transitional features which are left after editing them. Unencoded sounds as such present a challenge to most phonetic theories in particular, to all theories of coarticulation and coproduction.

INTRODUCTION

The results of the past decades of phonetic and phonological research show that speech is not perceived, produced, or neurally processed on a segmental basis. Normal human speech appears to be highly encoded, that is time-compressed to a degree in which the acoustic and motor properties are shared by all sounds forming larger speech units such as syllables (Liberman 1996). The acoustic cues of consonants are usually modulations of the formant frequency of their tautosyllabic vowels, and the spectral properties of the vowels are usually encoded in the spectral properties of the consonants of the same syllable. Furthermore, the articulatory targets for segments are never reached, and instead, the articulatory gestures are merged within the larger linguistic constituents, the primary one being the syllable. All natural speech results from complex encoding, which makes it extremely difficult to locate and isolate the invariant articulatory and acoustic properties of particular segments. Speech encoding has always been considered prime evidence for larger constituents like the syllable. However, the degree of speech encoding may differ for particular speech sounds, for various speech tasks, for different linguistic categories coded in speech, and for different parts of the speech code. For example, it has been shown that phonetic properties pertaining to stress (like duration) are encoded within the rime, but not within the onset. Further similar asymmetries have been discovered experimentally, others are being simulated in speech recognition research. There exists, however, at least one category of sounds for which it is safe to say that they are unencoded, or at least much less encoded than any other supralaryngeally articulated elements of the speech code. These sounds are clicks (Ladefoged & Maddieson 1996: 246-281). In our presentation we will refer to some "unencoded" properties of Xhosa clicks with special reference to their coarticulatory properties. We will argue that the fact that clicks are phonetically and phonologically dissociated from the rest of the speech code presents a challenge to theories of syllabic encoding. Furthermore, we will argue that the extreme richness of click accompaniments derives from the resistance of the anterior click articulations to interact with the rest of the speech code. Accompaniments are links between the clicks and the rest of the speech code.

In the oral version all coarticulatory properties of clicks will be discussed, in the written version we concentrate on some of the coarticulation-blocking properties of these consonants.

THREE EXTRAORDINARY PROPERTIES OF CLICKS

A milestone in the description of clicks has been reached with the seminal work Ladefoged & Traill (1994) (henceforth, L&T) on clicks and their accompaniments. While this work primarily focuses on the nature and extent of click articulations in general and in the Khoisan languages in particular, it also devotes some attention to various click articulations and their accompaniments found in a Bantu language, Xhosa (L&T 1994: 46-48). Despite the comprehensiveness of this work, a number of issues still remain unresolved, specifically in as far as it concerns a language such as Xhosa as well as its sister languages Zulu and Swati.

L&T give the impression that clicks are quite ordinary type of sounds, which are not only fairly easy to produce but also perceptually salient to a degree which should make them „highly favoured consonants in the worlds languages. (...). Indeed, we cannot explain why these easy to make and perceptually optimal consonants are found in so few languages“ (L&T: 62). We will argue that the possible reason for this restricted distribution of clicks as well as many other of their extraordinary features which have been reported in literature is due to the fact that they are phonetically „unencoded“. Before arguing this position in detail, however, let us review some of these extraordinary features of clicks.

Velic Suction

A lot of research has been devoted into the articulatory description of clicks in various languages and the picture that emerges from this research is relatively clear. The full-fledged production of any click consists of three distinct phases:

The Closure Phase, which in itself consists of two closures:
- the initiatory closure, which is velar for all clicks
- the articulatory closure, which is coronal for all oral clicks and labial for the labial click.

The Suction Phase, during which the air in the cavity between the two closures is rarefied by the downward movement of the tongue.

The Release Phase, which in itself consists of two releases:
-the articulatory release, which has been called influx in traditional descriptions and which is believed to determine the click type (L&T 34-35),
-the initiatory release, always velar or uvular, called efflux by traditional phonetics and referred to as the click accompaniement in L&T.

All these phases are timed in such a way, that an acoustically crisp (cf. Johnson 1993) and perceptually salient sound is produced (cf. Traill 1994). The salience is concentrated within the period of the noiseburst (the release phase) itself (cf. Traill 1994), because clicks appear not to coarticulate with their phonetic context as we will show (cf. Sands 1992). The schematic version of the articulatory phases and the resulting noisebursts of a Xhosa voiced alvelo-palatal click [!] are given in figure 1.

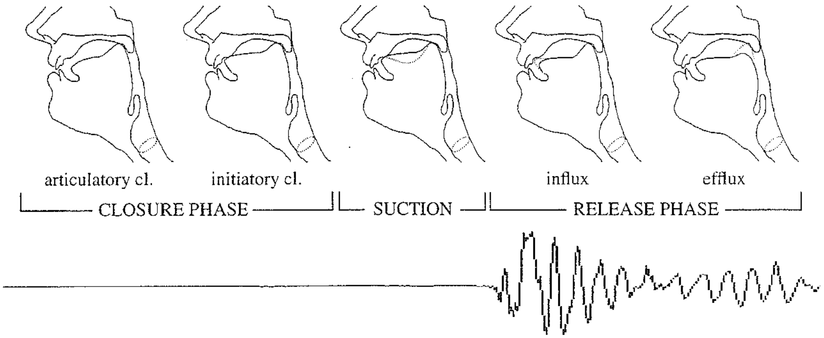

FIGURE 1. Articulatory and acoustic phases in the production of a click

Phonological opacity

Click consonants prove to be highly resistant to phonological processes induced by their environment. Out of approximately 36 clicks using languages on record - 8 Bantu, 12 Bushman, and 16 Hottentot - there is not a single one in which clicks mutate synchronically to other sounds. This is true both of the indigenous click languages of the Khoisan group and of the Bantu languages into which they entered through adoption, not mutation. Synchronically, clicks do not undergo phonological changes, neither in a particular phonetic, phonological or lexical context nor context independently as a whole class (cf. Dogil & Luschützky 1990: 33-36).

The resistance to context influence and the inability to influence the immediate context is also a feature of clicks in diachronic phonology. Traill & Vossen(1996) provide a complete list of diachronic changes of clicks in Khoisan languages and there does not appears to be a single one among them which is phonetically or phonologically conditioned.

> „*The changes we have been examining are context free in the sense that they affect a whole class of sounds no matter what the phonological or phonetic context.*" (Traill & Vossen 1996:30)

This context free behaviour is to be expected from unencoded sounds. In terms of feature geometry, sounds that do not interact with the context should be maximally and discretely specified. Indeed, all attempts at representing clicks in terms of feature geometry have assigned very complex structures to them. Consider Keyser & Stevens (1994: 222, 224); Halle (1995: 8-12); Sands (1993) for some more recent attempts. Note, however, that representing clicks in segmental phonology is in itself methodologically dubious. Standard view is that decisions about feature geometry should be grounded on generalizations regarding phonological processes. Since no process shows that clicks interact with any other part of the phonology, there is no reason to reflect them formally in the phonology. If they are phonological (segmental) at all, they appear to be highly opaque.

Phonetic unencoding

Recent acoustic phonetic research on clicks has shown that they have very distinctive noiseburst properties (cf. Traill 1994; L&T; Traill [forthcoming a,b]; Roux et al. 1995). Clicks also have quite unique temporal and intensity features. This allows the classification of [abrupt] and [noisy] and [grave] acoustic features (cf. Traill's recent work). Hence, most of the features relevant for the identification of a click are concentrated within its own noiseburst. Traill (1994) has also convincingly shown that clicks are perceived as soon as their noiseburst is present. Also, unlike other obstruent consonants, clicks are virtually never identified in syllables from which they had been edited. This points to the conclusion that they are in a sense unencoded. In the oral version we present evidence that this is supported by the measurements of the deltas of F2 and F3 in the vicinity of the click. Below, we provide experimental evidence for the other aspect of speech unencoding - the blocking of V-to-V coarticulation.

METHOD

Subjects

A homogeneous group of eight female native Xhosa speakers was chosen for the experiment. All speakers came from the same area (Cape Province), all were of approximately the same age (20 - 25 years old) and they all attended the same school - the Teachers Training College in Cape Town. They are studying to become professional Xhosa teachers. They have no speech, language or hearing impairments. The recordings were made by Phillip Lewis in a sound treated room of the Groote Schuur Hospital (Cape Town) using professional equipment.

Stimuli

Stimuli used in the experiment consisted of a set of nonsense words of the form [iCili] ~ [iCuli] with a variation in the C position. In the actual test the lateral [l], the four aspirated stops {bilabial, dental, palatal, velar}, the four ejectives, the bilabial implosive, the four murmured voiced stops {bilabial, dental, palatal, velar}, and the fifteen phonemic Xhosa clicks (see figure 2) were placed in this critical position between two vowels [i_i] and [i_u].

	clicks														
ACCOMP.	voiceless			aspirated			voiced-breathy			nasal			nasal-breathy		
<ORTH.>	c	q	x	ch	qh	xh	gc	gq	gx	nc	nq	nx	ngc	ngq	ngx

FIGURE 2. A table of Xhosa phonemic clicks. <c> [ǀ] is a noisy dental click. <q> [ǃ] is an abrupt alv.-pal. click. <x> [ǁ] is an alveolar/lateral click.

The vowels were chosen because they represent the extremes in terms of tongue position and lip rounding of the five elements of the Xhosa vowel system. The expectation of the theories of coarticulation is, that the vowel [u] will influence the vowel [i] in the [iCu..] context. The expected influence is to be observed both on the tongue position plane (the constriction should be further back - i.e. F2 formant should be lowered) and on the lip rounding plane (the lips are expected to be less spread and more protruded - i.e. F3 should be lowered).

Data collection

The tokens were digitized at a 16 kHz sampling rate and analyzed using LPC analysis routines of the Entropic/xwaves program (xspectrum). The analysis used the LPC BURG method, a 20 ms Hamming window, with 20 LPC coefficients.

For each talker and for each token F2 and F3 values were calculated. For this calculation the formants were measured at four different points: in the middle of the vowels preceding and following the critical consonant, and immediately before and after consonant closure.

The mid point of the vowels was defined using two criteria: the highest F1 value read from the LPC spectrum (corresponding to the greatest articulatory opening - the peak in sonority); the arithmetical mid value of the duration of the vowel's steady state. In most cases both criteria picked the same point.

The vocalic edges were defined using three criteria: they were the last/first analysis frames in which three formants were visible in the spectrogram; they coincided with the changing amplitude envelope in the vocalic portion of the waveform; they had a residual energy 10-25% of that found in the steady state of the vowel. In all cases at least two of these three criteria determined the edge points.

DATA ANALYSIS AND DISCUSSION

In order to test the tranparency/opacity of the consonant to its adjacent vowels we have calculated the differences between the F2's and F3's in the first [i]'s of [i_1 C i l i] and [i_2 C u l i]. In the case of transparency the difference F2/F3 [i_1 - i_2] is supposed to be large, i.e. , the back position of the tongue in [u] exerts influence on the front tongue position of [i_2] across the consonant, thus lowering F2. Similarly, in the case of consonant transparency, the lip rounding of [u] is transported to [i_2] across a consonant which results in the significant lowering of the third formant (F3). In cases where there is a difference in one formant only, one has to claim that the affected consonants are transparent only on one plane (i.e. backness or rounding/protrusion). If there is no difference in the formant values in F2/F3 [i_1 - i_2] we have to conclude that the consonants in question *block* the V-to-V coarticulation. Another interesting case of interaction and actually the one employed most often by our speakers, is the one where the value of F2/F3 [i_1 - i_2] is negative. This is the case when speakers dissimilate the vowels by choosing the particularly front and spread variaty of an [i] in front of a [Cu]. In theory, this process is supposed to be taking place in the vicinity of *unencoded speech sounds* which not only block the coarticulation (coproduction) on the vocalic plane, but also break up the signal into discontinuous and discrete parts. The results of the calculations for various classes of Xhosa consonants are given in figure 3.

As the graphic shows, lateral [l] is transparent to both backness and rounding parameters (cf. high positive difference for both F2 (dark bar) and F3 (dotted bar) between the [i C i] and [i C u] conditions). Similarly, there is a strong coarticulation in the backness feature (cf. F2 differences = dark bars) of voiced

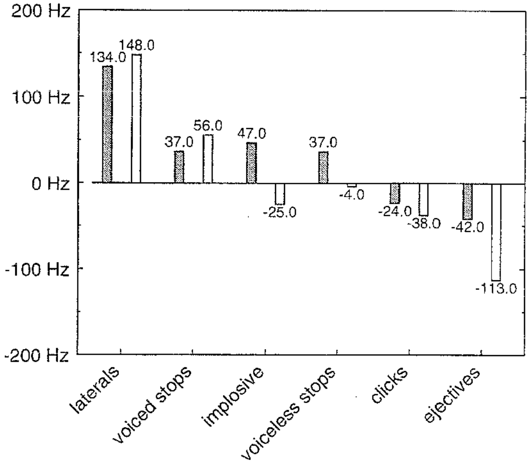

Figure 3. V-to-V coarticulation across consonant classes in Xhosa

and voiceless stops and the implosive. The voiced implosive in Xhosa is bilabial, and hence the dissimilatory influence upon the rounding/protrusion feature is to be expected. Notice that the lips are particularly active (and tense) in the production of the implosive bilabial, thus it is to be expected that the anticipatory [u] effects would be eliminated by this action. There is no simple explanation for the opacity of voiceless stops on the rounding/protrusion plane. Note the small negative value (-4.0 Hz) for F3 [i_1 - i_2] across these consonants. The only explanation that we can provide has to do with the durational properties of these consonants. Xhosa pulmonic voiceless stops are strongly aspirated, and these durational property may override the anticipatory tendency of lip rounding which is strong in voiced (much shorter) pulmonic stops.

Both clicks and ejectives are unencoded according to our criteria. A strong dissimilatory tendency in V-to-V relation is illustrated for both classes of sounds in figure 3. This tendency is even stronger in ejectives than in clicks. There is, however, an important difference. Whereas clicks are opaque irrespective of their place of articulation, ejectives block coarticulation in accordance with their articulatory (place) properties. This part of the analysis is based on A.I. rule estimations using the C4.5 machine learning technique (Quinlan, 1993). Unfortunately, the limits of space do allow us to discuss these results in detail (cf. Dogil & Roux, i.p.). This difference is supported by the analysis of ΔF's (difference between vowel mid-points and vowel-edges) for both types of sounds. Whereas ΔF's for ejectives are reminiscent of those for stops at the same place of articulation, ΔF's for clicks do not differ according to their place features.

The coarticulation-blocking properties of clicks are illustrated in figure 4. As can be seen the type of the accopmaniment (the property of the efflux) appears to be the decisive parameter. Voiceless and aspirated clicks are the most unencoded (i.e. dissimilating) ones, whereas the voiced and the nasal clicks are simply not transparent. The dissimilatory effect of clicks is stronger on the rounding/protrusion plane than on the backness/frontness plane. The place of articulation and the type of the articulator appear to play no role (cf. Dogil & Roux [i.p] for details of ΔF analysis). Necessary and sufficient for the coarticulation blocking effect is the presence of the noiseburst. The accompaniment (the efflux) apparently influences the degree of "unencodeing" of a click, the explanation of this influence, however, has to be left to a separate study.

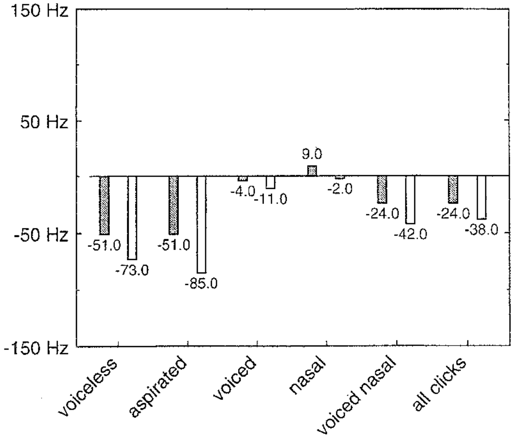

Figure 4. V-to-V coarticulation across clicks in Xhosa

REFERENCES

Dogil, G. & Luschützky, H.C. (1990) Notes on sonority and segmental strength, *Rivista di Linguistica* 2. 3-54.

Dogil, G. & J. Roux (in preparation) Notes on unencoded speech: clicks and their accompaniments in Xhosa.

Halle, M. (1995) Feature geometry and feature spreading, *Linguistic Inquiry* 26. 1-46.

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

Keyser, J. & K. Stevens (1994) Feature geometry and the vocal tract, *Phonology* 11. 207-236.

Ladefoged, P. & Traill, A. (1994) Clicks and their accompaniements, *Journal of Phonetics* 22. 33-64.

Liberman, A. (1996) *Speech: a special code*, (MIT Press: Cambridge, Mass.).

Quinlan, J. R. (1993) *C4.5 programs for machine learning*, (Morgan Kaufman: San Mateo)

Roux, J., Dogil, G., Wokurek, W. (1995) Click articulations in Xhosa: New perspectives through Wigner Distribution Analysis, *Proceedings of the XIIIth ICPhS Stockholm* 2. 574-577.

Sands, B. (1992) An acoustic study of Xhosa clicks, *Proceedinds of the XIIth ICPhS Aix-en-Provence* 4. 130-133.

Sands, B. (1993) Aperture theory and the representation of clicks in southern african languages, Abstract of a talk given to the Phonology Reading Group, University of Arizona.

Traill, A. (1994) The perception of clicks in !Xóo~, *Journal of the African Languages and Linguistics* 15. 161-174.

Traill, A. (forthcoming, a) Design features and the perception of clicks. *Proceedings of the Tüzing Khoisan Conference*, July 1994. (Quellen zur Khoisan Forschung. Köln: Rüdiger Köppe).

Traill, A. (forthcoming, b) Linguistic phonetic features for clicks: articulatory, acoustic and perceptual evidence, *Papers of the First World Congress of African Linguistics*, (Kwaluseni, Swaziland, edited by R.K. Herbert).

Traill, A. & Vossen, R. (1996) Sound change in the Khoisan languages: new data on click loss and click replacement, Unpublished manuscript, University of the Witwatersrand, Johannesburg.