

Click reduction in fluent speech: a semi-automated analysis of Mangetti Dune !Xung

Amanda Miller

Department of Linguistics
The Ohio State University
miller.5592@osu.edu

Micha Elsner

Department of Linguistics
The Ohio State University
melsner0@gmail.com

Abstract

We compare click production in fluent speech to previously analyzed clear productions in the Namibian Kx'a language Mangetti Dune !Xung. Using a rule-based software system, we extract clicks from recorded folktales, with click detection accuracy about 65% f-score for one storyteller, reducing manual annotation time by two thirds; we believe similar methods will be effective for other loud, short consonants like ejectives. We use linear discriminant analysis to show that the four click types of !Xung are harder to differentiate in the folktales than in clear productions, and conduct a feature analysis which suggests that rapid production obscures some acoustic cues to click identity. An analysis of a second storyteller suggests that clicks can also be phonetically reduced due to language attrition. We argue that analysis of fluent speech, especially where it can be semi-automated, is an important addition to analysis of clear productions in understanding the phonology of endangered languages.

1 Introduction

We compare click production in fluent speech to previously analyzed clear productions in the Namibian Kx'a language Mangetti Dune !Xung (hereafter !Xung). This language contains the four coronal click types recognized by the IPA (Association, 2006). Most content words contain an initial click, making clicks an important marker for lexical identity and useful for marking the beginning of words for speech processing (Beckman, 2013). Miller and Shah (2009) show that temporal cues, burst duration (BD) and rise time to peak intensity in the click burst (RT); and spectral cues, Cen-

ter of Gravity (COG) and Maximum Burst Amplitude (MBA), differentiate the clicks in clear productions. We extend this analysis to fluent, naturalistic speech in a corpus of folktales (Augumes et al., 2011). Using a semi-automated rule-based method to locate the clicks in the acoustic data, we are able to inexpensively align a large enough portion of the corpus for acoustic analysis. We show that the cues identified by Miller and Shah (2009) are less effective in differentiating clicks in running speech, providing quantitative evidence that this rare class of consonants is subject to phonetic reduction. Finally, we provide an analysis of the acoustic cues which differentiate the clicks, showing that the best cues for discriminating productions vary among speakers, but that in general spectral cues work better than measures of amplitude. Overall, our results demonstrate that clicks, which are known for being unique in their loudness, are not always so loud, and that even sounds that are known for their loudness undergo reduction just like other speech sounds.

2 Background

2.1 Click Burst Amplitude

It has long been noted that clicks are louder than pulmonic stop consonants. Ladefoged and Traill (1994) note that clicks in !Xóõ often have a peak-to-peak voltage ratio that is more than twice that of the onset of the following vowel (about a 6 dB difference in intensity), which Traill (1997) compares to Greenberg (1993) description of pulmonic stops as typically "... 40 dB less intense than the following vowel." This property of clicks should make them easy to recognize automatically, even with relatively unsophisticated methods. While Li and Loizou (2008) have shown that low amplitude pulmonic obstruents are degraded in noisy speech environments, clicks might be expected to differ in this regard due to their typically high amplitude

click	Alveolar, Lateral	Palatal	Dental
IPA Sym	!,	≠	

Table 1: Intensity of noisebursts for !Xóǀ clicks (loudest to quietest) (Traill, 1997).

bursts.

Previous work on click amplitudes has also noted a large degree of variability, which could make both click recognition and differentiation of different click types more difficult. Traill (1997) provides an intensity scale as in Table 2¹, but states that there is a great degree of variability. Miller-Ockhuizen (2003) shows similar results for Ju|’hoansi, and also comments on the great degree of variability. Traill and Vossen (1997) note that there is a large degree of variability in the amplitude of click bursts. None of these studies has numerically quantified the variability or determined to what degree it makes clicks confusable with non-clicks or with each other.

Traill (1997) argues that non-pulmonic stop consonants are enhanced versions of pulmonic stops, given the high amplitude bursts that are typically louder than the following vowel in clicks and in intermediate intensity bursts found in ejectives, building on Stevens and Keyser (1989)’s theory of consonant enhancement. The theory suggests that clicks should be easier to identify in the acoustic signal than pulmonic consonants. However, clicks with lower amplitude should also result in lower perceptibility in human speech recognition and lower identification in automatic speech recognition. What remains unknown from this work is whether or how often low-amplitude clicks are actually produced.

2.2 Click Burst Duration

Click amplitudes are useful cues for extracting clicks from the speech stream, but in order to differentiate between click types, listeners must also attend to other features. Previous work agrees that these features indicate different manners of articulation, although they differ in their theoretical account of the underlying phonological contrasts. Beach (1938) referred to this difference as affricate vs. implosive. Trubetzkoy (1969) recouched the manner contrast among clicks as fricative vs. occlusive. Both articulatory based

¹Traill’s scale also includes the bilabial click [ǀ], which is on average less intense than the others. This click does not occur in !Xung.

and acoustically based phonological features have been proposed to capture this contrast.

There are several acoustic cues that differentiate stops vs. affricates. Burst duration, rise time to peak amplitude and frication duration differences are all part and parcel of the manner contrast. Kagaya (1978) quantified the burst duration differences among Naro clicks, and showed that bilabial [ǀ], dental [ǃ] and lateral [ǁ] click types have long burst durations, while alveolar [!] and palatal [≠] click types exhibit short burst durations. Sands (1990), Johnson (1993) and Ladefoged and Traill (1994) showed that there are similar differences in Xhosa and !Xóǀ clicks. In addition to measuring click burst durations, Ladefoged and Traill (1994) measured Rise Time to Peak amplitude in the click bursts, following Howell and Rosen (1983), who showed that this measure differentiates pulmonic plosives from affricates. Ladefoged and Traill showed that the alveolar and palatal click types in !Xóǀ exhibit short rise times, while the bilabial, dental and lateral click types exhibit longer rise times to peak amplitude. Johnson proposed the feature [+/-noisy], focusing on the acoustic properties of the releases, and Ladefoged and Traill proposed the feature as [+/-abrupt] to describe this phonological contrast in terms of the speed of the anterior release.

In Mangetti Dune !Xung, these click features were studied by Miller and Shah (2009), who show that the palatal click burst duration preceding [u] in Mangetti Dune !Xung exhibits interspeaker variation. One of the four speakers’ productions that they studied exhibited longer burst durations for the palatal click type, suggesting that this speaker released the click less abruptly. Miller (to appear) explicitly compared the realization of clicks preceding [i] and [u], showing that the palatal click type in Mangetti Dune !Xung has two allophones. It is non-affricated (and thus presumably abruptly released) preceding [u] as in the other languages, but has a period of palatalization (palatal frication noise) following the click burst preceding [i]. Miller transcribes the palatalized allophone as [≠̠].

2.3 Click Discrimination

We know of one prior study using acoustic features for click discrimination. Fulop et al. (2004) applied discriminant analysis to the four coronal clicks in the Bantu language Yeyi on the basis

of the four spectral moments of the anterior click bursts, and showed that the classification for the laterals and palatals were much worse than the classification results for the alveolars and dentals. The alveolar clicks only displayed a 2.6% error rate, and the dental clicks an error rate of 24%, while the lateral and palatal clicks displayed an error rate of 93% and 67% respectively. The error rates given here represent measurements from isolated productions. In the present study, we give similar results for isolated productions in !Xung and compare these to results for productions from fluent speech. Like Fulop et al. (2004), we find relatively high degrees of confusion among the different clicks.

2.4 Click Reduction

Previous work on click reduction has distinguished two situations in which clicks are weakened: as an intermediate stage leading to click loss throughout a language, and as a prosodic phenomenon in ordinary speech. We find evidence of both these phenomena in our corpus data. Here, we review some prior work on them.

Traill and Vossen (1997) quantify a stage of “click weakening”, which they claim is an intermediate stage before click loss (the change from a click consonant to a pulmonic consonant). They describe click weakening as a process of acoustic attenuation that effects only the abruptly released clicks [!] and [≠]. They compare the same click types in !Xóó, a language that has not yet been described as undergoing any click loss, and Gǀana, a Khoe language where many of the alveolar clicks have been lost. They point out that the weakened Gǀana clicks are noisier, and have more diffuse spectra, than the strong !Xóó clicks of the same type. They quantify the amplitudes of the clicks in the different languages by providing difference measures of click intensity based on the peak amplitude of the click minus the peak amplitude of the following vowel, which provides a scale of click amplitude relative to the vowel across the different languages. Further, they provide palatograms of some of the strong, and weakened clicks, which show that “weakened articulations have larger cavities and this is a result of reduction in the degree of tongue contact.” They describe this weakening as a process of articulatory undershoot. They attribute the noisiness of the anterior releases in the weakened

clicks to more leisurely anterior releases, that lead to frication. They suggest that the affrication of the abruptly released alveolar and palatal clicks make them less perceptually distinct from the affricated dental and lateral clicks, and that full click loss would then resolve the perceptual ambiguity among the two classes of clicks.

Conversational reduction of clicks, meanwhile, is motivated not by language-wide change but by general articulatory concerns. Miller et al. (2007) provide qualitative evidence that nasal clicks have a stronger and longer duration of nasal voicing in their closures in weaker prosodic positions. Marquard et al. (2015) compared acoustic properties of voiceless oral plosives and clicks in three different phrasal positions (Initial, Medial and Final) in Nǀuu spontaneous speech. Their quantitative results showed that while the duration of pulmonic stop closures got shorter from initial, to medial, to final position, the clicks were shortest in initial position, and lengthened in medial and final positions. The clicks only showed reduction effects for Center of Gravity (lower COG values in phrase medial and phrase final positions, compared with phrase initial position), and in the acoustic energy level (degree of voicing) before the release burst, which is highest in phrase-final position, lower in medial position, and lowest in phrase-initial position. Neither study investigated the effects of reduction on the distinguishability of clicks.

3 Materials

The corpus used in the current study consists of three folktales told by two different speakers, totaling about 45 minutes of speech (Augumes et al., 2011)². One story, *Lion and Hare*, was told by one of the two oldest living speakers in Mangetti Dune, Namibia, Muyoto Kazungu (MK). Two additional stories, *Iguana* (BG1) and *Lion and Frog* (BG2) were told by Benjamin Niwe Gumi, (BG) who was a bit younger, but still a highly respected elder in the community. Our click identification tool does not require a transcript. The acoustic analyses of extracted clicks do require an orthographic transcript, since our tool assigns each detected click to its correct phonetic category by aligning the detections to the transcript. Two of the stories have existing ELAN transcripts in the archive, and the clicks of the third were transcribed by the first au-

²<https://elar.soas.ac.uk/Collection/MP1178567>

thor.

The laboratory data used was a set of words recorded in a frame sentence that were previously analyzed by Miller and Shah (2009) and Miller (to appear).

4 Click Detection

We present a simple rule-based tool implemented using the acoustic analysis program Praat (Boersma and Weenink, 2016) to automatically detect clicks in the audio stream. This method is intended to locate clicks as a general class; we discuss the problem of separating the clicks by type below (Sec. 5). Because clicks are relatively short in duration and high in amplitude, the tool searches the acoustic signal in 1 ms frames.

At each frame, a potential click is detected if the raw signal amplitude exceeds 0.3 Pascal and the Center of Gravity exceeds 1500 Hz. If the region of consecutive frames which passes these filters has a duration less than 20 ms, it is labeled as a click. For MK, the center of gravity cutoff is changed to 1000 Hz and the durations allowed to extend to 25 ms. (These parameters were tuned impressionistically.)

We explored a few other measurements for identifying clicks. A relative measurement of amplitude (checking that the frame has higher amplitude than the one 15 ms back) improves precision but at the expense of recall. Since we hand-corrected the output of our tool, we opted to emphasize recall (it is easier for a human analyst to reject click proposals than to find clicks that the tool has not marked). We also attempted to reject short vowel sounds by checking for detectable formants within the high-amplitude region, but this proved unreliable.

Following click detection with the tool, a human analyst corrected all three transcripts. This process took less than $\frac{1}{2}$ hour for BG, who consistently produced his clicks at higher amplitudes, but 2-3 hours for MK, who varied his click amplitudes more widely. The corrected transcripts are used as a gold standard for evaluating the tool's stand-alone performance.

5 Acoustic Analysis

We compute 4 acoustic features known to differentiate coronal clicks: Burst Duration (BD), Rise Time to Peak Amplitude in the Burst (RT), Center of Gravity (COG) and the Ratio of the Maxi-

mum Amplitude in the Burst to the Amplitude at 20 ms into the vowel. These features were used in a previous study (Miller and Shah, 2009) and shown to separate !Xung clicks preceding [u]. We use the same dataset of 248 click tokens studied by Miller and Shah (2009), extracted from single content words produced in the focused position of a frame sentence, and compare the results to those for 197 clicks extracted from the folktales. The Miller and Shah (2009) dataset includes COG values only for clicks produced before [u], so we restrict our analyses of the folktales to the clicks produced before non-low back vowels [u] and [o] to make the two sets as comparable as possible. The [u] data from Miller and Shah (2009) are all bimoraic monosyllabic words containing the long vowel [u:], though they vary in terms of tone and phonation type. In the texts, both monosyllabic bimoraic and bisyllabic words with two short vowels occur. The vowels following the clicks in the monosyllabic words in the stories are either a long monophthong like [u:] or [o:], or are one of the diphthongs that commences with a non-low back vowel: [ui, oe, oa]. In CVCV words, both vowels are short. All laryngeal release properties (voiced, aspirated, glottalized) of clicks and vowels with non-modal voice qualities were included, as these don't effect the vowel quality (only the voice quality of the vowel). Both nasal and oral clicks are also included. Uvularized clicks were excluded, as were epiglottalized vowels, as these affect the vowel quality, and it is unknown, but possible, that they might affect the C.O.G. of the click bursts.

For the detection of clicks, and for the acoustic analysis of detected clicks, we measured the Rise Time to Peak Amplitude (RT) in the burst as the duration from the onset of the click burst to the maximum RMS amplitude during the click burst proper (transient, not including separate frication noise or aspiration noise that follows the transient), following Ladefoged and Traill (1994). The click burst duration was measured as the duration of the transient itself. The center of gravity was measured using the standard Praat measure on a 22,050 Hz spectrum that was calculated using a Hanning window. The relative burst amplitude was measured as the maximum RMS amplitude found in the click burst (release of the anterior constriction) divided by the RMS amplitude of the following vowel at a point 20 ms from the start of the vowel. The 20 ms point was chosen as it

Transcript	Clicks	Prec	Rec	F
MK 1	250	38.1	17.6	24.1
BG 1	180	66.8	61.6	64.2
BG 2	202	65.3	70.8	67.9
All	632	59.6	47.2	52.7

Table 2: Number of clicks and detection results for three transcripts.

was far enough into the vowel to allow the vowel to reach a higher amplitude, but contained completely within the first mora of the vowel. This assured that the vowel being measured was [u] or [o] in all cases.

5.1 Results

Our evaluation (Table 2) scores a system-annotated click as correct if it occurs within 10 ms of a true click. (Small variations in this number affect the result relatively little, since click bursts are typically short, distinctive events.) On the two transcripts of speaker BG, results are relatively good (precision around 65, recall between 60-70), enabling rapid post-correction by a human analyst. Performance is much poorer for MK (precision 38, recall 17) and post-correction took over four times as long.

Precision errors for BNG generally corresponded to other short, loud speech sounds: coronal ejectives: [ts’], [tʃ’] and the highest amplitude part of [i] vowels. Errors for MK were more varied; MK produced many quieter click bursts which were less distinct from the surrounding speech, and it was harder to set cutoffs that would distinguish the clicks from pulmonic stops and vowel sounds. See Figure 1 for example spectrograms. We believe these very low-amplitude clicks are a consequence of MK’s linguistic background, a possibility we return to in more detail below (Sec. 7).

6 Acoustic Analysis of Clicks

Once the clicks have been extracted, we conduct an acoustic analysis of the four click types. The previous section focused on the task of distinguishing clicks from other sounds as an engineering application. Here, we build a model to discriminate between the four click types, in order to understand how much information they contribute for lexical identification in real speech processing. We conduct a linear discriminant analysis

Dataset	N Clicks	Disc. acc
Lab	248	75
Folktales	197	54
Lab (spkr JF)	75	87
Lab (spkr MA)	75	84
Lab (spkr TK)	74	92
Folktales (spkr BG)	142	73
Folktales (spkr MK)	55	56

Table 3: Linear discriminant analysis accuracies (leave-one-out) on folktale and laboratory clicks.

using the four acoustic features from Miller and Shah (2009), which were shown to differentiate among the four click types in clear productions. Here, we show that they are much less effective for fluent speech, suggesting that clicks, like other speech sounds, are reduced in fluent speech, blurring the primary acoustic cues that distinguish between them.

6.1 Features

Burst duration and Rise time to peak amplitude are both acoustic correlates of manner of articulation, indicating the click’s degree of frication. Longer burst durations and rise times to peak amplitude both indicate more frication, while affricates have an immediate high-amplitude burst right after the release of the initial constriction. The relative burst amplitude reflects the size of the cavity and the abruptness of the release burst. The fourth acoustic attribute that was measured, Center of Gravity (COG), correlates with the the size of the lingual cavity of the click, and therefore is determined by the place of articulation of both constrictions.

6.2 Discriminant Analysis

Using linear discriminant analysis in the R package MASS (Venables and Ripley, 2013), we find that these features indeed differentiate clicks in the single-word lab productions, but are less diagnostic in fluent speech. Accuracies (Table 3) are computed with leave-one-out cross-validation. The lab speech clicks are classified with 75% accuracy, while performance on the folktale clicks is reduced to 54%. This gap is exaggerated by the poor discriminability of clicks produced by MK, whose atypically quiet clicks were also difficult to detect. However, a similar result can be obtained by comparing individual speakers. When a model is learned for each speaker individually, the

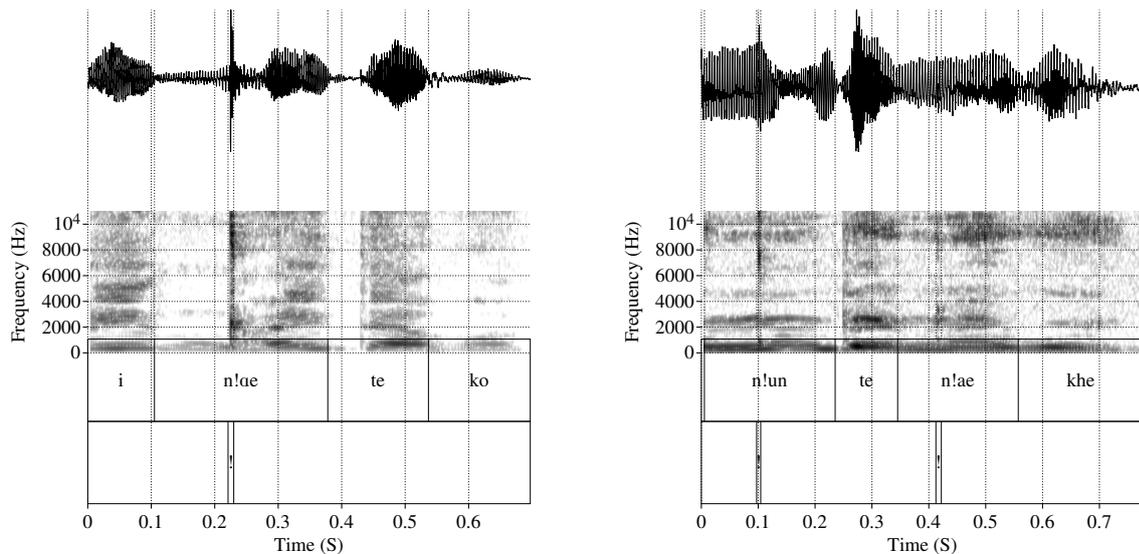


Figure 1: Spectrograms of the alveolar click in the word *n!ǃé* “lion” produced by BG (left) and MK (right) showing the difference in burst amplitudes.

three lab speakers’ clicks are discriminated with 84-92% accuracy, while the folktale speaker BG’s clicks are discriminated with only 73% accuracy. Thus, although intra-speaker variability decreases the accuracy of the classifier in both settings, it is still clear that the folktale clicks are harder to discriminate overall.

A visual explanation of the result is shown in Figure 2, where we plot the RT vs COG (the two most discriminative features for these speakers) for clicks from the folktale storyteller BG versus one laboratory elicitation speaker. The laboratory clicks show a clear separation among all four click types. Among the folktale clicks, only the dental [ʔ] is cleanly separable from the others.

An examination of the learned discriminant functions shows the relative importance of the four acoustic cues. Each discriminant function is a linear combination of the cues; in our data, the first discriminant function captures most of the variance between the clicks for all speakers except MK, whose clicks were poorly classified to begin with. Table 4 shows the coefficients of the first discriminant function for several datasets. For the other speakers, COG is the most discriminative property of clicks, but the second-most discriminative function varies among speakers. Amplitude is a good cue for two of the laboratory speakers, MA and TK, but not for JF or the folktale speaker BG; rise time is also a good cue for MA

and TK but neither of the others. Interestingly, MK’s atypical clicks are classified mainly based on their duration, a cue which was uninformative for the rest of the dataset. A small ablation analysis on BNG and MK’s data tells the same story; COG is responsible for most of the classification performance for BNG (70% with COG alone vs 73% with all features). For MK, it is less useful but still captures over half of classifier performance (35% vs 56%).

7 Discussion

We can infer from the evidence provided that !Xung clicks are subject to phonetic reduction in fluent speech. The primary temporal and spectral cues for click identification become highly variable and less informative in rapid production. Listeners presumably use top-down information like lexical context to make up for increased confusability. Thus, !Xung clicks behave much like other speech sounds in rapid production, despite their canonical loudness, which makes them stand out from the speech stream in clear speech.

Although clicks in fluent speech are harder to discriminate from one another, our results do support the widespread idea that clicks as a class are easy to pick out of the speech stream, at least for speakers who produce them in the canonical way. Despite relying on a few features and hand-tuned threshold parameters, our click detection script

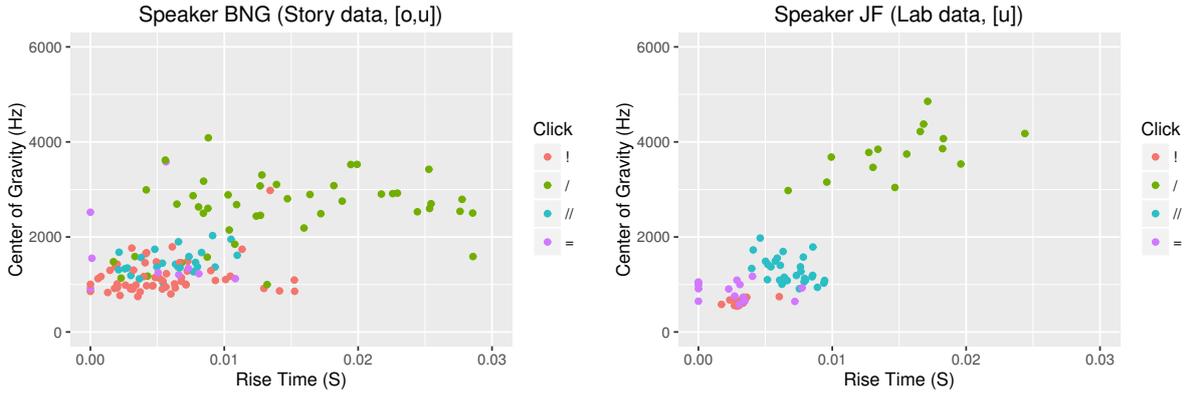


Figure 2: Folktale and laboratory clicks: RT vs. COG

	% var	Rise T	COG	Dur	Max amp	Spkr	Feats	Acc
Lab	92	0.41	1.63	0.03	-0.33	BNG	COG only	70
Lab (JF)	98	0.53	3.72	-0.25	-0.09	BNG	COG/rise	70
Lab (MA)	88	1.12	1.89	-0.17	-1.70	BNG	all	73
Lab (TK)	98	1.53	5.28	-0.03	-1.11	MK	COG only	35
Folk (BNG)	96	0.26	1.45	0.35	0.40	MK	COG/rise	47
Folk (MK)	67	0.03	-0.40	1.38	0.52	MK	all	56

Table 4: Left: Percent of variance captured and coefficients of four features in the first discriminant function learned in different datasets. Right: Ablation results for two speakers.

was able to automate enough of the acoustic analysis to save a substantial amount of transcriber time and effort. We expect that other non-pulmonic consonants like ejectives could also be detected with similar methods. These results are encouraging for corpus research in endangered languages.

The extremely low accuracy for click detections in the speech of MK should qualify this conclusion. There are a few possible reasons why MK’s clicks are much lower in amplitude, harder to detect and harder to discriminate than those of the other speakers. First, MK is the oldest speaker in the dataset (in his 70s). Second, MK spent a large portion of his life in Angola living among speakers of an unknown Bantu language which did not include clicks. This Bantu language was an important mode of communication for much of his life, and indeed, he occasionally code-switched into it during the storytelling session. Perhaps because of this L2 background, MK produced some phonemic click consonants as pulmonic stops (primarily [k,g] for [!,||], and [c,j] for [!, ≠]), and produced extremely variable amplitudes for many of the others.

Similar variability in click production is reported in language as a stage in click loss due to

language endangerment (Traill and Vossen, 1997). It seems, therefore, that MK’s clicks represent an initial stage of language loss and replacement with Bantu, which was reversed for the younger generation. BG’s speech represents this revitalization of !Xung and its replacement of Bantu as the prestige language in the community.

The implication for speech technology and corpus research is that detection methods may vary in their accuracy from community to community. Methods developed for robust language communities may need to be recalibrated when working with severely endangered languages or features undergoing rapid change. Within a single community, however, the accuracy of a tuned detector might serve as a measure of language loss by quantifying the degree to which the target segments have been lost.

Our results reveal new facts about the discriminative features for clicks. For example, although Traill (1997) provided a scale of click burst intensity, shown in (1) above, the variability of the amplitude of the alveolar click bursts relative to the following vowel is so high, that it is clear that the amplitude alone can not be very useful in discriminating the four click types. As mentioned

above, the relative perceptual weighting of the two temporal measures (click burst duration and rise time to peak amplitude) is completely unknown for clicks. Comparing our results to Fulop et al. (2004) Yeyi results, we can conclude that a combination of manner cues and place of articulation cues results in much better discriminability. Of course, we can not rule out the contribution of click reduction / loss to the poorer discriminability seen in the Yeyi results.

Machine learning can indicate how much information about click identity is carried by each of these cues, but this does not necessarily reveal which cues are important to human listeners. For instance, English fricatives and affricates are also differentiated by duration and rise time to peak amplitude (Howell and Rosen, 1983). Early studies assumed that Rise Time was the main acoustic feature of importance. However, Castleman (1997) showed that frication duration differences among the English contrast are more perceptually relevant than the rise time differences. While our results imply that COG is the most informative criterion for click identity, further perceptual experiments could tell whether this matches listeners' actual perceptual weightings. Of course, four manually selected features and a linear classifier do not tell the whole story of click discriminability. A more sophisticated model (King and Taylor, 2000) could discover features directly from the acoustic signal; however, we believe our features acceptably represent the major categories of cues.

8 Conclusion

Results suggest that phonetic studies of endangered languages must consider both clean productions and naturalistic speech corpora. It is important to discover not only the phonemic inventory of the language and the canonical landmarks that allow listeners to recognize speech sounds in clear speech, but also the range of phonetic variability displayed in fluent speech. In this study, investigation of connected speech led to the conclusion that the scale of click burst intensity is not very useful in distinguishing clicks, since the amplitude of alveolar click bursts is so variable. In studying natural data, rule-based extraction of particular segments may offer a low-cost alternative to developing a full ASR system for a language with little available data. The processed data could be used to supplement non-expert annotations (Liu

et al., 2016; Bird et al., 2014) in training a full-scale ASR system, or to bootstrap a learning-based landmark recognition system (Hasegawa-Johnson et al., 2005).

Acknowledgments

We thank Mehdi Reza Gholi Lalani, Muyoto Kazungu and Benjamin Niwe Gumi. This work was funded by ELDP SG0123 to the first author and NSF 1422987 to the second author.

References

- International Phonetic Association. 2006. The international phonetic alphabet (revised to 2005) [chart]. page 135.
- Christine Augumes, Amanda Miller, Levi Namaseb, and David Prata. 2011. Mangetti Dune !Xung stories: In !Xung and English. The Ohio State University, the University of Namibia, and the Mangetti Dune Traditional Authority. Deposited at ELAR.
- Douglas Martyn Beach. 1938. *The phonetics of the Hottentot language*. W. Heffer & Sons, Ltd.
- Jill N Beckman. 2013. *Positional faithfulness: an Optimality Theoretic treatment of phonological asymmetries*. Routledge.
- Steven Bird, Lauren Gawne, Katie Gelbart, and Isaac McAlister. 2014. Collecting bilingual audio in remote indigenous communities. In *COLING*, pages 1015–1024.
- Paul Boersma and David Weenink. 2016. Praat: doing phonetics by computer. Version 6.0.20 from <http://praat.org>.
- Wendy Ann Castleman. 1997. *Integrated perceptual properties of the [+/-continuant] distinction in fricatives and affricates*.
- Sean A Fulop, Peter Ladefoged, Fang Liu, and Rainer Vossen. 2004. Yeyi clicks: Acoustic description and analysis. *Phonetica*, 60(4):231–260.
- Steve Greenberg. 1993. Speech processing: Auditory models. In R.E. Asher and S.M.Y. Simpson, editors, *Pergamon Encyclopedia of Language and Linguistics*, Vol. 8, pages 4206–4227.
- Mark Hasegawa-Johnson, James Baker, Sarah Borys, Ken Chen, Emily Coogan, Steven Greenberg, Amit Juneja, Katrin Kirchhoff, Karen Livescu, Srividya Mohan, et al. 2005. Landmark-based speech recognition: Report of the 2004 Johns Hopkins summer workshop. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings (ICASSP'05). IEEE International Conference on*, volume 1, pages 1–213. IEEE.

- Peter Howell and Stuart Rosen. 1983. Production and perception of rise time in the voiceless affricate/fricative distinction. *The Journal of the Acoustical Society of America*, 73(3):976–984.
- Keith Johnson. 1993. Acoustic and auditory analyses of Xhosa clicks and pulmonics. *UCLA Working Papers in Phonetics*, 83:33–45.
- Ryohei Kagaya. 1978. Soundspectrographic analysis of Naron clicks: A preliminary report. *Annual Bulletin of Institute of Logopedics and Phoniatrics*, 12:113–125.
- Simon King and Paul Taylor. 2000. Detection of phonological features in continuous speech using neural networks. *Computer Speech and Language*, 14(4):333–353.
- Peter Ladefoged and Anthony Traill. 1994. Clicks and their accompaniments. *Journal of Phonetics*, 22(1):33–64.
- Ning Li and Philipos C Loizou. 2008. The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise. *The Journal of the Acoustical Society of America*, 124(6):3947–3958.
- Chunxi Liu, Preethi Jyothi, Hao Tang, Vimal Manohar, Rose Sloan, Tyler Kekona, Mark Hasegawa-Johnson, and Sanjeev Khudanpur. 2016. Adapting ASR for under-resourced languages using mismatched transcriptions. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 5840–5844. IEEE.
- Carina Marquard, Oliver Niebuhr, and Alena Witzlack-Makarevich. 2015. Phonetic reduction of clicks—evidence from N|uu. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*.
- Amanda Miller and Sheena Shah. 2009. The acoustics of Mangetti Dune !Xung clicks. In *Proceedings of INTERSPEECH*, pages 2283–2286.
- Amanda L Miller, Johanna Brugman, Bonny Sands, Levi Namaseb, Mats Exter, and Chris Collins. 2007. The sounds of N|uu: Place and airstream contrasts. *Working Papers of the Cornell Phonetics Laboratory*, 16:101–160.
- Amanda Miller-Ockhuizen. 2003. *The Phonetics and Phonology of Gutturals: A Case Study from Ju|’hoansi*. Outstanding Dissertations in Linguistics Series. Routledge.
- Amanda Miller. to appear. Palatal click allophony in Mangetti Dune !Xung: Implications for sound change. *Journal of the International Phonetic Association*.
- Bonny Sands. 1990. Some of the acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics*, (74):96.
- Kenneth N Stevens and Samuel Jay Keyser. 1989. Primary features and their enhancement in consonants. *Language*, pages 81–106.
- Anthony Traill and R Vossen. 1997. Sound change in the Khoisan languages: new data on click loss and click replacement. *Journal of African languages and linguistics*, 18(1):21–56.
- Anthony Traill. 1997. Linguistic phonetic features for clicks: Articulatory, acoustic and perceptual evidence. *African linguistics at the crossroads: Papers from Kwaluseni*, pages 99–117.
- Nikolai S. Trubetzkoy. 1969. *Principles of Phonology*. University of California Press.
- William N Venables and Brian D Ripley. 2013. *Modern applied statistics with S-PLUS*. Springer Science & Business Media.