ACOUSTIC PROPERTIES OF PARA-PHONEMIC SOUNDS: CLICKS IN AMERICAN ENGLISH

Betsy Pillion

University of Chicago
bpillion@uchicago.edu

ABSTRACT

Clicks in English are known to occur in transitional periods of talk, during word searches and as indicators of speaker affect. Less well established are the phonetic properties of these clicks, phonetic variation with respect to the click's discourse role, and phonetic variation between speakers. This study contributes to understandings of click acoustics by examining clicks used within the Buckeye Corpus of American English [9]. Percussive clicks varied significantly in intensity from those with a discourse role such as conveying affect or turn-management. Speakers also varied in the extent to which clicks of all discourse types were employed: male speakers clicked percussively at a significantly higher rate than female speakers, whereas female speakers were more likely to use turn-management and affect-conveying clicks. This research helps illuminate an understudied aspect of sound systems, and gives insight into the extent of intra- and inter-speaker variation in para-phonemic sounds.

Keywords: Clicks, paralinguistics, sociophonetics, para-phonemic sounds, corpus linguistics

1. INTRODUCTION

Clicks, a type of articulation created through the rarefication of air between two different points of articulation in the mouth, are known to occur in English in several contexts. English speakers use clicks to do conversational work, like indicating incipient speakership [8], indexing a new sequence [14], or searching for a word [13]. Additionally speakers use clicks to convey affect, such as disapproval [6]. Finally, speakers have been known to produce clicks incidentally, as the result of articulatory processes of speech [11] or breath intake before speech [8].

However, many questions remain about these clicks' acoustic or articulatory properties, how these properties might differ across these discourse purposes, and how these different clicks are used by different speaker populations. This study describes the properties of clicks within a corpus of American English, the Buckeye Corpus of American English. Speakers in the corpus are of a similar socioeconomic class, and are all speakers from the same restricted geographical area. Holding these features constant, this study provides insight into the extent of variation within this population, makes some preliminary claims about how gender may interact with para-phonemic click usage, and indicates a high degree of individual variation with respect to click usage. This study expands our understanding of the acoustic properties of para-phonemic clicks, and gives insight into their distribution of usage among speakers of American English.

2. BACKGROUND

Para-phonemic clicks are clicks that are produced in speech but are not used as phones within words of the language. Much of the literature looking at these clicks has focused on their use in English conversation. However, this phenomenon is by no means exclusive to English. Languages such as Wolof [5] have a complex system of para-phonemic clicks that convey negative affect, can answer 'yes' or 'no' to polar questions, and perform other conversational functions. Gil's [4] survey of linguists finds para-phonemic clicks that are used in 143 languages belonging to many unrelated language families.

Clicks in English can be used to convey affect, manage turns and discourse, or can be found as percussive sounds resulting from effects of other goal articulations. Affect conveyance may be the most well-known usage of click in English, typically described as negatively valanced represented orthographically by the sequence *tsk-tsk* or *tut-tut* [8]. However, clicks are more frequently found in discourse roles where they manage sequences such as indicating a speaker wishes to begin a turn, is searching for a word, or wishes to change a subject.

Articulatorily, clicks in English have been described as alveolar and dental [15], post-alveolar and alveolar lateral [8]. Aperliński's [1] experimental study of affective clicks in English found seven different articulations (bilabial, rounded bilabial, dental, alveolar, palato-alveolar, lateral, and palatal).

Clicks in this study were elicited as emotional responses to particular situations that could be experienced by the participant. As such, they represent a range of possible clicks that convey affect, but do not represent clicks used in turn-interactions or percussives. Moreno and Stuart-Smith [7] find a different range of articulations, alveolar-lateral, palato-alveolar, alveolar, dental-alveolar, dental, labio-alveolar, labiodental, and bilabial, with dental being the most common articulation.

While these studies give us information about the distribution of clicks within natural speech in English, and the acoustic properties of elicited clicks, we lack detailed information about phonetic properties of the clicks themselves as they occur in natural speech. Aperliński's [1] study associates particular stances or functional roles of clicks with articulations, and finds a range of variation. Wright [14] and Ogden [8] find dental, alveolar and bilabial clicks, but do not associate them with any particular conversational roles. In order to investigate the associations between these discourse roles and particular articulations or acoustic properties, we will address the full spectrum of click roles: stance display clicks, turn management clicks, and percussive clicks.

3. PREDICTIONS

We predict that clicks in this corpus resemble the articulatory types found in previous studies [1], [14], [8], and that speakers may use different click articulations in discourse roles like those displaying a stance, and those indicating a desire to begin speaking. Percussive clicks, those that arise as an articulatory side effect, are known to occur at the alveolar ridge and at the lips, but it is not known how or if percussive clicks differ from intentional clicks made at these places. The corpus's gender balanced design, combined with a geographically-restricted sample of speakers, allows us to ask some initial questions about how gender might factor into click usage, Ogden [8] finds that women click at a higher rate in the CallHome corpus, but does not include percussive clicks in the total click count.

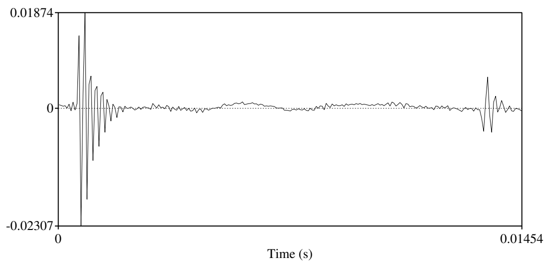
4. METHODOLOGY

Data from this study come from the Buckeye Corpus of Conversational Speech [9]. The corpus totals 40 speakers, all white, balanced for binary variables of gender (male, female) and age (young, old). Speakers were recorded during a modified sociolinguistic interview, totaling approximately one hour of recording per speaker. Two interviews were excluded due to audio problems, both males. The remaining speakers totaled 38 in number, approximately 38 hours of recordings. Speech from the interviewers was only partially audible and as a result was not analyzed.

4.1. Click Coding

The Buckeye Corpus's native transcription system is annotated at the lexical and phonemic levels. Clicks are not differentiated in this system, but some are labeled as vocal noises, along with in breaths and other non-word noises made by speakers. The corpus was annotated for clicks by the researcher in textgrids using Praat [2]. Clicks were demarcated into periods of noise and silence, where noise periods begin with a transient release and end with a transition to silence or another click transient.

Figure 1: Labial-dental click composed of two periods of noise. Speaker 39 - s3902b



Clicks were labeled for place of articulation, turn information and discourse type. Place of articulation was labeled by the researcher to be one of 11 articulations, 6 making up the bulk of the corpus. An example of a labial-dental click can be found below. Turn information was determined by the click's place within the speakers' utterances. Discourse type distinguished the different roles that the clicks played in the discourse. The click properties that were labeled can be found below in Table 1.

Table 1: Clicks were labeled according to their properties within the discourse, position within the turn structure, and perceived articulatory properties.

Discourse Type	Turn Position	Articulation
turn management affective percussive	initial medial final back channel aborted	dental alveolar labio-dental labial lateral palatal

Click articulations were determined auditorily by a team of phonetically trained labelers. Discourse roles were determined by surrounding conversational information. Stance display or affective clicks were used in response to negatively or positively valanced remarks. Clicks that occur during word searches, or when a speaker maintains their turn during a pause are labeled as turn management clicks. Additional turn management clicks include clicks preceding utterances that are not followed by in breaths. Percussive clicks occur as an acoustic side-effect of other articulatory goals by the speaker. This typically means clicks that occur as the articulators separate for an intake of breath before speaking, or when the tongue separates from the palate after swallowing. As these labelling methods are highly labeler dependent, additional work will need to be done to determine rates of intertranscriber reliability.

4.2. Acoustic Information

Acoustic information was extracted using a modified Praat script [3] from each of the periods of noise within the click. As click releases are made up of multiple bursts due to their complex articulations, many clicks were made up of multiple bursts and periods of silence. Periods of silence were analyzed only for length.

The signal was sampled at 16 kHz, and Discrete Fourier Transforms were averaged using time averaging [12]. These measures were taken across the center 80% of the duration of the click, with a window size of 10 milliseconds in two windows. The first four spectral moments were measured over these two windows within each noise period, with multiple noise periods within a click. Intensity and duration information were also collected.

5. ANALYSIS

5.1. Acoustic Properties

A one-way analysis of variance test was performed in R [10] using aov function within the stats package to assess the relationship of discourse type to maximum intensity of each noise period within a click. The ANOVA showed that the effect of discourse type was significant, $F(2, 7827) = 425$, $p < 0.0001$. The summary of this can be found below in Table 2. A post hoc Tukey HSD test indicated that the percussive and turn management clicks, and the percussive and affect conveying clicks varied significantly from one another ($p < 0.0001$), while turn management and affect conveying clicks also differed significantly from one another at a different alpha level ($p < 0.005$).

Table 2: ANOVA Statistical Summaries

Maximum Intensity by Discourse Type					
Source	Df	Sum Sq	Mean Sq	F val	P val
dis	2	53223	26611	425	<2e-16
Resid	7827	490060	63		
Maximum COG by Discourse Type					
Source	Df	Sum Sq	Mean Sq	F val	P val
dis	2	4.4e+8	2.2e+8	119.1	<2e-16
Resid	7827	1.4e+10	1.9e+6		
Maximum Standard Deviation by Discourse Type					
Source	Df	Sum Sq	Mean Sq	F val	P val
dis	2	5.3e+06	2.6e+6	12.64	3.32e-6
Resid	7827	1.6e+09	1.8e+6		

Maximum Intensity by Articulation					
Source	Df	Sum Sq	Mean Sq	F val	P val
art	11	2.1e+4	1906.3	28.53	<2e-16
Resid	7818	5.2e+5	66.8		
Maximum COG by Articulation					
Source	Df	Sum Sq	Mean Sq	F val	P val
art	11	1.7e+08	1.5e+7	8.237	1.57e-14
Resid	7818	1.5e+10	1.9e+6		

Similarly, ANOVAs were performed to assess the relationship of the first two spectral moments to discourse type. The ANOVAs showed that the effect of discourse type was significant $F(2, 7827) = 119.1$, $p < 0.0001$ with respect to center of gravity, and standard deviation $F(2, 7827) = 12.64$, $p < 0.0001$. Post hoc Tukey HSD tests indicated that percussive and turn-management discourse types' center of gravity means differed significantly at the $p < 0.0001$ level, while affect conveying and turn management differed at the $p = 0.005$ level, and percussive and turn management clicks differed at the $p = 0.01$ level. Percussive and turn management clicks' standard deviation differed at the $p < 0.0001$, while differences between turn management and affect conveying clicks, and percussive and affect conveying clicks were not significant. Skew and kurtosis were not evaluated due to the low sampling rate of the signal.

An ANOVA was performed to assess the relationship of articulatory label to maximum intensity. The test revealed a significant effect of articulatory label $F(11, 7818) = 28.53$, $p < 0.0001$. A post hoc Tukey HSD test indicated that the differences between lateral and labio-dental, palatal and lateral, lateral and labial, lateral and dental, labiodental and dental, labial and dental, labiodental and alveolar, and labial and alveolar were significant at the $p < 0.0001$ level. Differences between palatal and alveolar clicks were significant at the $p < 0.005$ level, and differences between palatal and dental clicks were significant at the $p = 0.05$ level. Other click articulation comparisons did not meet significance level.

Similarly, ANOVAs were performed to assess the relationship of articulation to the first two spectral moments. There was a significant effect of articulatory label $F(11, 7818) = 8.23$ $p < 0.0001$ with respect to COG and standard deviation $F(11, 7818) = 15.99$ $p < 0.0001$. Post hoc Tukey HSD tests found significant differences within COG between dental and alveolar, labio-dental and labial articulations at the $p < 0.0001$ level, palatal and alveolar $p = 0.01$, and near significant differences between labio-dental and alveolar, and palatal and lateral articulations ($p = 0.053$). Similarly, significant differences were found within standard deviation between labial and alveolar, palatal and labio-dental, lateral and labial, palatal and labial, lateral and labio-dental, labio-dental and alveolar, labio-dental and dental, and labial and dental at the $p < 0.05$ level. No other articulatory pairs were found to have significant differences.

5.2. Speaker Properties

A Pearson's Chi-Square test was performed to assess whether gender and discourse types were related. There was significant evidence of an association, ($\chi^2(2) = 187.06$, $p < 0.001$). Men produced 61% of percussive clicks (1441), while women produced 39% (912). Women produced 62% of turn management clicks (256) compared to 38% (158) by men, while women produced 69% (62) of affect conveying clicks compared to 31% (28) produced by men.

6. RESULTS

6.1. Statistical Analysis

Discourse types were found to be good predictors of maximum click intensity, maximum spectral center of gravity and maximum standard deviation. Similarly, a subset of articulatory labels were found to be good predictors of these attributes. Male speakers were more likely to produce percussive clicks, while female speakers were more likely to produce affect-conveying and turn-management clicks.

6.2. Click types and Frequencies

Percussive clicks were the most commonly found type of click, making up around 82% and 2353 tokens of the total 2857 clicks found within the corpus. Turn-management clicks consisted of approximately 14.5% of the total and 414 tokens, while affect conveying clicks were found the least frequently consisting of a mere 3% of the total clicks, making up 90 tokens. Speakers were consistent in employing percussive clicks most frequently during their interviews, with turn-management clicks being the second most frequently used, and finally affect-conveying clicks.

6.3. Individual Speakers

Individual variation with respect to click production was high. Speakers s1, s2, s9, and s36 produced 25 or less clicks during each interview, with s3 producing only 3 total clicks, all percussive. However, speakers s13, s19, s34, s12, s11, s22 and s35 all produced over 125 clicks, the majority of these also percussive, with speaker s35 producing the maximum of 167 clicks, 144 percussive. The average speaker clicked 75 times, the average male speaker clicked 90 times, the average female speaker clicked 61.5 times.

7. DISCUSSION

Para-phonemic clicks in English within the Buckeye Corpus have been shown to occupy a wide range of articulations, expanding on previous claims. Discourse type was found to be a good predictor of acoustic properties, while articulatory label was only a good predictor in a subset of the articulatory labels, indicating that discourse type may be a better indicator of the click's acoustic properties and in some way correlate with particular articulatory properties.

Speakers were consistent with few exceptions in employing percussive clicks more frequently than turn-management and affect-conveying clicks, contra Ogden [8] who found percussive clicks less frequently than turn-management and affect-conveying clicks. Similar to Moreno & Stuart-Smith [7], turn-managing clicks are found more frequently than affect-conveying clicks. Male and female speakers differed significantly in the ways that they employed clicks. Male speakers were found to produce percussives much more frequently than female speakers, making clicks available as a potential site of identity construction within this variety of English.

Relationships between contrast and meaning are often invoked in understandings of sound systems. We expect that for speakers to be able invoke meaning differences between sounds or groups of sounds, that there should be some kind of contrast in their acoustic properties. As such, we might expect that clicks that are used for different discourse roles have tendencies towards particular acoustic properties that contrast from one another. This paper adds to the growing body of work that addresses the properties of sounds outside of the main phonemic inventory of a language.

8. REFERENCES

[1] Aperliński, G. 2012. The paralinguistic use of clicks by speakers of English. Master's thesis

Adam Mickiewicz University.
[2] Boersma, P., Weenink, D. Praat: doing phonetics by computer.
[3] DiCanio, C. Time averaging for fricatives.
[4] Gil, D. 2013. *The World Atlas of Language Structures Online.* chapter Para-linguistic Usages of Clicks. Max Planck Institute for Evolutionary Anthropology.
[5] Grenoble, L., Martinović, M., Baglini, R. 2014. Verbal gestures in Wolof. *Selected Proceedings of the 44th Annual Conference on African Linguistics*. Cascadilla Press Somerville, MA.
[6] Ladefoged, P. 1982. *A Course in Phonetics*. Wadsworth Cengage Learning.
[7] Moreno, J., Stuart-Smith, J. forthcoming. Tut-tut: A study of click production across Scottish dialects.
[8] Ogden, R. 2013. Clicks and percussives in English conversation. *Journal of the International Phonetic Association* 43(3), 299–320.
[9] Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. V., Fosler-Lussier, E. 2007. Buckeye corpus of conversational speech (2nd release). Department of Psychology, Ohio State University (Distributor).
[10] R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.
[11] Scobbie, J. M., Schaeffler, S., Mennen, I. 2011. Audible aspects of speech preparation. *International Congress of Phonetic Sciences XVII* 1782–1785.
[12] Shadle, C. 2012. *The Oxford Handbook of Laboratory Phonology* chapter The Acoustics and Aerodynamics of Fricatives. Ch. 20, 511–526. Oxford University Press.
[13] Ward, N. 2006. Non-lexical conversation sounds in American English. *Pragmatics and Cognition* 14(1), 129–182.
[14] Wright, M. 2005. *Studies of the phonetics-interaction interface: clicks and interactional structures in English conversation*. PhD thesis University of York.
[15] Wright, M. 2011. The phonetics-interaction interface in the initiation of closings in everyday English telephone calls. *Journal of Pragmatics* 43, 1080–1099.