Volume 27
Issue 1 *Proceedings of the 44th Annual Penn Linguistics Conference*
Article 10

7-9-2021

NO NEED TO YELL: A Prosodic Analysis of Writing in All Caps

Maria Heath
University of Minnesota

NO NEED TO YELL: A Prosodic Analysis of Writing in All Caps

Abstract

Writing in all caps, while not unique to the internet, has become a common feature used in social media. English-speaking internet natives seem to have shared intuitions about what meaning it contributes to a text, even though it is not a feature taught in standard English orthographic education. This study employs production studies to determine how readers produce tweets written in all caps out loud, in order to provide evidence that there is a prosodic component to the interpretation of all caps. Though common discourse about all caps holds that it indicates yelling and anger, the data from this study shows that it's not just average loudness but also average pitch and syllable duration that tend to be increased in the production of all caps text versus text with standard capitalization. This combinaiton of prosodic features which can be associated with all caps also supports the conclusion that it can be used in contexts exemplifying a wide range of emotions, not just anger. This study lays groundwork for a better understanding of the interaction between orthography and prosody on social media and how production studies can be used to test such interactions.

NO NEED TO YELL: A Prosodic Analysis of Writing in All Caps

Maria Heath*

1 Introduction

In 2016, the National Oceanic and Atmospheric Association shared the following announcement: “LISTEN UP! BEGINNING ON MAY 11, NOAA’S NATIONAL WEATHER SERVICE FORECASTS WILL STOP YELLING AT YOU” (Buchanan 2016). They go on to explain that while their weather forecasts had historically been printed in all caps due to limits on the technology of the teleprinters which were used up until the 90s, technological advances had finally become widespread enough in the industry for them to be able to switch over to mixed caps for forecasts. One reason they cite for the change is that “in web speak, use of capital letters became synonymous with angry shouting”. In this paper, I will challenge this assumption that the use of all caps can be summed up as “angry shouting”, but also use this intuition as a starting point for exploring the use of all caps on social media.

The popularity of phrases like “yelling/shouting/screaming in all caps” demonstrates how commonly held an interpretation it is that all caps indicates high speech volume. However, I propose that the situation is more complex than that, and that understanding how all caps is used and interpreted on social media will provide a starting point for understanding the linguistic nuances of text-based social media language in general. Specifically, I claim that all caps is one of several patterns of nonstandard orthography[1] which is intended to represent prosodic information textually in CMC (Computer Mediated Communication). While this prosodic information may include loudness, as the general intuition for all caps suggests, I show that features like pitch and tempo may also be affected, such that the “meaning” of all caps may better be understood as the degree of arousal or emphasis of a range of emotions which include, but are not limited to, anger.[2] Therefore, even though all caps as an orthographic pattern doesn’t always mean “angry shouting”, the fact that it is associated with specific prosodic information does imply that its use may not be considered pragmatically congruous with all contexts, such as weather forecasts. Better understanding how nonstandard orthography such as all caps is interpreted in CMC will allow for more detailed linguistic analysis of social media language.

2 Background

This research explores the use and interpretation of text written in all caps on social media, as well as positioning the all caps orthographic pattern in a broader framework which can be used to explore a wide variety of nonstandard orthography on social media. Nonstandard orthography as a feature of CMC refers to intentional deviations from the formal textual norms of a language (English, in this case). In other words, nonstandard orthography, including informal uses of capitalization, punctuation, and spelling, is not the same as typos, which are errors of production. Nonstandard orthography on social media is systematic, productive, and incredibly creative, (Heath 2019, McCulloch 2019) which makes it an ideal channel for studying sociolinguistic norms in online spaces. In addition, as a naturally arising and evolving language form, we should expect it to follow familiar linguistic rules. In this project, I draw primarily on pragmatic and prosodic theory to frame an analysis

*Many thanks to Brian Reese and Claire Halpert for their help with all stages of this project, Ben Munson and Ivy Hauser for feedback and advice, the UMN Institute of Linguistics for funding my research, and Yash Travadi for checking the statistical methods.

[1]In this paper, the term “orthography” will be used in the sense which includes the typographic norms of a language such as punctuation and capitalization as well as spelling. The term “standard” is intended to refer to the formal orthographic norms associated with newsprint, books, and other long-form, non-interactive texts. “Nonstandard”, then, will refer to deviations from the formal norms, especially those which reflect systematic, conventionalized usage patterns which tend to arise organically on social media.

[2]Following Idemaru et al. (2019), the auditory terms “pitch” and “loudness” will be used when discussing perception or mental interpretation of acoustic properties, while the acoustic terms “F0” and “intensity” will be used when discussing measurements of phonetic features.

of the all caps orthographic pattern.

At the core of this analysis is Relevance Theory, a pragmatic framework which claims that people are by nature always looking for relevance in all forms of communication (Sperber and Wilson 2004). Under Relevance Theory, we assume that human communication is designed to maximize relevance by making the most efficient use of all available communicative resources. Specifically, Sperber and Wilson (2004) state that, "In relevance-theoretic terms, any external stimulus or internal representation which provides an input to cognitive processes may be relevant to an individual at some time." I propose that in text-based English CMC, the main communicative resource is orthography. Orthographic patterns such as all caps serve as the external stimuli, and familiar patterns may also carry internal representations drawn from their conventionalized use in online contexts, which must be learned through exposure.

The universal tendency for people to always look for relevance allows us "to predict and manipulate the mental states of others" to some extent (Sperber and Wilson 2004). In written texts, we have a conventionalized standard for how English should be written in formal contexts, which is the default from which any alterations deviate. Informal online writing uses intentional deviations from this standard to produce calculated effects in the interpretation, making non-standard orthographic patterns a type of "ostensive stimulus", or stimulus designed to attract the attention of the audience and therefore to convey its own optimal relevance to the communicative context. In this case, typing in all caps is a notable deviation from the standard, drawing attention to a text and potentially requiring more processing effort for those who are more used to reading standard mixed capitalization. In other words, because it is set apart visually from "normal" text, text written in all caps is visually salient and therefore reasonably assumed to carry some additional meaning relevant to the context. However, research analyzing what that meaning might be is limited, both in quantity and in robustness of theory.

There are three primary ways in which CMC scholars have talked about the interpretation of nonstandard orthography on social media. The first is corpus collection, in which large corpora of data are collected, typically from a single CMC platform or context, and orthographic data is analyzed quantitatively by frequency or context of usage, or specific examples are used as case studies exemplifying a usage (Haas et al. 2011, Lyddy et al. 2013, Zappavigna 2012, McSweeney 2018). Also common is the use of personal intuition as an experienced social media user, or "internet native", to interpret the function of orthographic patterns or specific examples (McCulloch 2019). This method relies completely on qualitative analysis, in contrast to the quantitative options provided by the corpus method, but is also able to provide a more thorough description and analysis of case studies. A third method which has been used to try to combine native user intuition with quantitative analysis is the use of judgment studies to elicit readers' interpretations of carefully constructed or controlled texts (Gunraj et al. 2016, Houghton, Upadhyay, and Klin 2018). However, while corpus studies and judgment studies provide quantitative data on how nonstandard orthographic patterns are used, they don't offer evidence for why the patterns are used the way they are or how social media users come to agree on these conventionalized meanings. Likewise, while researcher intuition can suggest possible answers to these questions, it is not backed up by quantitative data testing the generalizability of the intuition. In order to better understand how nonstandard orthography gets used and interpreted on social media, a wider variety of research methods designed to explicitly test theories of orthographic meaning are needed.

There are many different patterns of nonstandard capitalization that are used systematically in unique ways on social media, such as the capitalization of the entirety of a single word, using initial capitalization for sentence-internal words which are not proper nouns, alternation between upper and lower case within a single word, and so forth. This paper will focus on the specific pattern of all caps, in which the entirety of a sentence or post is capitalized. The all caps pattern is widely used across CMC platforms, and occasionally in other types of texts as well, making it an ideal candidate for study. In addition, many people share intuitions about what it means to write in all caps which, while not a conclusive factor in and of itself, offers a concrete starting point for study. Specifically, many people have described all caps as indicating yelling or shouting, from CMC academics to newscasters discussing celebrity tweets to young people explaining CMC etiquette to their parents. Oftentimes the affect of all caps text is described as angry, which aligns with the interpretation of it as indicating yelling. Contextually, however, it can be found in texts with a wide range of affective meanings, both positive and negative, which presents a puzzle.

A possible solution to this puzzle is that the all caps pattern doesn't correlate directly to a particular meaning, but rather to a way of producing the text subvocally. In other words, the inner voice with which many people read can have the same prosodic features as actual speech, with the same corresponding interpretations, and this prosody is what is indicated by the orthography. Writers have long struggled with how best to convey prosodic information in writing, and in an age where so much interpersonal communication happens via text, the demand for nuanced and emotive modes of written expression have continued to increase. Soffer 2010 lays this out in a discussion of "silent orality", describing the evolving relationship between text and speech. While early written English was primarily a record of speech and intended to be read aloud, orality was sidelined with the rise of silent reading of mass-produced texts, revived in the era of writing for audio media, and synthesized with silent reading in the modern era of interpersonal digital communication. In other words, the communicative needs of CMC text have pushed us back toward a mode of writing that better reflects the linguistic and paralinguistic cues of face-to-face communication, such as prosody.

The technology of CMC also limits the options for variation in writing. While text formatting options such as bold, italic, typeface, font size, font color, etc. are available in text editors, they are not often available (or easily accessible) on social media platforms. As such, the average social media user is limited to the orthographic resources available on a standard English keyboard. Capitalization, as an option available on any keyboard and visible on any platform, is an ideal resource for orthographic variation, especially since it has so few meaningful uses in formal text. It would be no surprise to find that creative uses of capitalization have been adapted to fill some of those prosodic gaps in writing.

The prosodic gaps typically left by standard text are largely emotive in nature, since punctuation already deals well with contour and phrasing. While the semantic content of a sentence can indicate the emotional valence of a text, the quality and degree of emotion is more difficult to convey using standard typographic features, not to mention the exaggerated prosodic cues of sarcasm, for example. In this study, I narrow my analysis to features of emotive arousal, since the physiological effects of high or low arousal states on prosody – specifically F0, intensity, and tempo – are well understood. Scherer & Bänziger (2004) note the following robust correlation between F0 (pitch) and arousal: "The F0 points for emotions with low arousal (such as sadness, happiness, and anxiety) are generally lower than the F0 points for emotions with high arousal (despair, elation, panic fear, and hot anger)." Intensity (loudness), on the other hand, is a bit more complicated, in that not all high-arousal emotions are correlated with higher intensity (e.g. surprise; Crumpton & Bethel 2015). However, no low-arousal emotions are correlated with high intensity (e.g. sadness; Crumpton & Bethel 2015). Likewise, while not all high-arousal emotions are expressed with fast tempos, low arousal emotions are consistently associated with slow tempos (Crumpton & Bethel 2015). With these points in mind, we can generally say that the higher the arousal level of the expressed emotion, the higher the F0 will be, and the more likely intensity is to be higher as well.[3]

Part of the reason why prosody is an attractive explanation for the role of non-standard orthography in meaning-making is that there are a wide range of prosodic features which can be described in the spatial-to-aural analogues of diagrammatic iconicity (Hiraga 2005, Nänny 2005) and which can be both segmental and suprasegmental. For example, pitch has a clear iconic correlate in the physical world if we consider the terms we use to describe it in English: high pitch and low pitch. Therefore, we might expect text that is "higher" in relation to the text around it (such as the consistently maximal height of capital letters) to possibly correlate to higher pitch, a diagrammatic relationship in which height is understood in two different modalities in order to refer to a particular type of relationship.

[3]It is worth noting that increases in F0, intensity, and syllable duration have previously been shown to be correlated as a result of the Lombard Effect, in which people tend to unconsciously alter not just one but a set of prosodic features in situations leading to increased vocal effort, such as when background noise is present. Even in artificial environments where a participant is speaking to themselves, the Lombard effect has been shown to play a role (Brumm & Zollinger 2011). However, these studies suggest that the correlated prosodic changes are intended to increase communicative effectiveness when there may be interference with perception of the speech. As such, it is unclear how much of a role the Lombard Effect might play on the results of this study, since the stimuli are never intended to be heard. This could be an interesting avenue to explore in future work, especially if the effect is robust enough that people might expect vocal effort to result in correlated prosodic alterations even in the absence of speech.

Additionally, we can apply size terms to both spatial dimensions and loudness. Terms such as "bigger" describe both a space taken up physically and sounds which are very loud. It has been shown that even children as young as 3 tend to associate largeness with loudness, an extension of a general principle by which both "loud" and "big" are on the "more" end of a dimensional scale from "less" to "more", and are thus iconically comparable (Smith and Sera 1992). "Length" is another word which has meanings both in the physical world and in aural terms. Poets regularly use the length of lines in a poem to represent largeness or length in time (Nänny 2005). As such, words that take up more horizontal space, which most capital letters do compared to their lowercase counterparts, may be perceived as taking longer to say.

In sum, while pragmatic frameworks such as Relevance Theory tell us that an intentional deviation from standard orthography must contribute something to the interpretation of a text, few studies in internet linguistics have attempted to systematically test what those interpretations are, and fewer still have made any claims about how readers arrive at the agreed-upon interpretations. The iconic potential of many common orthographic features of English gives us a clue regarding these interpretations, pointing to suprasegmental prosodic features as an ideal candidate for the meaningful use of non-standard orthography. The study of prosody gives us a baseline idea of what kinds of emotive interpretations we can expect certain prosodic features to add to a linguistic utterance, and while non-standard orthography cannot be perceived aurally, the theory of silent orality implies that we may still be able to draw on those prosodic meanings if we assume that readers bring their general prosodic competence to bear on silent reading.

3 Methodology

To test the prosodic interpretations of the all caps pattern, I ran a production study designed to elicit semi-natural verbal productions of tweets.[4] Production studies are frequently used to tease out acoustic differences between the verbal production of controlled scenarios in sociophonetics (Idemaru et al. 2019). While this methodology cannot provide a perfect representation of what may be happening in a participant's head while reading silently, the aim is to determine their verbal interpretation, which may be assumed to at least approximate their mental representation.

3.1 Procedure

For this study, 35 participants (18 female, 14 male, 3 non-binary) were recruited to participate in the 60-75 minute experiment. Ages of participants ranged from 18 to 56, with a median age of 21. All participants reported being native or fluent English speakers, most being native speakers of some variety of American English (including AAE) and 4 who reported being speakers of Indian English. 13 participants reported fluency in at least one other language in addition to English. All but 2 participants reported using social media for at least one hour a week, with the median social media use being reported as approximately 8 hours a week. By selecting what set of social media sites they considered to best represent when they first started using social media, 1 participant was determined to have been using social media since the 90s, 8 participants since the early 2000s, 16 from the late 2000s, and 10 since the 2010s.[5] Data from all 35 participants was used in the study.

The stimuli used for this study were initially drawn from a corpus of tweets collected using the Twitter API in June of 2019.[6] The tweets were slightly altered to make them appropriate for use in this study. Identifying information, links, emoji, and other extraneous material were removed to make sure the tweets were clean and easy to read. I specifically selected tweets written in all caps in order to make sure that the pattern of use was authentic. As a control, a second version of each

[4]This production study is a refined version of the study discussed in Heath 2019, which was a pilot study demonstrating the ability of this methodology to produce interpretable results.

[5]This method for determining native era of internet use was developed in McCulloch 2019 as a way to distinguish between age and years of experience on social media based on platform popularity through time.

[6]Twitter was chosen as a framework for this study for two main reasons. First, Twitter allows open access to its API, meaning that real examples of posts may be easily collected in large quantities from a public data source. Second, it can be expected that most people will be familiar with the type of language and formatting found in tweets due to the frequency with which they are shared across other social media platforms through screen shots, and even on the news as a way of reporting on celebrity opinions.

tweet was crafted which was the same in every regard except that the capitalization was changed to a standard pattern (sentence initial and proper nouns only).

In total, there were 30 test stimuli in the study, as well as 243 filler items of other non-standard orthographic patterns such as repeated letters and partial capitalization. Of the 30 test stimuli, each participant saw 15 of them in the all caps form and 15 in the standardized minimal pair form. An example of a minimal pair of stimuli is given in (1).

(1) a. THIS IS THE CUTEST THING I'VE SEEN ALL DAY
b. This is the cutest thing I've seen all day

There were four different versions of the stimuli list with different combinations of items in the all caps or standardized form, and each participant was given one version to read. Five practice items were also included at the beginning of each list to allow the participants to get used to the procedure, and these items were not used in analysis.

In order to make the presentation of the tweets feel as natural as possible, a fake tweet template was generated using prankmenot.com, into which the stimulus text for each item was inserted. Each generated tweet used a random username and handle generated by combining two random words from the NLTK Python library English word list (Bird et al. 2009), filtered to include only words between 5 and 10 characters long. A random number was also included, in the case of the username only. The profile picture for each stimulus was the same default egg used by Twitter when no photo has been selected (but in green rather than blue, for trademark reasons). In this way, the identity of the tweet writer was disguised from the reader, so that the produced prosody was not influenced by any assumed demographic characteristics of the writer. Figure 1 shows an example stimulus.

Figure 1: Example of production study stimulus.

The experiment was carried out in the University of Minnesota Multi-Sensory Perception (MSP) lab and the Center for Applied and Translational Sensory Science (CATSS) lab, each of which included a fully equipped sound-proof booth and computer set-up. Participants were recorded using a microphone on a boom which connected to a Zoom U-22 pre-amp outside of the booth. The recordings were made directly into Audacity in one continuous file and monitored for quality during the recording session.

Before beginning the experiment, participants were asked to sit in the sound booth and read the children's book The Monster at the End of This Book: Starring Lovable, Furry Old Grover (Stone & Smollin 1971) out loud while being recorded. This story is written in the voice of the Sesame Street character Grover as he tries to convince the reader to not continue reading the story. It uses a variety of stylistic orthographic choices to represent Grover's speech (font, size, color, shape, underlining, and capitalization of text). Reading this story gave participants a chance to acclimate themselves to the sound booth and to the microphone and gave the experimenter a chance to calibrate the microphone gain to the individual. It also encouraged participants to loosen up and read emotively, both due to the extreme orthography of the story and due to the nature of the task of reading a story intended for children.

After finishing the warm-up story, participants were directed to turn their attention to the computer monitor in the sound booth to begin the experiment. A PsychoPy script randomized and displayed the items from one of the lists of tweets one at a time on a computer monitor (Pierce et al. 2019). The experimenter gave participants instructions to read each tweet out loud in a way that conveyed the tweet-writer's intended meaning as closely as they were able to guess based solely on the text of the tweet, and the utterances were recorded continuously on the microphone. Participants were able to progress through the tweets using the right arrow key at their own pace. However, to encourage them to take the time to read each stimulus to themselves silently before reading it out loud, they were instructed not to read the next tweet aloud until they saw the signal to do so, a checkmark which appeared four seconds after the new tweet was first displayed. Participants were given an opportunity for a short break halfway through the recording, which was generally at about the 20-minute mark for the average reading speed.

After finishing the recording, participants were given an optional questionnaire (which all participants chose to fully participate in) asking for their age, gender, fluent languages, regional dialect, social media use habits, and a few questions about the nonstandard orthography of social media. Surveys were labeled to correspond to recording data, but no other identifying information about participants was associated with their responses or recordings.

3.2 Analysis

In order to account for the fact that there were multiple non-independent data points per participant and per item, I constructed a linear mixed effects model in R to test the effect of item condition on three critical test variables: Mean F0, Mean Intensity, and Mean Tempo. Mean F0 and Mean Intensity were calculated using F0/Intensity values at 0.01 second intervals across all items, such that individual differences in tempo would not affect the calculation of means. Mean Tempo was calculated by dividing the total duration of each item by the number of syllables in the item, such that there was one syllables-per-second observation for each speaker's production of each item. As a result, the Mean Tempo has a much smaller number of total observations than Mean F0 and Mean Intensity, resulting in larger standard errors.

I began with a maximal model for each test variable which included the following 8 fixed effects: *age* (speaker age, continuous), *gender* (speaker gender, 3 levels), *dialect* (speaker native dialect, 3 levels), *version* (version of item list seen by speaker, 4 levels), *media_time* (hours spent on social media each week by speaker, 8 levels), *media_era* (when speaker began using social media, 4 levels), *syllable* (number of syllables in item, continuous), and *condition* (whether the item included the non-standard pattern or not, 2 levels). Stepwise selection using AIC criterion was used to determine which of these fixed effects should be included in the final model, with the results in (2).

(2) F0(pitch) ~ gender + age + condition + media_time + syllable
intensity(loudness) ~ dialect + condition + media_time + syllable
duration(tempo) ~ age + version + condition + media_era + syllable

The initial models for each variable included random intercepts for both subject (speaker) and item (tweet). After using stepwise selection to determine the significant fixed effects, a likelihood ratio test was used to compare models with random slopes for condition to the ones with just random intercepts. In all cases, random slopes for condition were determined to provide significant improvement and were therefore included in the final models. The resulting models were used in the analysis of the respective critical test variables, and the results are reported below. In diagnosing the models, the Residuals vs Fitted Values plots were acceptable for intensity but showed a pattern for F0 and duration. Standard transformations did not greatly improve the distribution for either F0 or duration. The residuals were overall normally distributed, with pitch showing the most normalcy.

4 Results

The results of the mixed effects model for Mean F0 are given in Table 1 below. Condition shows a significant effect on Mean F0 ($p<0.05$), with the Standard condition lower than the capitalized condition by 17.6 Hz ± 2.84 (standard errors). Gender, Age, and some of the Media Time categories also showed a significant effect. Controlling for average speaker pitch reduces all of the speaker categories (Gender, Age, Media Time) to non-significance, but as the resulting data is not as clearly interpretable, it will not be reported here.

Fixed effects	Estimate	Std. Error	p-value	Significance
(Intercept)	189.556	13.20378	5.79E-14	***
Condition (Standard)	-17.6008	2.8396	1.70E-07	***
Gender (Female)	56.14162	6.36145	5.18E-09	***
Gender (Nonbinary)	71.85297	13.60116	2.03E-05	***
Age	-0.89936	0.37411	0.024243	*
Media Time (1 hour)	-33.6216	12.77237	0.014592	*
Media Time (1-3 hours)	0.41494	13.60777	0.975925	
Media Time (3-5 hours)	-25.8437	9.39056	0.011071	*
Media Time (8-12 hours)	-9.0504	8.21205	0.281309	
Media Time (12-16 hours)	-39.5589	10.01588	0.000598	***
Media Time (16-20 hours)	-27.0383	22.23739	0.23585	
Media Time (20+ hours)	-26.3733	18.0916	0.157833	
Syllables	-0.07972	0.19631	0.687871	

Table 1: Mean F0 across participants, estimates in Hz.

In Table 2, Condition (along with one category of Media Time) shows a significant effect on Mean Intensity ($p<0.05$) with the Standard condition lower than the capitalized condition by 3.49 dB ± 0.50.

Fixed effects	Estimate	Std. Error	p-value	Significance
(Intercept)	61.21725	1.609112	< 2e-16	***
Condition (Standard)	-3.49121	0.498478	4.84E-09	***
Dialect (AAE)	5.020676	2.710664	7.59E-02	.
Dialect (Indian Eng.)	-3.91003	2.106991	7.53E-02	.
Media Time (1 hour)	-9.73085	2.857237	0.00224	**
Media Time (1-3 hours)	-0.44426	2.865211	8.78E-01	
Media Time (3-5 hours)	-1.83199	1.975942	0.36271	
Media Time (8-12 hours)	-1.80438	1.702798	0.29942	
Media Time (12-16 hours)	-4.2386	2.082405	0.05251	.
Media Time (16-20 hours)	0.009928	3.818399	0.99795	
Media Time (20+ hours)	-5.81889	3.818499	0.14006	
Syllables	0.034306	0.072448	0.63955	

Table 2: Mean Intensity across participants, estimates in dB.

In Table 3, Condition is the only variable to show a significant effect on the Mean Duration of syllables in the utterance ($p<0.05$), with the Standard condition faster (shorter duration per syllable) by 0.0345 seconds ± 0.006.

Fixed effects	Estimate	Std. Error	p-value	Significance
(Intercept)	2.10E-01	4.98E-02	1.91E-04	***
Condition (Standard)	-3.45E-02	6.02E-03	8.34E-07	***
Age	1.24E-03	7.72E-04	1.19E-01	
Version (2)	2.23E-02	1.19E-02	7.03E-02	.
Version (3)	-8.63E-03	1.33E-02	5.22E-01	
Version (4)	1.62E-02	1.48E-02	2.82E-01	
Media Era (B)	-0.01047	0.031117	0.739108	
Media Era (C)	0.018166	0.03472	0.605101	
Media Era (D)	0.03748	0.035787	0.304247	
Syllables	-0.00093	0.001055	0.384514	

Table 3: Mean Duration across participants, estimates in seconds.

The box plots in Figure 2 provide a visual comparison of the production of the two conditions for F0, intensity, and duration (per syllable) respectively.

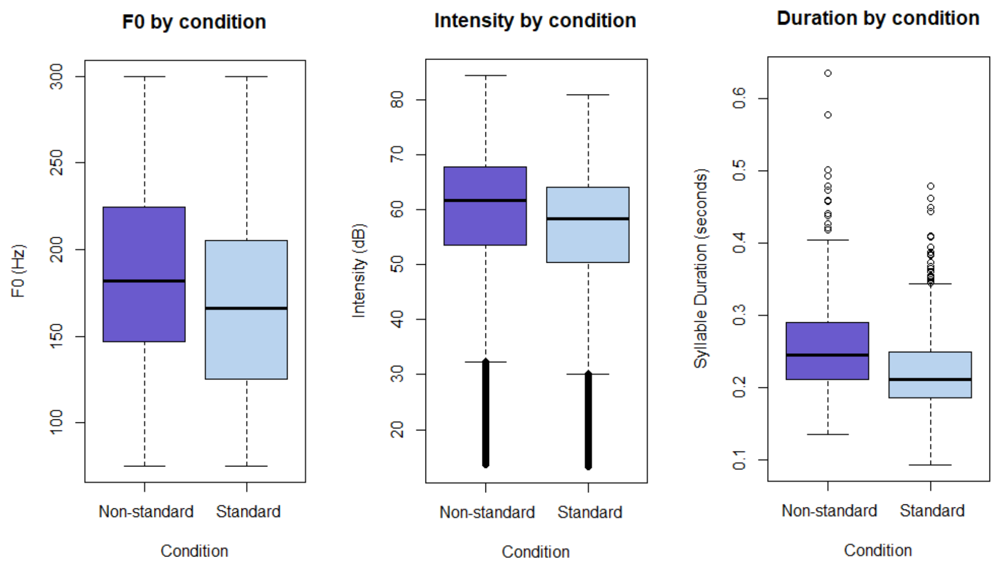

Figure 2: Graphs comparing standard and nonstandard conditions for each prosodic variable.

5 Discussion and Conclusion

F0, intensity, and syllable duration all appear to be influenced by whether a tweet was written in all caps or not. Though determining how much of an acoustic difference in speech prosody is perceivable is a difficult task, it seems that the differences for F0 and intensity are each above the just noticeable difference range (Jongman et al. 2017, Long 2014). In other words, tweets written in all caps would, when read out loud by these participants, most likely be perceived as higher pitched and louder. The difference in tempo might not have breached this threshold, though more research is needed to confirm this (Quené 2004). The fact that these prosodic differences were produced consistently despite the study being conducted in a lab setting which, for example, might make people less likely to feel comfortable raising their voices or showing strong emotion, strongly suggests that the differences in the way participants understood the intended prosody of the tweets based on capitalization was at least as large if not larger than the produced prosodic differences. Therefore, this data offers support for the claim that tweets written in all caps are understood by readers to indicate some combination of higher pitch, higher volume, and possibly longer average syllable duration (slower tempo).

In comparing these results to the expected features of emotive arousal, note that higher intensity and higher F0, both consistently found for all caps, are both common features of high arousal emotional states, which can include anger as well as excitement and joy. Although slower tempos do tend to be associated with low arousal emotions, which would seem to contradict the theory that all caps completely aligns with high arousal prosodic features, it is not always the case that high arousal emotions must be expressed with faster tempos, especially if all caps also has an association with emphasis due to its visual salience. Additionally, if the difference in tempo is not above the noticeable threshold, it may be evidence that tempo is not a consistent factor in the interpretation of all caps. Analysis of the prosodic interpretations of capitalization in other patterns besides all caps may shed light on the relationship between tempo, emotion, and emphasis.

In addition to offering support for the general intuition that writing in all caps can indicate yelling, these results provide a more nuanced picture of how readers interpret all caps writing. In particular, it's clear that yelling and anger are not the only interpretations of all caps. Instead, this orthographic pattern seems to be associated with a combination of prosodic cues which can be interpreted as angry yelling in certain contexts but may be interpreted with a different affect in other contexts. This might suggest that there is a direct relationship between the orthographic representation and the associated prosody, possibly cued by the iconic nature of the larger visual form of all caps text, and the emotive interpretations are deduced from the intuitive understanding of contextual prosody a reader is able to draw on.

The results of this study also replicate the pilot study results for this methodology, as discussed in Heath 2019. Even with this study having a different set of participants, a different set of stimuli, a more sensitive recording setup, and a more advanced statistical model, differences in the average F0 and intensity between the two conditions were very similar. Not only does this offer further support for the validity of the results, but it also offers support for the consistency of the production study methodology for testing prosodic inferences from text.

One shortcoming with the analysis is that it doesn't take into consideration the conventionalization of orthographic meaning over time and the way new readers acquire the conventional meaning. If the orthography is purely iconic, we would expect all readers to easily reach the same conclusion about its appropriate meaning and use with very little exposure. However, it's clear from the regular complaints of young people about older relatives using all caps inappropriately that this is not the case. Like with most iconicity, some cultural context is required to determine what variation is meaningful and what types of meanings are relevant options. Communal consensus has the final say, and future readers may learn of that consensus from context rather than through the iconic means that first created the opportunity.

Further evidence for communal consensus playing a role in learning the meaning associated with all caps may come from Deaf and Hard of Hearing (DHH) social media users. Though further research is needed to make generalizable claims about interpretations of nonstandard orthography by DHH readers, a preliminary interview with Deaf social media user Jonathan Ainsworth suggests that DHH readers reach similar conclusions about intended meanings of nonstandard orthography as hearing social media users, including for the all caps pattern (personal communication, Aug. 14, 2020). While American Sign Language (ASL) and other signed languages do have prosodic features that could potentially be iconically represented in English orthography, the fact that ASL speakers must rely on translation to English to communicate in writing makes the relationship between orthography and prosody more distant. This suggests that conventional interpretations and uses of nonstandard orthography can be acquired in the absence of prosodic iconicity. While this doesn't necessarily mean that patterns like all caps didn't originate as attempts at expressing verbal prosody, it does complicate the relationship between orthography and prosody.

In sum, I employ a novel methodology in this study to demonstrate that when social media users read a tweet written in all caps, averages in spoken prosodic features differ significantly between all caps text and regular text across participants. Beyond the relevance of this finding, this study affirms that production studies are a viable way of determining whether other nonstandard orthographic patterns are associated with prosodic features and of analyzing those features. The data and analysis discussed in this paper were a subset of a larger study that included a variety of other nonstandard orthographic features (such as all lowercase, single word capitalization, repeated letters, and periods following every word). By applying the analytical methods I develop here to these other patterns, this larger study will contribute to an even more nuanced understanding of the relationship between orthography and prosody in CMC. Further discussion with DHH social media users will also help with untangling the theoretic implications of this data. Overall, this research provides a starting point for the quantitative analysis of previously primarily qualitative intuition regarding the language of CMC. As linguists of all subdisciplines begin turning their attention to social media data, these types of analyses will form the foundation necessary for a complete linguistic understanding of text-based communicative spaces.

References

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media.

Brumm, Henrik and Sue Anne Zollinger. 2011. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 148:1173-1198.

Buchanan, Susan. 2016. National Weather Service will stop using all caps in its forecasts. National Oceanic and Atmospheric Administration. URL `https://www.noaa.gov/media-release/national-weather-service-will-stop-using-all-caps-in-its-forecasts`

Crumpton, Joe, and Cindy L. Bethel. 2015. A survey of using vocal prosody to convey emotion in robot speech. *International Journal of Social Robotics* 8:271–285.

Gunraj, Danielle, April Drumm-Hewitt, Erica Dashow, Sri Siddhi Upadhyay, and Celia Klin. 2016. Texting insincerely: The role of the period in text messaging. *Computers in Human Behavior* 55:1067–1075.

Hass, Christina, Pamela Takayoshi, Brandon Carr, Kimberley Hudson, and Ross Pollock. 2011. Young people's everyday literacies: The language features of instant messaging. *Research in the Teaching of English* 45:378–404.

Heath, Maria. 2019. Not-so-silent orality: Production studies as a test of orthographic-prosodic correlates. Ms., University of Minnesota. URL `https://drive.google.com/file/d/15K3m3h82MCjiA-3xJ9rMAyePDbVtBDvB/view`

Hiraga, Masako K. 2005. *Metaphor and Iconicity*. Palgrave Macmillan, London.

Houghton, Kenneth, Sri Siddhi Upadhyay, and Celia Klin. 2018. Punctuation in text messages may convey abruptness. Period. *Computers in Human Behavior* 80:112–121.

Idemaru, Kaori, Bodo Winter, Lucien Brown, and Grace Eunhae Oh. 2019. Loudness trumps pitch in politeness judgments: Evidence from Korean deferential speech. *Language and Speech.* 1–26.

Jongman, Allard, Zhen Qin, Jie Zhang, and Joan A. Sereno. 2017. Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America* 142:163–169.

Lyddy, Fiona, Francesca Farina, James Hanney, Lynn Farrell, and Niamh Kelly O'Neill. 2013. An analysis of language in university students' text messages. *Journal of Computer-Mediated Communication* 19:546–561.

Long, Marshall. 2014. Human perception and reaction to sound. In *Architectural Acoustics (Second Edition)*, ed. M. Long, 81-127. Oxford: Elsevier.

McCulloch, Gretchen. 2019. *Because Internet: Understanding the New Rules of Language.* New York: Riverhead Books.

McSweeney, Michelle. 2018. *The Pragmatics of Text Messaging: Making Meaning in Messages*. New York: Routledge.

Nänny, Max. 2005. Diagrammatic iconicity in poetry. In *Theory into Poetry: New Approaches to the Lyric*, ed. E. Müller-Zettelmann and M. Rubik, 229–252. Leiden: Brill..

Peirce, Jonathan et al. 2019. PsychoPy2: experiments in behavior made easy. *Behavior Research Methods* 51:195–203.

Quené, Hugo. 2004. What is the Just Noticeable Difference for tempo in speech? *LOT Occasional Series*, 2:149–158.

Scherer, Klaus, and Tanja Bänziger. 2004. Emotional expression in prosody: A review and an agenda for future research. In Proceedings of *Speech Prosody 2004*, 359–366.

Smith, Linda and Maria Sera. 1992. A developmental analysis of the polar structure of dimensions. *Cognitive Psychology* 24:99–142.

Soffer, Oren. 2010. ''Silent orality'': Toward a conceptualization of the digital oral features in CMC and SMS texts. *Communication Theory* 20:387–404.

Sperber, Dan, and Deirdre Wilson. 2004. Relevance theory. In *Handbook of Pragmatics*, ed. L. Horn and G. Ward, 607–632. Oxford: Blackwell.

Stone, Jon, and Michael Smollin (Illustrator). 1971. *The Monster at the End of This Book: Starring Lovable, Furry Old Grover*. New York: Penguin Random House.

Zappavigna, Michele. 2012. *Discourse of Twitter and Social Media*. Bloomsbury: Bloomsbury Publishing.

Institute of Linguistics
University of Minnesota
Minneapolis, MN 55455
heath167@umn.edu