Keystroke Patterns as Prosody in Digital Writings:
A Case Study with Deceptive Reviews and Essays

Ritwik Banerjee Song Feng Jun S. Kang
Computer Science
Stony Brook University
{rbanerjee, songfeng, junkang}
@cs.stonybrook.edu

Yejin Choi
Computer Science & Engineering
University of Washington
yejin@cs.washington.edu

Abstract

In this paper, we explore the use of keyboard strokes as a means to access the real-time writing process of online authors, analogously to prosody in speech analysis, in the context of deception detection. We show that differences in keystroke patterns like editing maneuvers and duration of pauses can help distinguish between truthful and deceptive writing. Empirical results show that incorporating keystroke-based features lead to improved performance in deception detection in two different domains: online reviews and essays.

1 Introduction

Due to the practical importance of detecting deceit, interest in it is ancient, appearing in papyrus dated back to 900 B.C. (Trovillo, 1939). In more recent years, several studies have shown that the deceiver often exhibits behavior that belies the content of communication, thus providing cues of deception to an observer. These include linguistic (*e.g.,* Newman et al. (2003), Hancock et al. (2004)) as well as paralinguistic (*e.g.,* Ekman et al. (1991), DePaulo et al. (2003)) cues. Recognizing deception, however, remains a hard task for humans, who perform only marginally better than chance (Bond and DePaulo, 2006; Ott et al., 2011).

Recent studies suggest that computers can be surprisingly effective in this task, albeit in limited domains such as product reviews. Prior research has employed lexico-syntactic patterns (Ott et al., 2011; Feng et al., 2012) as well as online user behavior (Fei et al., 2013; Mukherjee et al., 2013). In this paper, we study the effect of keystroke patterns for deception detection in digital communications, which might be helpful in understanding the psychology of deception and help toward trustful online communities. This allows us to investigate differences in the writing and revisional processes of truthful and fake writers. Our work thus shares intuition with HCI research linking keystroke analysis to cognitive processes (Vizer et al., 2009; Epp et al., 2011) and psychology research connecting cognitive differences to deception (Ekman, 2003; Vrij et al., 2006).

Recent research has shown that lying generally imposes a cognitive burden (*e.g.,* McCornack (1997), Vrij

et al. (2006)) which increases in real-time scenarios (Ekman, 2003). Cognitive burden has been known to produce differences in keytroke features (Vizer et al., 2009; Epp et al., 2011). Previous research has not, however, directly investigated any quantitative connection between keystroke patterns and deceptive writing.

In this paper, we posit that cognitive burdens in deception may lead to measurable characteristics in keystroke patterns. Our contributions are as follows: (1) introducing keystroke logs as an extended linguistic signal capturing the real-time writing process (analogous to prosody in speech analysis) by measuring the *writing rate*, *pauses* and *revision rate*. (2) showing their empirical value in deception detection, (3) providing novel domain-specific insights into deceptive writing, and (4) releasing a new corpus of deception writings in new domains.[1]

2 Related Work

Prior research has focused mainly on using keystroke traits as a behvioral biometric. Forsen et al. (1977) first demonstrated that users can be distinguished by the way they type their names. Subsequent work showed that typing patterns are *unique to individuals* (Leggett and Williams, 1988), and can be used for authentication (Cho et al., 2000; Bergadano et al., 2002) and intrusion detection (Killourhy and Maxion, 2009).

Keystroke pauses have been linked to linguistic patterns in discourse (e.g. Matsuhashi (1981), van Hell et al. (2008)) and regarded as indications of cognitive burden (*e.g.,* Johansson (2009), Zulkifli (2013)). In this paper, we present the *first* empirical study that quantitatively measures the deception cues in real-time writing process as manifested in keystroke logs.

3 Data Collection

As discussed by Gokhman et al. (2012), the crowd-sourcing approach to soliciting deceptive content simulates the real world of online deceptive content creators. We collected the data via Amazon Mechanical Turk.[2] Turkers were led to a separate website where keylogging was enabled, and asked to write truthful and deceptive texts ($\geq$ 100 words) on one of three top-

---

[1] Available at http://www3.cs.stonybrook.edu/~junkang/keystroke/
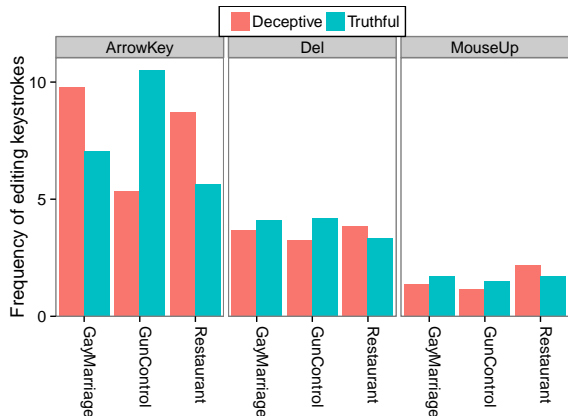[2] https://www.mturk.com/mturk

Figure 1: Number of keystrokes corresponding to the three types of edit patterns ($\mathcal{E}_3$): (a) use of arrow keys, (b) deletion (Delete and Backspace) and (c) text selection with mouse.
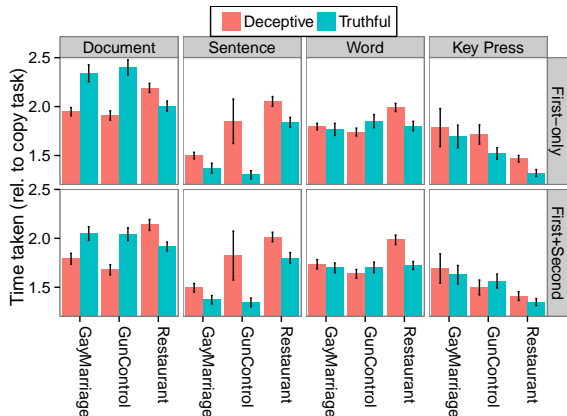

Figure 2: Average normalized timespan $\widehat{\delta}(e)$ for documents, sentences, words and key presses. The top row considers only the first text, while the bottom row considers both flows.

ics: *restaurant review*, *gay marriage* and *gun control*. Each Turker was required to agree to their typing being logged. Since copy/paste operations defeat our purpose of studying keystrokes in the typing process, they were disabled. This restriction also acts as a hindrance to plagiarism. All texts were reviewed manually, and those not meeting the requirements (due to the being too short, plagiarized content, etc.) were disregarded.

Writing task design: The task was designed such that each Turker wrote a pair of texts, one truthful and one deceptive, on the same topic. For restaurant reviews, they were asked to write a truthful review of a restaurant they liked, and a deceptive review of a restaurant they have never been to or did not like. For the other two topics – 'gun control' and 'gay marriage' – we asked their opinion: *support*, *neutral*, or *against*. Then, they were asked to write a truthful and a deceptive essay articulating, respectively, their actual opinion and its opposite.[3] The tasks further were divided into two 'flows': writing the truthful text before the deceptive one, and vice versa. Each Turker was assigned only one flow, and was not allowed to participate in the other. After completing this, each Turker was asked to *copy* their own typing, *i.e.,* re-type the two texts.

Finally, in order to get an idea of the cognitive burden associated with truthful and deceptive writing, we asked the Turkers which task was easier for them. Of the 196 participants, 152 answered "truthful", 40 answered "deceptive" and only 4 opted for "not sure".

What are logged: We deployed a keylogger to capture the mouse and keyboard events in the "text area". The events *KeyUp*, *KeyDown* and *MouseUp*, along with the keycode and timestamp were logged.[4] For the three topics *restaurant review*, *gay marriage* and *gun control* we obtained 1000, 800 and 800 texts, respectively.

In the remainder of this paper, k^{dn} and k^{up} denote the KeyDown and KeyUp events for a key k. For any

event e, its *timespan*, *i.e.,* the time interval between the beginning and end of e, is denoted by $\delta(e)$.

4 Feature Design

Keystroke logging enables the study of two types of information that go beyond conventional linguistic analysis. First, it captures editing processes (*e.g.,* deletions, insertions made by changing cursor position, etc.). Second, it reveals the temporal aspect of text generation (*e.g.,* duration, latency). Our exploration of these features and their application in deception detection is motivated by the similarities between text and speech generation. Editing patterns, for instance, can be viewed as attempts to veil *incoherence* in deceptive writing and temporal patterns like latency or pause can be treated as analogous to *disfluency*.

Different people, of course, have varying typing skills, and some may type faster than others. In order to control for this variation, we normalize all event timespans $\delta(e)$ with respect to the corresponding event timespan in the *copy* task: $\widehat{\delta}(e) = \delta(e)/\delta(e_{copy})$.

4.1 Editing Patterns

In this work, we treat keys that are used only for editing as different from others. Text editing is done by employing a small subset of available keys: deletion keys ('Backspace' and 'Delete'), arrow keys ($\leftarrow$, $\rightarrow$, $\uparrow$ and $\downarrow$) and by using the mouse for text selection (*i.e.,* the 'MouseUp' event). The three types of editing keystrokes are collectively denoted by

$$\mathcal{E}_3 = \langle |\textsc{del}|, |\textsc{mselect}|, |\textsc{arrow}| \rangle$$

where

(i) $|\textsc{del}|$ = number of deletion keystrokes
(ii) $|\textsc{mselect}|$ = number of 'MouseUp' events, and
(iii) $|\textsc{arrow}|$ = number of arrow keystrokes

The editing differences between truthful and deceptive writing across all three topics are shown in Fig. 1.

4.2 Temporal Aspects

Each event is logged with a keycode and a timestamp. In order to study the temporal aspects of digital writing, we calculate the timespan of different linguistic

---

[3]To prevent a change in opinion depending on task availability, Turkers were redirected to other tasks if their opinion was neutral, or if we had enough essays of their opinion.

[4]Printable (*e.g.,* alphanumeric characters) as well as non-printable keystrokes like (*e.g.,* 'Backspace'), are logged.

Topic	Features	Flow	
		First + Second	First-only
Restaurants	BoW	73.9	78.8
	BoW + $\mathcal{T}_6$	74.3	79.1
	BoW + $\mathcal{T}_6$ + $\mathcal{E}_3$	74.6	**80.3**[*]
Gun Control (Support)	BoW	86.5	80.0
	BoW + $\mathcal{T}_6$	86.8	82.5[*]
	BoW + $\mathcal{T}_6$ + $\mathcal{E}_3$	**88.0**[§]	83.5[*]
Gun Control (Oppose)	BoW	88.5	88.0
	BoW + $\mathcal{T}_6$	89.8	87.5
	BoW + $\mathcal{T}_6$ + $\mathcal{E}_3$	**90.8**[*]	89.1
Gay Marriage (Support)	BoW	92.5	92.0
	BoW + $\mathcal{T}_6$	93.8	92.5
	BoW + $\mathcal{T}_6$ + $\mathcal{E}_3$	**94.3**[*]	92.0
Gay Marriage (Oppose)	BoW	84.5	86.5
	BoW + $\mathcal{T}_6$	85.0	**87.0**
	BoW + $\mathcal{T}_6$ + $\mathcal{E}_3$	85.3	86.8

Table 1: SVM classifier performance for truthful vs. deceptive writing. Statistically significant improvements over the baseline are marked * ($p < 0.05$) and § ($p < 0.1$). $\mathcal{E}_3 = \langle |\text{DEL}|, |\text{MSELECT}|, |\text{ARROW}| \rangle$ denotes the editing keystrokes, and $\mathcal{T}_6$ is the set of normalized timespans of documents, words (plus preceding keystroke), all keystrokes, spaces, non-whitespace keystrokes and inter-word intervals: $\mathcal{T}_6 = \{\widehat{\delta}(\text{D}), \widehat{\delta}(k), \widehat{\delta}(\text{SP}), \widehat{\delta}(\neg\text{SP}), \widehat{\delta}(\neg\text{W}), \widehat{\delta}(k_{prv} + \text{W})\}$

units such as words, sentences and even entire documents. Further, we separately inspect the timespans of different parts of speech, function words and content words. In addition to event timespans, intervals between successive events (*e.g.,* inter-word and inter-sentence pauses) and pauses preceding or succeeding and event (*e.g.,* time interval before and after a function word) are measured as well.

5 Experimental Results

This section describes our experimental setup and presents insights based on the obtained results. All classification experiments use 5-fold cross validation with 80/20 division for training and testing. In addition to experimenting on the entire dataset, we also separately analyze the texts written first (of the two texts in each 'flow'). This additional step is taken in order to eliminate the possibility of a text being primed by its preceding text.

Deception cues in keystroke patterns: To empirically check whether keystroke features can help distinguish between truthful and deceptive writing, we design binary SVM classifiers.[5] Unigrams with `tf-idf` encoding is used as the baseline. The average baseline accuracy across all topics is 82.58% when considering both texts of a flow, and 83.62% when considering only the first text of each flow. The better performance in the latter possibly indicates that the second text of a flow exhibits some amount of lexical priming with the first.

The high accuracy of the baseline is not surprising. Previous work by Ott et al. (2011) reported similar per-

---

[5]We use the LIBLINEAR (Fan et al., 2008) package.

$\widehat{\delta}(\text{w})$		$\widehat{\delta}(k_{prv} + \text{w})$	
D > T	T > D	D > T	T > D
our	best	when	one
if	get	quality	other
when	well	even	get
were	your	on	service
it's	fresh	by	been
quality	not	me	their
dishes	my	has	not
the	one	also	with
i've	had	go	friendly
on	hat	we	great
they	of	had	an
we	other	is	our
friendly	very	at	are
has	love	which	really
at	service	from	but
wait	great	dishes	favorite
an	really	or	very
go	you	re	about
is	but	would	will
which	been	just	here

Table 2: Top 20 words in **restaurant reviews** with greatest timespan difference between deceptive and truthful writing.

formance of unigram models. The focus of our work is to explore the completely new feature space of typographic patterns in deception detection. We draw motivation from parallels between the text generation and speech generation processes. Prosodic concepts such as *speed*, *disfluency* and *coherence* can be realized in typographic behavior by analyzing timestamp of keystrokes, pauses and editing patterns, respectively.

Based on the differences in the temporal aspects of keystrokes, we extract six timespan features to improve this baseline. This set, denoted by $\mathcal{T}_6$, comprises of

(i) $\widehat{\delta}(\text{D})$ = timespan of entire document
(ii) $\widehat{\delta}(k_{prv} + \text{W})$ = average timespan of word plus preceding keystroke
(iii) $\widehat{\delta}(k)$ = average keystroke timespan
(iv) $\widehat{\delta}(\text{SP})$ = average timespan of spaces
(v) $\widehat{\delta}(\neg\text{SP})$ = average timespan of non whitespace keystrokes
(vi) $\widehat{\delta}(\neg\text{W})$ = average interval between words.

The improvements attained by adding $\mathcal{T}_6$ to the baseline are shown in Table 1. Adding the edit patterns ($\mathcal{E}_3$) (cf. § 4.1) further improves the performance (with the exception of two cases) by 0.7–3.5%.

Writing speed, pauses and revisions: To study the temporal aspect of language units across all topics, we first consider all texts, and then restrict to only the first of each 'flow'. The timespan measurements are presented in Fig. 2, showing the average duration of typing documents, sentences, words and individual keystrokes. The timespans are measured as the interval between the first and the last keystrokes. The sentence timespan, for instance, does not include the gap between a sentence end and the first keystroke marking the beginning of the next.

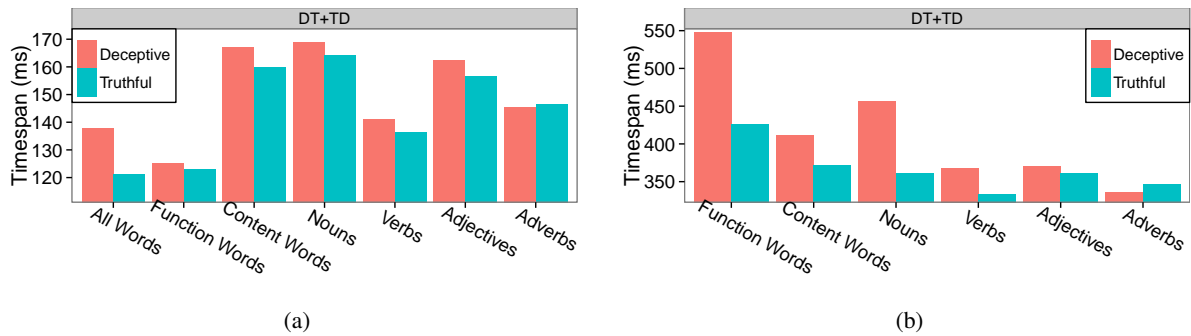The sentence timespans for "gay marriage" and "gun

Figure 3: Event timespans in restaurant reviews: (a) language units, and (b) language units including their preceding k^{dn}.

control" are lower in truthful writing, even though the document timespans are higher. This difference implies that the writer is spending a longer period of time to think before commencing the next sentence, but once started, the actual typing proceeds rapidly.

Apart from restaurant reviews, truthful writers have typed slower. This may be due to exercising better care while expressing their honest opinion.

For restaurant reviews, the document, sentence and word timespans are significantly higher in deceptive writing. This, however, is not the case for documents and words in the other two topics. We conjecture that this is because deception is harder to write for product reviews, due to their dependence on factual details. Gun control and gay marriage, on the other hand, are topics well discussed in media, and it is possible that the writers are aware of the arguments that go against their personal belief. The frequency of revisional occurrences (*i.e.,* keys used for editing) shown in Fig. 1, too, supports the thought that writing fake reviews may be harder than adopting a fake stance on well-known issues. Deceptive reviews exhibit a higher number of revisions than truthful ones, but essays show the opposite trend. Our findings align with previous studies (Ott et al., 2011) which showed that deception cues are domain dependent.

Writing speed variations over word categories: Next, we investigate whether there is any quantitative difference in the writing rate over different words with respect to the deceptive and truthful intent of the author. In an attempt to understand this, we analyze words which show the highest timespan difference between deceptive and truthful writings.

Table 2 presents words in the restaurant review topic for which deceptive writers took a lot longer than truthful writers, and vice versa. Some word categories exhibit common trends across all three topics. Highly subjective words, for instance (*e.g.,* "love", "best", "great") are words over which truthful writers spent more time.

Deceptive and truthful texts differ in the typing rate of first- and second-person pronouns. Deceptive reviews reveal more time spent in using 2nd-person pronouns, as shown by "you" and "your". This finding throws some light on how people perceive text cues. Toma and Hancock (2012) showed that readers per-

form poorly at deception detection because they rely on unrelated text cues such as 2nd-person pronouns. Our analysis indicates that people associate the use of 2nd-person pronouns more with deception not only while reading, but while writing as well.

Deceptive reviews also exhibit longer time spans for 1st-person pronouns (*e.g.,* "we", "me"), which have been known to be useful in deception detection (Newman et al., 2003; Ott et al., 2011). Newman et al. (2003) attributed the less frequent usage of 1st-person pronouns to *psychological distancing*. The longer time taken by deceptive writers in our data is a possible sign of increased cognitive burden when the writer is unable to maintain the psychological distance. Deceptive reviewers also paused a lot more around relative clauses, e.g., "if", "when", and "which".

In essays, however, the difference in timespans of 1st-person and 2nd-person pronouns as well as the timespan difference in relative clauses were insignificant ($< 50\,ms$).

A broader picture of the temporal difference in using different types of words is presented in Fig. 3, which shows deceptive reviewers spending less time on adverbs as compared to truthful writers, but more time on nouns, verbs, adjectives, function words and content words. They also exhibited significantly longer pauses before nouns, verbs and function words.

6 Conclusion

In this paper, we investigated the use of typographic style in deception detection and presented distinct temporal and revisional aspects of keystroke patterns that improve the characterization of deceptive writing. Our study provides novel empirically supported insights into the writing and editing processes of truthful and deceptive writers. It also presents the first application of keylogger data used to distinguish between true and fake texts, and opens up a new range of questions to better understand what affects these different keystroke patterns and what they exhibit. It also suggests new possibilities for making use of keystroke information as an extended linguistic signal to accompany writings.

Acknowledgements

References

Francesco Bergadano, Daniele Gunetti, and Claudia Picardi. 2002. User Authentication through Keystroke Dynamics. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):367–397.

Charles F Bond and Bella M DePaulo. 2006. Accuracy of Deception Judgments. *Personality and Social Psychology Review*, 10(3):214–234.

Sungzoon Cho, Chigeun Han, Dae Hee Han, and Hyung-Il Kim. 2000. Web-based Keystroke Dynamics Identity Verification Using Neural Network. *Journal of Organizationl Computing and Electronic Commerce*, 10(4):295–307.

Bella M DePaulo, James J Lindsay, Brian E Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper. 2003. Cues to Deception. *Psychological Bulletin*, 129(1):74.

Paul Ekman, Maureen O'Sullivan, Wallace V Friesen, and Klaus R Scherer. 1991. Invited Article: Face, Voice and Body in Detecting Deceit. *Journal of Nonverbal Behavior*, 15(2):125–135.

Paul Ekman. 2003. Darwin, Deception, and Facial Expression. *Annals of the New York Academy of Sciences*, 1000(1):205–221.

Clayton Epp, Michael Lippold, and Regan L Mandryk. 2011. Identifying Emotional States Using Keystroke Dynamics. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pages 715–724. ACM.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. *The Journal of Machine Learning Research*, 9:1871–1874.

Geli Fei, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Exploiting Burstiness in Reviews for Review Spammer Detection. In *ICWSM*, pages 175–184.

Song Feng, Ritwik Banerjee, and Yejin Choi. 2012. Syntactic Stylometry for Deception Detection. In *Proc. 50th Annual Meeting of the ACL*, pages 171–175. ACL.

George E Forsen, Mark R Nelson, and Raymond J Staron Jr. 1977. Personal Attributes Authentication Techniques. Technical report, DTIC Document.

Stephanie Gokhman, Jeff Hancock, Poornima Prabhu, Myle Ott, and Claire Cardie. 2012. In Search of a Gold Standard in Studies of Deception. In *Computational Approaches to Deception Detection*, pages 23–30. ACL.

Jeffrey T Hancock, L Curry, Saurabh Goorha, and Michael T Woodworth. 2004. Lies in Conversation: An Examination of Deception Using Automated Linguistic Analysis. In *Annual Conference of the Cognitive Science Society*, volume 26, pages 534–540.

Victoria Johansson. 2009. *Developmental Aspects of Text Production in Writing and Speech*. Ph.D. thesis, Lund University.

Kevin S Killourhy and Roy A Maxion. 2009. Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. In *Dependable Systems & Networks, 2009. DSN'09.*, pages 125–134. IEEE.

John Leggett and Glen Williams. 1988. Verifying Identity Via Keystroke Characteristics. *International Journal of Man-Machine Studies*, 28(1):67–76.

Ann Matsuhashi. 1981. Pausing and Planning: The Tempo of Written Discourse Production. *Research in the Teaching of English*, pages 113–134.

Steven A McCornack. 1997. The Generation of Deceptive Messages: Laying the Groundwork for a Viable Theory of Interpersonal Deception. In John O Greene, editor, *Message Production: Advances in Communication Theory*. Erlbaum, Mahwah, NJ.

Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie Glance. 2013. What Yelp Fake Review Filter Might be Doing. In *ICSWM*, pages 409–418.

Matthew L Newman, James W Pennebaker, Diane S Berry, and Jane M Richards. 2003. Lying Words: Predicting Deception from Linguistic Styles. *Personality and Social Psychology Bulletin*, 29(5):665–675.

Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding Deceptive Opinion Spam by Any Stretch of the Imagination. In *Proc. 49th Annual Meeting of the ACL: HLT*, pages 309–319. ACL.

Catalina L Toma and Jeffrey T Hancock. 2012. What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles. *Journal of Communication*, 62(1):78–97.

Paul V Trovillo. 1939. A History of Lie Detection. *Journal of Criminal Law and Criminology (1931-1951)*, 29:848–881.

Janet G van Hell, Ludo Verhoeven, and Liesbeth M van Beijsterveldt. 2008. Pause Time Patterns in Writing Narrative and Expository Texts by Children and Adults. *Discourse Processes*, 45(4-5):406–427.

Lisa M Vizer, Lina Zhou, and Andrew Sears. 2009. Automated Stress Detection Using Keystroke and Linguistic Features: An Exploratory Study. *International Journal of Human-Computer Studies*, 67(10):870–886.

Aldert Vrij, Ronald Fisher, Samantha Mann, and Sharon Leal. 2006. Detecting Deception by Manipulating Cognitive Load. *Trends in Cognitive Sciences*, 10(4):141–142.

Putri Zulkifli. 2013. *Applying Pause Analysis to Explore Cognitive Processes in the Copying of Sentences by Second Language Users*. Ph.D. thesis, University of Sussex.