

RESEARCH

Open Access

Automated deception detection of 911 call transcripts

Mary B Burns^{1*} and Kevin C Moffitt²

Abstract

This study is a successful proof of concept of using automated text analysis to accurately classify transcribed 911 homicide calls according to their veracity. Fifty matched, caller-side transcripts were labeled as truthful or deceptive based on the subsequent adjudication of the cases. We mined the transcripts and analyzed a set of linguistic features supported by deception theories. Our results suggest that truthful callers display more negative emotion and anxiety and provide more details for emergency workers to respond to the call. On the other hand, deceivers attempt to suppress verbal responses by using more negation and assent words. Using these features as input variables, we trained and tested several machine-learning classification algorithms and compared the results with the output from a statistical classification technique, discriminant analysis. The overall performance of the classification techniques was as high as 84% for the cross-validated set. The promising results of this study illustrate the potential of using automated linguistic analyses in crime investigations.

Keywords: Automated Linguistic Analysis; Deception Detection; 911 Calls

Introduction

In part due to a natural “truth bias,” humans (including those with special training) can generally detect deception at a rate only slightly better than chance, at around 54% [1,2]. This inability to accurately separate truth from deception can have serious consequences; this is particularly true in the case of law enforcement. Not only do criminal investigators have to be concerned with deception that is not detected (false negatives), but they must also take into account the serious outcome of labeling truth tellers as deceivers (false positives). Thus, there is a critical need for more reliable and accurate methods of identifying deception, especially in the earliest contact between a suspected perpetrator and law enforcement. In this study, we evaluate linguistic cues extracted from transcriptions of 911 homicide^a calls as potential indicators of deception.

As a dataset, 911 calls may be ideal for deception detection research because they occur in a real-world setting, are relatively unrehearsed, occur soon after the crime in question, and are part of emotionally charged situations [3,4]. In contrast, in laboratory settings where deception is

often sanctioned, the consequences of lying and getting caught are minor, and incentives to deceive are artificial (e.g., [5,6]). Thus, as Mann et al. [7] suggest, these low-stakes laboratory environments may not induce feelings of guilt or elicit behavior found in real settings; as a result, they may adversely affect researchers’ ability to accurately judge credibility, and therefore diminish the external validity of the results. Thus, it is important to add real-world scenarios to the collection of deception research data.

Deception studies using real-world person-of-interest statements begin to address the research gap identified above [8–10]. Person-of-interest statements are written explanations of crimes by a person who has not been formally charged with a crime but who is “of interest” to law enforcement in an investigation. Studies have been conducted using pre-polygraph interviews, which also involve real-world data [11]. However, relative to 911 calls, the time between the crime and the statements in pre-polygraph interviews is much longer, giving deceivers time to rehearse their responses. Additionally, such datasets may be influenced by investigative procedures or the deceiver’s contact with acquaintances between the incident and the statement [4,12]. Thus, 911 calls may be considered a less biased dataset because the interaction with authorities occurs much sooner after the crime and the individual’s

* Correspondence: mary.burns1@montana.edu

¹Jake Jabs College of Business & Entrepreneurship, Montana State University, P.O. Box 173040, 59717-3040 Bozeman, MT, USA

Full list of author information is available at the end of the article

words have been less affected by outside influences. Therefore, this dataset provides researchers an unusual opportunity for an unfiltered look at deception.

A subset of 911 calls reporting homicides are available on the Internet because they are high-profile crimes. The calls can be corroborated for ground truth by examining the associated court outcomes. Establishing ground truth is one of the most difficult aspects of research into deception detection [13]. Because initial homicide reports, subsequent investigations, legal proceedings, and judgments are covered widely by news media, it is possible to substantiate ground truth to a high degree of certainty for these types of 911 calls. Although 911 homicide calls represent an extremely small fraction of total 911 calls, they can serve as a proxy for calls made reporting other high-profile crimes such as arson, bomb threats, and sexual assault.

Data mining techniques, including text mining, linguistic feature mining, and classification by text features, can be used to analyze the “caller side” of the transcripts of these calls. Text mining involves looking for hidden patterns or cues in texts, while linguistic feature mining refers to dissecting texts with respect to specific linguistic categories, such as words associated with positive affect. Text mining is a multidisciplinary research area that combines approaches used in the fields of computer science, linguistics, mathematics, communication, and psychology. It focuses on using computing power to process unstructured human language in spoken or written form [5]; furthermore, text mining has been used to process text data to discover linguistic cues or features in order to classify documents, including fraudulent versus non-fraudulent financial statements [14,15] or deceptive versus truthful statements in instant messages, email exchanges, or person-of-interest statements [6–10,16].

In this study, we applied linguistic feature mining to evenly matched (i.e., deceptive matched with truthful) transcripts of 911 homicide calls via Linguistic Inquiry and Word Count (LIWC). LIWC 2007, a general-purpose psychosocial linguistic dictionary comprising 4,500 words and word stems, has been used by researchers to quantify linguistic cues for deception [16–19].

This paper contributes to research streams in security informatics, deception detection, and crime analysis in the following ways:

- 1) Using a truthful/deceptive matched convenience data set of fifty 911 calls, we identify linguistic cues based on deception theories that may be used to discriminate between deceptive and truthful 911 calls;
- 2) We extract useful information from unstructured text comprised of transcribed 911 calls to demonstrate that the largely unexploited data of 911

calls can be analyzed for further investigative work; and

- 3) The classification results achieve up to 84% accuracy.

The remainder of this paper proceeds as follows: we first review deception theories and previous research involving automated or manual deception detection; we then examine advantages and disadvantages of using 911 calls as a data source and discuss linguistic cues of deceptive and truthful 911 callers as we develop our research question and hypotheses; next we describe our methodology, and finally we present and discuss the results, research contributions, limitations of the study, and future research directions.

Literature review

Deception is defined as purposefully concealing the truth, either by omission or commission. In the present study, our general hypothesis is that 911 callers who deceive exhibit systematic differences in the words they use compared to 911 callers who are telling the truth. Possible underlying causes for the differences in deceptive speech are described by deception theories, including four factor theory [20], interpersonal deception theory (IDT) [21], information manipulation theory [22], and reality monitoring [23]. Deception detection researchers rely on theory to identify strategically employed clues that can discriminate between those who deceive and those who do not. In this study, we rely on four factor theory and IDT because they best fit the interpersonal context of a 911 call.

Four factor theory [20] delineates four processes or factors that underlie deceivers’ behaviors. *Control*, the first process, describes how deceivers control or suppress their behavior to try to conceal their deception. For example, in 911 calls, deceivers will manage the linguistic and paralinguistic features of their interaction with the dispatcher in order to appear as truthful as possible and not to induce suspicion. The second factor, *arousal*, refers to various autonomic arousal responses of the deceiver’s central nervous system that coincide with the deceptive behavior or story. *Felt emotion*, the third factor, encompasses various emotions that deceivers experience, specifically guilt, anxiety, and/or satisfaction in pulling the wool over others’ eyes (i.e. “duping delight”). For instance, because of the negative feelings associated with guilt, deceivers try to disassociate themselves from their crime by referring to others rather than to the self through a greater use of through the use third-person pronouns. Anxiety may also impair the quality of the control that deceivers use to conceal their deception. Finally, due to the fourth factor, *cognitive processing*, deceivers have an increased cognitive load as they fabricate and maintain lies. This factor ties

into a proposition of IDT, specifically that high cognitive load may be detrimental to a liar's performance and increase the chances of detection. Because of this, deceivers in 911 calls may shorten their responses and use a smaller set of words.

According to IDT [21,24], deception is goal-oriented and strategic. Arising out of interpersonal communication and deception research, IDT predicts the behaviors of senders and receivers in an interactive context^b. The theory acknowledges the "superordinate role" of the context and relationship within which the interaction occurs. For example, the situational factors of a 911 homicide call will influence how deceptive exchanges play out, and consequently, the hypotheses regarding these exchanges. IDT proposes that the behavior of the deceiver will vary systematically with the spontaneity of the interaction (i.e., lying to a 911 operator requires more dexterity than lying in a written letter) and the immediacy of the context. In 911 calls, the interaction is spatially non-immediate, but temporally immediate. IDT also predicts that deceivers are strategic in managing the information they send, their image, and their behavior; as a result, however, they experience nonstrategic byproducts including dampened affect, noninvolvement, and performance decrements. In the current study, callers should experience increased cognitive loads relative to most contexts, making deception even more difficult. The increased stakes and spontaneity of the conversation may impair deceptive performance.

These theories lend support to the feasibility of using linguistic analysis to carry out deception detection. To discern these deceptive patterns in communication, many recent studies involving automated linguistic cue analyses have leveraged a general-purpose, psychosocial dictionary such as LIWC [5,8-10,25-27]. LIWC contains predefined categories composed of words related to a particular construct, such as Anxiety or Negative Emotion. Depending on the context in which deception occurs, deceivers have been found to display elevated uncertainty and affect, share fewer details, provide more spatiotemporal details, and use less diverse and less complex language than truth tellers [17,26,28]. Researchers have also documented cases in which deceivers use more words, group references and use more informal, non-immediate language than truth tellers [6,9,29].

Researchers who have conducted manual content analyses have also documented linguistic markers of deception [30]; for example, perpetrators of homicide or kidnapping may use past tense (vs. present or future) when discussing the victim, and deceivers may try to distance themselves from the crime by using "they" rather than "I" in statements [4]. Law enforcement researchers [3,4] were successful in manually coding verbal indicators to classify 911 homicide callers as "guilty" or

"innocent." In these studies, the key verbal indicators of guilt included extraneous information, inappropriate politeness, a lack of plea for help, and evasion. In an earlier study, Olsson [31] analyzed emergency calls made to report fires in London, UK. He found that hoax calls could be discriminated from truthful ones based on how the caller described or implied his/her relationship to the emergency and the urgency and cooperation of the caller.

The original mode of communication (i.e., written vs. spoken language) is particularly important with respect to linguistic cues that distinguish between truth tellers and deceivers. As noted in IDT, deception is strategic and the spontaneity of the interaction is important [21,24]. Thus, for example, a team tasked with writing the text to include with financial statements has months in which to develop a document that may include strategic misrepresentation or obfuscation via increasing word count and employing words of more than three syllables [14,15]. On the other hand, a person calling a 911 operator after committing a crime has far less time for strategic wordsmithery and may use fewer, simpler words to suppress or control verbal cues for deception.

Research question and hypotheses

Contrary to media hype regarding highly publicized crimes, most 911 calls are not full of drama, excitement, and/or rich descriptions of the crime, crime scene, or victim(s). The 911 operator is trained to elicit information to deploy the right emergency resources as quickly as possible. Although a dramatic outpouring of emotion may be observed in some calls, most 911 calls comprised a great deal of mundane information gathering and information passing to clarify names, addresses, and directions; give life-saving instructions (such as the steps to perform cardiopulmonary resuscitation (CPR)); and ask questions that only require a one-word answer (e.g., "is the victim breathing?").

There are several key advantages to using automated linguistic analysis techniques to detect deception in transcripts of 911 homicide calls. First, 911 calls represent the initial contact between a caller and an emergency response team, including law enforcement, leaving callers little time to prepare or to settle down for the encounter. Moreover, as noted above, the caller's statement has not yet been corrupted by contact with criminal investigators or lawyers [4]. Furthermore, because callers do not perceive 911 operators to be members of law enforcement, deceptive callers may exhibit less controlled behavior and more cues of deception; consequently, 911 callers may become more engaged in interpersonal communication and be less guarded because 911 operators interview rather than interrogate. Due to the temporal immediacy of the crime in relation to the 911 call, there

may also be more active stress on the caller, in turn causing the caller to display more deceptive cues. Another advantage for comparison of truthful and deceptive responses in emergency calls is that 911 operators use a structured interview style that was similar across calls.

On the other hand, there are a number of issues with 911 calls that make them a less than ideal source for linguistic analysis. First, these calls can be very short. In the dataset used for this study, calls ranged from less than thirty seconds to over ten minutes in length. On average, the calls were about three to four minutes. Another problem is that there can be a lot of dead air time while the 911 operator puts the caller on hold to coordinate with the rest of the emergency team; such gaps cannot be used in linguistic analyses. Third the sound quality of the calls can make them difficult to transcribe. Fourth, dispatchers, operators, callers, and rescue workers may talk simultaneously. Fifth, certain sounds such as sobs, shrieks, or gasps of pain cannot be easily transcribed and/or analyzed using current linguistic feature dictionaries. Finally, a caller may be poised to offer what could prove to be valuable clues to future investigators, but 911 operators may have to interrupt them to elicit the best information for timely dispatching of the right resources. Consequently, to accomplish their primary mission, the operators who perform both call-taking and dispatching functions may restrict “telling” cues or information [32].

The strengths of the dataset ultimately outweigh the potential weaknesses of the dataset. Thus, we analyzed cues in transcribed 911 calls using the same approach that other researchers have adopted to analyze unstructured texts with a view to distinguish between deceivers and truth tellers. Based on IDT, four factor theory, and previous research, we considered that deceivers may try too hard to cover up what they perceive to be deceptive cues. For example, in a face-to-face context, deceivers may try to limit fidgeting and/or posture shifts, thereby displaying more “stillness” during a lie in an attempt to inhibit overt signs of deception [28]. In the circumstances of this study, where callers are not visible, deceptive 911 callers may restrict their verbal responses, exhibiting low affect or shortened responses when compared to truth tellers. This may be observed in a higher rate of negation and assent words (e.g., “no” and “yes”) that may be used by deceivers to limit and control answers. Deceivers may also feel emotions such as guilt, anxiety, and/or duping delight [33]. To lessen these feelings, they may attempt to distance themselves from the situation. The use of first-person singular constructs implies that the speaker “owns” the statement. However, because liars try to distance themselves [34] from the crime or bad situation, they include fewer self-references in retelling stories [17]. Thus, we should find that

deceptive 911 callers use more third-person plural to share the blame.

Four factor theory posits that cognitive effort is required to not only lie, but also to maintain the lie. Vrij, Fisher, Mann, and Leal [35] expand upon this claim and outline all of the tasks a deceiver undertakes that increase cognitive load, including developing a plausible lie, self-monitoring for credibility, monitoring the listener, and remembering the details of the lie while concealing the truth. In short, they argue that lying requires strategic intent and more cognitive effort than truth telling. To mitigate these effects, a liar may rehearse his or her story to keep the facts and details in order. However, because deceptive 911 callers have often had little opportunity to rehearse between the time of the crime and the time of the call, they may face extreme cognitive overload. Thus, these callers may compensate by supplying shorter, controlled statements, sharing fewer details with emergency responders (such as those that would be helpful in locating the physical address of the victim), and asking the 911 operators to “hold on” or “wait” when the operator gives instructions.

Based on these theories and building on previous research on deception detection using linguistic cue analysis, our research question is as follows:

Can automated linguistic analysis techniques accurately classify deceptive versus truthful callers in transcripts of 911 homicide calls?

To define the types of cues that can be examined for deception or truthfulness in 911 calls, we suggest that truth tellers will exhibit more *immediacy* through greater use of first-person singular and first-person plural words. On the other hand, we expect deceivers will show more *non-immediacy*, a distancing from what is said, by referencing others in the third-person singular or plural. To *control* verbal output, or to suppress reactions or answers, deceivers will tend to answer more frequently with shorter, simpler “yes” or “no” answers. Because deceivers tend to suppress reactions, we expect that truth tellers will display more *felt emotion*. On the other hand, we anticipate that deceivers will use more swear words because instances of swearing can be perceived as more credible [36]. Thus, deceivers may include swear words as a way to appear to be emotionally connected to an incident while suppressing their true emotions. Relative to deceivers, truth tellers have a lower *cognitive overload* that allows them to give more location details, such as house numbers and generic information about location. Truth tellers want to provide many clues to get rescue teams to their location as quickly as possible, and will therefore provide specific addresses and phone numbers more clearly, as well as giving more

details about the location, such as whether it is a house or apartment building. Finally, due to cognitive overload deceivers may be reluctant, or find it difficult to follow life-saving instructions given by the dispatcher even though not doing so would seem suspicious.

Formally stated, the hypotheses in this study are as follows:

Deceptive 911 callers will display:

- (a) higher use of third-person plural, (b) higher use of third-person singular, (c) more assent terms, (d) more negation terms, (e) more emotionally-charged swearing, (f) more inhibition words, and
- (g) lower use of first-person plural, (h) lower use of first-person singular, (i) less negative emotion, (j) less anxiety, (k) lower use of numbers, and (l) lower use of generic location details than truthful 911 callers.

Methodology

Our dataset represented a convenience sample obtained from publicly available sources found on the Internet. The majority of the calls came from Dispatch Magazine On-Line (911dispatch.com). This resource contains a tape library of public domain 911 and other emergency calls that have been collected since 2006. The website contains documentation about the calls that enables the user to determine the outcome of the case.

Because we do not have control over the chain of custody of these 911 calls, we cannot state to what degree they were edited. For the most part, personally identifying information has been redacted, but we have no reason to believe that the calls were edited otherwise. However, the inability to state this conclusively is a potential limitation of this convenience dataset. Still, the archive presents a unique opportunity to access this type of real-world data.

The final dataset of 50 transcribed 911 calls was equally split between truthful and deceptive callers. The size of the dataset was restricted by the number of publicly available deceptive calls for which ground truth could be established. To establish ground truth, we corroborated subsequent arraignment, prosecution, and/or admission of guilt via news articles about the crimes based on the information at the source website and other websites as necessary. Once we had identified 25 deceptive calls, we randomly chose 25 calls from our set of transcribed truthful calls to create a matched set. After transcribing the calls and removing the 911 operators' side of the conversations, we analyzed the caller side of the transcripts using LIWC. LIWC normalizes the data by dividing category counts by the number of words in each document.

As summarized in Table 1, the various constructs that we examined comprised one to several linguistic cues as defined in LIWC. The *immediacy* of truth tellers was

Table 1 Constructs and corresponding LIWC categories

Constructs	LIWC categories
Immediacy	1st person plural, 1st person singular
Non-immediacy	3rd person singular; 3rd person plural
Control	Assent; Negate
Felt Emotion	Negative emotion; Anxiety
Lack of felt emotion	Extreme swearing
Cognitive overload	Inhibition
Lack of cognitive overload	Numbers; Leisure (location-related)

anticipated to be observable in greater use of First-Person Plural and First-Person Singular terms (LIWC categories). Conversely, we expected deceivers to show more *non-immediacy* by referencing others in the Third-Person Singular or Third-Person Plural (LIWC categories). Deceivers would also tend to answer with more Negation terms (LIWC category) or more Assent terms (LIWC category) answers as part of *control*. We expected that truth tellers would display more *felt emotion* via the LIWC categories of Negative Emotion and Anxiety than deceivers. In contrast, we anticipated that deceivers, who would attempt to fake an emotional connection while suppressing their true emotions, would use more swear words (LIWC category = Sexual (includes emotionally-charged swear words)). Also due to *cognitive overload*, deceivers might be reluctant to initiate CPR or other first aid efforts (LIWC category = Inhibition). Meanwhile, the lack of *cognitive overload* would enable truth tellers to give more location details such as house numbers (LIWC category = Number) or location-related words such as “garage” or “apartment” (the LIWC category that includes these location words = Leisure).

The first step for identifying significant linguistic cues was to run a one-tailed independent sample t-test for each linguistic cue to compare truth tellers with deceivers. We considered each 911 call transcript to be an independent observation, since it represented one call placed by a unique caller.

Next, we trained various machine-learning classification algorithms on the cues and tested their classification accuracy using 10-fold cross-validation as a bootstrap technique to increase the validity of the results. The following machine-learning algorithms were selected to classify the 911 calls because of their theoretically diverse foundations: logistic model tree induction, naïve Bayes, neural network, and random forest. Classification by one statistical technique, discriminant analysis, was also performed. Table 2 reports the results for each classification method.

Each machine-learning algorithm builds a model based on a different set of theoretical premises. Logistic tree

Table 2 Results of Classification Algorithms

Classification methods	Logistic model tree induction		Naïve Bayes		Random forest		Neural network		Discriminant analysis	
	Training	Cross-valid	Training	Cross-valid	Training	Cross-valid	Training	Cross-valid	Training	Cross-valid
Overall performance	98%	82%	90%	78%	100%	70%	82%	74%	96%	84%
Truth performance	100%	82%	88%	76%	100%	64%	84%	64%	96%	88%
Deception performance	96%	84%	92%	80%	100%	76%	80%	84%	96%	80%

model induction is a classifier for building logit models using regression functions as base learners [37]. A simple naïve Bayes is a probabilistic classifier based on Bayes’ theorem. A neural network is a “black box” that performs a classification using hidden layers. A random forest is a type of decision tree that applies decision rules to divide an overall dataset into smaller classification sets. Using these theoretically diverse algorithms, we reduce the likelihood of relying on the results of one algorithm that over-learns the data and fails to generalize to a broader population.

Results and discussion

Table 3 includes the original hypothesis for each construct and associated variable (LIWC category), as well as the results from each one-tailed independent samples t-test. In this way, it gives a comparison of linguistic differences in the deceptive and truthful conditions.

On average, compared to truthful calls, deceptive 911 calls exhibited greater use of “they” ($t_{(50)} = 1.802, p < .05$). They also involved more negation ($t_{(50)} = 2.031, p < .05$) and assent terms ($t_{(50)} = 1.905, p < .05$). Furthermore, they displayed a higher rate of inhibition terms ($t_{(50)} = 2.428, p < .05$). In contrast, truthful callers used more numeric words ($t_{(50)} = -2.417, p < .05$) and leisure (location-related) words ($t_{(50)} = -2.109, p < .05$). The transcripts for truthful calls contained more negative emotion words ($t_{(50)} = -1.915, p < .05$), terms for anxiety ($t_{(50)} = -1.975, p < .05$), and leisure (location-related) terms ($t_{(50)} = -2.109, p < .05$). Table 2 reports the results of the classification algorithms for both a training set and a cross-validation set.

The overall performance of the classification techniques was very strong and ranged from 70% to 84% for the cross-validation tests. The results yielded predictive models with much higher accuracy than that of unaided humans, which, as mentioned above, is 54% [1,2]. The best classification method used was discriminant analysis, followed by logistic model tree induction.

Table 4 lists each variable name (LIWC category) with examples from the transcripts that conform to the results. In part, the accurate classification performance of this study may be due to high motivation exhibited by the callers. DePaulo et al. [28] point out that cues to

deception are more evident when individuals are striving to carry out deception successfully.

The results suggest that truthful callers display more negative emotion and anxiety than deceivers, who tend to display flat affect. Although we had hypothesized that deceivers would use more swear words as an attempt to appear more credible through a faked emotional response, we actually discovered that truth tellers used more extreme swearing (the mean for truthful swearing = .0652; the mean for deceptive swearing = .0000). Emotionally charged swearing was another way for truth tellers to convey negative emotion or frustration during the calls. This finding corresponds to previous research that demonstrated that the primary reason that people swear is to express negative emotions or frustration [38,39].

Honest callers also tended to refer to others in third-person singular. To aid emergency responders, truth tellers used more numbers related to addresses and/or phone numbers and used names of locations, such as “apartment” or “garage”.

Deceivers used third-person plural at a higher rate, perhaps to distance themselves from an incriminating situation. However, contrary to our hypotheses, they also demonstrated more immediacy than truth tellers by using both first-person singular (the mean for truthful first-person singular = .2964; the mean for deceptive first-person singular = 1.1760) and first-person plural pronouns (the mean for truthful first-person plural = 9.4136; the mean for deceptive first-person plural = 10.6788).

Deceivers’ use of negation and agreement words may have represented their need to suppress or contain their own verbal responses and/or affect. Finally, deceivers tended to tell the 911 operator to “wait” or “hold [on]” (inhibition terms) at a higher rate than truth tellers. This occurred in the 911 calls when the operator asked them to do something they were reluctant to do, such as CPR.

The aim of this research was to expand the understanding of how we can analyze 911 call transcripts for crime analysis and solving. Thus, this study makes three major contributions. First, it has advanced deception detection research by applying linguistic feature mining methods to a unique corpus, namely, transcripts of 911 homicide calls, which represents extremely raw and

Table 3 Analysis of constructs/variables (LIWC categories) in transcripts

Constructs	Associated variables (LIWC categories)	Predicted	Actual	Truthful mean	Truthful Std Dev	Deceptive mean	Deceptive Std Dev
<i>Immediacy</i>	1st person plural	T > D	D > T	.2964	.50586	1.1760	2.15572
	1st person singular	T > D	D > T	9.4136	3.42226	10.6788	4.67963
<i>Non-immediacy</i>	3rd person singular	D > T	T > D*	5.4756	3.84711	3.7344	3.06803
	3rd person plural	D > T	D > T*	.6280	1.16957	1.2652	1.32602
<i>Control</i>	Negation	D > T	D > T*	3.7160	2.18237	4.9892	2.25060
	Assent	D > T	D > T*	4.8160	3.52826	6.9232	4.25887
<i>Felt emotion</i>	Negative emotion	T > D	T > D*	1.5980	1.48383	.8736	1.17341
	Anxiety	T > D	T > D*	.3632	.63672	.0904	.26723
	Extreme swearing (Sexual)	D > T	T > D	.0652	.15075	.0000	.0000
<i>Cognitive overload</i>	Number	T > D	T > D*	5.5628	5.71633	2.5760	2.34458
	Leisure	T > D	T > D*	.7564	.65370	.3560	.68811
	Inhibition	D > T	D > T*	.1748	.27091	.4648	.53210

*Significant at p-value <= 0.05.

largely unrehearsed human communication. We determined that deceivers use language and linguistic cues differently than truth tellers in a high-stakes, real-world corpus of 911 homicide calls. Thus, this study represents a successful proof of concept of using automated linguistic analysis to classify deceptive versus truth telling 911 homicide calls accurately, quickly, and objectively in comparison to manual methods of content analysis, which involve extensive training, time-consuming analyses, and subjectivity. Through analysis of 911 calls, law enforcement could detect deceptive, guilty perpetrators earlier in the investigative process and use that information for crime analysis. Second, we extracted useful information from unstructured text comprising transcribed 911 calls to demonstrate that these largely unexploited data can be analyzed for further investigative work. Third, this study provided strong classification results of up to 84%

accuracy (cross-validated). These results approach the highest reported accuracy of field-based polygraph tests which is 92% [40,41]. Combined with speech-to-text software, automated deception detection techniques could be used to monitor 911 calls in real time. Although the mission of the 911 operators would remain the same, automated monitoring of 911 calls could enable law enforcement and rescue workers to focus efforts and resources more successfully on post-hoc crime solving and analysis.

Limitations and conclusion

Despite these contributions discussed above, the convenience sample of 50 archived 911 homicide calls downloaded from the Internet represents a limited dataset. The size of the current dataset was restricted by the number of publicly available deceptive calls for which

Table 4 Variables (LIWC Categories) with Examples from Transcripts

Variable Name	Direction	Truthful Transcript	Deceptive Transcript
3rd person singular	T > D	She's right on the floor. She's not breathing.	
3rd person plural	D > T		Yes, they said, they said if they heard anything they were going to my house.
Negation	D > T		No, nothing , he's gone.
Assent	D > T		Okay, they're here.
Negative emotion	T > D	There was a fight . It was terrible .	
Anxiety	T > D	I found out about an hour ago and I've been in a panic ever since.	
Number	T > D	Five seventeen West Doty Street	
Leisure	T > D	I see her in her garage right now.	
Inhibition	D > T		[conversation while the operator is trying to give CPR instructions] Hold on, I have to throw up, please hold on.

Note: Bold-faced type in transcript columns indicates words associated with each respective variable.

ground truth could be established. Furthermore, the 911 calls were taken from the Internet, not the original source. As a result, some information—such as names—may have been blocked out to protect the privacy of the caller when the 911 call was released to the public. An additional limitation arises from the fact that we did not control the chain of custody over the 911 calls. Thus, we are not able to establish whether or to what degree the 911 calls had been edited. Moreover, only certain states currently release 911 calls to the public, so the geographic origin of the calls could be skewed. Finally, the calls were restricted to those placed by English-speaking callers located in the U.S. To counter these limitations in the future, audio files and/or transcripts should be collected directly from law enforcement in subsequent research to validate our results, and sampling from diverse geographical locations should be carried out.

Accurate, credible assessment decisions are critical for law enforcement personnel, who may not detect deception, may act on false positives, or may use incorrect information for crime analyses. Moreover, other clues for deception, such as vocalic cues, should be studied in conjunction with linguistic cues for a practical decision support tool that includes combined analyses. In the future, an integrated system for deception detection that can be used in real time in high-stakes 911 calls could be added to law enforcement's overall crime-solving and crime-analysis strategy. Text mining tools could also be used to analyze 911 calls linguistically as part of a portfolio of crime-solving techniques to enable law enforcement and rescue workers to focus their efforts and resources more successfully.

This study's findings provide critical knowledge about how deceivers communicate during typically unrehearsed verbal exchanges and expand the usefulness of deception models from a low-stakes, laboratory setting into a high-stakes, real-world environment where there are serious consequences not only for the deceiver, but also for law enforcement. The next step for this research will be to validate these results using a larger real-world dataset. Additionally, when we use a larger dataset, we can conduct testing of revised hypotheses for 1st person plural ($D > T$), 1st person singular ($D > T$), and swearing ($T > D$) to establish if new results are significant in the revised directions. Current and future quantitative models and decision aids could assist law enforcement in detecting deception.

Endnotes

^aIn this paper, the term "homicide" includes not only murders, but also accidents that were later deemed by investigators to be homicides, reports of kidnapping that masked an underlying homicide or neglectful death, and murder-suicides.

^bA unique characteristic of this dataset that may not be true in others is that 911 deceivers are treated implicitly as truthful by the dispatcher. A 911 dispatcher operates under the assumption that the caller is telling the truth. Therefore, a deceptive caller may never change his story based on his perception that the operator is suspicious of it.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

MB conceived of the study, participated in its design, obtained funding for the project, gathered the audio data files and transcribed the files into text, performed the statistical analysis and interpretation of the data, and drafted the manuscript. KM participated in the design of the study, participated in obtaining funding for the project, performed the parsing of the data into linguistic cue categories, participated in the overall analysis and interpretation of the data, and helped to draft the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

We are pleased to acknowledge the generous support from a Center for Identification Technology Research (CITeR) Grant for funding the data collection and analysis phases of this project.

Author details

¹Jake Jabs College of Business & Entrepreneurship, Montana State University, P.O. Box 173040, 59717-3040 Bozeman, MT, USA. ²Accounting and Information Systems, Rutgers Business School, Rutgers, The State University of New Jersey, 100 Rockefeller Road, 08854 Piscataway, NJ, USA.

Received: 1 August 2013 Accepted: 7 July 2014

Published online: 12 August 2014

References

1. CF Bond, BM DePaulo, Accuracy of deception judgments. *Personal. Soc. Psychol. Rev.* **10**, 214–234 (2006)
2. M Aamodt, H Custer, Who can best catch a liar? *Forensic Examiner* **15**, 6–11 (2006)
3. SH Adams, T Harpster, 911 homicide calls and statement analysis. *FBI Law Enforce. Bull.* **77**, 22–31 (2008)
4. T Harpster, SH Adams, JP Jarvis, Analyzing 911 homicide calls for indicators of guilt or innocence. *Homicide Stud.* **13**, 69–93 (2009)
5. L Zhou, JK Burgoon, DP Twitchell, T Qin, JF Nunamaker Jr, A comparison of classification methods for predicting deception in computer-mediated communication. *J. Manag. Inf. Syst.* **20**, 139–165 (2004)
6. L Zhou, JK Burgoon, JF Nunamaker Jr, DP Twitchell, Automating linguistics based cues for detecting deception in text based asynchronous computer mediated communication: an empirical investigation. *Group Decis. Negotiation* **13**, 81–106 (2004)
7. S Mann, A Vrij, R Bull, Suspects, lies, and videotape: an analysis of authentic high-stake liars. *Law Hum. Behav.* **26**, 365–376 (2002)
8. CM Fuller, DP Biros, M Adkins, JK Burgoon, JF Nunamaker Jr, S Coulon, Detecting Deception in Person-Of-Interest Statements, in *Lecture Notes in Computer Science*, vol. 3975 (Springer, Berlin/Heidelberg, 2006), pp. 504–509
9. CM Fuller, DP Biros, D Delen, Exploration of Feature Selection and Advanced Classification Models for High-Stakes Deception Detection, in *Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS)* (Waikoloa, Big Island, HI, 2008)
10. CM Fuller, DP Biros, RL Wilson, Decision support for determining veracity via linguistic-based cues. *Decis. Support. Syst.* **46**, 695–703 (2008)
11. ML Jensen, E Bessarabova, B Adame, JK Burgoon, SM Slowik, Deceptive language by innocent and guilty criminal suspects: the influence of dominance, question, and guilt on interview responses. *J. Lang. Soc. Psychol.* **30**, 357–375 (2011)
12. VA Sandoval, Strategies to avoid interview contamination. *FBI Law Enforce. Bull.* **72**, 1–12 (2003)

13. A Nijholt, RC Arkin, S Brault, R Kulpa, F Multon, B Bideau, DR Traum, H Hung, E Santos Jr, D Li et al., Trends and controversies. *IEEE Intell. Syst.* **27**, 60–75 (2012)
14. S Humpherys, KC Moffitt, MB Burns, JK Burgoon, WF Felix, Identification of fraudulent financial statements using linguistic credibility analysis. *Decis. Support. Syst.* **50**, 585–594 (2011)
15. KC Moffitt, MB Burns, *What does that mean?* Investigating Obfuscation and Readability Cues as Indicators of Deception in Fraudulent Financial Reports, in *Fifteenth Americas Conference on Information Systems* (San Francisco, CA, 2009), p. 2009
16. JT Hancock, LE Curry, S Goorha, M Woodworth, On lying and being lied to: a linguistic analysis of deception in computer-mediated communication. *Discourse Process.* **45**, 1–23 (2008)
17. ML Newman, JW Pennebaker, DS Berry, JM Richards, Lying words: predicting deception from linguistic styles. *Personal. Soc. Psychol. Bull.* **29**, 665–675 (2003)
18. JW Pennebaker, ME Francis, RJ Booth, *Linguistic Inquiry and Word Count* (Erlbaum Publishers, Mahway, NJ, 2001)
19. YR Tausczik, JW Pennebaker, The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* **29**, 24–54 (2010)
20. M Zuckerman, BM DePaulo, R Rosenthal, Verbal and Nonverbal Communication of Deception, in *Advances in Experimental Social Psychology*, ed. by L Berkowitz, vol. 14 (Academic, New York, NY, 1981), p. 60
21. DB Buller, JK Burgoon, Interpersonal deception theory. *Commun. Theory* **6**, 203–242 (1996)
22. SA McCormack, Information manipulation theory. *Commun. Monogr.* **59**, 1–16 (1992)
23. M Johnson, C Raye, Reality monitoring. *Psychol. Rev.* **88**, 67–85 (1981)
24. JK Burgoon, DB Buller, Interpersonal deception: III: effects of deceit on perceived communication and nonverbal behavior dynamics. *J. Nonverbal Behav.* **18**, 155–184 (1994)
25. CM Fuller, DP Biros, JK Burgoon, M Adkins, DP Twitchell: An analysis of text-based deception detection tools. In *AMCIS 2006 Proceedings*. 2006: Paper 418
26. T Qin, JK Burgoon, JP Blair, JF Nunamaker Jr, Modality Effects in Deception Detection and Applications in Automatic Deception Detection, in *Proceedings of the 38th Hawaii International Conference on Systems Sciences (HICSS)*, 2005
27. L Zhou, DP Twitchell, T Qin, JK Burgoon, JF Nunamaker Jr, An Exploratory Study into Deception Detection in Text-Based Computer-Mediated Communication, in *Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS '03)*, 2003
28. BM DePaulo, JJ Lindsay, BE Malone, L Muhlenbruck, K Charlton, H Cooper, Cues to deception. *Psychol. Bull.* **129**, 74–118 (2003)
29. L Zhou, JK Burgoon, DP Twitchell, A Longitudinal Analysis of Language Behavior of Deception in E-Mail, in *Intelligence and Security Informatics: First NSF/NIJ Symposium, ISI 2003, Tucson, AZ, USA, June 2-3, 2003 Proceedings*, ed. by H Chen, R Moore, D Zeng, J Leavitt, vol. 2665/2003 (Springer, Berlin / Heidelberg, 2003), pp. 102–110. *Lecture Notes in Computer Science*
30. LN Driscoll: A validity assessment of written statements from suspects in criminal investigations using the scan technique. *Police Stud: Int. Rev. Police Dev.* **17**, 77–78 (1994)
31. J Olsson, *Forensic Linguistics: An Introduction to Language, Crime, and Law* (Continuum International Publishing Group, London, 2004)
32. SJ Tracy, When questioning turns to face threat: an interactional sensitivity in 911 call-taking. *West. J. Commun.* **66**, 129–157 (2002)
33. P Ekman, Mistakes when deceiving. *Ann. N. Y. Acad. Sci.* **364**, 269–278 (1980)
34. A Vrij, *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd edn. (Wiley, Chichester, West Sussex, England, 2008)
35. A Vrij, R Fisher, S Mann, S Leal, A cognitive load approach to lie detection. *J. Investig. Psychol. Offender Profiling* **5**, 39–43 (2008)
36. E Rassin, S van der Heijden, Appearing credible? Swearing helps! *Psychol. Crime Law* **11**, 177–182 (2005)
37. M Sumner, E Frank, M Hall, Speeding up Logistical Model Tree Induction, in *9th European Conference on Principles and Practice of Knowledge Discovery in Databases*, 2005, pp. 675–683
38. E Rassin, P Muris, Why do women swear? An explanation of reasons for and perceived efficacy of swearing in Dutch female students. *Personal. Individ. Differ.* **38**, 1669–1674 (2005)
39. T Jay, K Janschewitz, The pragmatics of swearing. *J. Politeness Res. Lang. Beh. Cult.* **4**, 267–288 (2008)
40. PE Crewson, *Comparative analysis of polygraph with other screening and diagnostic tools*. No. DODPI01-R-0003 (Research Support Service, Ashburn, VA, 2001)
41. CR Honts, DC Raskin, A field study of the validity of the directed lie control question. *J. Police Sci. Adm.* **16**, 56–61 (1988)

doi:10.1186/s13388-014-0008-2

Cite this article as: Burns and Moffitt: Automated deception detection of 911 call transcripts. *Security Informatics* 2014 **3**:8.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com