Which Spoken Language Markers Identify Deception in High-Stakes Settings? Evidence From Earnings Conference Calls

Judee Burgoon[1], William J. Mayew[2], Justin Scott Giboney[3], Aaron C. Elkins[1], Kevin Moffitt[4], Bradley Dorn[1], Michael Byrd[1], and Lee Spitzley[1]

Abstract

Quarterly conference calls where corporate executives discuss earnings that are later found to be misreported offer an excellent test bed for determining if automated linguistic and vocalic analysis tools can identify potentially fraudulent utterances in prepared versus unscripted remarks. Earnings conference calls from one company that restated their financial reports and were accused of making misleading statements were annotated as restatement-relevant (or not) and as prepared (presentation) or unprepared (Q&A) responses. We submitted more than 1,000 utterances to automated analysis to identify distinct linguistic and vocalic features that characterize various types of utterances. Restatement-related utterances differed significantly on many vocal and linguistic dimensions. These results support the value of language and vocal features in identifying potentially fraudulent utterances and suggest important interplay between utterances that are unscripted responses rather than rehearsed statements.

[1]University of Arizona, Tucson, AZ, USA
[2]Duke University, Durham, NC, USA
[3]University at Albany, State University of New York, Albany, NY, USA
[4]Rutgers, The State University of New Jersey, Newark, NJ, USA

Corresponding Author:
Judee Burgoon, Eller College of Management, University of Arizona, McClelland Hall 418, Tucson, AZ 85721, USA.
Email: judee@email.arizona.edu

Recent technological advances have enabled the creation of tools for automating analysis of spoken language to detect deception and other forms of misrepresentation (Moffitt, Giboney, Ehrhardt, Burgoon, & Nunamaker, 2012; Tausczik & Pennebaker, 2010). From spoken language, researchers can now systematically assess whether various word patterns or vocal patterns provide meaningful markers, or "tells," of deception. In this article, we investigate whether, in high-stakes settings, strategic spoken language tells are differentially indicative of deception relative to nonstrategic spoken language tells.

We focus on high-stakes settings (i.e., ones that entail serious harm to targets of deceit and adverse consequences to deceivers if caught) for three reasons. First, high-stakes settings have high significance for those involved, where substantial economic and reputation loss can occur. Second, many studies test deception in a laboratory where stakes cannot be raised to high levels and hence, despite the importance of high-stakes settings, evidence in the literature related to such contexts is sparse (for exceptions, see Burns & Moffitt, 2014; Fuller, Biros, Burgoon, & Nunamaker, 2013; ten Brinke & Porter, 2012). Third, and most important, in high-stakes settings, it is unclear whether strategic or nonstrategic spoken language will better identify deception. On the one hand, in high-stakes settings deception comes with a high cost, which arguably should heighten negative emotions related to fear of detection or guilt and increased cognitive demands while prevaricating. Linguistic and vocal patterns resulting from such emotional and cognitive burdens we label as nonstrategic, as they would reflect inadvertent and uncontrolled reactions to the act of deception.

On the other hand, undertaking high-stakes deception is itself a choice and may only be undertaken by individuals who believe they can avoid detection, or have mastered strategies to in fact do so. Such individuals may, for example, prepare extensively before speaking to lower the cognitive burden that can accompany deception, or may undergo voice training in an attempt to sound vocally like the antithesis of someone engaging in deception. We label linguistic and vocal patterns from spoken language strategic if they reflect such ex ante planning and investment to avoid deception detection.

To empirically examine spoken language indicators of high-stakes deception, we exploit the earnings conference call setting. Earnings conference calls are ideal for our examination for four reasons. First, earnings conference calls are regularly scheduled quarterly public broadcasts whereby corporate executives speak about the financial performance of their firms. At over an hour in length on average, earnings conference calls provide a rich and expansive set of ecologically valid linguistic and vocalic data for the study of deception (Mayew & Venkatachalam, 2012).

Second, the earnings conference call setting uniquely allows for the identification of instances of both deception and high-stakes. Subsequent to the earnings conference call, researchers can observe whether the financial performance being discussed was later found to be deceptive and can identify the related economic consequences (Hobson, Mayew, & Venkatachalam, 2012; Larcker & Zakolyukina, 2012). In our empirical analysis, we study over 1,000 utterances from a chief executive officer (CEO) and chief financial officer (CFO) of a firm where financial reports were

subsequently revealed to be overstated, and where the CEO and CFO were replaced amid being named in multiple securities class action lawsuits that settled in favor of the plaintiffs.

Third, recent research by Braun, Van Swol, and Vang (2015) exploits variation in scripted and unscripted communication to identify deception among politicians. Earnings conference calls also have such exploitable variation, as they begin with a presentation portion followed by a question and answer (Q&A) session. Relative to the Q&A session, the presentation is more rehearsed and scripted. Relative to the presentation, the Q&A is more similar to a live interrogation by financial analysts, which arguably should intensify the emotion profile of the speaker. Together, these prepared and unprepared segments of the call provide natural settings where spoken language is potentially more likely to be strategic and nonstrategic (Bloomfield, 2008), respectively, which allows for a comparative analysis of the role of strategic and nonstrategic spoken language in detecting deception.

Finally, identifying the set of telltale signs of deception that might be derived from a public speech corpus of conference calls is important to regulators interested in the early detection of corporate fraud. In 2011, the Federal Bureau of Investigation investigated 726 cases of corporate fraud, with some cases involving more than one billion dollars in losses to public investors (Federal Bureau of Investigation, 2011). Given the frequent occurrence and potential economic damage of corporate fraud, US securities regulators are now considering increasing their focus on identifying financial fraud by using new tools, including automated linguistic analysis (Eaglesham, 2013).

Literature Review and Hypothesis Development

The Role of Preparation on Strategic and Nonstrategic Behavior

To develop predictions about how strategic and nonstrategic aspects of spoken language might differentially identify deception in the high-stakes earnings conference call setting, we begin with a brief overview of relevant theoretical perspectives on deceptive communication. As expressed in Ekman and Friesen's (1969) leakage hypothesis, the act of deception is thought to prompt feelings of guilt, stress, and fear of detection that lead to outward behavioral displays from which a person's true state or deceptive intent can be "read." Zuckerman, DePaulo, and Rosenthal (1981) formalized etiologies of such leakage, including arousal, negative affect, and cognitive load, with much of the subsequent literature focusing on observable physical and behavioral markers (e.g., sweating, blinking). More recently, attention has turned to cognitive load, with investigations centering on cognitive difficulties and effects on working memory associated when constructing and telling lies (e.g., Sporer & Schwandt, 2006). All the aforementioned research focuses on uncontrolled and unmonitored aspects of behavior, which Buller and Burgoon (1994) labeled collectively as nonstrategic behavior. The nonstrategic label was utilized to distinguish inadvertent signals of deceit from more premeditated, deliberate, and voluntary communication that they labeled as strategic behavior.

Regarding strategic behavior, researchers have forwarded numerous theories, including information manipulation theory (IMT and IMT2; McCornack, 1992; McCornack, Morrison, Paik, Wisner, & Zhu, 2014) and interpersonal deception theory (Burgoon & Buller, 2015). Cognitions or actions that are deliberately performed during communication represent strategic behavior (Buller & Burgoon, 1994, 1996). An example of such behavior is the use of words and phrases that convey credibility in an attempt to convince the receiver that the deceiver is truthful. IMT centers on the ways in which deceivers intentionally manipulate the contents of their communication to evade detection. Interpersonal deception theory characterizes deceit as a strategic behavior that produces compound displays of actions directed toward a variety of communication goals. While deception experts do not always agree on what indicators are effective signals of deception (Fiedler & Walka, 1993; McCornack, 1997; McCornack et al., 2014), there is ample evidence from experiments, case studies, and meta-analyses to indicate that deceivers manage their communication deliberately and that both nonstrategic and strategic vocal and linguistic cues may be effective measures of deception (Buller & Burgoon, 1994; Grazioli, Johnson, & Jamal, 2006; Hauch, Blandón-Gitlin, Masip, & Sporer, 2014; Juslin & Scherer, 2005; C.-C. Lee, Welker, & Odom, 2009). Collectively, the literature makes it clear that to understand deception markers in spoken language, researchers must consider both nonstrategic and strategic forces.

To gain an understanding of how strategic and nonstrategic spoken language markers are associated with deception in our earnings conference call setting, we begin by first considering whether spoken language features naturally vary between the presentation portion and the question and answer portion of calls. Earnings conference calls begin with remarks prepared in advance and often read from cue cards by company officials—usually the CEO and CFO. This is followed by a less formal Q&A phase in which analysts pose questions to the management team. We refer to this latter portion of the call as unprepared remarks.[1]

Prior research has shown that there is valuable information contained in both portions of these calls from the point of view of capital market participants listening to the dialog; however, generally more information is gleaned from the Q&A. The relative informational potency of the Q&A likely stems from the ability of analysts to ask particular questions that assist in the comprehension of the firm's financial prospects (Matsumoto, Pronk, & Roelofsen, 2011). In addition, the live interrogation by financial analysts can arouse management and elicit emotions that contain more information (Mayew & Venkatachalam, 2012).

Given these institutional features of earnings conference calls, one might reasonably be concerned that spoken language could be differentially influenced by strategic and nonstrategic forces depending on the portion of the call, regardless of deception. The carefully prepared remarks should be regarded as relatively more strategic than the Q&A. The Q&A remarks, being neither fully anticipated nor rehearsed, should be expected to include relatively more nonstrategic behavior. Preparation, then, is one significant factor that may affect how much strategic and nonstrategic behavior is displayed. We therefore forward the following hypotheses regarding characteristics of spoken language overall, ignoring deception for the moment:

Hypothesis 1a: The linguistic components of prepared spoken utterances differ from those of unprepared utterances on (1) length, (2) level of detail, (3) complexity, (4) hedging and uncertainty language, (5) comprehensibility, (6) personalism and immediacy, and (7) positive and negative affect.

Hypothesis 1b: The vocalic components of prepared spoken utterances differ from those of unprepared utterances on (1) pitch, (2) voice quality, (3) tempo, (4) loudness, (5) response latencies, and (6) pitch slope.

The strategic and nonstrategic role of the particular features noted in Hypotheses 1a and 1b are discussed under each of the succeeding hypotheses that explicitly accommodate the role of deception. Turning now to the role of preparation in deception, the differences between strategic and nonstrategic behavior may be attenuated when deceivers have time to prepare. Preparation of deceitful spoken utterances before delivery to an audience gives deceivers time to strategize their messages and behaviors (Anolli, Balconi, & Ciceri, 2003). Preparation reduces cognitive effort, or the amount of resources a deceiver needs to commit to crafting a spoken utterance, thereby reducing many nonstrategic behaviors associated with deception (Sporer & Schwandt, 2006). The overall effect is that prepared spoken utterances are more similar to truthful messages than "unprepared" spoken utterances (Lancaster, Vrij, Hope, & Waller, 2013).

Now consider how preparation may play out in our current conference calls context. When executives speak about financial topics that are later revealed to be the reason for a financial restatement (i.e., an admission that certain aspects of the firm's financial situation were misrepresented), such utterances, which we label fraudulent utterances, may be influenced by countervailing strategic and nonstrategic forces. Discussing topics about which executives know they are being deceptive may increase arousal, negative affect, and cognitive load, causing them to inadvertently exhibit nonstrategic behavior. On the other hand, executives should attempt to behave strategically to attempt to appear credible. Furthermore, if deceptive topics are discussed in the presentation portion of the call, arguably executives have more potential to strategically craft their message in advance than during the more extemporized Q&A.

Which forces will prevail depends in large part on the specific behaviors being displayed. To systematically assess this complex situation, we therefore consider the vocalic and linguistic features of spoken language in related sets and make specific predictions for each.

Vocalics During Fraud-Related Utterances

Vocalic characteristics that research has associated with deception are voice quality, fundamental frequency (pitch), tempo, intensity (loudness), response latencies, and pitch slope (Elkins & Derrick, 2013; Sporer & Schwandt, 2006). High pitch, fast tempo, and loudness are all indicators of arousal and negative affect, while longer response latencies and slower tempos are indicators of high cognitive load. All usually are nonstrategic behaviors as they are often uncontrolled, inadvertent signals that,

when present, undermine a speaker's credibility. Fraudulent utterances, then, should be characterized more so by nonstrategic forces relative to strategic forces. Additionally, we expect these effects to be particularly salient in the Q&A due to the relative inability to prepare while fielding analysts' pointed questions in real time. We therefore hypothesize the following:

> **Hypothesis 2a:** Relative to nonfraudulent utterances, fraud-related utterances (related to restatement topics) are (1) higher pitched, (2) lower in voice quality, (3) louder, (4) with longer response latencies, (5) an upward pitch slope, and (6) differential tempo, especially during unprepared remarks.

Of course, the underpinnings for Hypothesis 2a, which are derived primarily from low-stakes settings, may not generalize to our high-stakes setting. The credible null and/or alternative hypothesis to Hypothesis 2a is that, due to strategy, we may not observe such effects and/or may in fact observe the opposite.

Linguistics During Fraud-Related Utterances

As the literature is much more developed on linguistic markers of deception than the vocal markers of deception forwarded in Hypothesis 2a, we develop linguistic hypotheses in categories.

Utterance Length and Specificity

The length of an utterance and the amount of information in it are indicative of how forthcoming a speaker is. Specificity concerns the amount of detail present in an utterance. Terms related to place and time, to the five senses, and to numbers all make an utterance more specific. Utterance length and specificity are features that commonly distinguish deceptive from nondeceptive discourse (Burgoon & Qin, 2006; Zhou, Burgoon, Nunamaker, & Twitchell, 2004). Compared with vocal features, utterance length and quantity of information are more controllable and can be manipulated intentionally by speakers. They are therefore strategic resources.

One deception strategy for avoiding detection is reticence—saying less and concealing incriminating information (Buller & Burgoon, 1994; Van Swol, Braun, & Malhotra, 2012). Multiple meta-analyses and recent experiments have found that deception is often associated with briefer, less forthcoming utterances (DePaulo et al., 2003; Hartwig & Bond, 2011; ten Brinke & Porter, 2012). From a cooperative discourse perspective (Grice, 1989), reticence would also be associated with a paucity of specific details. Amount of detail (information) is addressed in information manipulation theory (McCornack, 1992) and information management theory (Burgoon, Buller, Guerrero, Afifi, & Feldman, 1996) under the rubric of quantity (e.g., quantity of such details as space, time, sight, and sound references) and in the Criteria-Based Content Analysis system (Steller & Köhnken, 1989) under contextual embedding.

However, interactive experiments have found that when deceivers have the opportunity to plan, their utterances are actually longer than those of truth tellers, whether measured as words, verbs, sentences, or talk time (Burgoon, Wilson, Hass, & Schuetzler, in press; Van Swol et al., 2012; Zhou et al., 2004). Greater loquacity may be part of a strategy to be more persuasive and dominant. Thus, quantity of verbiage can be deployed in contradictory ways—by saying very little or saying a lot.

We hypothesize that both truthful and fraud-relevant utterances should be similar in quantity during the prepared utterances because one can control in advance how much or how little to divulge. During unprepared utterances, our conjecture is that fraud-relevant utterances should reflect more reticence than nonfraudulent utterances because reticence reduces the risk of disclosing adverse information and prolonging conversation.

Hypothesis 2b: Relative to nonfraudulent utterances, fraudulent utterances are (1) shorter in length and (2) less specific, but only during unprepared remarks.

A credible null and alternative hypothesis to Hypothesis 2b exists, however. If investors interpret short Q&A answers as an implicit indication of bad news (Hollander, Pronk, & Roelofsen, 2010), managers may instead opt to go on the offensive during fraud relevant utterances, saying more in an effort to be convincing.

Complexity

During deception, liars often have to create and maintain a false story. Generating more linguistically complex sentences consumes more cognitive resources than generating simple sentences and stories (Newman, Pennebaker, Berry, & Richards, 2003; Vrij, Fisher, Mann, & Leal, 2006). The construction of sophisticated and complex utterances requires deliberate thought, making complex language a marker of strategic activity and simpler language potentially indicative of nonstrategic discourse, with the caveat that speakers can simplify their discourse as they are speaking but cannot easily make it more complex. Because of the increased cognitive load and the human mind's finite processing capacity (Vessey, 1991), liars will have difficulty simultaneously maintaining a false story and producing linguistically complex utterances (Zhou et al., 2004). In other words, the more difficult it is for deceivers to concoct a believable response, the more they must resort to simpler language.

However, such difficulty may not exist during a formal presentation due to advance preparation. Research suggests prepared statements of formal language are more likely to include bigger words and sentences with qualifying and clarifying phrases and clauses, regardless of veracity (Anolli et al., 2003). This implies that during the unprepared Q&A portion of earnings calls, executives are especially likely to experience cognitive taxation and thus to produce simpler discourse. That is, simpler language is more likely to mark deception in the Q&A.

Hypothesis 2c: Relative to nonfraudulent utterances, fraudulent utterances are less complex, especially during unprepared remarks.

Hedging/Uncertainty-Certainty

Hedging and uncertainty refers to words, phrases, and constructions that introduce ambiguity, abstruseness, or vagueness in a statement (Bachenko, Fitzpatrick, & Schonwetter, 2008; Sherblom & van Rheenen, 1984). Hedging words indicate uncertainty (Palomares, 2008) and/or noncommitment to what is being said (e.g., "I guess" and "maybe"; Duran, Hall, McCarthy, & McNamara, 2010). Additional contributors to vagueness are indefinite pronouns (e.g., "it") and demonstratives (e.g., "this" or "that one") that point to ambiguous referents, weak modal verbs (e.g., "could," "might"), and verbs, adjectives, and adverbs that are semantically uncertain (e.g., "approximately," "variable," "depend"). In financial contexts, there are additional words that represent uncertainty (e.g., "indefinite," "speculate"; Loughran & McDonald, 2011). A strategy of injecting hedging and uncertainty can be used as a mechanism to lessen culpability when securities litigation is brought against management. Such a strategy is particularly beneficial when questions are directly posed to management, as would occur in the question and answer session. We therefore predict:

Hypothesis 2d: Relative to nonfraudulent utterances, fraudulent utterances include more hedging and uncertainty language, especially during unprepared remarks.

Comprehensibility

Comprehensibility is closely related to hedging and uncertainty. It refers to language that is equivocal, obfuscates, and decreases intelligibility. Incomprehensible language is difficult to understand. Some studies find that more-difficult-to-read financial reports are associated with poor future performance (Courtis, 1998; Li, 2008), consistent with managers attempting to disguise bad news. Regarding financial fraud, Goel, Gangolly, Faerman, and Uzuner (2010) found that fraudulent firms had harder-to-read financial reports. Using words that are difficult to understand during the presentation portion of a call may attract subsequent pointed analyst questioning. However, when used as part of answering analyst questions, lowering comprehensibility may lessen the analyst's ability to digest information or unravel the deception:

Hypothesis 2e: Relative to nonfraudulent utterances, fraudulent utterances include less comprehensible language, especially during unprepared remarks.

Immediacy and Personalism

Language can reflect the extent to which the speaker personalizes and takes ownership of the utterance or attempts to disassociate oneself from what is being said. Such distancing can help a deceiver lower the perceived costs of deception, and this can be

achieved through pronouns, verb tense, passive voice, and modifiers in a statement (Wiener & Mehrabian, 1968). Deceivers tend to use fewer self-reference words and to reference other people more frequently (Bond & Lee, 2005; Hancock, Curry, Goorha, & Woodworth, 2008; Newman et al., 2003; Van Swol et al., 2012). In fact, pronouns are typically among the top 10 features that distinguish between truths and lies (C.-C. Lee et al., 2009), and first-person singular pronouns ("I," "me," "myself") typically signal that the communicator takes ownership of the statement (Newman et al., 2003). Furthermore, a passive rather than active voice, and use of past rather than present tense verbs, can lower immediacy as a mechanism to distance oneself from a deceptive act (Mehrabian, 1972).

As verbal nonimmediacy was originally conceived, future tense word usage was also thought to distance oneself from the present. However, in our context, future-oriented language will commonly refer to financial projections (Matsumoto et al., 2011). Safe Harbor provisions in securities legislation provide protection for management with respect to forward looking projections, which provides an additional vehicle for avoiding ownership. Linguistic features that put distance between the speaker and the deceptive act thus might be more pronounced for deceptive utterances, particularly during analyst questioning, because analysts can directly ask penetrating questions that attempt to link executive actions to firm outcomes. In such cases, distancing would be particularly important for a deceiver. Formally we hypothesize:

> **Hypothesis 2f:** Relative to nonfraudulent utterances, fraudulent utterances include (1) fewer self-referencing pronouns, (2) more pronouns referencing others, (3) more passive voice, and (4) more future tense, especially during unprepared remarks.

As a caveat to the verb tense prediction, future-oriented statements combined with other assertive linguistic forms might instead convey confidence and correlate with certainty markers. In our analysis, we examined how closely future verb tense aligned with certainty and immediacy measures to assess where it best fit in forming indices.

Affect

A final category of language studied in the deception literature relates to the speaker's emotional or affective state. Negative affect stems from guilt and is nonstrategic (Bond & Lee, 2005; Vrij, Edward, Roberts, & Bull, 2000). However, strategically, deceivers attempt to mask these negative emotions by using positive language (Buller & Burgoon, 1996). Greater use of both negative emotion words and positive emotion words by deceivers has been confirmed in a recent meta-analysis (Hauch et al., 2014), suggesting deceivers use more emotion words in general. To the extent live questioning by analysts enhances the emotion profile of the executive, we expect this effect to be exacerbated in the Q&A portion. Formally we hypothesize:

> **Hypothesis 2g:** Relative to nonfraudulent utterances, fraudulent utterances include more affect language, especially during unprepared remarks.

Method

Sample

Analyses were conducted on a corpus of 1,114 statements made by the CEO and CFO of one company during six quarterly earnings conference calls for which the authors had already collected the audio recordings. The unit of analysis was the utterance level, where utterances were defined as a sentence, question, or partial sentence if a complete sentence was not provided. Of these utterances, 884 were spoken by the CEO and 230 were spoken by the CFO. Additionally, 648 (466) of the utterances were from the presentation (Q&A) portion of the call.

We chose this particular firm for analysis because, in addition to the audio recordings being available, (1) the firm restated the financial statements discussed during the conference calls due to an overstatement of originally reported financial earnings; (2) multiple class action lawsuits alleging false and misleading statements and breach of fiduciary duty were filed by investors naming both the CEO and CFO as defendants and the class action periods overlap the fiscal periods of the conference calls; (3) the lawsuits were not dismissed, but rather settled in favor of the plaintiffs; (4) both the CEO and CFO left the company subsequent to the allegations; and (5) the Securities and Exchange Commission initiated a formal investigation. Collectively, these conditions help ensure that the stakes were high when the executives were speaking and the executives were likely actively aware of the deception before it was revealed publicly. Additionally, resulting lawsuits, media coverage, and regulatory investigation provided sufficient detail to identify the precise nature of the alleged financial deception, thereby allowing for a classification of whether any particular utterance spoken by the CEO or CFO pertained to issues related to the allegations and subsequent restatement of financial results.

Ideally, one would prefer a large sample of utterances taken from multiple companies and executives to achieve high generalizability. However, preparing conference calls for linguistic and vocalic analysis is a costly and time-consuming undertaking, especially at the utterance level. Conference calls generally last an hour, and each call must be segmented into utterances by matching the audio with the transcript. All utterances must be tagged by speaker, and a human coder with financial expertise must discern whether each sentence utterance pertains to the financial topic that was the basis for the eventual restatement of financial reports. Studying only two executives from one firm limits the generalizability of our findings. However, focusing on one firm is consistent with the analysis a financial statement auditor or analyst would conduct—that is, they would look for variation in vocalics and linguistics for the same individual over time for the purpose of identifying fraudulent topics. Moreover, a within-firm analysis removes the potential for unobserved heterogeneity across executives with respect to linguistic speaking style and physiology of the speaking apparatus, while also controlling for history effects that could differentially affect a sample of multiple firms.

Data Segmentation and Annotation

A major decision point in analyzing speech is the unit of analysis. Should the unit of analysis be the entire call, the phases of the call (prepared presentation remarks vs.

unprepared Q&A remarks), speaker turns at talk, or single utterances? In the Larcker and Zakolyukina (2012) analysis, the authors studied a large firm-quarter panel data set using the entire call as the unit of analysis in their statistical analysis and concluded, "It may be the case that there are better verbal cues for identifying deception in answers to questions related to specific accounts that were actually manipulated" (p. 537). As such, we have opted for the most granular unit of analysis for which we can still assess the topic being discussed, which is the utterance level. This unit of analysis is intended to increase the statistical power of the study, as called for by Bloomfield (2012).

To identify whether each utterance pertained to a restatement topic or not, an independent coder with financial statement expertise and no knowledge of our hypotheses manually coded each utterance with a 1 (related to the restatement topic) and 0 otherwise.[2] Press releases, court documents from related class action lawsuits, and mandatory filings with the Securities and Exchange Commission by the company were reviewed by the coder to identify topics pertaining to the misreporting that resulted in a financial restatement.

Linguistic and Vocalic Feature Extraction and Data Reduction

For linguistic analysis, transcripts of each conference call were obtained from Thomson Reuters StreetEvents (www.streetevents.com), utterances were extracted that matched the related .wav file and were submitted to Structured Programming for Linguistic Cue Extraction (SPLICE; Moffitt et al., 2012; http://splice.cmi.arizona.edu/). Following the existing literature on linguistic analysis in high-stakes settings, we also reduced linguistic features to a smaller set of orthogonal composite variables similar to Fuller et al. (2013) via principal components factor analysis with varimax rotation. Extraction criteria included eigenvalues greater than 1.0, at least two features with primary loadings on a given factor, secondary loadings .20 or more below the primary loading, inspection of the scree test, and interpretability of the factor. All retained factors met these criteria and corresponded to theoretically meaningful factors in the literature. We list our linguistic variables sequentially below and summarize them in Table 1.

Utterance length was measured in two ways, as the total number of words in each utterance segment and as the duration of each utterance in seconds. These two measures were highly correlated ($r = .86$) with a standard score coefficient alpha reliability of .92.

Specificity was measured by a composite of (1) spatiotemporal terms, (2) determiners, (3) cardinal numbers, (4) prepositions, (5) subordinating conjunctions, and (6) adjectives, which together had a coefficient alpha reliability of .80.

Complexity was measured by a composite of (1) polysyllabic words (three syllables or longer), (2) singular or mass nouns, (3) plural nouns, (4) coordinating conjunctions, (5) subordinating conjunctions, (6) prepositions, (7) commas, and (8) average sentence length. The multisyllabic words represented the degree of semantic complexity through the use of big vocabulary words. The number of nouns reflects the likelihood

Table 1. Variable Definitions.

Hypothesis	Variable name	Variable definition
H1a andH2b	*Utterance Length*	The total number of words in each utterance.
H1a and H2b	*Specificity*	Amount of detail present in an utterance, measured by a composite of the number of spatiotemporal terms (total and past), determiners, cardinal numbers, subordinating conjunctions, prepositions and adjectives, all standardized and averaged together.
H1a and H2c	*Complexity*	Syntactically and linguistically complex utterances, measured by a composite of polysyllabic words (three syllables or longer), singular or mass nouns, plural nouns, coordinating conjunctions, commas, and average sentence length, standardized and averaged.
H1a and H2d	*Hedging/Uncertainty*	Words, phrases, and constructions that introduce fuzziness or vagueness in a statement, measured by a sum of the number of hedging words, indefinite terms, uncertainty quantifiers, and demonstratives from SPLICE to which were added the Loughran-McDonald dictionaries of uncertainty terms and weak modals applicable to financial texts.
H1a and H2d	*Certainty*	Words, phrases, and constructions that convey certainty, measured by the SPLICE certainty dictionary, future tense words, and the Loughran-McDonald dictionary of strong modal verbs (words that demonstrate confidence in a statement like "should" and "must") and intense adverbs.
H1a and H2e	*Comprehensibility*	Language that is equivocal, obfuscates, and decreases the ability to understand what is said, measured with the Gunning-Fog index.
H1a and H2f	*Immediacy/ Personalism*	Language that reflects the extent to which a speaker personalizes an utterance, takes ownership of it, and creates psychological closeness measured by present-tense gerunds, spatiotemporal terms in future-tense utterances, and first-person plural pronouns. The inverse, nonimmediacy, refers to attempts to depersonalize an utterance, distance oneself from what is being said, and create psychological distance from an interlocutor through such measures as past tense and passive voice.
H1a and H2g	*Affect*	The speaker's emotional or affective state, measured using pre-scaled terms in the Whissel, Fournier, Pelland, Weir, and Makarec (1986) dictionary of affect in language. Separate measures of negative and positive affect were included.

(continued)

Table 1. (continued)

Hypothesis	Variable name	Variable definition
H1b and H2a	*Pitch and Voice Quality*	A composite audio measure comprising mean, standard deviation, variance, maximum and range for fundamental frequency (pitch), and mean voice quality.
H1b and H2a	*Vocal Intensity*	A composite audio measure comprising the intensity (loudness) mean, inverted intensity standard deviation, and tempo.
H1b and H2a	*Response Latency*	The amount of time between one speaker ending solo speaking and the next speaker beginning solo speaking (also known as turn switch pauses).
H1b and H2a	*Pitch Slope*	Upward or downward inflection pattern of fundamental frequency.

of high nominalization being present, which makes language more wordy and difficult to comprehend. Commas are a proxy for pausality. Pausality is a measure of amount of punctuation in written discourse, which produces compound and complex sentences. Because the utterances analyzed consisted of sentences, natural end-of-sentence punctuation like periods and question marks were not used for identifying pausality. Conjunctions also signal the presence of compound and complex sentences. The coefficient alpha reliability for the complexity composite was .81. Because complexity included in its composite average sentence length, the measure was not further adjusted for number of words.

Hedging and uncertainty were measured with indefinite pronouns, which refer to other unspecified referents (e.g., "it"); demonstratives, which produce a vague referent for the subject of the sentence (e.g., "this," "that one," and "these," and hedging terms from SPLICE to which were added the Loughran and McDonald, 2011, list of weak modals and words specific to the financial realm that indicate uncertainty). The weak modal verbs and adverbs includes terms such as "could," "might," and "possibly" that add tentativeness. The uncertainty list includes 285 words, such as "approximate," "depend," "fluctuate," and "variable." Reliability for this composite was .75. Because hedging and uncertainty were somewhat correlated with the quantity of words, analyses were conducted on the original composite and the rate of hedging and uncertainty language (i.e., adjusted for number of words).

Certainty was an additional composite, consisting of the SPLICE certainty dictionary, which includes intense adverbs like "definitely" that convey certainty; future tense words; plus the Loughran-McDonald dictionary of strong modal verbs like "should," and "must." This measure produced a reliability of .80.

Comprehensibility was measured by the Gunning FOG readability index, a formula that weights number of words and average sentence length to project how difficult to read or comprehend a sample of discourse is.

Immediacy and personalism were initially investigated together because of the role of pronouns in both. Pronouns (first-person singular, first-person plural, second-person, and third person), spatiotemporal terms in past tense sentences, past and future tense verbs, gerunds, and passive voice were examined to see if composites could be formed. Two composites with acceptable reliabilities were *immediacy*, consisting of future tense verbs and spatiotemporal terms in future-tense utterances, with a reliability of .84, and *nonimmmediacy*, consisting of past tense verbs and spatiotemporal terms in past-tense sentences, with a reliability of .70. The pronouns failed to load with either of these composites or with one another, and they had insignificant correlations with one another, so they were analyzed as individual measures.

Affect was measured with the Whissel et al. (1986) dictionary of terms with emotional connotations. Each term in the dictionary has a prescaled value ranging from positive/pleasant negative/unpleasant. This dictionary has been used in numerous text-based deception experiments (e.g., Fuller et al., 2013; Hancock, Woodworth, & Porter, 2013; Zhou et al., 2004).

For vocalic analysis, a unique .wav file was created for each utterance made by the CEO and CFO. For each audio file we measured the mean, standard deviation, variance, maximum, and range for fundamental frequency, or F_0 (pitch), tempo (measured as the number of words per segment), voice quality (based on the signal to noise ratio), and amplitude (or loudness) using Praat software for signal processing (Boersma & Weenink, 2006). To reduce the vocal features to a few parsimonious dimensions, principal components analysis with varimax rotation was conducted. It produced three orthogonal dimensions accounting for 72% of the variance. The first factor was composed of all of the pitch (fundamental frequency) and voice quality features. These were combined into a composite labeled *pitch and voice quality* and had a coefficient alpha reliability of .93. The second factor, labeled as *vocal intensity*, composed of the intensity mean and the inverted intensity standard deviation (representing less, rather than more variance in amplitude), with a reliability of .81. The third factor consisted of the slope of the pitch and response latency. These latter two features were negatively and weakly correlated and so were analyzed separately, as was response latency, which did not load cleanly on any factor.

Results

Preliminary Analyses

Because the current corpus had only one male speaker (the CEO) and one female (the CFO), role and gender were conflated and separate analyses could not be conducted for each. Nevertheless, analysis of the two speakers provides a context for understanding other vocal and linguistic differences and confirms that these measures were able to make discriminations between the two different speakers. Compared with the female CFO, the male CEO had lower pitch, talked faster, especially during the Q&A portion of the call, and maintained a consistently louder voice with less variance. This combination of pitch, tempo, and intensity connotes dominance, as would be expected

due to both sexual dimorphism and the CEO's higher rank in the organization (Mayew, Parsons, and Venkatachalam, 2013). Linguistically, the CEO also used more personal pronouns, more complex syntax, and more complex semantics, resulting in language that was more precise but also more difficult to comprehend.

Hypothesis Tests

The hypothesized effects were initially tested in fully saturated three-way factorial multivariate or univariate analyses of variance with the three independent variables of speaker (male CEO/female CFO), preparation (prepared remarks/unscripted Q&A), and fraud (fraudulent restatement-related/nonfraudulent) and vocalic and linguistic composite measures as dependent variables. Inspection of the means showed that the phatic communication utterances (greetings, thanks) were far different than other utterances, so they were omitted from these analyses to avoid confounding effects. Where appropriate, reduced models were tested that excluded main effect variables and interaction terms that yielded *F* values less than 1.0 and therefore did not contribute to the efficiency of the design (see Snedecor & Cochran, 1967). Univariate analyses were interpreted to provide more insight into which linguistic factors were most responsible for the multivariate effects.

Hypothesis 1a: Effects of Preparation on Linguistic Features. Hypothesis 1a predicted that prepared remarks are significantly different than unprepared remarks along several linguistic dimensions. Length, duration and specificity were initially analyzed together in a three-way factorial multivariate analysis of variance (MANOVA), which produced significant multivariate effects for preparation, fraud, and fraud by preparation.[3] Due to high multicollinearity with the quantity variables, specificity was analyzed in a separate analysis of variance (ANOVA) and converted to a rate, adjusted for number of words. The main effect for preparation was no longer significant, but the fraud by preparation interaction remained statistically significant, $F(1, 1,010) = 4.41$, $p = .035$. Compared with unprepared remarks, prepared utterances had more words, were longer in duration, and were more specific (e.g., more adjectives, cardinal numbers, determiners). However, the greater specificity was due entirely to the fraudulent-topic utterances having more details than the nonfraudulent utterances.

A reduced model MANOVA (with two interaction terms removed) on the hedging/uncertainty and certainty composites produced significant effects for preparation, Wilks' $\Lambda = .935$, $F(3, 1,008) = 23.41$, $p < .001$, and preparation by speaker role, Wilks' $\Lambda = .980$, $F(3, 1,008) = 7.03$, $p < .001$, as well as fraud and preparation by fraud, discussed under Hypothesis 2d. The univariate analyses produced significant main effects for preparation on both measures, hedging/uncertainty, $F(1, 1,010) = 29.78$, $p < .001$, certainty $F(1, 1,010) = 6.07$, $p = .014$.

The preparation by role interaction was also significant on both measures, hedging/uncertainty, $F(1, 1,010) = 4.75$, $p = .029$, certainty, $F(1, 1,010) = 14.51$, $p < .001$. Hedging and uncertainty were highest during the unscripted Q&A on fraudulent topics and lowest during prepared remarks on fraudulent topics. Curiously, certainty was also

highest and lowest under the same conditions. This can be explained by the fact that the measures were formed from different dictionaries and features. The CEO also did more hedging than the CFO during the Q&A.

An ANOVA conducted on the comprehensibility measure produced two significant interactions: preparation by role, $F(1, 1,008) = 4.27$, $p = .04$, and preparation by fraud, $F(1, 1,008) = 4.22$, $p = .04$. Whereas nonfraudulent statements were somewhat more incomprehensible than fraudulent statements during the prepared remarks, the fraudulent remarks were less comprehensible than nonfraudulent remarks during the Q&A.

A reduced model ANOVA with insignificant interaction terms omitted on the complexity measure produce a significant preparation effect, $F(1, 1,012) = 9.49$, $p = 002$, and a preparation by fraud interaction, discussed under Hypothesis 2c. Unscripted Q&A remarks were less complex than prepared remarks.

A MANOVA on the immediacy and nonimmediacy measures produced significant effects on preparation, Wilks' $\Lambda = .978$, $F(2, 1,010) = 11.33$, $p < .001$, and preparation by role interaction, Wilks' $\Lambda = .976$, $F(2, 1,010) = 12.22$, $p < .001$. However, when the two measures were adjusted for utterance length, no effects remained significant, indicating the results were a function of the number of words in the utterances. The four pronoun measures failed to produce any significant effects. Finally, although the MANOVA on positive and negative affect produced significant effects for preparation by fraud, once the affect measures were adjusted for utterance length, the effects were no longer significant.

In summary, prepared utterances fit a strategic profile in that they had more words, were longer in duration, and were more specific, but specificity was confined to the fraudulent topic utterances. The hedging/uncertainty, certainty, and comprehensibility measures were affected by the combination of preparation and fraud and are discussed under Hypothesis 2. The remaining measures only showed preparation effects if not adjusted for utterance length. (See Appendix A for all main effect means and standard deviations. Where measures are adjusted for utterance length, both the unadjusted standardized score and the rate per number of words in the utterance are displayed.) Therefore, Hypothesis 1a was partially supported but preparation also interacted with fraud.

Hypothesis 1b: Effects of Preparation on Vocalic Features. Hypothesis 1b predicted differences between prepared and unprepared remarks along vocalic dimensions. Though we posed a nondirectional hypothesis, we anticipated that compared with prepared remarks, unprepared remarks would exhibit higher pitch and lower voice quality (combined into the pitch and voice quality composite), faster tempo and louder amplitude (combined into the composite of vocal intensity), longer response latencies, and upward pitch slopes. These expectations were predicated on the assumption that vocal indicators of executives would reveal arousal and cognitive taxation occurring during live interrogation by financial analysts.

A fully saturated three-way MANOVA on the four vocal measures produced a significant main effect for speaker role (to be expected, given the gender differences) and the hypothesized main effect for preparation, Wilks' $\Lambda = .980$, $F(4, 1,097) = 5.17$,

p < .0001. The univariate analyses revealed that the preparation main effect occurred on pitch and quality, $F(1, 1,005) = 4.95$, $p = .026$, intensity, $F(1, 1,005) = 33.08$, $p <$ .001, and on pitch slope, $F(1, 1,005) = 3.91$, $p = .044$. (See Appendix A for means and standard deviations for preparation main effects.) Prepared remarks were higher pitched with a higher range and variance, a downward slope, and more variability in voice quality. Unscripted utterances were louder and faster but more monotone than prepared remarks. Unscripted remarks did not have a longer response latency, as might have been expected if they were experiencing undue cognitive taxation. However, there were also several interaction effects qualifying the main effects. These are discussed under Hypothesis 2.

As a collection Hypothesis 1b is partially supported in that some differences on vocal dimensions were observed. However, the results on vocal indicators of emotional arousal and cognitive load do not support a consistent picture of high emotional arousal and cognitive taxation in the Q&A. One possible explanation for our inability to systematically observe nonstrategic features in the Q&A is that the executives we studied were able to utilize strategies that diminish arousal and cognitive taxation. Such strategies recently forwarded in the literature include executives "staying on script" in the Q&A (J. Lee, 2014), and selecting relatively favorable analysts with whom to dialog to minimize public scrutiny (Mayew, 2008).

Hypothesis 2 was decomposed into a series of subhypotheses, most of which predicted interaction effects between fraud and preparation, indicating particularly salient effects during the Q&A. The full set of main effect and interaction means related to fraud appear in Appendix B. The specific results are discussed by subhypothesis.

Hypothesis 2a: Effects of Fraud on Vocalics. Hypothesis 2a predicted that fraudulent utterances would show more vocal signs of cognitive taxation and emotional arousal than nonfraudulent utterances, especially during the Q&A portion of the earnings call. A reduced model MANOVA that omitted nonsignificant interactions and response latency (which had no significant effects) produced a near-significant main effect for fraud with significant univariate effects on pitch and quality, $F(1, 1,006) = 4.10$, $p = .043$. The main effect for tempo was qualified by a preparation by fraud interaction, $F(1, 1,006) = 4.16$, $p = .042$. Consistent with Hypothesis 2a, pitch was higher and voice quality more varied during fraud-related than nonfraudulent statements (see Figure 1). Tempo was slowest during prepared fraud-related statements. The slower tempo fits with greater cognitive effort but not greater arousal. This is potentially consistent with the strategic use of slow tempo to be more deliberate and persuasive during the initial presentation in an attempt to minimize potential follow-up questions in the Q&A with respect to fraud. The other vocal features did not vary between fraud-related and nonfraudulent topics. Thus, results were mixed on vocal changes due to the combination of fraud topic and preparation.[4]

Hypothesis 2b: Effects of Fraud on Length and Specificity. Hypothesis 2b postulated that fraudulent and nonfraudulent remarks are similar during the prepared portion of conference calls but that during the Q&A period, fraudulent utterances are shorter and less
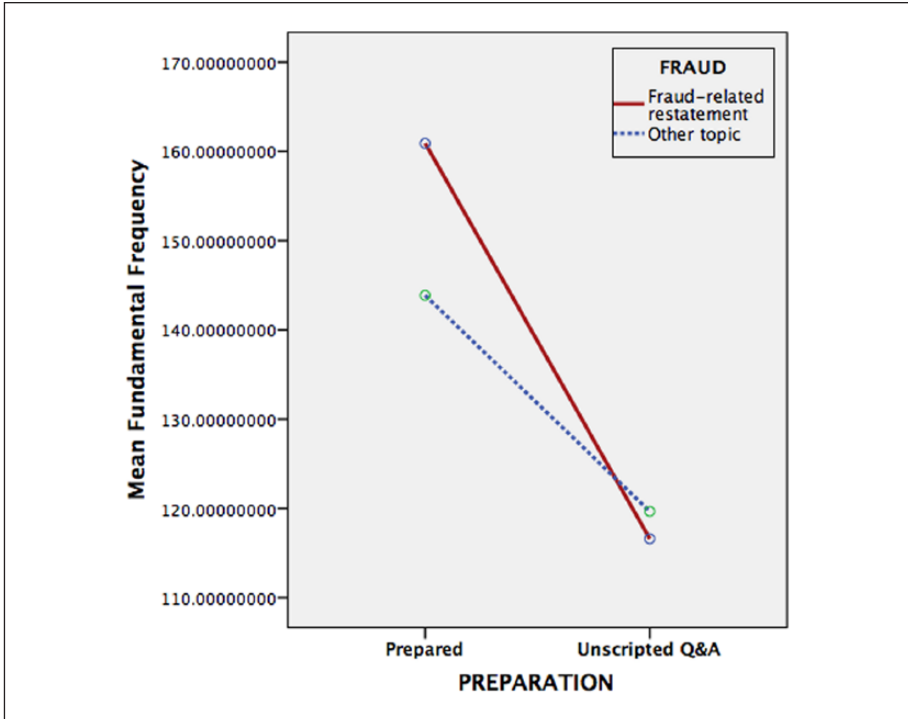
Figure 1. Effects of fraud and preparation on fundamental frequency (pitch).

specific than nonfraudulent utterances. A reduced model MANOVA on the combined quantity measures of word count, utterance duration, and specificity produced significant effects for preparation (noted in Hypothesis 1), fraud, Wilks' $\Lambda = .988$, $F(3, 1,006) = 4.04$, $p = .007$, speaker role, Wilks' $\Lambda = .946$, $F(3, 1,006) = 19.14$, $p < .001$, a preparation by fraud interaction, Wilks' $\Lambda = .991$, $F(3, 1,006) = 3.58$, $p = .014$, and a fraud by role interaction, Wilks' $\Lambda = .987$, $F(3, 1,006) = 4.29$, $p = .005$. The fraud main effect showed that managers had longer statements, both in terms of words and duration, when giving fraudulent than nonfraudulent answers, and maintained that length during the Q&A for fraud-related remarks but shortened them for other topics. This is constituent with Braun et al. (2015), who found lying politicians used longer words in both scripted and interactive settings. Specificity was highest during the prepared fraud-related remarks and declined during the Q&A. The same pattern held when specificity was adjusted for utterance length.

Because these results (longer and more specific fraud-related utterances) were contrary to Hypothesis 2b, we conducted a follow-up analysis on the specific features comprising the specificity composite. Almost all features were higher for fraudulent as compared with nonfraudulent statements. Fraudulent utterances had far more cardinal numbers during the prepared remarks than the Q&A, speculatively indicative of
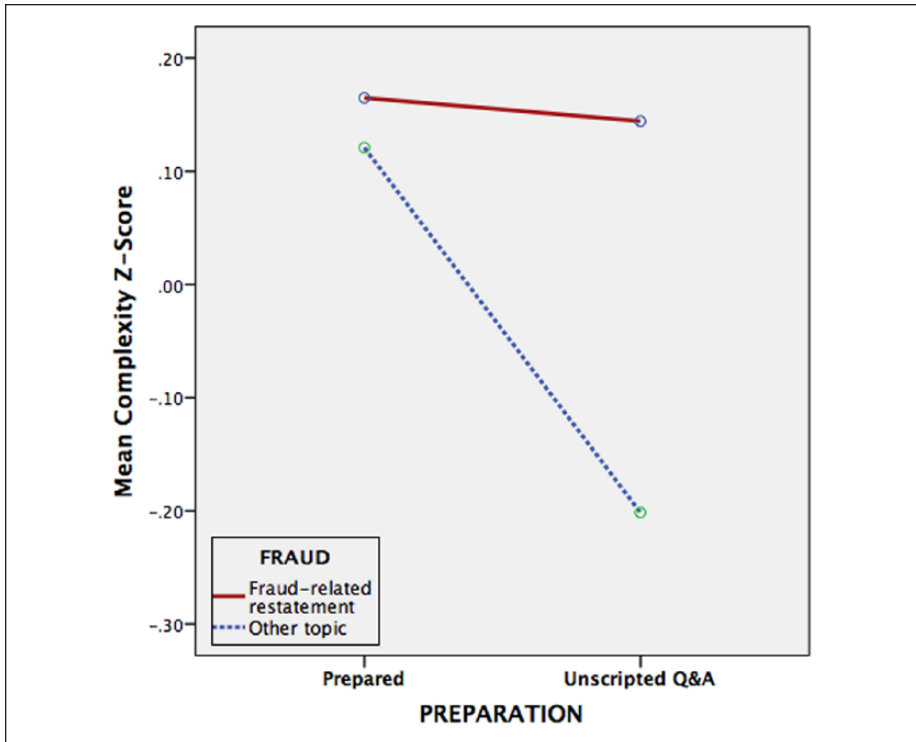
Figure 2. Effects of fraud and preparation on mean complexity score.

attempts to forestall questions on specifics by providing an excessive amount of data and consistent with the use of slower tempo noted above to sound deliberate to avoid questions. Executives compensated for the relatively fewer details during the Q&A by using more words and speaking longer. These patterns form a nuanced strategy for giving the appearance of being forthcoming and complete while responding on fraud-related topics.

Hypotheses 2c to 2e: Effects of Fraud on Complexity, Hedging/Uncertainty-Certainty, and Comprehensibility. Hypothesis 2c through Hypothesis 2e predicted that fraudulent utterances would be less clear than nonfraudulent utterances, especially when not prepared in advance. The *complexity* composite produced significant univariate effects for fraud, $F(1, 1,012) = 12.25$, $p < 001$; and a fraud by preparation interaction, $F(1, 1,012) = 7.35$, $p = 007$. On average, fraudulent utterances were more complex than nonfraudulent utterances during both portions of the call. This is perhaps due to fraudulent topics by nature involving more complex economic activity. Nonfraudulent remarks declined in complexity during the Q&A (see Figure 2), and this decline in complexity for nonfraudulent remarks is contrary to the predicted pattern of

the interaction in Hypothesis 2c under notions of cognitive load. Conceivably by remaining linguistically complex in terms of syntax and vocabulary when answering deceptively during the Q&A, managers were pursuing a strategy of obfuscation. It is also possible that, given the complexity of fraud topic utterances in the Q&A were similar to the presentation, management is simply attempting to "stay on script" (J. Lee, 2014).

The two measures of hedging/uncertainty and certainty were analyzed together in a reduced model MANOVA. The MANOVA produced significant effects for preparation, Wilks' Λ = .935, $F(3, 1,008)$ = 23.41, $p < .001$; fraud, Wilks' Λ = .978, $F(3, 1,008)$ = 7.49, $p < .001$; preparation by fraud, Wilks' Λ = .978, $F(3, 1,008)$ = 7.71, $p < .001$; and preparation by speaker role, Wilks' Λ = .980, $F(3, 1,008)$ = 7.03, $p < .001$. The *hedging/uncertainty* and *certainty* composites had significant univariate effects for fraud, Wilks' Λ = .978, $F(3, 1,008)$ = 7.49, $p < .001$; and preparation by fraud, Wilks' Λ = .978, $F(3, 1,008)$ = 7.71, $p < .001$ (in addition to the earlier discussed preparation effects). The univariate preparation by fraud effect was significant on both measures, hedging/uncertainty, $F(1, 1,010)$ = 12.55, $p < .001$ and certainty, $F(1, 1,010)$ = 7.09, $p = .008$.

Both hedging/uncertainty and certainty language were lowest for prepared fraudulent remarks and highest for fraudulent Q&A (see Figure 3). This may seem contradictory, inasmuch as one would expect higher scores on the hedging and uncertainty composite to be accompanied by lower scores on the certainty composite and vice versa. However, the dictionaries underpinning each composite are not identical, and hence not inversely related. It is therefore feasible to include hedging language on particularly vulnerable topics, when asked in the Q&A, but compensate for such hedging by adding in certainty language so as to appear confident. Such a strategy would be consistent with previous results. Given we observe such deliberate delivery of numeric facts when discussing fraud in the presentation, when faced with questions covering the fraud topic in the Q&A, managers are perhaps able to add certainty claims by referencing the presentation to stay on script and appear confident.

A supplementary analysis showed that all the individual components of each composite followed the same pattern, with the exception of future tense words (which were most prevalent with nonfraudulent Q&A remarks and least prevalent with fraudulent Q&A remarks). These results indicate that while a significant amount of hedging, indefinite and uncertain language was present in the fraud-relevant unprepared remarks, some certainty terms were also included in these remarks, making it quite difficult to use any particular portion of the uncertainty composite alone to sharply discriminate between fraudulent and nonfraudulent utterances.

Comprehensibility (Gunning-FOG index) produced, in addition to a main effect on preparation, an interaction of preparation with fraud, $F(1, 1,012)$ = 7.07, $p = .008$. The most difficult-to-understand "foggiest" utterances were scripted (both fraudulent and nonfraudulent); the most easily understood were the unscripted nonfraudulent utterances. If the unscripted nonfraudulent utterances represent the degree of clarity achievable during spontaneous discourse, then the greater incomprehensibility of the fraudulent responses may be a sign of intentional efforts to obfuscate. Given the ability
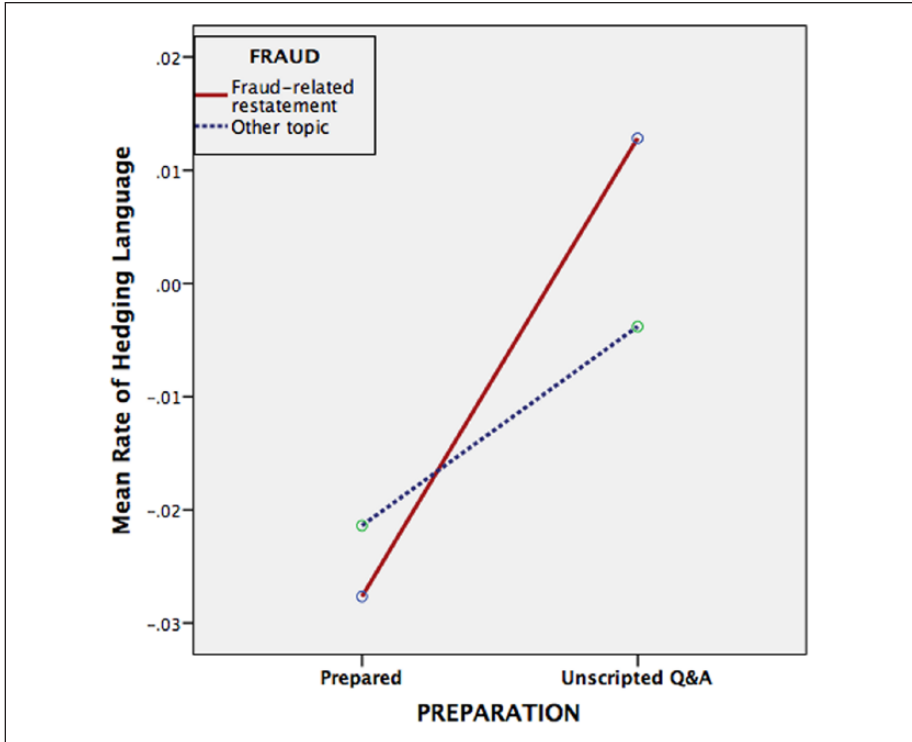
Figure 3. Effects of fraud and preparation on the rate of hedging language, adjusted for utterance length.

to prepare in advance, the finding that the foggiest utterances were scripted is perhaps surprising. One explanation is that the presentation, being carefully scripted, is more likely to precisely articulate technical accounting and business terminology, which results in high FOG scores (Loughran & McDonald, 2014). If true, the nature of the obfuscation in the Q&A may stem again from management staying more on script (i.e., repeating items from the presentation which are less comprehensible overall) when discussing fraud related topics in the Q&A relative to nonfraud-related topics.

Hypothesis 2f: Effects of Fraud and Preparation on Immediacy and Personalism. Hypothesis 2f predicted that fraudulent utterances would include (1) fewer self-referencing pronouns, (2) more pronouns referencing others, (3) more passive voice, and (4) more future tense language than unrelated statements, but primarily during Q&A responding. The immediacy and nonimmediacy measures containing some of these indicators were analyzed together in a reduced MANOVA. Significant multivariate effects emerged for fraud, Wilks' $\Lambda = .953$, $F(2, 1,010) = 24.77$, $p < .001$. Fraudulent statements scored lower on immediacy and higher on nonimmediacy, as measured by
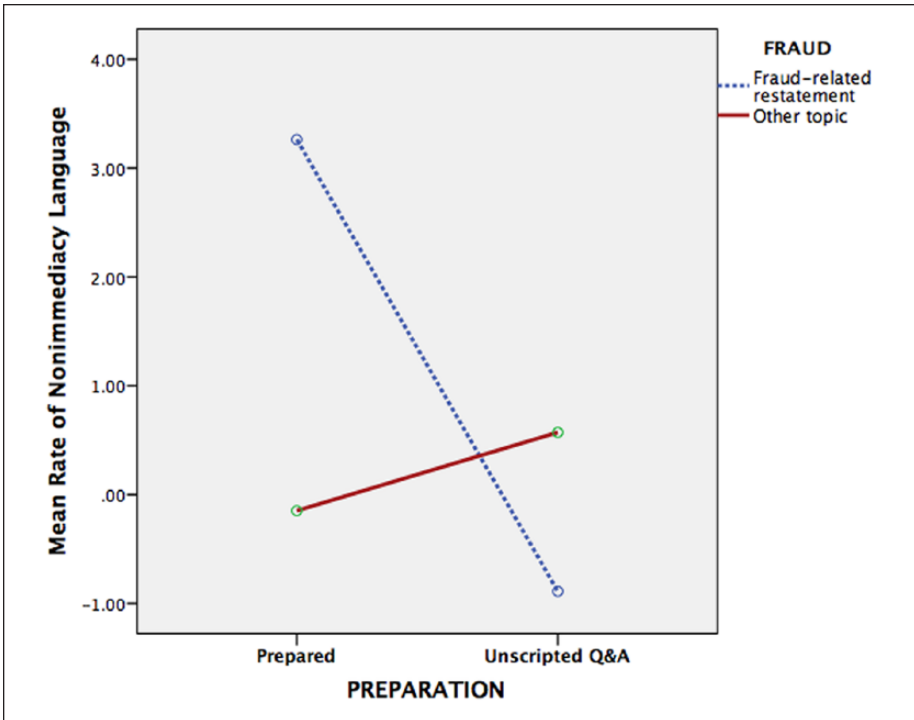
Figure 4. Effects of fraud and preparation on rate of nonimmediacy language, adjusted for utterance length.

spatiotemporal language, verb tense, and active or passive voice (see Appendix B). Univariate results revealed that nonimmediacy was more common in fraudulent than nonfraudulent and prepared than unprepared remarks (see Figure 4). Immediacy was higher during nonfraudulent than fraudulent and prepared than unprepared remarks. These results support Hypothesis 2f overall, although nonimmediacy was not more pronounced in the Q&A.[5] None of the pronoun analyses yielded significant results.

Hypothesis 2g: Effects of Fraud on Affect Language. Hypothesis 2g predicted that utterances related to fraudulent topics include more affect terms than unrelated statements, but primarily during unscripted responding. A reduced MANOVA produced significant effects for fraud, Wilks' $\Lambda = .994$, $F(2, 1,011) = 3.19$, $p = .042$, and the fraud by preparation interaction, Wilks' $\Lambda = .989$, $F(2, 1,009) = 8.08$, $p < .001$. Univariate effects showed that the main effects were overridden by the interaction such that positive affect was highest on unscripted fraud-related remarks, and negative affect was least present during prepared fraud-related remarks. These results support Hypothesis 2g for positive affect, but instead of more negative affect, there was less negative language during prepared fraudulent statements—perhaps an attempt to present a rosier picture than

reality supported. The positive language was most prevalent during the prepared remarks. This makes sense inasmuch as such remarks typically discuss in detail reported financial performance, which was ultimately revealed to be overstated.

Discussion

The spoken and written language transacted in the public sphere serves a critical role in helping resolve the information asymmetry that naturally exists between principals and the agents who ostensibly work on their behalf. For example, language is the vehicle through which citizens learn of actions undertaken by elected politicians (Moen, 1988), economists learn of monetary policy undertaken by the Federal Reserve chairman (Ehrmann & Fratzscher, 2007), and shareholders learn of business deals undertaken by chief executive officers (Kimbrough & Louis, 2011). Agents may not always communicate truthfully, so language also serves as a tool that enables monitoring of agents by outsiders. The news media scrutinizes and assesses the veracity of statements made by politicians, the Federal Reserve Board does so for the Federal Reserve chairman, and auditors and financial analysts do so for corporate management. Deeper understanding of language in its written and spoken forms is therefore key to uncovering the unintentional from intentional, and truthful from deceptive, layers of information and meaning that constitute public discourse.

This exploratory study of spoken language in a high-stakes, real-world context is meant to offer four substantive and methodological contributions to that goal. First, we offer a more extensive linguistic and vocalic feature set than used previously in research to detect financial fraud, deception, or other forms of misrepresentation. Second, we show the utility of a multimodal combination of linguistic and vocalic indicators to evaluate fraud allegations at the utterance level in earnings conference calls. To our knowledge, no studies to date have used both of these modalities at the utterance level. Third, we identify which linguistic and vocalic features are potential indicators of fraud conditional on the fraud discussion occurring in the presentation monolog and the iterative question and answer dialog. Finally, we provide a unique methodology for studying deception that utilizes actual frauds in a naturally occurring laboratory. Subsequent to conference calls, Securities and Exchange Commission filings provide information on whether the financial reports previously discussed were actually misrepresented. This allows a researcher to identify situations where corporate executives were very likely conducting deceptive discussions in a high-stakes setting.

Summary of Results

Prepared remarks differed in substantial ways from unprepared ones. The results underscore the importance of considering these characteristic differences before attributing any deception or misrepresentation to one's responses. For example, utterances might be longer because they were part of the prepared presentation *or* because they

Table 2. Summary of Fraud-Related Hypotheses.

Hypothesis	Construct	Supported?
H2a	*Pitch and Voice Quality*	Supported
H2a	*Vocal Intensity (Tempo, Loudness)*	Supported on tempo during prepared remarks
H2a	*Pitch Slope and Response Latency*	Not supported
H2b	*Utterance Length*	Not supported; fraud-related utterances longer than nonfraudulent utterances
H2b	*Specificity*	Not supported; fraud-related utterances more detailed than nonfraudulent utterances
H2c	*Complexity*	Not supported; nonfraudulent utterances were less complex in the Q&A, while fraudulent utterances exhibited similar high levels of complexity in both the presentation and Q&A
H2d	*Hedging-Uncertainty*	Supported
H2d	*Certainty*	Not supported
H2e	*Comprehensibility*	Partially supported: both fraud-related and nonfraudulent were less comprehensible during the presentation, but during the Q&A fraud-related utterances were less comprehensible
H2f	*Immediacy-Nonimmediacy*	Supported for fraud overall, but not fraud by preparation
H2f	*Personalism (Pronoun Use)*	Not supported
H2g	*Affect*	Supported for positive affect but not negative affect

concerned restatement topics. The type of discourse establishes critical context for interpreting other features. Tests of the hypotheses specifically related to fraud are summarized in Table 2.

Our hypotheses were guided by theory forwarding both nonstrategic and strategic markers of deception in speech. As a collection, no clear pattern emerges to suggest that strategic or nonstrategic are better markers of fraud in the high-stakes setting we study. However, we did find that fraud-related utterances differed systematically from nonfraudulent utterances. The quantity and specificity measures revealed that, contrary to our expectations but consistent with recent evidence in the political arena by Braun et al. (2015), fraudulent utterances were longer and more laden with details than nonfraudulent ones. That pattern persisted from the prepared remarks into the Q&A, whereas utterance length and specificity declined in the Q&A for the nonfraudulent responses. These results bolster an interpretation of the management team going on the offensive when being deceptive, and staying as consistently on the offensive in the Q&A as in the presentation. Such consistency between the presentation and Q&A is

plausibly consistent with the notion of management staying on script in the Q&A particularly for fraud related topics.

Specificity was one of the most variable features in that it differed by speaker, topic, and preparation. Whereas the CFO used very specific language on fraud-relevant utterances during the prepared remarks, she became the least specific when responding about restatement topics during the Q&A. Whether such a shift was strategic or not may be informed by the concomitant hedging/uncertainty and complexity results. Fraud-related utterances had more hedging and uncertainty language and more complexity throughout. At the same time, the Q&A language surrounding restatement topics curiously was also the most certain. Speculatively, this implies speakers were offsetting any appearance of weakness by being very assertive in their restatement-related responses when interrogated. Such a result would be consistent with cognitive load initially playing a role but being overcome, perhaps via reference to prepared remarks in an effort to stay on script.

Results with respect to the Gunning-FOG index, which was meant to measure comprehensibility, were inconsistent with our predictions. We anticipated that less comprehensible language would be used in the Q&A as an obfuscation technique. Recently, the FOG index has been questioned for its applicability to financial domains because many common business terms, like "depreciation" are multisyllable, in turn generating high FOG scores (Loughran & McDonald, 2014). Noting this possibility, the high FOG scores we observed in the presentation, regardless of fraud, plausibly reflects the firm simply discussing the verbatim accounting results. In the Q&A, nonfraud utterances have lower FOG scores but fraud utterances had FOG scores of similar magnitude as the presentation. Together, this again suggests the possibility that when discussing fraudulent topics in the Q&A, management attempted to stay on script (J. Lee, 2014).

Another explanation may be that some ambiguity and verbalism is offset by other responses that are clear and comprehensible, producing indeterminant assessments with the FOG index. Future research could profitably compare and contrast alternative measures of complexity and ambiguity and attempt to identify the extent to which fraudulent versus nonfraudulent responses show high variability across a given call and across multiple calls. Inconsistencies in the degree of concision and clarity might render these measures problematic for gauging when managers are dissembling. Alternatively, the presence of such inconsistencies might help to identify statements that warrant more in-depth investigation.

The immediacy and nonimmediacy language conformed to hypotheses. Fraudulent statements used more distancing language (i.e., were more nonimmediate) and less immediate language than did nonfraudulent statements. The manipulation of such linguistic features as verb tense and spatial references is one way for deceivers to disassociate themselves from their statements and actions.

Positive affect language was also much higher during the Q&A section than the prepared remarks. Fraud-related responses were marked by more positive and fewer negative emotion words, suggesting a desire to put a positive spin on what was being reported. To the extent high voice pitch distinguishes arousal from excitement, which is one type of positive affect, our findings that both high voice pitch and positive affect

words mark fraud utterances suggest potential congruence between our linguistics and vocalics. The fact that speaking tempo was faster during the Q&A restatement topics and the CFO's pitch was lower coincides with a characterization of the speakers adopting a more assertive demeanor when responding to the analysts regarding restatement topics.

Surprisingly, the personalism measures failed to differentiate fraudulent from nonfraudulent utterances. Personal pronouns have garnered significant attention in the deception literature but in this investigation showed neither individual nor combinatorial effects.

Implications

Our investigation complements and builds on an emerging literature that considers strategic and nonstrategic language use by corporate executives. With respect to strategic language, a recent analysis (Stapleton & Hargie, 2011) showed how bank executives modified their language to avoid accountability for the banking crisis while maintaining professional credibility during an official government inquiry. We also observe credibility building and lack of accountability in our setting when discussing fraud topics. We find fraudulent utterances contained features connoting credibility: voice pitch was lowered and utterances were longer and more specific in terms of words. However, immediacy was lower during fraudulent utterances, consistent with executives distancing themselves from the fraud.

Our investigation also builds on an emerging literature that studies speech of public corporate executives. Other recent research (Hobson et al., 2012; Larcker & Zakolyukina, 2012) provides an initial investigation of conference call language differences between fraudulent and nonfraudulent firms, and finds some evidence of nonstrategic deception markers. Bloomfield (2012) suggests that this budding literature can move forward with more powerful research designs. As such, we conducted our study at the utterance level, a more granular level than the firm level analysis commonly utilized in the extant literature (Larcker & Zakolyukina, 2012). An utterance level analysis allows for the coding of whether an utterance was a restatement topic or not, which should be particularly important for auditors and financial analysts who are interested in knowing when to allocate scarce effort resources as part of providing professional services.

Finally, the collection of evidence we provide is arguably most consistent with a combination of nonstrategic markers being subsequently offset with strategic responses in real time. For example, fraudulent utterances were higher pitched (a nonstrategic marker) which was possibly offset by management providing longer and more detailed utterances (a strategic marker). Q&A responses contained both higher uncertainty and certainty words when discussing fraudulent topics. If initial uncertainty is nonstrategic and is subsequently offset by strategic certain words (perhaps via reference to the presentation), it is again plausible that corporate managers speaking in public are well equipped to handle and offset nonstrategic deception markers. Whether the effects we observe are actually a manifestation of "staying on script" (J. Lee, 2014) and whether in turn staying on script is a way to identify fraud in the cross section of firms are important questions for future research. But if managers do have the ability to offset

nonstrategic forces via such scripting, it may provide an explanation for why Larcker and Zakolyukina (2012) are able to predict fraud in the cross section of firms equally well in the presentation and the Q&A (Bloomfield, 2012).

Limitations

Due to the nature of corporate fraud, we can only conceptually detect fraud in situations where the CEO or CFO knows that the communications are false. While it is of benefit to perform this type of analysis at a more granular level than across firms, this leads to a limitation in the data sets available. We were only able to apply our analysis to the conference call recordings from one firm. To expand the data set, we would need not only to collect additional recording sets, but also identify additional frauds or restatements that have become public knowledge and where executives are reasonably believed to be complicit in the fraud. Cost constraints prevented the examination of a larger speech corpus.

Another limitation is that we study only two individual executives, the CEO and the CFO. This limits the vocal and linguistic samples available and affects the generalizability of our findings. We only use the conference calls from a single company, so there are relatively few restatement topics in the data that we used. Additionally, we could only code statements as being related to the restatement, not as actually being deceptive or a lie. A given statement that is related to the restatement topic may or may not itself be a deceptive statement. The statement may be several steps removed from actual deception or lying. For example, an individual may have been aware of an issue generally, but not aware of the association between a specific fact and the issue. A more thorough knowledge of the exact timeframes and actual mechanisms by which the fraud was perpetrated than is available publically would be necessary to evaluate the statements at a more precise level. Statements related to the restatement topic may be systematically different from nonrelated statements simply due to the nature of the topic. For example, statements related to revenue recognition might inherently contain more complex language than statements related to inventory levels. Similarly, more or fewer pronouns of a given type may occur in the context of a particular topic than in the context of another.

A third limitation is that, during the prepared remarks section of the call, the call participant is reading from a script. This script has likely been revised and edited by legal counsel. It is unlikely that cognitively motivated linguistic indicators would be present in the same manner in such prepared statements as they would be in more extemporaneous remarks. However, if staying on script is indeed a phenomenon at work, the prepared remarks may play an important role in identifying deception for reasons not grounded in theories of cognitive taxation.

Future Directions

Fraud remains an ongoing issue in need of attention, and this investigation sought cues indicative of fraud in a relatively uninvestigated, but potentially powerful setting.

Using data derived from public company disclosures of a firm that experienced a financial restatement and subsequently was litigated and investigated by the Securities and Exchange Commission, we conducted linguistic and vocalic analysis to create a model for classifying statements in this data set as restatement-related or not restatement-related. The fact that many of the linguistic and vocal features studied here were significantly associated, but not necessarily in the predicted direction, with a type of discourse or with statements with possible fraudulent intent stands as evidence of the robustness of these features as revealing markers of discourse, demeanor, and fidelity of report. That many of the effects were differentially salient during the Q&A phase also confirms intuition that unscripted responses can, at times, provide more information than those prepared and rehearsed.

Whether our results generalize to other firms, other speakers, and during different time periods is an important area for future inquiry. Future work in this area could also expand the examination of indicators to broader data sets and different levels of granularity. For better or worse, new frauds are exposed on a regular basis leading to an expansion of available data. As additional candidate scenarios and features become available, richer analysis is possible. For example, textual data—ranging from improved (and more accessible) call transcripts to online chat and social media content—are increasingly available. A small number of firms such as Yahoo! and Netflix have recently experimented with video broadcasts of earnings conference calls, suggesting the potential for analysis of kinesics. Researchers, to the extent possible, should also work to partner with law enforcement and other investigating agencies (such as the Securities and Exchange Commission and for profit enterprises like Business Intelligence Advisors www.biadvisors.com) in order to accelerate the development of automated tools for deception detection.

Appendix A

Standardized Means (and Standard Deviations) for Preparation Main Effects on Vocalic and Linguistic Features

Variable	Preparation	
	Unprepared remarks (Q&A)	Prepared remarks
Pitch and voice quality	−0.38 (0.61)	0.27 (0.96)
Vocal intensity	0.43 (0.75)	−0.31 (0.90)
Tempo (words per second)	3.16 (0.94)	2.79 (.59)
Response latency (seconds)	0.76 (0.88)	0.59 (0.50)
Pitch slope	−1.15 (97.07)	32.16 (90.19)
Utterance length (# words)	17.93 (10.08)	19.48 (7.51)
Specificity	−0.16 (0.58)	0.20 (0.75)
Specificity rate	0.57 (2.13)	1.03 (6.80)
Complexity	−0.15 (0.70)	0.13 (0.64)

(continued)

Appendix A (continued)

Variable	Preparation	
	Unprepared remarks (Q&A)	Prepared remarks
Comprehensibility (FOG score)	14.70 (9.97)	20.96 (10.38)
Hedging/Uncertainty	0.32 (1.09)	−0.23 (0.86)
Hedging/Uncertainty rate	−0.00 (0.07)	−0.02 (0.05)
Certainty	0.01 (1.00)	−.01 (1.00)
Certainty rate	0.31 (5.09)	0.64 (11.14)
Immediacy	−0.07 (0.94)	0.05 (1.04)
Immediacy rate	0.44 (6.97)	0.98 (10.04)
Nonimmediacy	−0.17 (0.90)	0.12 (1.05)
Nonimmediacy rate	0.36 (7.84)	0.63 (11.34)
Self-references	0.16 (1.23)	−0.12 (0.78)
Other-references	0.31 (1.40)	−0.22 (0.44)
Negative affect	−0.035 (0.86)	0.03 (1.09)
Negative affect rate	−0.42 (3.97)	−0.46 (6.60)
Positive affect	−0.09 (0.85)	0.06 (1.09)
Positive affect rate	0.22 (6.17)	−0.28 (9.75)

Appendix B

Means (and Standard Deviations) for Fraud by Preparation Interaction

	Fraud-related restatements			
	Nonfraudulent topic		Fraudulent (restatement) topic	
	Prepared remarks	Unprepared remarks	Prepared remarks	Unprepared remarks
Pitch and voice quality	0.15 (0.90)	−0.38 (0.63)	0.65 (1.06)	−0.39 (0.51)
Vocal intensity	−0.25 (0.83)	0.43 (0.72)	−0.51 (1.10)	0.41 (0.90)
Tempo (wps)	2.90 (0.53)	3.14 (0.96)	2.44 (0.63)	3.31 (0.80)
Response latency (seconds)	0.58 (0.49)	0.76 (0.87)	0.62 (0.51)	0.74 (0.96)
Pitch slope	−30.39 (86.66)	0.50 (101.30)	−38.17 (101.41)	−10.71 (67.31)
Utterance length (# words)	19.20 (7.52)	17.10 (9.72)	20.45 (7.41)	22.67 (10.87)
Specificity	0.05 (0.61)	−0.21 (0.53)	0.72 (0.93)	0.13 (0.74)
Specificity rate	0.53 (4.38)	0.58 (2.25)	2.76 (11.63)	0.49 (1.24)
Complexity	0.12 (0.65)	−0.20 (0.68)	0.16 (0.63)	0.14 (0.74)
Comprehensibility	21.68 (10.55)	14.49 (10.02)	18.53 (9.42)	15.88 (9.64)
Hedging/Uncertainty	−0.20 (0.87)	0.26 (1.03)	−0.34 (0.81)	0.65 (1.33)

(continued)

Appendix B (continued)

	Fraud-related restatements			
	Nonfraudulent topic		Fraudulent (restatement) topic	
	Prepared remarks	Unprepared remarks	Prepared remarks	Unprepared remarks
Hedging/Uncertainty rate	−0.02 (0.05)	0.00 (0.07)	−0.03 (0.05)	0.01 (0.06)
Certainty	0.10 (1.04)	−0.04 (0.98)	−0.38 (0.76)	0.28 (1.07)
Certainty rate	1.07 (12.18)	0.07 (4.03)	−0.81 (6.29)	1.72 (8.96)
Immediacy	0.15 (1.07)	−0.09 (0.89)	−0.28 (0.85)	0.04 (1.19)
Immediacy rate	1.36 (10.33)	0.34 (6.24)	−0.31 (8.93)	1.03 (10.28)
Nonimmediacy	−0.03 (0.97)	−0.19 (0.84)	0.85 (1.19)	0.16 (1.33)
Nonimmediacy rate	−0.15 (9.25)	0.57 (8.20)	3.26 (16.35)	−0.89 (5.14)
Self-references	−0.07 (0.81)	0.14 (1.21)	−0.30 (0.63)	0.33 (1.33)
Other references	−0.21 (0.46)	0.29 (1.39)	−0.26 (0.34)	0.37 (1.46)
Negative affect	0.11 (1.18)	−0.05 (0.85)	−0.25 (0.63)	0.02 (0.93)
Negative affect rate	−0.49 (6.35)	−0.29 (3.66)	−0.37 (7.40)	−1.17 (5.41)
Positive affect	0.08 (1.08)	−0.15 (0.79)	−0.01 (1.15)	0.27 (1.09)
Positive affect rate	−0.25 (10.29)	0.09 (5.58)	−0.36 (7.67)	1.02 (8.85)

Authors' Note

Declaration of Conflicting Interests

Funding

Notes

1. The formal prepared remarks, though given orally, originate in written form and have been carefully vetted and edited in advance. The extemporaneous remarks during the Q&A session may be a blend of both spontaneous answers and rehearsed answers to anticipated

analyst questions. Thus, the Q&A portion is not completely spontaneous but rather relatively more so than the presentation portion of the conference call.

2. We only used one coder due to the high cost of employing a judge with the requisite level of financial expertise to read the relevant Securities and Exchange Commission filings and court documents pertaining to the restatement, and apply judgment as to whether a given conference call utterance pertained to a restatement topic. To ensure reliability given the cost of expertise in this domain, we employed a second coder with similar financial expertise to classify 20% of the utterances, randomly chosen. We observed high interrater reliability (Cronbach's α = .9023).

3. Full statistical results are available from the first author.

4. A role by fraud interaction also emerged on the same measures. The variations on pitch were slight but the CFO slowed down her pace during the fraud-related statements.

5. Because pronouns have received a lot of attention in other linguistic research, we conducted a supplementary analysis of them. None the pronoun measures produced significant results either when analyzed individually or with self-references and other references summed.

References

Anolli, L., Balconi, M., & Ciceri, R. (2003). Linguistic styles in deceptive communication: Dubitative ambiguity and elliptic eluding in packaged lies. *Social Behavior and Personality*, *31*, 687-710.

Bachenko, J., Fitzpatrick, E., & Schonwetter, M. (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. In *Proceedings of the 22nd International Conference on Computational Linguistics—Volume 1* (pp. 41-48). Cambridge: MIT Press.

Bloomfield, R. (2008). Discussion of "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and Economics*, *45*, 248-252.

Bloomfield, R. (2012). Discussion of detecting deceptive discussions in conference calls. *Journal of Accounting Research*, *50*, 541-552.

Boersma, P., & Weenink, D. (2006). Praat [Computer software]. Retrieved from http://www.fon.hum.uva.nl/praat/

Bond, G. D., & Lee, A. Y. (2005). Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, *19*, 313-329.

Braun, M., Van Swol, L. M., & Vang, L. (2015). His lips are moving: Pinocchio effect and other lexical indicators of political deceptions. *Discourse Processes*, *52*, 1-20.

Buller, D. B., & Burgoon, J. K. (1994). Deception: Strategic and nonstrategic communication. In J. A. Daly & J. M. Wiemann (Eds.), *Strategic interpersonal communication* (pp. 191-223). Hillsdale, NJ: Erlbaum.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*, 203-242.

Burgoon, J. K., & Buller, D. B. (2015). Interpersonal deception theory: Purposive and interdependent behavior during deceptive interpersonal interactions. In D. O. Braithwaite & P. Schrodt (Eds.), *Engaging theories in interpersonal communication* (pp. 349-362). Thousand Oaks, CA: Sage.

Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W. A., & Feldman, C. M. (1996). Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Communication Monographs*, *63*, 50-69.

Burgoon, J. K., & Qin, T. (2006). The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology*, *25*, 76-96.

Burgoon, J. K., Wilson, D., Hass, M., & Schuetzler, R. (in press). Interactive deception in group decision-making: New insights from communication pattern analysis. In D. Magnusson, D. McNeill, J. K. Burgoon, & M. Casarrubea (Eds.), *Discovering hidden temporal patterns in behavior and interaction: T-pattern detection and analysis with THEME*. New York, NY: Springer.

Burns, M. B., & Moffitt, K. C. (2014). Automated deception detection of 911 call transcripts. *Security Informatics*, *3*, 1-9.

Courtis, J. K. (1998). Annual report readability variability: Tests of the obfuscation hypothesis. *Accounting, Auditing and Accountability Journal*, *11*, 459-472.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74-118.

Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, *31*, 439-462.

Eaglesham, J. (2013, May 27). Accounting fraud targeted: With crisis-related enforcement ebbing, SEC is turning back to main street. *Wall Street Journal*. Retrieved from http://online.wsj.com/news/articles/SB10001424127887324125504578509241215284044

Ehrmann, M., & Fratzscher, M. (2007). Communication by central bank committee members: different strategies, same effectiveness? *Journal of Money, Credit and Banking*, *39*, 509-541.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and cues to deception. *Psychiatry*, *32*, 88-106.

Elkins, A. C., & Derrick, D. C. (2013). The sound of trust: Voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation*, *22*, 897-913.

Federal Bureau of Investigation. (2011). *Financial crimes report to the public*. Retrieved from http://www.fbi.gov/stats-services/publications/financial-crimes-report-2010-2011

Fiedler, K., & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research*, *20*, 199-223.

Fuller, C. M., Biros, D. P., Burgoon, J., & Nunamaker, J. (2013). An examination and validation of linguistic constructs for studying high-stakes deception. *Group Decision and Negotiation*, *22*, 117-134.

Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings? *Journal of Emerging Technologies in Accounting*, *7*, 25-46.

Grazioli, S., Johnson, P. E., & Jamal, K. (2006). A cognitive approach to fraud detection. *Journal of Forensic Accounting*, *7*, 65-68.

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*, 1-23.

Hancock, J. T., Woodworth, M. T., & Porter, S. (2013). Hungry like the wolf: A word-pattern analysis of the language of psychopaths. *Legal and Criminological Psychology*, *18*, 102-114.

Hartwig, M., & Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*, 643-659.

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2014). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*. Advance online publication. doi:10.1177/1088868314556539

Hobson, J. L., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, *50*, 349-392.

Hollander, S., Pronk, M., & Roelofsen, E. (2010). Does silence speak? An empirical analysis of disclosure choices during conference calls. *Journal of Accounting Research*, *48*, 531-563.

Juslin, P. N., & Scherer, K. R. (2005). *Vocal expression of affect*. New York, NY: Oxford University Press.

Kimbrough, M. D., & Louis, H. (2011). Voluntary disclosure to influence investor reactions to merger announcements: An examination of conference calls. *Accounting Review*, *86*, 637-667.

Lancaster, G. L. J., Vrij, A., Hope, L., & Waller, B. (2013). Sorting the liars from the truth tellers: The benefits of asking unanticipated questions on lie detection. *Applied Cognitive Psychology*, *27*, 107-114.

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, *50*, 495-540.

Lee, C.-C., Welker, R. B., & Odom, M. D. (2009). Features of computer-mediated text-based messages that support automatable, linguistic-based indicators for deception detection. *Journal of Information Systems*, *23*, 5-24.

Lee, J. (2014). *Scripted earnings conference calls as a signal of future firm performance* (Unpublished Working Paper). St. Louis, MO: Washington University in St. Louis.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*, 221-247.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, *66*, 35-65.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance*, *69*, 1643-1671.

Matsumoto, D., Pronk, M., & Roelofsen, E. (2011). What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *Accounting Review*, *86*, 1383-1414.

Mayew, W. J. (2008). Evidence of management discrimination among analysts during earnings conference calls. *Journal of Accounting Research*, *46*, 627-659.

Mayew, W. J., Parsons, C. A., & Venkatachalam, M. (2013). Voice pitch and the labor market success of male chief executive officers. *Evolution and Human Behavior*, *34*, 243-248.

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *Journal of Finance*, *67*, 1-44.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs*, *59*, 1-16.

McCornack, S. A. (1997). The generation of deceptive messages: Laying the groundwork for a viable theory of interpersonal deception. In J. O. Greene (Ed.), *Message production: Advances in communication theory* (pp. 91-126). Mahwah, NJ: Lawrence Erlbaum.

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information Manipulation Theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, *33*, 348-377.

Mehrabian, A. (1972). *Nonverbal communication*. Piscataway, NJ: Transaction.

Moen, M. (1988). The political agenda of Ronald Reagan: A content analysis of State of the Union messages. *Presidential Studies Quarterly*, *18*, 775-785.

Moffitt, K. C., Giboney, J. S., Ehrhardt, E., Burgoon, J. K., & Nunamaker, J. F. (2012). Structured programming for linguistic cue extraction (SPLICE). In M. L. Jensen, T. O. Meservy, J. K. Burgoon, & J. F. Nunamaker (Eds.), *Proceedings of the HICSS-45 Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium* (pp. 103-108). Retrieved from http://www.hicss.hawaii.edu/Reports/HICSS45_RapidScreeningTechnologiesDeceptionDetectionandCredibilityAssessmentSymposium.pdf

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*, 665-675.

Palomares, N. A. (2008). Explaining gender-based language use: Effects of gender identity salience on references to emotion and tentative language in intra- and intergroup contexts. *Human Communication Research*, *34*, 263-286.

Sherblom, J., & van Rheenen, D. D. (1984). Spoken language indices of uncertainty. *Human Communication Research*, *11*, 221-230.

Snedecor, G. W., & Cochran, W. (1967). *Statistical methods*. Ames: Iowa State University Press.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: a meta-analytic synthesis. *Applied Cognitive Psychology*, *20*, 421-446.

Stapleton, K., & Hargie, O. (2011). Double-bind accountability dilemmas: Impression management and accountability strategies used by senior banking executives. *Journal of Language and Social Psychology*, *30*, 266-289.

Steller, M., & Köhnken, G. (1989). Criteria-based content analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York, NY: Springer-Verlag.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24-54.

ten Brinke, L., & Porter, S. (2012). Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception. *Law and Human Behavior*, *36*, 469-477.

Van Swol, L. M., Braun, M. T., & Malhotra, D. (2012). Evidence for the Pinocchio effect: Linguistic differences between lies, deception by omissions, and truths. *Discourse Processes*, *49*, 79-106.

Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, *22*, 219-240.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*, 239-263.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, *10*, 141-142.

Whissel, C., Fournier, M., Pelland, R., Weir, D., & Makarec, K. (1986). A dictionary of affect in language: IV. Reliability, validity, and applications. *Perceptual and Motor Skills*, *62*, 875-888.

Wiener, M., & Mehrabian, A. (1968). *Language within language: Immediacy, a channel in verbal communication*. New York, NY: Irvington.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. P. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decision and Negotiation*, *13*, 81-106.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, *14*, 1-59.

Author Biographies

Judee Burgoon is a professor of communication, family studies, and human development and director of research, Center for the Management of Information, College of Management, University of Arizona. Her research and publications center on interpersonal interaction, deception, nonverbal and verbal communication, and human–computer interaction.

William J. Mayew is an associate professor of accounting at the Fuqua School of Business at Duke University. He studies managerial communication of firm performance, including both voluntary disclosures made outside of the financial reporting system and mandatory financial disclosures in financial reports. His research uses theories from economics and psychology to better understand the determinants and consequences of voluntary and mandatory managerial communication.

Justin Scott Giboney is an assistant professor in information technology management at the University at Albany. His research focuses on system security, deception detection, expert systems, and meta-analytic processes.

Aaron C. Elkins is a postdoctoral researcher in the Department of Management Information Systems at the University of Arizona. He investigates how the voice, face, body, and language reveal emotion, deception, and cognition for human–computer interaction applications. He also researches how decision makers will use and are affected by these artificial intelligence–based technologies.

Kevin Moffitt is an assistant professor of accounting in the Department of Accounting and Information Systems at the Rutgers Business School, Rutgers University. He is affiliated with the Continuous Auditing & Reporting Laboratory at Rutgers and publishes in the areas of text mining, fraud detection, and big data in accounting.

Bradley Dorn is a doctoral student in the Department of Management Information Systems at the University of Arizona. His research interests include privacy, online behavior, the design and development of information systems, human–computer interaction, and computational linguistics.

Michael Byrd is a doctoral student in the Department of Management Information Systems at the University of Arizona. His research interests include fraud and deception detection, behavioral and linguistic indicators of deception, human–computer interfaces, and judgment and decision-making bias.

Lee Spitzley is a doctoral student in the Department of Management Information Systems at the University of Arizona. His research interests include corporate financial fraud detection, deception detection, and linguistic indicators of deception.