Analyzing Speech to Detect Financial Misreporting

Jessen L. Hobson*
William J. Mayew†
Mohan Venkatachalam†

*Department of Accountancy, University of Illinois at Urbana-Champaign
† Duke University – Fuqua School of Business

First Draft: October 18, 2009
This Draft: March 3, 2011

Abstract

We examine whether vocal markers of cognitive dissonance are useful for detecting financial misreporting, using both laboratory generated data and archival data. In the laboratory, we incentivize misreporting for personal gain, thereby generating an endogenous distribution of truth tellers and misreporters. All subjects are interviewed about their reported performance of a private task, much like managers are interviewed by analysts and auditors following an earnings report. Recorded responses to a series of automated and pre-scripted questions are analyzed using a vocal emotion analysis software that purports to capture negative emotions stemming from cognitive dissonance. We find the cognitive dissonance scores generated by the software discriminate between truth tellers and misreporters at the rate of 17% above chance levels. For the archival data, we use speech samples of CEOs during earnings conference calls and find that vocal dissonance markers are positively associated with the likelihood of adverse accounting restatements, even after controlling for financial accounting based predictors. The diagnostic accuracy levels are 8% better than chance and of similar magnitude to models based solely on financial accounting information. Our results from using both lab generated data and archival data provide some of the first evidence on the role of vocal cues in detecting financial misreporting.

We thank Dan Ariely, Bob Ashton, Rob Bloomfield, Larry Brown, Brooke Elliott, Kevin Jackson, Rick Larrick, Bob Libby, Charles Lee, Laureen Maines, Maureen McNichols, Mark Nelson, Chris Parsons, Mark Peecher, Madhav Rajan, Doug Stevens and Mark Zimbelman, for helpful comments and discussions. We also appreciate suggestions from workshop participants at the 2009 BYU Accounting Research Symposium, Cornell University, Florida State University, George Washington University, Georgia State University, Harvard University, Indiana University, Lancaster University, Oklahoma State University, Queen's University and Stanford University. Ozlem Arikan, Zhenhua Chen, Virginia Chung, Katie French, Mickey Hartz, Chris Kalogeropoulos, Hui Liu, Hongling Ma, John Montgomery, Kristina Moultrie, Simeon Tzolov, Belinda Wen, Christina Woytalewicz, and Yuepin Zhou provided excellent research assistance.

Analyzing Speech to Detect Financial Misreporting

1. Introduction

Detecting financial misreporting is an increasingly important concern for auditors, regulators, investors, and the various constituents that interact with corporations. High profile accounting scandals such as Enron, WorldCom and Tyco have cost market participants several billions of dollars and eroded confidence in both the published financial statements and the auditors who perform the attestation function. Despite the swift legislative response through the Sarbanes-Oxley Act of 2002, cases of financial fraud continue to surface. The continuing wave of corporate fraud calls into question the ability to uncover financial misstatements by auditors who review and provide an opinion on the financial statements (PCAOB 2007, 2010). Even sophisticated market participants such as institutional investors and analysts have been remarkably unsuccessful at detecting financial fraud (Dyck, et al. 2008). Therefore, developing a cost effective framework to predict financial misstatements can be enormously useful to investors, auditors, analysts and regulators. In this paper, using both laboratory generated data and archival data we empirically examine whether nonverbal vocal cues elicited from speech are useful in detecting deception in financial reporting.

Research in social psychology (e.g., Zuckerman and Driver 1985; Vrij, et al. 2000; DePaulo, et al. 2003) suggests emotions and cognitive processes of deceivers may result in many different markers that can help identify deception, such as verbal linguistic cues (e.g., speech content), nonverbal cues (e.g., tone of voice, facial expressions or gestures), and physiological changes (e.g. heart rate). While prior research finds some support for each of these classes of markers to detect deception, Bond and DePaulo (2006) show that individuals display only a modest accuracy in correctly identifying deceit (see also Vrij 2008). Part of the challenge that individuals face in detecting deception is the identification of behavioral cues (markers) associated with deception.

Experimental research in linguistics (Newman, et al. 2003) suggests that the choice of word usage (verbal cues) captures emotions and motives that mark deception. Using Linguistic Inquiry and Word Count (LIWC) software, Newman, et al. (2003) provide an automated framework for extracting verbal deception makers. Numerous studies in finance and accounting have now applied similar linguistic analysis to detect financial misreporting and to assess fraud risk (see for example Burns, et al. 2010, Loughran and McDonald 2010, Loughran and McDonald 2011, Larcker and Zakolyukina 2010, Purda and Skillicorn 2010 and Humpherys, et al. 2011).[1] The explosive growth in this area of research stems from both the availability of software programs to systematically code language and easy accessibility of large volumes of corporate text files. The general finding from this body of work is that across different software programs (e.g. LIWC, custom dictionaries, Naïve Bayes learning models) and different corporate texts (e.g. 10-K MD&A, earnings conference calls), linguistic deception markers have predictive ability.

We extend this line of research by examining the predictive ability of nonverbal deception cues, in particular vocal cues, to detect financial misreporting. Vrij, et al. (2000) suggest that verbal and nonverbal measures are complementary mechanisms in detecting deception, but little is known about the information contained in nonverbal cues in the capital market setting. Experimental work by Elliott, et al. (2010) finds that communication of restatements via online video, a venue containing both vocal and visual nonverbal cues, impacts investor perceptions of managerial trustworthiness and investment decisions. Archival work by Mayew and Venkatachalam (2011) finds that vocal emotion cues exhibited by managers during earnings conference calls have information content. Neither of these studies, however, speak to whether nonverbal vocal cues might help detect financial deception and assess fraud risk.

We provide evidence on the predictive ability of vocal deception markers in two distinct yet related ways. First, we use a dataset of endogenously determined misreporters and truth-tellers generated in a laboratory setting. In particular, we follow the experimental design of existing research (Mazar, et al. 2008) that allows us to i) generate a sample of truth-telling and misreporting subjects and ii) invoke a

[1] A much broader set of earlier research establishes the informational role of linguistic cues more generally. For research that uses newspaper articles see Kothari, et al. (2009), Tetlock (2007) and Tetlock, et al. (2008). For company press releases, see Demers and Vega (2008), Engelberg (2008), Henry (2008), Davis, et al. (2008). For earnings conference calls see Matsumoto, et al. (2010). For SEC filings and IPO prospectus, see Li (2008, 2009), and Feldman, et al. (2009).

particular emotional state in misreporting subjects. We then employ a vocal emotion analysis tool to identify markers of deception-related emotional states at the precise moments when subjects should be feeling such emotions. A distinct advantage of using laboratory created data is that we can assess the predictive ability of vocal deception markers in a more controlled environment as well as assess construct validity of the software generated vocal deception markers. These are important first steps when using a tool that has been used sparingly in the financial reporting domain.

Second, we generate vocal markers of deception for a sample CEOs for whom we are able to obtain speech samples from their interactions with analysts and investors during earnings conference calls. We then examine whether the vocal deception markers are predictive of future adverse financial restatements. The advantage of using archival data is that it lends external validity to our laboratory findings and allows us to i) quantify the predictive ability of vocal cues and, ii) compare with other predictors of financial misreporting. Thus, our use of both laboratory generated data and archival data to test the role of vocal deception markers in predicting financial misreporting integrates the comparative advantages of both the experimental and archival research methods to examine an important and salient question facing the accounting profession today (Libby, et al. 2002; Sprinkle 2003).

To generate data in the laboratory, we pay college students to answer SAT questions and to self report their score, with payment increasing in the reported score. Thus, we generate an endogenous distribution of misreporters and truth-tellers by incentivizing participants to misreport. Consistent with Mazar, et al. (2008), we posit that, on average, individuals in our setting who engage in misreporting will experience negative emotions due to cognitive dissonance (Festinger 1957; Festinger and Carlsmith 1959). Cognitive dissonance is a state of psychological arousal and discomfort occurring when an individual takes actions that contrast with a held belief, such as misreporting while holding a self-belief of honesty (Mazar, et al. 2008; Graham 2007). We measure vocal dissonance cues, which serve as our marker of deception, by interviewing participants after they report their test scores. Participants responded in a video-recorded interview to a common set of automated and preset questions about their SAT answers and reported score. Then, using automated vocal emotion analysis software based on

Layered Voice Analysis (LVA) technology, we process the audio files to generate nonverbal deception markers and estimate logistic regressions to determine whether these markers are useful in predicting misreporting. Estimating this prediction model requires identification of misreporters and truth-tellers in the participant pool. Because it is not practical to directly observe deceivers from truth-tellers in our laboratory setting, we solicit participants' voluntary admission of whether they overstated their SAT score after they completed the laboratory session. This *ex post* admission of overstatement serves as our primary measure of misreporting.

Logistic regression analysis reveals that voice based cognitive dissonance markers successfully classify misreporters at levels approximately 17% greater than chance. The predictive power of vocal dissonance markers is robust to different specifications of the model, subsets of the data and an alternative misreporting measure. We also find that the effects of vocal cues are most pronounced early in the questioning, consistent with cognitive dissonance theory that suggests dissonance levels will diminish as individuals resolve their cognitive conflict. Finally, we document that the vocal dissonance markers are positively associated with belief revision in participants, a common measure used in the cognitive dissonance literature to measure dissonance resolution. This provides additional construct validity to the dissonance marker generated by the LVA software.

Turning to archival data, we measure the levels of cognitive dissonance for a broad sample of CEOs who speak in quarterly earnings conference calls held during calendar year 2007. Logistic regressions reveal a positive association between vocal dissonance markers and whether the financial statements associated with the conference call are adversely restated in the future. Diagnostic accuracy levels are 8% better than chance and of similar magnitude to models based on financial statement predictors of restatements. We also find vocal dissonance markers to be incrementally associated with adverse restatements even after controlling for financial statement based predictors. Overall, we conclude that analyzing speech is useful for assessing the likelihood of financial misreporting.

This study makes several contributions to accounting literature and practice. First, it provides evidence that elements of voiced speech are helpful in detecting financial misreporting. In our use of vocal emotion analysis software, we add a new tool to the current approach of predicting financial misstatements that is almost exclusively based on quantitative financial measures or linguistic features. We acknowledge that identifying deception using speech is part of a large body of research on detecting deception (see DePaulo, et al. 2003 for a summary), and we are by no means the first to undertake such an endeavor. However, to our knowledge, we are the first to investigate the predictive ability of vocal cues for actual misreporting in a capital market setting using ecologically valid speech samples from corporate executives. In this way, we begin to fill the void noted by Hirschberg (2010) of a relative lack of evidence on vocal cues for deception detection. Yet, we caution the reader that our analysis is a first step, since we identify only one vocal marker of deception. There may be other powerful nonverbal markers of deception, but we leave such an exploration for future research.

Second, our evidence suggests that investors, analysts, auditors and other parties that rely on communications with management pay particular attention to both the questions they ask of the management and the answers that they receive. In other words, our evidence highlights the importance of interactions between executives and investors during earnings conference calls, road shows, financial press appearances, shareholder presentations, etc. Also, auditors responsible for attesting to a firm's financial statements could potentially use speech analysis of audit inquiries as an additional input into the assessment of fraud risk. The Public Company Accounting Oversight Board (PCAOB) has recently emphasized the importance of examining quarterly earnings calls as a means of detecting fraud (PCAOB 2010) and our research provides evidence in support of this assertion.

Third, this paper adds to prior work (Mayew and Venkatachalam 2011; Han and Nunes 2010) that uses LVA based software as a tool to measure vocal emotions by providing construct validity to one of the measures produced by the software. The results are consistent with LVA software capturing emotions resulting from researcher induced cognitive dissonance, suggesting the software might be useful for measuring the construct of cognitive dissonance more generally in other fields such as finance (e.g., Goetzmann and Peles 1997) and marketing (e.g., Koller and Salzberger 2007). Admittedly, cognitive dissonance reflects only one element of the myriad emotions experienced by deceivers. As such, we view our results as a starting point for future research that considers other emotions known to exist during deceptive speech with both existing and emerging technologies that capture emotional content in speech.

2. Prior Research and Hypothesis Development

Deception is a deliberate attempt to mislead. Financial misreporting is a particular type of deception intended to deceive a company's stakeholders. Unlike deception in everyday life in which ground truth and the context is generally well established (e.g., crimes such as theft or murder), financial misreporting encompasses a broader range of deception ranging from outright falsification to subtle manipulation of numerous line items in the financial statements (e.g., earnings management), to concealment of information (e.g., withholding bad news). Unfortunately, detecting deception that involves subtle deceit and exaggerations is particularly challenging (Vrij 2008). Furthermore, public revelation of the specific context of financial misreporting is often delayed until established by an external agency such as the SEC. Thus, timely prediction of misreporting is useful for providers of capital, auditors and information intermediaries alike.

Prior archival work in financial accounting has primarily explored the predictive ability of financial variables (Beneish 1997; Dechow, et al. 2010) and nonfinancial performance measures (Brazel, et al. 2009) in detecting financial misreporting and assessing fraud risk.[2] While quantitative information contained in the financial statements represents a significant component of the overall communication of a firm's strategic decisions and outcomes, managers supplement mandated disclosures with press releases, conference presentations and earnings conference calls to elaborate on previously disclosed information or provide timely new information to market participants and contracting parties. With the passing of Regulation FD in 2000 and subsequent innovations in technology over the last decade, vast audiences can access corporate communications in the form of audio, and more recently, video broadcasts. Such communications, particularly spontaneous and interactive communications of firms with analysts during

[2] Experimental work in managerial accounting has also investigated misreporting, primarily with a focus on how the design of management control systems reduce lying and/or promote honesty. Salterio and Webb (2006) provide a review of this literature. Our investigation differs from this stream of literature in that our interest lies in understanding whether a tool for capturing vocal markers of cognitive dissonance is useful in detecting misreporting as opposed to designing a control system that prevents misreporting in the first place.

earnings conference calls, expose market participants to richer information sets that include both verbal and nonverbal cues.

Market participants and regulators seem to appreciate the potential for these verbal and nonverbal cues to serve as indicators of misreporting and fraud risk. Anecdotal evidence suggests that equity research firms employ former CIA (Central Intelligence Agency) agents to identify verbal and nonverbal clues to deception during corporate earnings conference calls (Javers, 2010). The Public Company Accounting Oversight Board (PCAOB) recently issued Auditing Standard No. 12, which explicitly mandates that auditors consider "observing or reading transcripts of earnings calls" as part of the process for identifying and assessing risks of material misstatement (PCAOB 2010). Unfortunately, the precise verbal and nonverbal cues identified and used by equity research firms is proprietary, and the authoritative auditing standards are silent on what specifically an auditor should "observe" from an earnings call. To begin to fill this void, we examine whether vocal markers of cognitive dissonance can assist in predicting financial misreporting.

For both practical and theoretical reasons we focus on vocal cues, in particular, vocal markers of cognitive dissonance. Research in psychology (e.g., Zuckerman, DePaulo and Rosenthal 1981, Horvath 1979) suggests several ways to identify a deceiver from a truth teller, including cues from physiological traits (e.g., blood pressure, heart rate, brain activity), kinesics (e.g., facial expressions, body movements), word usage (lexical features) and vocal profiles. We focus on nonverbal vocal deception markers, as opposed to other deception markers, for the following reasons. First, capturing physiological changes of corporate executives during earnings conference calls via standard methods such as brain fMRI (functional magnetic resonance imaging) or skin conductance tests is neither possible nor practical. Second, verbal linguistic markers of deception in the financial reporting context is the subject of numerous recent studies (Burns, et al. 2010; Loughran and McDonald 2010, 2011; Larcker and Zakolyukina 2010; Purda and Skillicorn 2011; Humpherys, et al. 2011). Third, while software programs exist for measuring changes in facial expression (Meservy, et al. 2005; Jensen, et al. 2008), automated kinesics is still in a state of infancy. Moreover, video broadcasts of corporate earnings are still quite rare.

Finally, audio broadcasts of earnings conference calls are increasingly common and commercial software products such as Layered Voice Analysis (LVA) have emerged for the measurement of vocal deception markers. We are, as a result, able to answer the call by Hirschberg (2010) for more research on the ability of vocal cues to detect deception in real world settings, which has historically been hampered by lack of software for systematic measurement and/or real world speech corpus of sufficient audio quality.

Conditional on our interest in vocal cues, we then focus exclusively on vocal markers of cognitive dissonance. Research in psychology suggests that emotions are conveyed through an individual's voice (Juslin and Laukka 2003; Juslin and Scherer 2005; Scherer 1986), and deceivers commonly experience various emotions including fear, anxiety, guilt and shame. Some of these emotions stem from the prospect of being caught, while other emotions stem from cognitive dissonance. Cognitive dissonance is a feeling of psychological discomfort felt when one's actions and beliefs are discrepant (Festinger 1957). DePaulo, et al. (2003) indicate that liars generally feel guilt and shame because they have done something they consider wrong. We focus on measuring emotions stemming from cognitive dissonance through the vocal channel for three reasons. First, Javers (2010) notes that former CIA agents hired by equity research firms search earnings conference calls specifically for markers of cognitive dissonance. Second, we have access to a commercial software product that purports to capture emotions related to cognitive dissonance. This software has recently been used in archival research to study the information content of executive emotion profiles during earnings conference calls (Mayew and Venkatachalam, 2011). Third, recent experimental research by Mazar, et al. (2008) directly links misreporting and cognitive dissonance. Mazar, et al. (2008) discuss the aversive feeling experienced by an individual during or after a dishonest action and argue that this aversion results because individuals generally view themselves as being honest and value this self-concept. They find that, in a setting where subjects are given incentives to misreport performance for personal gain, simple reminders of the emotional costs of deviating from the self-concept of honesty (i.e. cognitive dissonance costs) substantially dampen individuals' propensities to misreport.

Based on the discussion above, we hypothesize:

H1: The probability of financial misreporting is positively associated with the extent of cognitive dissonance markers contained in the vocal wave.

Our generic empirical design to test for the association predicted by H1 entails i) obtaining speech samples from a distribution of misreporters and truth-tellers, ii) measuring the level of cognitive dissonance contained in the vocal wave for each observation, and iii) assessing the predictive ability of dissonance markers for misreporting by estimating a logistic regression of the following form:

$$\Pr(\textit{Misreporting}) = f(\textit{Vocal Dissonance Markers}) \qquad (1)$$

In the subsequent sections, we discuss precisely how we operationalize these two constructs. Two conditions must hold for us to observe evidence consistent with H1: (1) misreporters must feel cognitively dissonant when they misreport, and (2) the LVA software must be able to identify vocal markers of cognitive dissonance stemming from misreporting without significant measurement error. Regarding the first condition, if corporate executives are inherently overconfident, they may never believe they are misreporting and in turn may not experience dissonant feelings. On the other hand, former Satyam Chairman B. Ramalinga Raju, in his letter admitting fraud, stated that he was carrying a "tremendous burden on his conscience." That a CEO would reference a burden on his conscience suggests that dissonance may occur even at the highest levels of corporate management. However, *ex ante* one might challenge that our predictions under H1 might not hold in an empirical archival setting.

Regarding the second condition, we are unaware of any systematic archival or experimental evidence that directly assesses the construct validity of LVA's measure of cognitive dissonance.[3] Mayew and Venkatachalam (2011) present indirect archival evidence suggesting that LVA captures cognitive dissonance. In theory, if managers hold beliefs that they are both competent and in control of their firm,

[3] Gamer, Rill, Vossel and Godert (2006) experimentally investigate LVA based cognitive dissonance levels as part of an overall assessment of all LVA metrics provided in an early version of the LVA software and find them to be higher for participants in the guilty condition than for those in the innocent condition, but not statistically different. Audio files constructed in Gamer et al. (2006) restricts experimental subjects to monosyllable verbal responses of "yes" and "no" to questions rather than free flowing responses, which may have significantly reduced the predictive power of the metrics generated by the LVA software (Palmatier 2005). See Mayew and Venkatachalam (2011) for a literature review of the studies investigating LVA based metrics.

dissonance would be higher in settings of poor financial performance and when the firm operates in more volatile environments. Consistent with this intuition, they find a negative (positive) correlation between the LVA based dissonance measure and firm performance as measured by return on assets or prior stock returns (stock price volatility). While this evidence is potentially consistent with the software capturing cognitive dissonance, it does not speak to cognitive dissonance arising from misreporting. Moreover, although corporate executives may feel dissonant when misreporting through speech training they be able to mask their emotions during conference calls in a manner undetectable by the software.

3. Generation and Analysis of Laboratory Generated Data

We first test H1 on data generated in a laboratory setting. A laboratory setting gives the LVA software its best chance to work because we can follow existing experimental work to (1) generate a distribution of misreporters and truth tellers, (2) infuse emotions associated with cognitive dissonance into misreporting subjects, and (3) obtain speech samples in a controlled setting that helps minimize background noise that can potentially contaminate speech samples. The laboratory setting also offers the opportunity to capture other measures of cognitive dissonance, which, if correlated with the LVA voiced based dissonance measure, would provide construct validity for the LVA dissonance metric. Naturally, finding results consistent with H1 on laboratory generated data does not ensure predictive power in an archival setting, but a lack of observing results on laboratory data would arguably make an archival investigation moot.

3.1 Design Overview

To estimate equation (1) we first need a distribution of truth tellers and misreporters. We have two options for generating such a distribution in the laboratory. First, we could sanction misreporting by randomly assigning subjects between a control condition in which truthful reporting is ensured and a treatment condition in which subjects are mandated to report deceptively (e.g., Frank and Ekman 1997; Newman, et al. 2003). Alternatively, we could monetarily reward misreporting in a constant manner across participants and allow subjects to endogenously choose whether they misreport or not for personal gain (e.g., Evans, et al. 2001; Mazar, et al. 2008). Sanctioning deceptive reporting is not an appropriate experimental research design for our purposes because emotions stemming from cognitive dissonance are unlikely to be present in individuals who are authorized to be deceptive (Harrigan, et al. 2005, p. 345). This is because authorizing deception provides an implicit affirmation to the subjects that deception is acceptable and hence, will not violate their own self-belief about honesty. Thus, the self-concept and the behavior are no longer at odds, a feature necessary to encounter cognitively dissonant feelings (Cooper 2007). Further, sanctioning deception lacks external validity. We therefore believe that a quasi experimental design (Cook and Campbell 1979), where we allow truth-tellers and misreporters to arise endogenously, is important for our research objective.

To amplify dissonance in misreporting subjects, we utilize a design feature of Mazar, et al. (2008), who show that binding cognitive dissonance costs can be generated in a laboratory setting. In particular, they show that it is possible to increase the emotional costs of deviating from personal honesty norms (i.e., costs from cognitive dissonance) by reminding subjects of their personal moral codes via subject recitation of the Ten Commandments. This simple moral code reminder alters the endogenous choice to misreport for personal gain. We capitalize on this feature in our design by first incentivizing misreporting for personal gain, thereby generating an endogenous sample of truth tellers and misreporters.[4] However, instead of curbing the misreporting *ex ante* with dissonance costs, we infuse reminders of moral codes *after* participants have reported their score on a private task. The purpose of this shift in timing of the moral code reminder is to exacerbate cognitive dissonance in participants who have misreported. Through the moral code reminder, we attempt to stimulate the emotional burden of cognitive dissonance, the emotional markers of which the LVA software claims to capture from voice. We describe the research design in more detail below.

[4] A byproduct of our quasi-experimental design is that we are unable to pinpoint the relative causal power of the different aspects of our experimental design on the incidence of misreporting. That is, we only run one "cell" and in that cell we provide both monetary incentives to misreport and a moral code reminder to exacerbate dissonant feelings. However, we note that this type of causal analysis is not germane to our goal of generating an endogenous sample of misreporters and truth tellers.

3.1.1 Design Timeline and Procedure

Fifty-nine undergraduate volunteers from two large public U.S. universities participated in a two part experiment (see timeline of events in Figure 1). Participants were 37% female, with the median age of 20 years, in their sophomore year, and have completed three (one) math (English) college courses (see Panel A of Table 1). The first part of the experiment was an online portion containing initial, general instructions, Scholastic Aptitude Test (SAT) background instructions and examples, and a self-timed, five-minute SAT test.[5] Participants were given 4 points for each correct answer, -1 point for each incorrect answer, and 0 points for each skipped question. Responses were graded automatically through the online interface, which revealed to the participant how many questions had been answered correctly. After receiving this feedback, subjects were asked to predict how many SAT questions they could answer correctly if they were to take this SAT test again using similar questions. The answer to this question captures the subjects' beliefs about their ability to answer SAT questions before entering the laboratory and we label this *BELPRE*. As a whole, the purpose of this online portion was to i) re-acquaint the student with SAT questions, and ii) initiate a prediction of self assessed ability. Panel A of Table 1 reveals that average subjects scored 11.63 points and believed they could answer 6.00 SAT questions correctly if given an additional 5 minutes.

After completing the online portion of the experiment, participants completed the laboratory portion of the experiment.[6] The laboratory portion consisted of four activities: (1) taking, scoring and reporting the results from a timed, five-minute SAT test, (2) filling in answers to a set of questions on a midpoint questionnaire, (3) answering a set of interview questions while being video recorded, filling in

[5] At one of the universities, students traditionally take the ACT exam to qualify for admission. As such, at that university, we labeled all materials ACT instead of SAT.

[6] Aside from answering the timed SAT questions, participants were able to complete the online portion of the experiment at their own time and pace. We made sure that participants had finished the online portion before starting the laboratory portion, and that they completed the online portion only once. The duration between online completion and the laboratory portion ranged from 14 days to one hour (median number of days equals 1). The number of days is not significantly correlated with participants' scores on either the laboratory SAT questions or the online SAT questions.

answers to a set of exit questionnaire questions, predicting performance on a future SAT test, and finally, (4) being paid and debriefed.

The laboratory portion of the experiment proceeds as follows. First, after some brief initial instructions by a student administrator, participants took a timed, five-minute SAT test. This test had questions that were similar, though not identical, to those in the online portion of the study. Next, the participants self graded their answers and reported an overall score of their performance on a separate score sheet, which ultimately determined their payoff. Participants were informed they would be permitted to retain their test sheets and only needed to hand in a sheet containing their reported score for determining payoffs. The student administrator left the experiment room both when the SAT test was taken and when the SAT test was self-scored. During this time, the participants had both the test form and the answer sheet. The purpose of informing participants that they would not be turning in their original testing sheets, of having the student administrator leave the room, and for using a student for administration instead of an experimenter was to lower perceptions of monitoring, and in turn invoke misreporting in subjects.

Second, participants answered a midpoint questionnaire. The purpose of this questionnaire was to obtain demographic information and make the participants cognizant of their own personal moral code. To this end, we asked participants to write down as many of the Ten Commandments as they could remember. Participants are likely to be aware that the Ten Commandments represent a moral code regardless of their personal religious beliefs (Mazar, et al. 2008).[7] This moral code reminder was intended to invoke emotions associated with cognitive dissonance.

Next, an experiment administrator separately videotaped each participant's answers to interview questions in a separate interview room. All instructions and interview questions were prerecorded and sequenced with a PowerPoint presentation. The experiment administrator operated the video equipment, played and advanced the prerecorded audio in the PowerPoint, and prompted the interviewee to expand

[7] We find that the Ten Commandments were widely known to our subject pool. On average participants correctly recalled five of the Ten Commandments. More than 90% of participants correctly recalled at least two of the Ten Commandments.

their answers if the answers were overly short. Thus, the interviewer's interaction with each participant was minimal and the interviewer did not alter interview questions. This minimized differences between the participant's interactions with the experiment administrator and helps remove perceptions by subjects that the questioning was in itself a strategic interaction. Naturally, in real world settings, questions and answer dialogs between corporate executives and analysts are likely strategic and dynamic, with subsequent analyst questions being conditioned on answers to preceding questions. While our experiment lacks this realism, we crafted interview questions to parallel the type of questions commonly asked in an earnings conference call. The prerecorded interview had seven questions. The first question was innocuous and calibrated the participant's voice for the vocal emotion analysis software. The remaining six questions pertained to reported performance of the participant, with questions ranging from general to specific, similar to the progression of questions earnings conference calls.[8] For example, the first question asked the subject to verbally repeat the score they reported on the score sheet (much as managers repeat reported earnings from the press release when beginning discussions on a quarterly earnings conference call). The second question asked whether the reported score was better or worse than expected, and the third question asked about what the most difficult portions of the test were and why. Importantly, no question directly asked whether the subject was able to achieve their reported score via misreporting, and as such we never explicitly solicited for participants to make an untrue statement. This is an externally valid feature of our design, in the sense that in capital market settings, executives are rarely, if at all, asked questions about misreporting *ex ante*. In addition, it is likely that executives minimize the number of outright lies they tell, due to litigation costs and other concerns. We do suspect,

[8] The calibration question is as follows: "To help us calibrate our equipment and make sure we are ready, please do the following three things: Describe the room you are sitting in. Spell the following words letter for letter: Dictionary and Abbreviation. Read the following numbers aloud: 1,965; 818; 11,757;" The six interview questions were as follows. (1) "First, please restate the score that you wrote down on the answer sheet." (2) "Was your performance on the SAT questions you just answered better or worse than how you have done in the past? Please explain your answer in detail." (3) "Which types of SAT questions did you find most difficult to answer? Why were these questions so difficult? What strategies did you use to answer these difficult questions? Please explain in detail." (4) "Many of our participants score below 10 points on these SAT questions. Describe as completely as possible how you were able to achieve the score you reported." (5) "Overall, how do you feel about the SAT score you just reported? Please explain in detail." (6) "How would you respond to someone that told you they thought the SAT score you just reported was too high? Please explain your answer in detail."

however, that executives are more prone to avoid telling the whole truth, especially when doing so would be detrimental to the executive (Kothari, et al. 2008; Roychowdury and Sletten 2009).

Subsequent to the interview, participants were taken back to the original room and answered a final questionnaire containing additional demographic and manipulation check questions. Also, participants predicted the number of SAT questions they could answer correctly in a hypothetical future session, the answer to which we label *BELPOST*. This prediction question had the exact same wording as the prediction question at the end of the online survey (*BELPRE*). Changes in belief (*BELREV*) about the ability to successfully answer SAT questions is simply the difference between *BELPOST* and *BELPRE*. *BELREV* is an attitude-change measure from the induced compliance cognitive dissonance paradigm (Cooper 2007; Harmon-Jones and Mills 1999; Elliott and Devine 1994). Specifically, one way participants can resolve the cognitive dissonance resulting from overstating their SAT score is to modify their belief about how competent they actually are at answering SAT questions. Being unable to change their misreporting behavior post interview, participants may instead modify their beliefs in order to resolve their cognitive dissonance. Thus, we would expect that participants who have overstated their SAT score would have a relatively higher second prediction score, and hence higher values of *BELREV*. This measure serves as an alternative indicator for the presence and resolution of cognitive dissonance.

Finally, participants were individually taken to a separate room to be paid and debriefed. Participants were paid $5 for completing the online survey, and $10 for coming to the laboratory portion of the experiment. As mentioned before, each participant was compensated based upon the *self-reported* number of points scored on the SAT test. The average participant reported scoring 24.69 points and each participant received $0.50 for each point, yielding an average payout of $12.35. In addition, all participants were entered into a random drawing for one of two $500 prizes. Excluding these $500 prizes, participants earned $27.35 on average.

The laboratory portion of the experiment took an average of seventy five minutes. The videotaping portion of the experiment took an average of five minutes. After receiving their payment, participants were informed of the research purpose and that the researchers had intended that some participants would overstate their true SAT score. Once assured that overstatement was in fact something the researchers had expected to see, participants were asked whether they had overstated their score. Response to this question formed our main variable of interest, *MISREP*, which is coded as one if the participant admitted overstating the SAT score and zero otherwise. Providing this debriefing information to the subjects after payment was a mandatory condition in our design to ensure that the subjects were not harmed from our experimental infliction of cognitive dissonance costs.

Table 1 Panel A reveals that 32.2% of our participants admitted to misreporting. This is comparable to the proportion of lying reported in Chow, et al. (1988), Waller (1988) and Webb (2002) of 34%, 24% and 24%, respectively. We face tradeoffs by relying on *MISREP* as our measure of misreporting. Confession does not ensure that all misreporters are identified, as it is possible that some participants misreported but did not confess.[9] This introduces noise in the dependent variable in equation (1) that would bias against finding an association between vocal cues and *MISREP* in statistical tests. Alternatively, it is possible that only misreporters that felt dissonance chose to confess. This issue can be viewed in two ways. First, if one takes as given that the LVA software sufficiently captures cognitive dissonance, then using confession to identify misreporting biases towards finding results consistent with H1. On the other hand, if one questions whether the LVA software sufficiently captures vocal dissonance markers, using confession as our proxy for true misreporting is ideal because it gives LVA its best shot at detecting dissonance in a misreporting context.

To avoid this measurement error in our dependent variable altogether, we could have directly identified misreporting by overt or covert (deceptive) means.[10] However, if participants perceived that we

[9] Such Type II errors are externally valid in the sense that all financial misreporting is not detected (Dechow, et al. 2010). Subjects could also claim to misreport even though they had not misreported, although we view such errors as unlikely.

[10] Deceiving the subjects to obtain an error free measure of misreporting could be accomplished in a myriad ways, including but not limited to, secretly video recording subjects and secretly marking the test sheets and providing a waste basket where we *ex post* "dumpster dive" to identify the true performance of each subject. Alternatively, one could follow the protocol in Zhong et al. (2010) and embed secret codes into the score sheets as a mechanism to identify true task performance. The implicit benefit in these setups is that the researcher may retain the ability to verify the incidence and magnitude of misreporting, while convincing subjects the risk of identification of true performance is small. It is unlikely that we would have been able to capture these benefits with our subjects. In pilot tests, we found that in order to get subjects to misreport, we had to do all three of the following: state that

would be able to verify their misreporting even with a low probability, it would have been very difficult to induce misreporting. Hence, to increase the incidence of misreporting, as part of our research design we ensured that misreporting was unverifiable *ex post*.

3.2 Vocal Measurement Of Emotions Stemming from Cognitive Dissonance

To generate speech samples for analysis, we replay each video and manually isolate only the audio of each participant's answers to the interview questions. The resulting audio files were then analyzed using a commercial version of the LVA software developed for business applications called the Ex-Sense Pro R (version 4.3.9) Digital Emotion Analyzer. This software has been used in archival work to measure emotion profiles of corporate executives in the capital markets (Mayew and Venkatachalam 2011).[11] LVA is comprised of a set of unique proprietary signal processing algorithms to identify different types of stress, cognitive processes, and emotional reactions. The algorithms measure features of the speech waveform to create a foundation for indentifying the speakers' emotional profile. Because the waveforms are inherently person specific, the software measures deviations from a calibrated baseline for the speaking subject. Measuring deviations from a calibration benchmark is important because vocal parameters can vary across subjects due to innate differences in the physical generation of vocal waves.

The Ex-Sense Pro R software produces four "fundamental" voice based measures, labeled Emotional Stress Level, Cognition Level, General Stress Level and Thinking Level. We restrict our attention to Cognition Level because it is purported to measure cognitive dissonance.[12] The software also produces other measures deemed "conclusion" variables (e.g., Lie Stress), which are proprietary

---

participants could keep their test forms and score sheets, use a student administrator (because a peer was viewed as less of a monitoring threat) and have the student administrator repeatedly leave the room (again to lower the perception of monitoring). Despite these measures, participants in the sample analyzed here frequently stated they still had doubts that somehow their private task performance would be revealed. As such, we weighed the ethical costs of deceiving the subjects (see Bonetti 1998; Hey 1998; McDaniel and Starmer 1998) against the potential benefit of increased precision in the misreporting variable, and determined it was not cost beneficial.

[11] LVA based software products have been used in a variety of contexts for measuring other emotions such as embarrassment, stress associated with post traumatic stress disorder, and for detecting deception in experimental and field settings. See Mayew and Venkatachalam (2011) for a review of this literature and more generally for understanding the process of extracting voice based emotion markers.

[12] Ex-Sense Pro R user manual states: "Cognition Level reflects a situation when two or more non-complimentary logical processes are "processed" in the brain, for example, a logical conflict between what the mouth is saying and what the brain thinks. This is also referred to as cognitive dissonance (Festinger 1957)."

combinations of the fundamental measures and are meant to indicate when a speech segment may represent untruthful statements. Because our laboratory setting is specifically designed to evoke cognitive dissonance and in the interview we purposefully do not ask direct questions to which the answers will necessarily be untruthful, we do not consider the other fundamental or conclusion variables produced by the software. Moreover, the literature has found little evidence that the built in LVA conclusion variables perform better than chance levels, but suggests the more primitive fundamental level variables offer better predictive ability (Elkins 2010; Elkins and Burgoon 2010).

Discussions with the software developer suggest Cognition Level values greater than 120 are indicative of dissonance levels that require attention (see also Mayew and Venkatachalam 2011). Hence, we measure cognitive dissonance, *COGDIS*, as the number of utterances yielding Cognition Level values greater than 120 divided by the total number of utterances. An utterance is the voice wave segment automatically isolated by the software that occurs roughly over a two-second interval. Panel A of Table 1 reveals an average *COGDIS* in our sample of 0.217 and a standard deviation of 0.088, similar to those reported in Mayew and Venkatachalam (2011).

3.3 Association between Voice Based Dissonance Markers and Belief Revision Dissonance Markers

Before proceeding to estimate equation (1), we begin by examining whether the LVA based marker of cognitive dissonance, *COGDIS*, is associated with the belief revision measure, *BELREV*. Recall that *BELREV* captures the revision in a participant's beliefs in order to resolve dissonant feelings. Belief revision is a classic *ex-post* indicator that cognitive dissonance was present (Cooper 2007). If both *BELREV* and *COGDIS* capture the latent construct of cognitive dissonance, they should be positively correlated. Consistent with this intuition, the Spearman correlation is positive and statistically significant ($\rho = 0.333$, $p = 0.010$). The Pearson correlation is also positive but marginally significant in a one tailed test ($\rho = 0.192$, $p = 0.072$ one tailed).

While both measures are ways to operationalize cognitive dissonance, they differ in one critical aspect. *COGDIS* is measured at small intervals throughout the entire speech sample, whereas *BELREV* measures the revision in beliefs of participants from the beginning to the end of the laboratory session. The intuition for *BELREV* is that participants experiencing dissonance from misreporting attempt to remove the uncomfortable feelings associated with dissonance by upgrading their beliefs about their own ability to answer test questions. Prior research does not indicate *when* this belief revision occurs other than finding that the revision has occurred prior to the completion of the laboratory study (e.g., see references in Cooper 2007). In theory, if the LVA software is capturing emotions associated with cognitive dissonance as they occur, voice samples analyzed in between the invocation of dissonance and the point of belief revision should better capture the latent cognitive dissonance construct. Empirically, this would imply a stronger association between *COGDIS* and *BELREV* when *COGDIS* is measured between the point of dissonance invocation and belief revision than between belief revision and the end of the interview.

To test this conjecture, we define early (late) vocal based dissonance, *E_COGDIS* (*L_COGDIS*) as *COGDIS* from the first (second) half of the interview. The first half of the interview immediately follows the moral code reminder, which is our invocation of dissonance, and hence is more likely to overlap with the period where cognitive dissonance is most likely to be present. If true, the association between *E_COGDIS* and *BELREV* should be stronger than the association between *L_COGDIS* and *BELREV*. Panel B of Table 1 is consistent with this intuition. The Spearman (Pearson) correlation between *E_COGDIS* and *BELREV* is 0.440 (0.320) with a p-value of 0.001 (0.014). In contrast, the respective correlations between *L_COGDIS* and *BELREV* are still positive but much lower in magnitude at 0.098 (0.062) and not statistically significant. These results are suggestive that the LVA software captures cognitive dissonance as it occurs. Further, it suggests that in our prediction model for misreporting, allowing *COGDIS* to vary by whether it was measured early or late in the interview will offer a more powerful specification.

3.4 Predictive Power of Voice Based Dissonance Markers for Identifying Misreporting

We test H1by estimating logistic regressions with the following empirical counterpart of equation (1) using robust standard errors:

$$\Pr(MISREP) = \alpha_0 + \alpha_1\, COGDIS + \varepsilon \tag{2a}$$

$$\Pr(MISREP) = \beta_0 + \beta_1\, E\_COGDIS + \beta_2\, L\_COGDIS + \nu \tag{2b}$$

where equation (2b) explicitly allows the dissonance markers to vary early and late in the interview. In model (2a), we expect $\alpha_1 > 0$ if vocal dissonance markers can identify misreporting. In model (2b), because we expect cognitive dissonance to be mitigated through attitudinal change, we predict that *E_COGDIS* exhibits a stronger positive relation with misreporting than *L_COGDIS*. That is, we expect that $\beta_1 > 0$ and $\beta_1 > \beta_2$ in equation (2b).

Column A of Table 2 provides the results of estimating equation (2a). The coefficient on *COGDIS* is positive and marginally significant (p-value = 0.09 one tailed). To assess predictive ability, we use the area under the Receiver Operator Characteristic (ROC) curve, a technique originally used in signal detection theory (see Hosmer and Lemeshow 2000). ROC curves help assess the overall discriminatory ability of predictor variables as well as facilitate comparison among alternative predictor variables. In simple terms, ROC curve is a graphical plot of the probability of detecting a true signal (Type I error, also called sensitivity) against a false signal (1-Type II error, also called specificity). To plot the curve, it is necessary to estimate the Type I and Type II errors for various "cutoff points" used for classifying the continuous predicted probabilities from the logistic regression in a binary fashion. For example, the predicted probabilities in equation (2a) would help identify the proportion of Type I and Type II errors for each cutoff point ranging from (0 to 1). The area under the ROC curve (AUC) is a summary of the overall diagnostic accuracy, with values of 0.500 representing chance levels and 1.000 representing a perfectly accurate prediction model. Column A of Table 2 reveals that the AUC for prediction equation (2a) is 0.602, but not statistically different from chance level of 0.50 (p-value = 0.424).

Column B presents the results of estimating equation (2b), where we allow the vocal measure of dissonance to vary dynamically. We find that the coefficient on *E_COGDIS* is statistically significant at 1% significance level, whereas *L_COGDIS* is not statistically different from zero. Also, *E_COGDIS* is more positive than *L_COGDIS* (p value of F test = 0.013). More important, the area under the ROC curve for equation (2b) is 0.670, which is statistically greater than chance levels by 17% (p value = 0.027). Collectively, the evidence is consistent with the LVA software capturing markers of dissonance in the voice and with such markers being predictive of misreporting at better than chance levels. The results also suggest that, in our laboratory setting, cognitive dissonance is short lived. The predictive dissonance signals come from the early portion of the interview, which represents a speech sample with an average length of about a minute and a half.

3.5 Robustness of Vocal Cues as Predictors of Misreporting

To ensure the robustness of the predictive ability of vocal cues, we conduct a number of additional tests, the results of which are presented in the remaining columns of Table 2. First, to ensure that our results are not due to outlier covariate patterns, we re-estimate equation (2b) after removing three observations that appeared to represent outlier covariate patterns from visual inspection of plots of the following logistic regression diagnostics (Hosmer and Lemeshow 2000): standardized Pearson residuals, deviance residuals, leverage, change in Pearson Chi-Square and change in deviance.[13] Estimation of equation (2b) with the remaining 56 observations reveals a positive and statistically significant coefficient on *E_COGDIS* of 10.293 ($p < 0.01$), implying the results are not driven by outliers (see Column C of Table 2). The AUC is 0.681, which is of similar magnitude to that reported in Column (B), and significantly better than chance (p value = 0.018).

Second, since our data was generated at two separate universities, we investigate whether the results are sensitive to the university where the data was generated. We re-estimate equation (2b) separately for the 24 (35) observations generated at the first (second) university, and provide the results in

[13] Specifically, we removed observations with absolute standardized Pearson residuals greater than 2, absolute deviance residuals greater than 2, leverage greater than 0.1, change in Pearson Chi-Square greater than 4 and change in deviance greater than 3.5.

Column D (E) of Table 2. Despite the small sample sizes in each of the regressions, the results are consistent with the overall sample results, as we find a positive and significant coefficient on *E_COGDIS* of 17.881 (12.990) at the first (second) university. The AUCs are 0.752 and 0.708, respectively, and both reject chance levels of 0.500 (p values of 0.019 and 0.048, respectively). If one views these two laboratory samples as independent because they were generated at separate universities on separate dates, the results confirm that the predictive ability of the LVA software can be replicated.

Third, we re-estimate equation (2) including all subject demographic variables, including age, year in school, number of math and English classes, ability as proxied by the number of points scored during the online portion of the experiment, and gender. If experience or ability enable certain subjects to better hide their emotions, the predictive ability of vocal cognitive dissonance markers for misreporting may be understated. In Column F of Table 2, we observe a positive and significant coefficient on *E_COGDIS* of 9.057 ($p < 0.05$), despite the inclusion of demographic variables.

Fourth, we consider a different proxy for misreporting. Given our laboratory design, we are unable to observe the "true" score as we rely on subject confessions to identify misreporters. As such, it is possible that our dependent variable *MISREP* is measured with error. To ensure robustness, we use an alternative proxy for misreporting by comparing the self reported score with a benchmark score. We use the actual scores on the pretest (the online portion prior to the lab experiment) as a benchmark for the true "expected" score. Admittedly, this is a crude and noisy benchmark, but it has the advantage of capturing subject specific performance in a setting that is devoid of incentives.[14] Thus, our alternative misreporting proxy (*USCORE*) is the difference between the self-reported score (*SCORE*) and the score obtained in the online portion of the experiment (*SURVEY*). That is, *USCORE = SCORE – SURVEY*. In our sample, the average value of *USCORE* is 13.059 as reported in Panel A of Table 1, suggesting that participants performed better in the laboratory portion of experiment on average. Results of estimating an OLS regression with USCORE as the dependent variable are reported in Column G of Table 2. The coefficient

[14] Recall that all participants received a flat fee of $5 for completing the online portion of the experiment.

on *E_COGDIS* is positive and statistically significant, buttressing our earlier findings and suggesting that the measurement error in *MISREP* may not be large.[15]

Finally, to ensure the inferences from Columns B through D are not impacted by the correlation between *E_COGDIS* and *L_COGDIS*, we re-estimate each column and include only *E_COGDIS*. In untabulated results, we find that the coefficient on E_COGDIS is positive and significant in every case, and observe no qualitative difference in the magnitude or statistical significance of the AUCs. Despite these robustness checks, we acknowledge that given our lack of random assignment of subjects in our quasi experimental laboratory design, it is possible that some omitted and unknown factor(s) drive the association between measured vocal dissonance markers and confessed misreporting. This would inhibit a generalization of our laboratory results to our setting of interest, which is financial misreporting in a capital market setting. As a result, we next turn to the archival setting.

4. Empirical Analysis Using Archival Data

4.1 Design Overview

To assess whether the aforementioned findings generalize to the archival setting, we expand our conceptual prediction model outlined in equation (1) as follows:

Pr(*Misreporting*) = *f*(*Vocal Dissonance Markers, Controls for Non-Misreporting Dissonance Drivers, Financial Statement Based Predictors of Misreporting, CEO characteristics*) (3)

Relative to equation (1), equation (3) adds three additional conceptual components to the prediction model of misreporting. First, we control for factors other than the act of misreporting that may cause cognitive dissonance in executives. Recall that cognitive dissonance induces negative emotions due to a disjoint between beliefs and actions.[16] Suppose CEOs believe they are honest, competent, and in

[15] We do not use *USCORE* as the primary variable in the manuscript because *USCORE* attains some negative values (which have no theoretical analog) and is a noisy measure because the online portion of the experiment differs in several necessary ways from the in-lab portion of the experiment.

[16] Note that the LVA software measures cognitive dissonance after taking into account the baseline vocal characteristics in the calibration phase of the speech analysis. Therefore, any dissonance that is felt by an executive due to other factors and therefore, inherent in his vocal characteristic is likely to be differenced away by the LVA software because LVA calibrates each executive's speech for their unique vocal characteristics at the beginning of the speech. Nevertheless, for completeness, we include other drivers of dissonance in the model.

control of their firms. In such a case, dishonest reporting, actions that result in poor performance, and actions that result in volatile firm outcomes would contradict, respectively, these three held beliefs and in turn inflict cognitive dissonance. The LVA software does not distinguish between sources of cognitive dissonance, but rather measures the extent of dissonance present from whatever source. Mayew and Venkatachalam (2011) provide evidence consistent with higher levels of dissonance in poorly performing firms, and in firms that are smaller and more volatile. Therefore controlling for these non-misreporting sources of dissonance should yield a more powerful specification when attempting to predict misreporting from vocal dissonance markers. In our empirical specifications, we control for performance via return on assets, unexpected earnings and prior year stock returns. We control for uncertain environments with firms size and stock return volatility.

The second addition pertains to existing financial statement based predictors of misreporting. Since our objective is to assess whether, and to what extent, vocal dissonance markers predict misreporting, adding known predictors of misreporting provides a benchmark against which we can compare the predictive ability of vocal dissonance markers. Moreover, we can assess whether vocal dissonance markers provide incremental predictive ability to financial information. We use two summary metrics from the recent accounting literature to assess the predictive ability of financial statement data in our setting. The first metric is the F-Score, developed by Dechow, et al. (2010). The second metric is a commercially available summary metric, called accounting risk, that recent work shows to be a potent predictor of misstatements (Correia 2010; Price, et al. 2010).

The third addition pertains to executive characteristics. If older or more seasoned executives are better able to control their emotions and have lower incentives to misreport due to lessened career concerns, we may observe both lower dissonance levels and lower levels of misreporting for experienced executives. That is, the hypothesized positive association between vocal dissonance and misreporting may be driven by executive age and tenure. So, we control for both of these factors in our empirical specification.

4.2 Executive voice data

An ideal dataset that mimics our laboratory setting would contain a sample of executives who are deceptive, and a matched sample of executives in firms with similar economic characteristics but who are not deceptive. Achieving this ideal is difficult for two reasons. First, audio files of executives speaking during earnings conference calls are publicly available for relatively short periods of time, perhaps due to litigation risk concerns. Although transcripts of the conference calls are available for a large cross section dating back to the passage of regulation FD, the related audio files are re-streamed over the internet for periods typically ranging from one fiscal quarter to one fiscal year from the earnings call date. Data providers such as ThomsonReuters StreetEvents, who provide subscribers with restreaming access for periods specified by the firms, do not allow downloading of the audio files. Reseachers, therefore, face enormous data collection costs because the only way by which they can analyze the audio files is to restream audio files while publicly available. Adding to the costs, the researcher must manually isolate and extract the voice of the executive of interest from the conference call dialog in order to conduct an audio analysis for each executive's voice.

Data constraints notwithstanding, a second challenge is that we cannot ensure that managers will discuss a deceptive topic either voluntarily in the presentation or when probed by analysts in the Q&A (Hollander, et al. 2009), although it is unlikely that managers can avoid speaking about major economic factors impacting their firms altogether. Assuming a sufficient number of managers speak directly on the specific issues that are found to be *ex post* deceptive, a researcher could conceivably use restatements to identify the key deceptive topics and the related portions of the conference call dialog. The difficulty then would lie in isolating the specific moments when dissonance would manifest itself in voice.

Given these data collection challenges, we begin with a sample of audio files collected by Mayew and Venkatachalam (2011). Specifically, the sample in Mayew and Venkatachalam (2011) comprises 1,647 quarterly earnings conference calls spanning the period January 1 through December 31, 2007, and represent fiscal quarters from Q4 of 2006 through Q3 of 2007. These calls represent the set of quarterly earnings calls available on Thomson Reuters StreetEvents for which basic firm data is available from

CRSP, Compustat and I/B/E/S.[17] From this initial sample, we remove observations where the CEO does not speak during the Q&A section and where we are unable to obtain financial statement based predictors (i.e., F Score and accounting risk). Our final sample consists of 1,572 conference call observations. For each of these conference calls we analyze the CEO's voice during the first 5 minutes of conversation during the Q&A portion of the call and obtain the archival analog of the vocal based dissonance metric, *COGDIS*.[18]

4.3 Data and Descriptive Statistics for Archival Data

4.3.1 Misreporting Data

To identify executives who overstated performance as in the experiment, we use the Audit Analytics database to identify which of the firms in our sample restated their financials such that it resulted in a downward adjustment to earnings or equity. Specifically, we query the Audit Analytics database for such adverse restatements for a period of about 3 years following the end of calendar year 2007 (when our voice data acquisition ends), i.e., January 2008 to January 2011.[19] We are able to identify 113 firm quarters from our sample for which the Audit Analytics database reported a restatement announcement during this period. Audit Analytics identifies none of these adverse restatements as resulting from clerical errors, as frauds or as undergoing SEC investigations as of our query date. We define *RESTATE* as an indicator variable that takes a value of one for firm quarters with subsequent adverse restatement and zero otherwise.

4.3.2 Remaining Data

We construct our remaining variables, including the Dechow, et al. (2010) F-Score (*FSCORE*), using data from Compustat, CRSP, I/B/E/S, and Execucomp as needed. Additional executive demographic information not available in Execucomp is hand collected when necessary. The commercial

[17] See Mayew and Venkatachalam (2011) for a more detailed discussion of the data collection procedures.
[18] To calibrate the speech of each CEO, we use the opening moments of the CEO speech during the presentation portion of the conference call.
[19] Such a delay is unavoidable since it takes time for restatements to be identified. On average, among all restatements provided by Audit Analytics, the length of time between the beginning of a restatement period and the actual restatement announcement is 2.4 years.

misstatement predictor accounting risk (*ACCT_RISK*), developed and sold by Audit Integrity, LLC, identifies the risk of financial report misrepresentation due particularly to overstated (understated) revenue and assets (expenses and liabilities). *ACCT_RISK* is based exclusively on financial statement information (Correia 2010), is available on a quarterly basis, and is a parsimonious summary metric that has been shown to perform as well as or better than other accounting based prediction models in the literature (Price, et al. 2010; Correia (2010).[20] A drawback of this measure is that we cannot state precisely which accounting metrics are the critical drivers behind its predictive ability. *ACCT_RISK* ranges from 0 to 100, with low risk receiving higher *ACCT_RISK* scores. We modify *ACCT_RISK* by subtracting it from 100 so that higher values capture a higher likelihood of an adverse restatement.

4.3.3 Descriptive Statistics

Panel A of Table 3 provides descriptive statistics for the archival sample. We find that 6.9% of our sample observations, representing 53 unique firms, report an adverse restatement. Cognitive dissonance measure, *COGDIS*, has a mean (standard deviation) of 0.179 (0.076), which is similar to the 0.217 (0.088) reported in Table 1 Panel A for the laboratory setting. The market capitalization of the median firm in our sample is \$1.299 billion ($e^{7.169}$), which is substantially larger than the median market capitalization of an average Compustat firm of \$212 million in fiscal year 2006. In this respect, our analysis differs from other papers investigating the determinants of misstatements (e.g., Dechow, et al. 2010; Price, et al. 2010), which commonly use all available Compustat data. While our sample firms are larger, Panel C of Table 3 reveals that the proportion of sample firms across industries is similar to the Compustat population, with the exception of slight over (under) representation in pharmaceuticals (insurance/real estate).

In Panel B of Table 3, we observe several important bivariate correlations. First, our financial statement predictors of adverse restatements, *FSCORE* and *ACCT_RISK*, are positively correlated as expected, indicating that both variables capture a common construct. Both variables are also positively correlated with *RESTATE*, although only *ACCT_RISK* is statistically significant. Regarding our variable

[20] We thank Jack Zwingli of Audit Integrity for providing the *ACCT_RISK* data for our academic use.

of interest, we find a positive and significant association between *COGDIS* and *RESTATE*. Several variables, however, are associated in the same direction with both *COGDIS* and *RESTATE,* suggesting the potential for correlated omitted variables to confound a univariate assessment. To draw more definitive conclusions and to quantify the predictive ability of vocal dissonance cues for misreporting, we turn to multiple regressions in the next section.

4.4 Multiple Logistic Regression Results

All of the specifications in Table 4 are logistic regressions where the dependent variable is *RESTATE.* In the first two columns, we estimate baseline models that include only vocal dissonance markers, which are analogous to the models estimated in the laboratory setting reported in Columns (A) and (B) of Table 2. The main difference is that in Column (B) of Table 4, we identify conditions where dissonance is more likely, by identifying high scrutiny settings in the cross section. In the laboratory, we identified settings where dissonance was more likely by partitioning the speech samples based on proximity to our cognitive dissonance manipulation. Such an analog does not exist in the archival data since we do not know the location during the audio file where a topic that invokes dissonance is discussed. We therefore follow Mayew and Venkatachalam (2011) and identify high (low) scrutiny settings by whether the firm missed (met or exceeded) analysts quarterly expected earnings. Formally, we define high (low) scrutiny dissonance, $COGDIS^{HS}$ ($COGDIS^{LS}$), as *COGDIS* when the firm missed (met or exceeded) the most recent I/B/E/S summary consensus median estimate, and zero otherwise. The intuition for this partition is that analysts become more scrutinizing when reported earnings do not meet analysts expectations (Graham, et al. 2005), and this enhanced scrutiny is more likely to increase demand for topics managers would rather not speak about, which in turn induces dissonant feelings on the manager.

Column (A) of Table 4 reveals that *COGDIS* is positive and marginally significant, and the area under the ROC curve of 0.552 marginally rejects chance levels (p = 0.085). In Column B, when we allow *COGDIS* to vary by whether the conference call was highly scrutinizing or not, we find a positive and significant coefficient on vocal dissonance markers in the high scrutiny setting (coefficient on

$COGDIS^{HS}$= 3.774, p < 0.05), but insignificant results in the low scrutiny setting (coefficient on $COGDIS^{LS}$= 1.198, p > 0.10). The predictive ability of the model improves to 8.4% above chance levels, as the AUC is equals 0.584 (p = 0.005). As Columns A and B together suggest, the predictive ability of vocal dissonance markers is most pronounced in settings where dissonance is more likely to be present, just as we observed in the laboratory setting. Further, these results are consistent with contemporaneous market reactions to managerial affect in Mayew and Venkatachalam (2011), who show that effects are most salient in high scrutiny setting.

To put these results in context, the AUC in Column B of 0.585 is an order of magnitude lower than analogous result of 0.670 reported from the laboratory generated data. This may result from a number of factors including, but not limited to (1) undergraduate participants feeling more dissonance than that felt by corporate executives, (2) audio files being of less quality in the archival setting, and (3) a larger fraction of type II errors in the dependent variable in the archival setting. We also note that the AUC we document in the archival setting is of similar magnitude to the AUC of 0.564 reported in Larcker and Zakolyukina (2010), who model restatements solely on linguistic deception markers from CEO (CFO) speech during earnings conference calls.[21]

To provide further comparison for the Column B results, in Column C of Table 4 we model adverse restatements solely as a function of our two financial statement based predictors, *FSCORE* and *ACCT_RISK*. Consistent with the univariate results we observe a positive coefficient on both variables, but only the coefficient on *ACCT_RISK* is statistically significant (p<0.05). The resulting AUC is 0.588, which is statistically better than chance (p < 0.01). Thus focusing on either vocal dissonance cues alone or financial accounting predictors alone provides better than chance predictive ability of approximately 9%. A test of equality of the AUC in Columns B and C cannot be rejected (p = 0.950).

[21] We consider the role of linguistic predictors as a robustness test.

The relative lack of statistical significance on the Dechow, et al. (2010) *FSCORE* stems from at least three sources.[22] First, our sample is less than 2% of the sample used in Dechow, et al. (2010) and our sample has a greater concentration of large firms. Second, Dechow, et al. (2010) construct accounting variables based on annual data and conduct the prediction analysis of misstatements at the firm year level, whereas the analysis we conduct is on a firm quarter basis and as such our accounting change variables are based on quarterly seasonal changes. Finally, the determinant model in Dechow, et al. (2010) predicts AAERs, which represent the most egregious forms of adverse restatements.

In Column D, we include both vocal dissonance markers and financial statement predictors to assess whether each of the predictors is incrementally predictive. We observe positive and significant coefficients on $COGDIS^{HS}$ and on *ACCT_RISK*, with magnitudes qualitatively similar to the coefficients in Columns B and C, respectively. The AUC increases only slightly to 0.595 that is better than chance levels, but not statistically different at better than the 10 percent level from the predictive ability of either vocal dissonance markers or financial statement predictors in isolation. This implies that both vocal cues and financial information are incrementally useful in predicting restatements and that neither subsumes the other in terms of predictive ability.

In Columns E-G, we examine whether the results reported in Columns B-D are affected by the inclusion of control variables in equation (3). That is, we include controls for dissonance drivers that may stem from sources other than misreporting, such as firm performance (proxied by current year market adjusted stock returns, return on assets and unexpected quarterly earnings) and uncertainty (proxied by return volatility and firm size). The inclusion of unexpected quarterly earnings also helps control for potential main effects associated with scrutiny, as we extract the sign from unexpected earnings to partition *COGDIS*. We also control for CEO characteristics by including CEO age and CEO tenure at the firm, and industry fixed effects. The inclusion of these control variables does not change our inferences,

[22] In unreported analysis, we redefine *FSCORE* to include off balance sheet measures, non-financial measures and stock market based variables using the weights from Dechow, et al. (2010) but our inferences are unchanged. We also include the individual determinants of *FSCORE* in the regression instead of *FSCORE*, but observe positive and significant coefficient on only security issuances and the use of operating leases. Our inferences on $COGDIS^{HS}$ and *ACCT_RISK* are, however, unchanged.

and the coefficients on both vocal dissonance markers and financial statement predictors are of similar size and significance in Columns E-G as in Columns B-D. Of the control variables, the only significant predictor is firm size, which is inversely related with the restatements.

4.5 Further Robustness Tests

In our main analysis, we do not formally consider linguistic markers of deception in our sample because, despite a substantial amount of research in the area, there is not yet a consensus on an accepted linguistic measure to detect deception. Regardless, we provide some evidence on the robustness of the predictive ability of vocal cues to the inclusion of linguistic cues. We use the word list and weighting scheme in Newman et al. (2003) who examine the predictive ability of linguistic style by conducting experiments across different topics and contexts. Using a computerized text analysis program called LIWC, Newman et al. (2003) document that deceivers show lower cognitive complexity, used fewer self-referential words and more negative emotion words. The general prediction equation derived from Newman et al. (2003) takes the following form: *LIWCgpe* = 0.26*(*FP*) + 0.25*(*TP*)-0.217**NEGEMO* + 0.419**EXCL* - 0.259**MOTION*, where *FP* (*TP*) is the proportion of total words that are first (third) person pronouns, *NEGEMO* is the proportion of negative emotion words, *EXCL* is the proportion of exclusive words, and *MOTION* is the proportion of motion verbs. Higher values of *LIWCgpe* indicate less likelihood of deception. In untabulated results, a logistic regression with *MISREP* as the dependent variable and *LIWCgpe* as the independent variable yields an insignificant coefficient on *LIWCgpe*. When including *LIWCgpe* in the specification reported in Column (G) of Table 4, we find (results not tabled) that the coefficient on $COGDIS^{HS}$ is statistically positive and similar (coefficient = 3.392; p = 0.01) in magnitude to that reported in Column (G). The coefficient on *LIWCgpe* is negative as expected, and marginally significant in a one tailed test (p-value = 0.08).

One reason for the weak significance on *LIWCgpe* may be that the weighting scheme from Newman et al. (2003) does not generalize to a capital market setting. To explore this further, we include the five primitive variables that comprise *LIWCgpe* in the logistic model estimated in Table 4 Column (G) instead of the summary variable *LIWCgpe*. Including these primitive linguistic variables does not affect our inferences. Also, the only primitive linguistic variable that loads significantly is *MOTION*.

Collectively, the evidence is consistent with the hypothesis that vocal dissonance markers are robust predictors of misreporting in an archival setting. Vocal dissonance markers alone predict misreporting at better than chance levels, predict at levels similar to financial statement predictors alone, and are incrementally associated with restatements after controlling for financial statement based predictors. Further, the vocal markers are predictive even though extracted from one executive (the CEO) albeit for a short time - the first 5 minutes of the question and answer session.

5. Conclusions

We empirically document that vocal dissonance markers are useful for identifying misreporting. In a laboratory setting, we generate a sample of misreporters and truth-tellers and find that the vocal dissonance markers produced by the LVA software we use (1) are positively correlated with an alternative belief revision dissonance marker, and (2) can identify misreporters at better than chance levels. Extending these findings to the archival setting, we find that vocal markers of cognitive dissonance in CEO speech can also predict whether a firm's quarterly financial reports will be adversely restated at better than chance levels. The predictive ability of vocal dissonance markers is incremental to accounting based predictors of restatements. The effects in both the laboratory setting and archival setting are most pronounced in speech samples where cognitive dissonance is more likely to be present in the speaker.

The results in this paper provide some of the first archival evidence to suggest that important nonverbal clues to detect financial misreporting are available in earnings conference calls. These results should be informative to investors, analysts and auditors who attempt to use earnings conference calls as an information source for assessing the risk of misreporting. However, we caution the reader of the following limitations. First, while we attribute our findings to cognitive dissonance in subjects it is possible that some unknown emotional factor(s) correlated with this construct accounts for our results. Second, consistent with Mayew and Venkatachalam (2011), the predictive power of vocal dissonance markers is only salient in settings where analysts are highly interrogating, which implies more powerful vocal emotion analysis software may be necessary in settings where analysts are not scrutinizing management.

In spite of these limitations, we view our results as a starting point for expanding our understanding of how nonverbal cues play an informational role in capital markets. Numerous questions remain such as understanding whether emotions stemming from cognitive dissonance matter more or less than other emotions associated with deception, whether vocal dissonance markers incrementally inform about misreporting relative to linguistic cues, and to what extent humans can detect dissonance markers independent of voice analysis software. We view these issues as important areas for future inquiry.

Appendix 1

Variable Definitions – Laboratory Setting

MISREP	Indicator variable that equals 1 if the participant admitted to misreporting score in debriefing, 0 otherwise.
SCORE	Self reported score as stated in response to the first interview question.
SURVEY	Number of points scored during self-timed 5 minute SAT questions online. Four points are given for every correct answer, -1 point for every incorrect answer, and 0 points for every skipped answer.
USCORE	Unexpected score calculated as *SCORE - SURVEY*
BELPRE	Belief regarding number of correctly answered SAT questions that would be obtained in an additional 5 minute SAT test reported after online SAT question timed exam, but before laboratory portion of experiment.
BELPOST	Belief regarding number of correctly answered SAT questions that would be obtained in an additional 5 minute SAT test reported during debriefing after laboratory portion of experiment.
BELREV	*BELPOST – BELPRE*
COGDIS, E_COGDIS, L_COGDIS	LVA Ex-Sense Pro-R voice based measure of cognitive dissonance, measured as the number of voice segments registering greater than 120 on the Cognition Level measure during the entire interview session, divided by the total number of voice segments in the total interview session. Voice segments are approximate 2-second voice wave intervals. The prefix *E_(L_)* represents measurement during the first (second) half of the interview.
TIME_MIN	Then length of time in minutes of the entire vocal wave file used to generate *COGDIS.*
WC	Total number of words spoken in response to non-calibration questions during the interview session.
SCHOOL	Level of education, that equals 2 if subject is a Sophomore, 3 if a Junior, and 4 if a Senior.
MATH	Number of Math courses taken by participant.
ENGLISH	Number of English courses taken by participant.
AGE	Age of participant in years
FEMALE	Indicator variable that equals 1 if the participant was female, zero otherwise.

Variable Definitions – Archival Setting

RESTATE	Indicator variable that equals 1 if the firm's quarterly financial statements were restated (i.e. the fiscal quarter end falls between *RES_BEGIN_DATE* and *RES_END_DATE*) and the restatement had an adverse impact on the financial statements (*RES_ADVERSE = 1*) per Audit Analytics. At the time of our data extraction in September 2010 from Audit Analytics, the most recent restatement filing date was September 17, 2010.
ACCT_RISK	Accounting Risk, defined as the amount of accounting risk the company faces as of the firm's fiscal quarter end. Accounting risk is a financial statement based predictor of the risk that the financial statements are misreported and are provided by commercial vendor Audit Integrity, LLC. Values range from 0 – 100, with higher values indicator more accounting risk.
FSCORE	Scaled probability of misstatement, estimated as the predicted probability of misstatement scaled by the unconditional probability of misstatement from Dechow, et al. (2010) Table 7 Panel A Model 1. The predicted probability is equal to ($e^{\text{predicted\_value}}$ /(1+$e^{\text{predicted\_value}}$)) where the predicted_value = -7.893*RSST Accruals + 0.790*Change in Receivables + 2.518 * Change in Inventory + 1.979*% Soft Assets + 0.171*Change in Cash Sales – 0.923*Change in Return on Assets + 1.029 * Actual Issuance. Variable definitions in the prediction equation are quarterly versions of the annual definitions used in Dechow, et al. (2010), where changes are derived from the seasonal quarter. The unconditional probability is 494+(494+132,967) = 0.003701. All input variable for calculating predicted_value are winsorized at the 1% and 99% level.
COGDIS	LVA Ex-Sense Pro-R voice based measure of cognitive dissonance, measured as the number of voice segments registering greater than 120 on the Cognition Level measure from management speech during the quarterly earnings conference call, divided by the total number of voice segments from management speech during the quarterly earnings conference call. Voice segments are approximate 2-second vocal wave intervals.
$COGDIS^{HS}$	Cognitive dissonance in high interrogation settings, measured as *COGDIS* when unexpected earnings is less than zero, and zero otherwise. Unexpected quarterly earnings (*UE*) is less than zero for 515 of the 1,647 firm quarter observations.
$COGDIS^{LS}$	Cognitive dissonance in low interrogation settings, measured as *COGDIS* when unexpected earnings is greater than or equal to zero, and zero otherwise. Unexpected quarterly earnings (*UE*) is greater than or equal to zero for 1,132 of the 1,647 firm quarter observations.
RET	Current year market adjusted buy and hold stock return as estimated from CRSP, where market adjustment is based on the CRSP value weighted index. Buy and hold return is calculated for the trading days spanning the four fiscal quarters ending at quarter *t* for firm *i*.
VOL	Stock return volatility, measured as the standard deviation of daily stock returns over the half year period (trading days -127 to ,-2 relative to the conference call date).
UE	Unexpected earnings at fiscal quarter end, measured as the difference between actual I/B/E/S earnings per share and I/B/E/S analyst summary consensus median earnings per share scaled by price per share two days before the conference call
ROA	Return on assets, measured as income before extraordinary items divided by total assets at the beginning of the quarter.
AGE	Age of the CEO in years as of fiscal quarter end, as identified by Execuomp or hand collected as necessary.
TENURE	Number of years the CEO has been employed by the firm as of the fiscal quarter end, as identified by Execuomp or hand collected as necessary.

References

AICPA. 2002. Statement on Auditing Standards No. 99: Consideration of Fraud in a Financial Statement Audit. New York: AICPA.

Barth, M. E., W. H. Beaver, J. R. M. Hand, and W. R. Landsman. 2005. Accruals, accounting-based valuation models, and the prediction of equity values. *Journal of Accounting, Auditing and Finance*, 20 (4), 311-345.

Beneish, M.D. 1997. Detecting GAAP violation: Implications for assessing earnings manageement among firms with extreme financial performance. *Journal of Accounting and Public Policy* 16: 271-309.

Bloomfield, R. 2008. Discussion of "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and Economics* 45(2-3): 248-252.

Bond, C. F., and B. M. DePaulo. 2006. Accuracy of deception judgments. *Personality and Social Psychology Review* 10 (3):214-234.

Bonetti, S. 1998. Reply to Hey and Starmer & McDaniel. *Journal of Economic Psychology* 19: 397-401.

Brazel, J., K. Jones and M. Zimbelman. 2009. Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research* 47(5): 1135-1166.

Burgoon, J. and T. Qin. 2006. The dynamic nature of deceptive verbal communication. *Journal of Language and Social Psychology* 25 (1): 76-96.

Burns, M., K. Moffitt, W. Felix and J. Burgoon. 2010. Using lexical bundels to discriminate between fraudulent and non-fraudulent financial reports. Working paper, University of Arizona.

Chow, C., J. Cooper, and W. Waller. 1988. Participative budgeting: Effects of a truth-inducing pay scheme and information asymmetry on slack and performance. *The Accounting Review* 63 (1): 111-122.

Cook, T., and D. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Cooper, J. 2007. *Cognitive Dissonance Fifty Years of a Classic Theory*. Los Angeles, CA: Sage Publications.

Correia, M. 2010. Political connection, SEC enforcement and acounting quality. *Working Paper* London Business School.

Davidson, B. and D. Stevens. Can a code of ethics improve management behavior and investor confidence? Some intuition and experimental evidence. Working paper, Florida State University.

Davis, A., J. Piger and L.Sedor. 2008. Beyond the numbers: Managers' use of optimistic and pessimistic tone in the earnings press release. Working Paper, University of Oregon and University of Washington.

Dechow, P.M., W. Ge, C.R. Larson, and R.G. Sloan. 2010. Predicting material accounting misstatements. *Contemporary Accounting Research* (Forthcoming).

Dechow, P, W. Ge and C. Schrand. 2011. Understanding earnigns quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50 (2-3): 344-401.

Demers, E. and Vega, C. 2008. Soft information in earnings announcements: News or noise? Working Paper. INSEAD.

DePaulo, B. M., J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. 2003. Cues to deception. *Psychological Bulletin* 129 (1):74-118.

Dyck, A., A. Morse, and L. Zingales. 2008. Who Blows the Whistle on Corporate Fraud? Working Paper, University of Chicago.

Ekman, P. 1992. *Telling Lies*. New York: Norton.

Elkins, A. 2010. Evaluating the credibility assessment capability of vocal analysis software. *Forty-Third Annual Hawaii International Conference on System Sciences*, January 5-8, 2010, Koloa, Kauai, Hawaii. http://www.hicss.hawaii.edu/Reports/FullProceedings.pdf,

Elkins, A., and J. Burgoon. 2010. Validating vocal analysis software to assess credibility in interpersonal interaction: A multilevel factor analytic approach. Working Paper, University of Arizona.

Elliot, A. J., and P. G. Devine. 1994. On the motivational nature of cognitive dissonance: dissonance as psychological discomfort. *Journal of Personality and Social Psychology* 67 (3): 382-394.

Elliott, B., F. Hodge and L. Sedor. 2009. Using online video to announce a restatement: Influences on investor trust and investment decisions. Working paper, Univesity of Illinois and University of Washington.

Engelberg, J. 2008. Costly information processing: Evidence from earnings announcements. Working Paper, Northwestern University.

Eriksson, A., & Lacerda, F. 2007. Charlatanry in forensic speech science: a problem to be taken seriously. *International Journal of Speech, Language and the Law*, 14(2): 169-193.

Evans, J., R.L. Hannan, R.Krishnan and D. Moser. 2001. Honesty in managerial reporting. *The Accounting Review* 76 (4): 537-559.

Feldman, R., S. Govindaraj, J. Livnat, and B. Segal. 2009. Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies* (forthcoming).

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Festinger, L., and J. M. Carlsmith. 1959. Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology* 58 (2):203-210.

Frank, M.G., and P. Ekman. 1997. The ability to detect deceit generalizes across different types of high stake lies. *Journal of Personality and Social Psychology*. 72 (6): 1429-1439.

Gamer, M., H. Rill, G. Vossel and H.W.Godert. 2006. Psychophysiological and vocal measures in the detection of guilty knowledge. *International Journal of Psychophysiology* 60 (1):76-87.

Goetzmann, W.N. and N. Peles. 1997. Cognitive dissonance and mutual fund investors. *The Journal of Financial Research* 20 (2): 145-158.

Graham, R. 2007. Theory of cognitive dissonance as it pertains to morality. *Journal of Scientific Psychology* (July): 20-23.

Graham, J.R., C. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40: 3-73.

Han, Y., and J. Nunes. 2010. Please read the signal but don't mention it: How acknowledging identity signals leads to embarrassment. *Advances in Consumer Research* Volume 37, eds. Margaret C. Campbell, Jeff Inman, and Rik Pieters, Duluth, MN: Association for Consumer Research.

Harrigan, J.A., R. Rosenthal, and K.R. Scherer. 2005. *The new handbook of Methods in Noverbal Behavior Research*. Oxford University Press, Oxford.

Harmon-Jones, E., and J. Mills. 1999. *Cognitive Dissonance: Progress on a Pivotal Theory in Social Psychology*. Washington, DC: American Psychological Association.

Hey, J.D. 1998. Experimental economics and deception: A comment. *Journal of Economic Psychology* 19: 397-401.

Henry, E. 2008. Are investors influenced by how earnings press relesases are written? *Journal of Business Communication* 45(4): 363-407.

Hirschberg, J. 2010. Deceptive speech: Clues from spoken language. In F. Chen and K Jokinen (Eds.): *Speech Technology*, 79-88.

Hollander, S., M. Pronk and E. Roelofsen. 2009. Does silence speak? An empirical analysis of disclosure chocies during earnings conference calls. *Journal of Accounting Research* 48(3): 531-563.

Horvath, F. S. 1979. Effect of different motivational instructions on detection of deception with the psychological stress evaluator and the galvanic skin response. *Journal of Applied Psychology* 64 (3):323-330.

Hosmer, D., and S. Lemeshow. 2000. *Applied logistic regression*. New York: Wiley-Interscience Publication.

Humpherys. S.L, K. Moffitt, M. Burns, J. Burgoon and W. Felix. 2011. Identification of fraudulent financial statements using lingustic credibility analysis. *Decisision Support Systems* 50 (3): 585-594.

Javers, E. 2010. Broker, Trader, Lawer, Spy: The Secret World of Corporate Espionage. HarperCollins Publishers, New York, NY.

Jensen, M. T. Meservy, J. Burgoon, J. Nunamaker. 2008. Video-based deception detection. In H. Chen and C.C. Yang (Eds.): *Intelligence and Security Informatics Techniques and Applications* Hillsdale, NJ: Springer Berlin/Heidelberg: 233-259.

Juslin, P. N., and P. Laukka. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* 129 (5): 770-814.

Juslin, P. N., and K. R. Scherer. 2005. Vocal expression of affect. In *The new handbook of methods in nonverbal behavior research*, edited by J. A. Harrigan, R. Rosenthal and K. R. Scherer. Oxford: Oxford Press, 65–135.

Koller, M. and T. Salzberger. 2007. Cognitive dissonance as a relevant construct throughout the decision-making and consumption process - an empirical investigation related to a package tour. *Journal of Customer Behaviour* 6 (3), 217-227.

Kothari, S.P., Xu Li, and James Short, 2009. The effect of disclosures by management, analysts, and financial press on cost of capital, return volatility, and analyst forecasts: A Study Using Content Analysis. *The Accounting Review* 84 (5):1639-1674.

Kothari, S.P., S. Shu and P. Wysocki. 2008. Do managers withhold bad news? *Journal of Accounting Research* 47 (1): 241-276.

Larcker, D., and A. Zakolyukina. 2010. Detecting deceptive discussion in conference calls. Working Paper, Stanford University.

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

Li, F. 2009. The determinants and information content of the forward-looking statements in corporate filings - A naïve bayesian machine learning approach. Working Paper, University of Michigan.

Libby, R., R. J. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting Organizations and Society* 27 (8):777-812.

Loughran, T. and B. McDonald. 2010. Barron's red flags: Do they actually work? *Journal of Behavioral Finance*, Forthcoming.

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

Matsumoto, D.A., M. Pronk and E. Roelofsen. 2010. Managerial disclosure vs. analyst inquiry: An empirical investigation of the presentation and discussion portions of earnings-related conference calls. *The Accounting Review,* forthcoming.

Mayew, W. J., and M. Venkatachalam. 2011. The power of voice: Managerial affective states and future firm performance. *The Journal of Finance* , forthcoming.

Mazar, N., O. Amir, and D. Ariely. 2008. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research* 45 (6):633-644.

McDaniel, T., and C. Starmer. 1998. Experimental economics and deception: A comment. *Journal of Economic Psychology* 19: 403-409.

Meservy, T., M. Jensen, J. Kruse, J. Burgoon and J., Nunamaker. 2005. Automatic extraction of deceptive behaviorl cues from video. *Lecture Notes IN Computer Science* 3495: 219-227

Newman, M. L., J. W. Pennebaker, D. S. Berry, and J. M. Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin* 29 (5):665-675.

Palmatier, J.J. 2005. Assessing Credibility: ADVA technology, voice and voice stress analysis. In R.J. Montgomery and W.J. Majeski. *Corporate Investigations,* Tucson, AZ: Lawyers & Judges Publishing Company, Inc: 37-60.

Price, R., N. Sharp, and D. Wood. 2010. Detecting and predicting accounting irregularaties: A comparison of commercial and academic risk measures. *Working Paper,* Rice Unviersity, Texas A&M University and Brigham Young University.

Public Company Accounting Oversight Board. 2007. Observations on auditors' implementation of pcaob standards relating to auditors' responsibilities with respect to fraud. In *PCAOB Release No. 2007-001*.

Public Company Accounting Oversight Board. 2010. Auditing standards related to the audiotr's assessment of and response to risk and related amendments to PCAOB standards, In *PCAOB Release No. 2010-004*.

Purda, L. and D. Skillicorn. 2010. Reading between the lines: Detecting fraud from the language of financial reports. Working paper, Queen's University.

Roychowdhury, S. and E. Sletten. 2009. Managerial incentives and the informativeness of earnings announcements. Working Paper, MIT.

Salterio, S. E., and A. Webb. 2006. Honesty in Accounting and Control: A discussion of "the effect of information systomes on honesty in managieral reporting: A behavioral perspective." *Contemporary Accouning Research* 23 (4): 919-932.

Scherer, K. R. 1986. Vocal Affect Expression - A review and a model for future-research. *Psychological Bulletin* 99 (2):143-165.

SEC Release No. 33-8177 and 34-47235. 2003. Disclosure required by sections 406 and 407 of the Sarbanes-Oxley act of 2002, January 23, 2003.

SEC Release No. 34-48745. 2003. NASD and NYSE rulemaking: Relating to corporate governance, November 4, 2003.

Sprinkle, G. B. 2003. Perspectives on experimental research in managerial accounting. *Accounting Organizations and Society* 28 (2-3):287-318.

Tetlock, P. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62 (3): 1139-1168.

Tetlock, P., Saar-Tsechansky, M., and Macskassy, S. 2008. More than words: Quantifying language to measure firm's fundamentals. *Journal of Finance*, 63 (3): 1437-1467.

Vrij, A., K. Edward, K.P. Roberts, and R. Bull. 2000. Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*, 239-264.

Vrij, A. 2008. *Detecting Lies and Deceit: Pitfalls and Opportunities.* Chichester, UK, Wiley.

Weil, J. 2004. Behind wave of corporate fraud: A change in how auditors work. *The Wall Street Journal*, A1, March 25 2004.

Waller, W. 1988. Slack in participative budgeting: The joint effect of a truth-inducing pay scheme and risk preferences. *Accounting, Organizations and Society* 13 (1): 87.98.

Webb, A. 2002. The impact of reputation and variance investigations on the creation of budget slack. *Accounting, Organizations and Society*. 27 (4/5): 361-378.

Zhong, C., V. Bohns and F. Gino. 2010. Good lamps are the best police: Darkness increases dishonesty and self-interested behavior. *Psychological Science* 21 (3): 311-314.

Zuckerman, M., B. M. DePaulo, and R. Rosenthal. 1981. Verbal and nonverbal communication of deception. In *Advances in experimental social psychology*, edited by L. Berkowitz. New York: Academic Press, 1-59.

Zuckerman, M., and R. Driver. 1985. Telling lies: Verbal and nonverbal communication of deception. In A. W. Siegman and S. Feldstein (Eds.). *Multichannel Integrations of Nonverbal Behavior*, Hillsdale, NJ: Lawrence Erlbaum: 129-14.

Figure 1
Timeline of Events for Generation of Laboratory Data

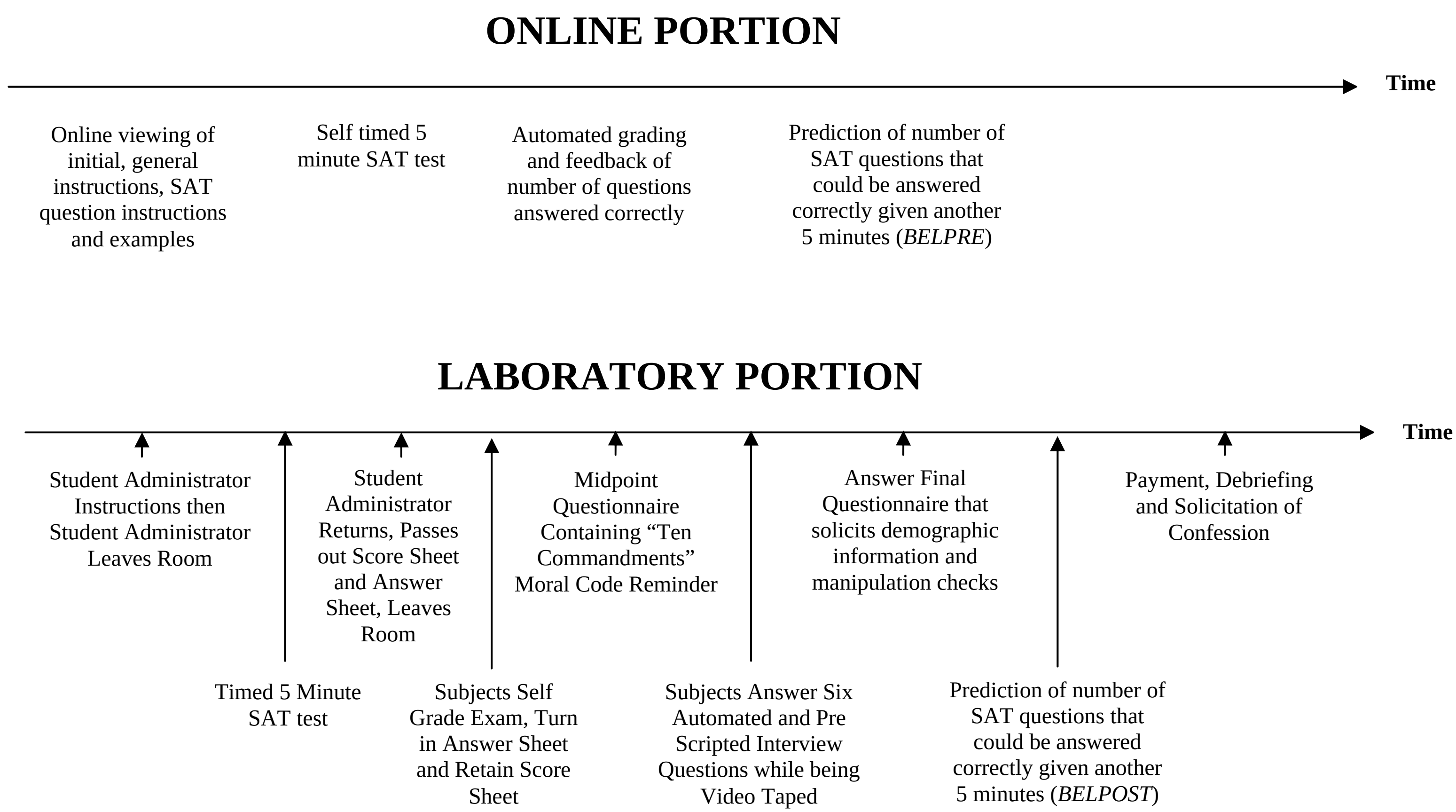

Table 1
Descriptive Statistics and Correlations for Laboratory Generated Data

Panel A: Descriptive Statistics (N=59)[a]

Variable[b]	Mean	Std. Dev	Median	Min	Max
MISREP	0.322	0.471	0.000	0.000	1.000
SCORE	24.686	11.559	23.000	6.000	70.000
SURVEY	11.627	11.884	8.000	-4.000	61.000
USCORE	13.059	15.967	14.000	-37.000	54.000
BELPRE	6.000	2.205	5.000	2.000	13.000
BELPOST	8.415	3.412	8.000	4.000	22.000
BELREV	2.415	3.467	2.000	-6.000	16.000
COGDIS	0.217	0.088	0.220	0.020	0.484
E_COGDIS	0.214	0.095	0.216	0.000	0.438
L_COGDIS	0.220	0.106	0.200	0.000	0.571
TIME_MIN	2.716	0.791	2.667	1.400	4.733
WC	333.492	123.399	305.000	126.000	628.000
SCHOOL	2.763	0.916	2.000	1.000	4.000
MATH	2.661	1.469	3.000	0.000	9.000
ENGLISH	1.441	0.992	1.000	0.000	3.500
AGE	19.864	0.937	20.000	18.000	22.000
FEMALE	0.373	0.488	0.000	0.000	1.000

Panel B: Correlation between Vocal Dissonance Markers and Belief Revision Dissonance Markers[c]

Variable[b]	***COGDIS***	***BELREV***	***E_COGDIS***	***L_COGDIS***
COGDIS		**0.333 (0.010)**	**0.877 (0.000)**	**0.847 (0.000)**
BELREV	0.192 (0.144)		**0.440 (0.001)**	**0.098 (0.462)**
E_COGDIS	**0.875 (0.000)**	**0.320 (0.014)**		**0.538 (0.000)**
L_COGDIS	**0.895 (0.000)**	**0.062 (0.812)**	**0.567 (0.000)**	

Notes: [a]The full sample is 59 observations gathered from two different Universities. [b] Variable definitions are listed in Appendix 1. [c]Spearman (Pearson) correlations are presented above (below) the diagonal. Coefficients in bold represent statistical significance at < 10% level. Two sided p-values presented in parenthesis below the correlations coefficients.

Table 2
Association Between Confessed Misreporting and Vocal Dissonance Cues using Laboratory Generated Data[a]

Variable[b]	Predicted Sign	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Intercept	(?)	-1.707**	-1.968**	-2.200**	-1.951*	-3.978*	-8.840	0.089
		(0.794)	(0.921)	(0.988)	(1.071)	(2.197)	(9.529)	(0.152)
COGDIS	(+)	4.330*						
		(3.232)						
E_COGDIS	(+)		9.202***	10.293***	17.881**	12.990**	9.057**	1.923***
			(3.542)	(3.517)	(8.459)	(6.398)	(4.044)	(0.649)
L_COGDIS	(+)		-3.739	-4.344	-8.071*	-1.367	-4.085	-0.816
			(2.748)	(3.618)	(4.638)	(4.878)	(0.134)	(0.527)
SCHOOL	(?)						-0.232	
							(0.725)	
MATH	(?)						-0.147	
							(0.203)	
ENGLISH	(?)						0.159	
							(0.364)	
AGE	(?)						0.367	
							(0.552)	
FEMALE	(?)						0.374	
							(0.661)	
SURVEY	(?)						0.029	
							(0.023)	
Dependent Variable		*MISREP*	*MISREP*	*MISREP*	*MISREP*	*MISREP*	*MISREP*	*USCORE*
Model Type		Logit	Logit	Logit	Logit	Logit	Logit	OLS
Pseudo R^2 or Adj. R^2		0.024	0.082	0.096	0.181	0.180	0.119	0.103
# of observations		59	59	56	24	35	59	59
P value: *E_COGDIS* = *L_COGDIS*		N/A	0.013	0.012	0.040	0.070	0.022	0.007
AUC[c]		0.602	0.670	0.681	0.752	0.708	0.675	N/A
Z-Stat for Test AUC = 0.500[c]		0.800	2.216	2.363	2.339	1.975	2.163	N/A
P Value for Test AUC = 0.500[c]		0.424	0.027	0.018	0.019	0.048	0.030	N/A

Notes: ***, **, * Statistically significant at 1%, 5% and 10% levels in a two tailed test (one tailed test when predicted). Robust standard errors are presented in parentheses below the coefficient estimates. [a]The full sample is 59 observations gathered from two different Universities. In Column C, 3 observations representing outlier covariate patterns were removed. In Column D (E) data relating to only the 24 (35) observations obtained from the first (second) university was used. [b]Variable definitions are listed in Appendix 1. [c]The Receiver Operating Characteristic (ROC) curve analysis is used to quantify the accuracy of the logistic prediction equation at classifying participants as having misreported or not. The ROC curve is a graph of the sensitivity versus 1 – specificity of the prediction test. This area measures the global performance of the test. The greater the area under the ROC curve (AUC), the better the performance. The test statistic for testing whether the AUC is statistically different from zero is (AUC – 0.50)/((standard error (AUC)). AUC and standard error (AUC) were obtained from STATA's roctab command. This test statistic is approximately normal (Zhou, et al. 2002), and is therefore reported as a Z statistic with two sided p-values.

TABLE 3
Descriptive Statistics, Correlations and Industry Composition for Archival Data[a]

PANEL A: Descriptive Statistics (N=1,572)

Variable[b]	Mean	Std. Dev	Median	Min	Max
RESTATE	0.069	0.254	0.000	0.000	1.000
COGDIS	0.179	0.076	0.172	0.000	0.472
RET	-0.020	0.383	-0.064	-0.770	1.388
FSCORE	1.280	0.863	1.119	0.146	4.487
ACCT_RISK	45.045	27.137	42.000	1.000	100.000
lnMVE	7.281	1.547	7.169	3.952	11.457
VOL	0.021	0.009	0.020	0.008	0.051
UE	-0.001	0.013	0.000	-0.090	0.031
ROA	0.004	0.044	0.010	-0.206	0.116
AGE	54.339	7.261	55.000	37.000	83.000
TENURE	6.087	6.035	5.000	0.000	44.000

PANEL B: Pearson (Spearman) Correlations above (below) Diagonal[c]

	Variable[b]	1	2	3	4	5	6	7	8	9	10	11
1	***RESTATE***		**0.050**	0.030	0.038	**0.077**	**-0.057**	0.028	-0.038	-0.003	0.006	-0.018
2	***COGDIS***	**0.046**		**-0.052**	-0.022	0.038	**-0.114**	**0.093**	**-0.054**	**-0.084**	0.016	-0.004
3	***RET***	0.023	**-0.073**		**-0.091**	**-0.085**	**0.210**	**-0.094**	**0.171**	**0.266**	0.022	-0.035
4	***FSCORE***	0.023	-0.017	**-0.049**		**0.146**	**0.086**	**-0.121**	**-0.152**	**-0.066**	0.000	0.029
5	***ACCT_RISK***	**0.075**	0.029	**-0.066**	**0.095**		0.034	0.029	-0.028	**-0.092**	**-0.061**	**-0.065**
6	***lnMVE***	**-0.055**	**-0.108**	**0.267**	**0.115**	**0.044**		**-0.552**	**0.087**	**0.356**	**0.114**	0.003
7	***VOL***	**0.046**	**0.055**	**-0.171**	**-0.172**	0.002	**-0.574**		**-0.192**	**-0.360**	**-0.091**	-0.016
8	***UE***	**-0.057**	**-0.075**	**0.123**	**-0.117**	-0.006	**0.045**	-0.023		**0.256**	0.020	0.026
9	***ROA***	**-0.044**	**-0.080**	**0.288**	**-0.080**	**-0.131**	**0.344**	**-0.193**	**0.195**		**0.048**	0.038
10	***AGE***	0.005	0.019	0.019	0.041	**-0.054**	**0.120**	**-0.109**	-0.034	**0.045**		**0.310**
11	***TENURE***	-0.004	-0.026	-0.014	0.035	**-0.054**	0.013	-0.006	-0.020	**0.042**	**0.221**	

PANEL C: Industry Composition

Industry[d]	Sample Firms[a]		All Compustat Firms[e]	
	N	**%**	**N**	**%**
Chemicals	28	1.78	411	1.82
Computers	232	14.76	2,908	12.85
Extractive	55	3.50	904	3.99
Financial	207	13.17	3,050	13.48
Food	21	1.34	401	1.77
Insurance/RealEstate	115	7.32	2,306	10.19
Manf:ElectricalEqpt	51	3.24	767	3.39
Manf:Instruments	108	6.87	1,062	4.69
Manf:Machinery	28	1.78	544	2.40
Manf:Metal	19	1.21	473	2.09
Manf:Misc.	8	0.51	214	0.95
Manf:Rubber/glass/etc	9	0.57	371	1.64
Manf:TransportEqpt	27	1.72	340	1.50
Mining/Construction	24	1.53	622	2.75
Pharmaceuticals	114	7.25	900	3.98
Retail:Misc.	78	4.96	933	4.12
Retail:Restaurant	17	1.08	286	1.26
Retail:Wholesale	26	1.65	781	3.45
Services	171	10.88	2,064	9.12
Textiles/Print/Publish	79	5.03	845	3.73
Transportation	100	6.36	1,388	6.13
Utilities	49	3.12	658	2.91
Not assigned	6	0.38	405	1.79
Total	1,572	100.00	22,633	100.00

Notes: [a]The number of observations equals 1,647 observations available for voice analysis from Mayew and Venkatachalam (2011) less 13 observations where CEO does not speak, less 52 observations where *ACCT_RISK* is not available, less 10 observations where we cannot calculate *FSCORE* due to missing data in Compustat. [b]Variable definitions are listed in Appendix 1. [c]Coefficients in bold represent statistical significance at < 10% level. [d]Industry definitions follow Barth, et al. (2005). [e]Population is derived from all observations on the annual Compustat database for fiscal year 2006 where SIC code is populated. All continuous variables are winsorized at the 1% and 99% levels to mitigate the effects of outliers.

Table 4
Logistic Regression Estimation of the Association Between
Income Decreasing Restatements and Vocal Dissonance Cues[a]

Variable[b]	Predicted Sign	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Intercept	(?)	-3.069***	-3.004***	-3.250***	-3.586***	-2.298	-2.350	-2.787*
		(0.335)	(0.335)	(0.341)	(0.502)	(1.463)	(1.486)	(1.463)
Vocal Dissonance Markers								
COGDIS	(+)	2.542*						
		(1.627)						
$COGDIS^{HS}$	(+)		3.774**		3.521**	3.619**		3.491**
			(1.669)		(1.588)	(1.780)		(1.754)
$COGDIS^{LS}$	(+)		1.198		1.177	1.133		1.071
			(1.790)		(1.789)	(1.817)		(1.843)
Financial Statement Based Predictor								
FSCORE	(+)			0.107	0.095		0.121	0.139
				(0.135)	(0.136)		(0.157)	(0.156)
ACCT_RISK	(+)			0.011**	0.010**		0.012**	0.011**
				(0.005)	(0.005)		(0.005)	(0.005)
Non Misreporting Dissonance Drivers								
RET	(?)					0.439	0.492	0.525
						(0.314)	(0.327)	(0.320)
lnMVE	(?)					-0.193**	-0.247***	-0.221***
						(0.097)	(0.095)	(0.095)
VOL	(?)					-5.847	-6.949	-9.319
						(16.783)	(16.915)	(16.962)
ROA	(?)					-0.645	1.722	1.741
						(8.237)	(3.887)	(3.926)
UE	(?)					-0.478	-6.493	0.111
						(8.237)	(7.691)	(8.641)
CEO Characteristics								
AGE	(?)					0.005	0.008	0.007
						(0.022)	(0.022)	(0.022)
TENURE	(?)					-0.012	-0.009	-0.008
						(0.024)	(0.025)	(0.024)
Industry Fixed Effects		No	No	No	No	Yes	Yes	Yes

Table 4 (continued)

Pseudo R^2	0.005	0.014	0.013	0.025	0.050	0.054	0.064
# of observations[a]	1,572	1,572	1,572	1,572	1,572	1,572	1,572
AUC[c]	0.552	0.584	0.588	0.595	0.673	0.664	0.679
Z-Stat for Test AUC = 0.500[c]	1.723	2.838	3.103	3.170	7.049	6.393	6.851
P Value for Test AUC = 0.500[c]	0.085	0.005	0.002	0.002	<0.001	<0.001	<0.001

Notes: Dependent Variable = *RESTATE*. ***, **, * Statistically significant at 1%, 5% and 10% levels in a two tailed test (one tailed test when predicted). Robust standard errors clustered by firm are presented in parentheses below the coefficient estimates. [a]The number of observations equals 1,647 observations available for voice analysis from Mayew and Venkatachalam (2011) less 13 observations where CEO does not speak, less 52 observations where *ACCT_RISK* is not available, less 10 observations where we cannot calculate *FSCORE* due to missing data in Compustat. [b]Variable definitions are listed in Appendix 1. [c]The Receiver Operating Characteristic (ROC) curve analysis is used to quantify the accuracy of the logistic prediction equation at classifying participants as having misreported or not. The ROC curve is a graph of the sensitivity versus 1 – specificity of the prediction test. This area measures the global performance of the test. The greater the area under the ROC curve (AUC), the better the performance. The test statistic for testing whether the AUC is statistically different from zero is (AUC – 0.50)/((standard error (AUC)). AUC and standard error (AUC) were obtained from STATA's roctab command. This test statistic is approximately normal (Zhou, et al. 2002), and is therefore reported as a Z statistic, with two sided p-values. All continuous variables are winsorized at the 1% and 99% levels to mitigate the effects of outliers.