DRAFT VERSION

What You Say and How You Say It:
Textual And Audio Disclosures in Earnings Calls

ABSTRACT

I investigate textual and audio disclosures during the earnings conference calls. I use machine learning model to analyze text sentiment. I analyze audio features of each sentence and match it with its text sentiment. I find that both text sentiment and audio features are informative for the investors.

Keywords: textual analysis; audio analysis; machine learning; BERT model; earnings calls.

JEL Classification: C52, C55, G14, M41.

INTRODUCTION

Earnings conference calls are an important medium for information dissemination. Firms are known to exercise discretion in what and how they deliver information. In this study, I use machine learning to analyze sentence sentiment and then simultaneously analyze audio features of each sentences in earnings conference calls. I use corresponding stock returns and trading volume to measure the information content of textual and audio aspects of earnings conference calls and find that textual sentiment and audio emotion of a sentence helps to explain abnormal stock returns and abnormal trading volume. In particular, I find that negative text sentiment is negatively correlated with the abnormal return and positively correlated with the abnormal trading volume. Furthermore, the interaction between textual analysis and audio emotion explain stock returns and trading volume.

I examine earnings calls using text-only, audio-only, and a joint approach. I analyze the presentation and M&D sessions separately since, the informational environment of the M&D sessions in earnings conference calls is "dynamic" and less structured (Blau et al. 2015) and M&D section is more informative for investors (Matsumoto et al. 2011). First, research shows that investors react to the textual sentiment of earnings calls (Kearney and Liu 2014; Loughran and Mcdonald 2016; Wang et al. 2016). Loss aversion bias suggests that investors have a have a stronger inclination towards avoiding losses as compared to their desire for acquiring gains (Kahneman and Tversky 1979). Consistent with the loss aversion bias, I find that investors respond more strongly to negative sentiment than positive sentiment.

Second, the psychology literature shows that voice is an important means of communication because it conveys a wide range of information such as our emotions (Scherer et al. 1991; Juslin and Laukka 2001; Kreiman and Sidtis 2011). The voice is considered as a significant source of information for listeners and it influences how we perceive and interact with others (Scherer et al. 1991; Goudbeek and Scherer 2010). Research show that managers' vocal cues contain value relevant information (Mayew et al. 2012; Price et al. 2017; Mayew et al. 2020). Mayew et al. (2012) find that positive and negative emotional states of managers are informative about the firm's financial future. Price et al. (2017) examine investor reaction to the emotional content of managers' vocal cues. They examine both transcripts and audio files of quarterly earnings conference calls to identify raw emotion found in the voice patterns of executives. They find a strong positive association between executive emotion and investors' initial reaction. Mayew et al. (2020) measure voice pitch changes by comparing the average pitch during a dialog with that in a presentation to assess the impact of the dialog capital market reaction. Corive et al. (2001), Ten Bosch (2003) show that prosody features such as pitch and energy have high relation to the emotional states of speech. However, it is unkown how investors react to these vocal cues. During the presentation section, managers may speak with a more monotone voice, therefore audio features will be less pronounced when compared to the M&D section. Unlike the presentation section, I expect to see stronger investor reactions to vocal cues during the M&D section, where the interactions are unscripted. Additionally, managers may be heavily coached to improve their speech skills, leading them to watch for their speech characteristics during the M&D session (Wigglesworth 2020). Therefore, managers may be mindful of their speech during the M&D session and may convey negative news with softer tones to minimize the impact on capital markets. However, I find that managers' vocal cues such as energy are informative for investors.

Third, I will use a multimodal approach and analyze the textual and audio contents of earnings calls simultaneously. The multi-modal approach of analyzing both textual and audio features leads to a more comprehensive understanding of investor behavior. This is because vocal cues can either support or contradict the message in the text. My research has the potential to provide richer information to investors and researchers than they have had before about the true meaning of a message. For example, an investor might review a transcript of an earnings call and observe that the CEO of a company appears confident about the future. However, the CEO's pitch may show uncertainty or hesitation. Similarly, managers might strengthen or weaken the information in negative or positive sentences through the use of vocal cues. A joint textual and audio approach could help the investor better understand the CEO's underlying message, and may thus impact company's stock price (Goodell et al. 2023). It is also conceivable that investors can see through the manipulation by the management. In this case, vocal cues will not be informative during the presentation or M&D section. However, especially during the M&D section, the management might want to convey additional information beyond what can be found on the text, leading to a greater impact on investor behavior. Therefore, investors may pay attention to these vocal cues and supplement the information received from the textual data.

Then the market should react to vocal cues and textual sentiment separately during both the presentation and M&D sections.

I investigate the capital market response to verbal and nonverbal signals to empirically measure the informativeness of textual and audio data. First, I analyze sentiment of each sentence in earnings calls using finetuned Bidirectional Encoder Representations from Transformers or BERT (Devlin et al. 2018) model. I pre-train the model by using corporate annual filings (10-Ks of firms) and earnings conference call transcripts to reduce the computing costs and the time to develop an effective model. Second, I extract audio features, such as such as pitch and energy, of each sentence.

I gather earnings conference call transcripts and audio files of 45,373 firms between 2017 and 2020. I divide the transcripts and audio files into sentences, tag their timings, and examine textual sentiment and audio features of each sentence. I then connect these sentences with the NYSE TAQ database and calculate abnormal returns and volume during which the sentence is uttered.

This study introduces several contributions to the literature. First, to the best of my knowledge, this study is the first to granularly analyze textual and audio data simultaneously in earnings calls. The findings are conclusive in univariate analysis as well as in multivariate analysis that control for firm characteristics such as size, growth, profitability, leverage, and number of analysts. Second, I contribute to the growing research on the importance of vocal cues/audio features in earnings calls and their impact on these calls and investor behavior. Research has shown that audio features such as pitch can influence investors' behavior (Hobson et al. 2012; Mayew et al. 2012; Mayew et al. 2020). I find that some of the audio features are highly correlated with both abnormal stock returns and trading volume. The findings suggest that these nonverbal aspects of speech can be just as important as the words themselves in influencing investor perceptions and decision-making. Finally. I build upon Matsumoto et al.'s (2011) research, which found that the M&D section is more informative for investors. I add additional evidence to support the idea that investors pay more attention to the M&D section by showing that investors pay attention to the sentiment of financial text and audio features.

The remainder the my paper is organized as follows. In Section 2, I discuss prior linquistics and accounting literature on sentiment analysis and textual analysis In this section, I summarize machine learning applications in textual analysis. Specifically, I discuss the Bidirectional Encoder Representation from Transformers (BERT) language model (Devlin et al. 2018), which classifies earnings call sentiment in this study. I also discuss prior literature that examines the information content and sentiment of the qualitative disclosures. Section 3 develops hypotheses. Section 4 describes the sample and the research design. Section 5 discusses empirical results. Section 6 concludes and presents suggestions for future research.

SECTION 2

LITERATURE REVIEW

2.1 Textual Analysis in Accounting Research

Textual analysis is a powerful tool that allows researchers to automate the process of extracting textual data. By applying natural language processing (NLP) techniques, researchers can quickly extract key insights and trends from a wide range of written sources including social media and news articles. One main benefit of textual analysis is that it allows researchers to process large amounts of text in a short period. Another benefit of textual analysis is that it uncovers attitudes and emotions. The field of sentiment analysis examines and analyzes these expressions of sentiment and opinion in large amounts of text. Textual analysis has become increasingly common in social science research, because researchers have recognized its potential to help understand social phenomena and patterns of behavior (Pang and Lee 2008; Ravi and Ravi 2015; Mohammad 2016).

Studies of textual analysis largely use conference calls and annual reports (Kearney and Liu 2014; Loughran and Mcdonald 2016; Guo et al. 2016). These studies focus on detecting sentiment, such as overall positive or negative sentiment. The sentiment analysis refers to the task of automatically detecting and classifying the affective content into a small number of classes representing different sentiments. This can be done at various levels of granularity, from individual words to complete documents, and can focus on specific aspects or entities discovered in the text. The sentiment analysis is often viewed as a binary problem, in which the goal is to detect the presence of affective content or to distinguish between positive and negative sentiment. This is sometimes referred to as polarity detection (Pang and Lee 2008).

Many studies ignore grammar or word order, assuming that a text is a collection of independent words. This approach, which is known as a bag-of-words structure, ignores the context in which words are used (Loughran and Mcdonald 2016). Recently, language models were developed to augment such analysis. The machine learning based algorithms, like Bidirectional Encoder Representations from Transformers or BERT (Devlin et al. 2018), consider the context and sequence of words. BERT is a transformer model that uses attention mechanisms to process text. These algorithms are more effective in finance and accounting studies than the bag-of-words models in sentiment classification, language translation, and question answering (Devlin et al. 2018). In finance and accounting, BERT has been used for tasks such as sentiment analysis and predicting stock prices.

Conference calls can be used for various purposes, such as presenting financial results, company strategies, or any material developments. In a conference call, there are typically two sessions: the management discussion (MD) session and the question and answer (M&D) session. During the MD session, the management team presents the company's quarterly results, usually following a scripted presentation. This allows the management to present information in a structured and controlled manner, highlighting the most important points and company’s achievements. In the M&D session, financial analysts and other stakeholders ask questions and request clarifications about the company's performance and future plans. This session is usually more spontaneous and unscripted, as the analysts and stakeholders are free to ask any question. The management team must be prepared to respond to these questions in real-time and provide additional information as needed. Conference calls allow companies to communicate important information to their stakeholders, while also providing an opportunity for stakeholders to ask questions and get more detailed information about the company's performance.

Matsumoto et al. (2011) show that both segments of the conference call contain information beyond what was included in the accompanying press release. The M&D segment is more informative than the presentation segment, and this higher level of information is correlated with the number of analysts following the call. Additionally, when a company performance is poor, managers disclose more information during the MD segment, but the M&D segment still contains more information. Following Matsumoto et al. 2011, I separately analyze the MD and M&D sections.

Manual and computational analysis of content methods have been used by researchers to draw inferences from textual data. Accounting and finance research examined various aspects of text such as the readability, tone, and the amount of disclosure (Loughran and Mcdonald 2016). Frankel et al. (2010) found that the tone of conference call, as indicated by the use of positive or negative language in the conference call, improves when company performance is better than expected by analysts. Li (2008) found that certain features of annual reports, specifically the frequency of certain words in the Management Discussion and Analysis (MD&A) section, are related to the persistence of a company's earnings. For profitable companies, a higher frequency of causation words (such as "because") is associated with less persistent earnings, while a higher frequency of positive emotion words is associated with more persistent earnings. On the other hand, companies that report losses and a higher frequency of positive emotion words A have less persistent earnings. The study also found that for profitable firms, a higher frequency of future tense verbs relative to past or present tense verbs is associated with lower earnings persistence. Kothari et al. (2009) used General Inquirer software to classify disclosures as favorable or unfavorable. The study found that, for favorable disclosures, the company's cost of capital, stock return volatility, and the dispersion of analyst forecasts decreased significantly. In contrast, for unfavorable disclosures, the cost of capital, stock return volatility, and analyst forecast dispersion all increased significantly. Brown and Tucker (2011) developed a score to measure changes in a company's MD&A disclosure from one year to the next. The score was based on an algorithm commonly used by search engines to determine document similarity. The study found that companies with larger economic changes modify their MD&A more than those with smaller economic changes, suggesting that companies meet a minimum requirement for MD&A disclosure. Cho and Muslu (2021) found that when peer firms' MD&A narratives are more optimistic, the firm increases capital investments and inventory in the following year. Additionally, the impact of the peer firms' MD&A narratives on the subject firm's investments is dependent on the content of the narratives, with more optimistic discussions about the industry and investments having a stronger effect on the subject firm's capital investments. These findings suggest that the information contained in MD&A disclosures impact firms' investment decisions based on the proprietary information they provide. Muslu et al. (2015) examine the information contained in the forward-looking disclosures in the MD&A. They found that firms are more likely to make forward-looking MD&A disclosures when their stock prices are less efficient in terms of reflecting future earnings information. These disclosures can help to improve the informational efficiency of the stock prices of such firms, although they are not able to completely correct for the lower efficiency. The effects of these disclosures are more pronounced for operations-related disclosures, disclosures made before 2000, and disclosures made by firms that are experiencing losses.

Davis et al. (2012) argue that managers' tone during earnings releases signal expectations about the future of the firm. Price et al. (2012) use textual analysis to assess the tone used during quarterly earnings conference calls. They examine the relation between conference call tone and future stock price reaction. They show that the tone of the conference call is associated with future stock prices. Rennekamp et al. (2022) separate the analyst and managers conversation and found that more engaging conversations during the M&D section are positively correlated with absolute stock returns. This suggests that these more interactive conversations are more informative for the market and help to shape the price of the stock. Jiang et al. (2019) examine manager sentiment based on the aggregated textual tone of corporate financial disclosures. They investigate the relationship between managers' sentiment and stock prices. They find that manager sentiment (especially negative tone) is a strong predictor of future aggregate stock market returns. Bochkay et al. (2020) used a large sample of earnings calls to create a dictionary of extreme language. The entries in this dictionary were ranked by human annotators based on their positive or negative extremity. The study found that the use of extreme language explains market reactions and future operating performance. This was true both for positive and negative extreme language. The study also found that the market reactions to extreme language were more pronounced in firms with weaker information environments and higher information processing costs. Overall, the study suggests that extreme language in earnings calls has a significant impact on trading volume and stock price reactions.

A major disadvantage of using a classical textual analysis such as bag of words is that word order is not taken into consideration. These models ignore the context and in turn the meaning of words. Technological advancements enabled context analysis. Guo, Shi, and Tu (2016) discuss methods used in textual analysis such as machine learning models (Naïve Bayes, Support Vector Machines, Neural Network). They compare the forecast performance of each algorithm. They find that neural network methods outperform classical textual analysis methods and many other machine learning techniques in classifying news category. Li (2010) uses Naive Bayesian machine learning algorithm and finds that tone of forward-looking statements in the MD&A section are correlated with various factors such as current performance, accruals, firm size, MTB ratio, return volatility, MD&A Fog, and firm age. Chatterjee et al. (2019) show that a deep learning-based approach significantly outperforms traditional machine learning models in recognizing emotions. They propose a deep learning-based approach to detect emotions such as happy, sad and angry in textual dialogues. They combine both semantic and sentiment-based analysis for more accurate emotion detection.

Devlin et al. (2018) developed and published BERT which is a machine learning model developed by Google that is used for natural language processing (NLP) tasks. It is based on the Transformer architecture and is trained to generate a language model by reading and understanding the context of words in a given text. BERT is bidirectional, meaning it considers the context of a word based on all of the words that come before and after, rather than just those that come before. This allows BERT to more accurately understand the meaning of language and perform tasks such as translation and text classification. BERT has been widely adopted and is used by Google in almost every English language search query. Original implementation of transformer architecture is developed by Vaswani et al. (2017). It has many encoder layers and self-attention heads, which are mechanisms that enable to learn contextual relationships between words in a text. An advantage of transformer architecture is transfer learning in which a model trained on one task is used as the starting point to solve a different but related task (Weiss et al. 2016). The goal of transfer learning is to leverage the "knowledge" gained from the original task, which is often trained on a large dataset, to solve the new task more efficiently and effectively, without the need for a large dataset of labeled examples specific to the new task. This is particularly useful when there is a shortage of labeled data or when it is expensive or time-consuming to obtain.

Huang et al. (2022) developed Finbert which is an algorithm developed using BERT and a corpus of financial texts that has been adapted for the finance domain. The authors evaluated for its performance in sentiment classification, a widely used NLP task in financial texts, using a sample of analyst report sentences labeled by researchers. FinBERT has been found to have significantly higher out-of-sample accuracy than other popular algorithms such as the LM dictionary, NB, SVM, RF, CNN, and LSTM, and to be particularly effective when the training sample is small, likely due to its ability to consider contextual information in financial text and texts containing financial words not frequently used in general texts. Finally, FinBERT has been shown to better capture the textual sentiments of earnings conference calls, as indicated by its association with market reactions to the calls. Finetuning, or training a pre-trained model on a dataset specific to a task, is a powerful technique to improve model performance (Huang et al. 2022). BERT algorithm consider the context and sequence of words unlike other approaches such as bag-of-words structure. Huang et al. (2022) show that FinBERT, a variation of BERT, effectively summarizes contextual information in financial texts (for example, analyst reports) to sentiment groups and outperforms other algorithms including naive Bayes, support vector machine, random forest, convolutional neural network, and long short-term memory.

2.2 Audio Analysis in Accounting Research

In the psychology literature, researchers show that investors' decision making is affected by their emotions (Lerner and Keltner 2000; Han et al. 2007). Emotions could change the behaviors of the investors. For example, fear could lead to risk-averse choices while anger could lead to investors taking excessive risks (Lerner et al. 1998; Lerner and Keltner 2000; Han et al. 2007). According to appraisal theory of emotion, emotions are determined by our appraisals of stimuli (Lazarus 1991) and in turn influence the likelihood of specific courses of action (Lazarus 1991; Scherer 1999). According to Lazarus (1991), anxiety is characterized by appraisals of facing uncertain threats and results in actions tendencies to reduce uncertainties. Sadness is characterized by appraisals of facing loss and results in tendencies to change the circumstances.

In psychological studies, vocal parameters such as pitch and intensity indicate emotion and sentiment (Murray and Arnott 1993). The voice is a key means of communication because it allows us to convey a wide range of information through vocal signals. These signals can convey relatively enduring features such as age and gender, as well as transitory states such as health and power (Kreiman and Sidtis 2011). Research has shown that the voice is an important source of information for listeners, and that it can influence how we perceive and interact with others. For example, a person's voice may convey their level of confidence or dominance, which can affect how others respond to them. Additionally, changes in the voice can signal changes in a person's emotional or physical state, such as when a person's voice becomes strained or hoarse when they are sick or under stress. Overall, the voice is a powerful tool for communication and can have a significant impact on how we interact with others (Scherer et al. 1991; Juslin and Laukka 2001; Goudbeek and Scherer 2010). (Banse and Scherer 1996; Juslin and Laukka 2001) show that emotion-specific patterns of acoustic features result in considerable differentiation depending on emotional states, especially for the negative emotions such as anger, fear and sadness.

Hobson et al. (2012) and Price et al. (2017) show that investors react to vocal cues present in earnings conference calls. Hobson et al. (2012) examine vocal markers of cognitive dissonance, using vocal emotion analysis software, in detecting financial misreporting. They find that CEO speech can be used to detect irregular restatements. In addition to emotions, researchers show that investor sentiment can affect expected return and stock price volatilities. Lee et al. (2002) and Brown and Cliff (2004) shows that investors sentiment affects asset valuations. They found that there is a positive relationship between excess return and investor sentiment. By using textual analysis techniques, Tetlock (2007), Loughran and Mcdonald (201) and Smales (2015) show that investors react to sentiment in financial text. Mayew and Venkatachalam (2012) examine managerial affective states during conference calls using vocal emotion analysis software and find that positive and negative affects displayed by managers through their vocal cues are informative about the firm's financial future. The also the commonly used audio features (Owren and Bachorowski 2007) (mean fundamental frequency (F0), standard deviation of fundamental frequency (F0Std), Jitter , Shimmer , and mean harmonic-to-noise (HNR) ratio) to further analyze the audio files. The study also shows that audio features such as jitter and shimmer are linked to indicators of negative emotions and decreased stock prices, which may indicate deceptive behavior. Mayew et al. (2013) investigate the connection between the pitch of male's voice and their success in the corporate world. The results show that men with deeper voices tend to oversee larger companies, earn more compensation, and stay in their positions for longer periods of time.

Mayew et al. (2020) measure voice pitch changes by comparing the average pitch during a dialog with that in a presentation to assess the impact of the dialog capital market reaction. By comparing the pitch during the dialog to a baseline established by the presentation, which is assumed to capture the normal speaking voice of the manager, it may be possible to observe any changes in pitch that occur as a result of the dialog. They find that stock prices are influenced by the language and vocal tone used by analysts and managers. Specifically, positive language and a positive vocal tone were found to be associated with positive stock price changes, while negative language and a negative vocal tone were associated with negative stock price changes. This suggests that the language and vocal tone used by analysts and managers can have an impact on market perceptions and investor behavior. Price et al. (2017) examine investor reaction to the emotional content of managers' vocal cues using both transcripts and audio files of quarterly earnings conference calls. Price et al. (2017) use a bag of words approach, which does not consider context or a sentence level analysis. Their audio analysis produces an aggregate sentiment index. (Burgoon et al. 2016) examine audio recordings of earnings calls to pinpoint specific linguistic and vocal characteristics. They found that statements related to restatements vary significantly in many vocal and linguistic aspects. Their findings show the usefulness of language and vocal characteristics in identifying potentially deceitful statements and indicate a significant connection between spontaneous responses and rehearsed statements.

Qin and Yang (2019) uses a multimodal approach to predict financial risk levels by considering both vocal features of CEOs, such as pitch, intensity, and jitter, and word embeddings from the textual data of each sentence in earnings conference calls. Word embeddings are a method for representing words as continuous vectors in a low-dimensional space, capturing their lexical and semantic properties. This allows for the representation of words and their relationships in a more meaningful way, which can be useful in tasks such as language translation and text classification (Arora et al. 2017).

2.3 Audio Features

Human speech contains both linguistic and emotional content. The emotional content can play a significant role in how the speech is perceived and understood by the listener. For example, a speaker may convey sadness through their tone of voice, or excitement through their inflection. This emotional content can also affect the way the words are spoken. As a subtopic for audio analysis, the speech emotion recognition detects and classifies emotions through spoken language (El Ayadi et al. 2011). However, emotions in speech can be hard to identify reliably primarily because emotions are subjective. There is no standard way to measure and categorize emotions (Akçay and Oğuz 2020). However, audio features such as pitch and energy can provide valuable insights into the emotional state of a speaker (El Ayadi et al. 2011; Akçay and Oğuz 2020). Therefore, I focus on audio features of speech to understand the responses of investors, instead of attempting to identify emotions.

Drawing from prior literature on the informational role of vocal cues in financial markets (Hobson et al. 2012; Mayew and Venkatachalam 2012; Price et al. 2017), I investigate whether and how the audio features exhibited by managers during the MD and the M&D sessions influence the market participants (i.e., investors and analysts). I expect to show that certain audio properties such as pitch and energy has positive or negative contemporaneous stock return effects. I analyze both the MD and M&D sessions separately since, the informational environment of the M&D sessions in earnings conference calls is "dynamic" and less structured (Blau et al. 2015). Non-verbal communication, including voice, contains information about an individual's belief (Caffi and Janney 1994). As documented in the psychology literature (Zuckerman et al. 1981), the tone of voice can convey non-verbal cues and information. Researchers show that acoustic features such as the pitch, timing, voice quality highly correlate with the underlying emotional states (Cahn 1990).

I argue the multi-modal approach of analyzing both textual and audio features lead to a better understanding of investor behavior. A primary advantage is that it allows for a more comprehensive understanding of investor behavior. Textual data provides a written record of what was said during the earnings call. Audio features, such as pitch, provide valuable insights into the speakers' context, emotions, and intentions (El Ayadi et al. 2011; Akçay and Oğuz 2020). This is because vocal cues, such as tone and emotion, can either support or contradict the message being communicated. Capturing the dependency between the text and audio can be useful in understanding the actual meaning of a message. Prior studies have not analyzed textual and audio aspects of earnings conference calls in a granular way. I fill this void. My research has the potential to provide richer information to investors and researchers than they have had before. For example, an investor might review a transcript of an earnings call and observe that the CEO of a company appears confident about the future. However, the CEO's pitch may show uncertainty or hesitation, implying that the CEO is less confident than what the written transcript suggests. This multi-modal approach could help the investor better understand the CEO's genuine emotions and goals, and may thus impact company's stock price (Goodell et al. 2023).

I posit that an independent analysis of the MD and M&D sections of earnings calls in addition to a multi-modal textual and audio analysis can provide valuable insights into the information content of the conference calls. We can gain a more comprehensive understanding of a firm's outlook, investment decisions, and strategies. The MD section typically presents prepared information. On the other hand, the M&D section, also known as conference calls or earnings calls, allows investors and analysts to seek clarification and ask questions. This section helps understand investor behavior, as it provides an opportunity for market participants to engage with executives about their interests and concerns . Research has shown that firms use these calls to discuss and clarify their earnings news, and investors react to this information (Matsumoto et al. 2011). In particular, both the MD and M&D segments contain additional information beyond what was included in the accompanying press release. The M&D section tends to be more informative than the MD section.

SECTION 3

HYPOTHESIS DEVELOPMENT

Through the multi-modal approach, I analyze textual sentiment of each sentence using machine learning and match with audio features. This approach can be more accurate than unimodal approaches, which use a single mode of communication (D'Mello and Kory 2015). This is because text alone is often insufficient to predict sentiments, especially in cases of sarcasm or ambiguity (Cambria et al. 2017). Poria et al. (2018) demonstrated the improved performance of multi-modal sentiment analysis. The overlap between different sentiments was reduced when using multi-modal features, compared to when using only unimodal features. Multi-modal analysis can more accurately represent sentiment and its relation with audio features present in the earnings conference calls than an analysis that focuses on a single mode.

Poria et al. (2016) examine videos on the web and use both feature and decision-level fusion methods to extract emotion information from video files on Youtube. In their analysis, (Poria et al. 2016) also examine figures of text and visual data. Their model significantly increases the accuracy of the system in recognizing emotion in video files. However, their model also use the bag of words technique for textual analysis. Price et al. (2017a) examine investor reaction to the emotional content of managers' vocal cues. They examine both transcripts and audio files of quarterly earnings conference calls to identify raw emotion found in the voice patterns of executives. They find a strong positive association between executive emotion and investors' initial reaction. They use bag of words approach for textual analysis and use aggregate sentiment for their audio analysis.

Various studies focused on identifying the types of features that are most effective for analyzing affect (emotion or feeling) in speech (Vogt and André 2005). Some common features appear to be particularly useful for this purpose include pitch, energy (Murray and Arnott 1993; Datcu and Rothkrantz 2015; Dellaert et al.?). These features are often used to extract information from speech signals for use in affect recognition tasks.

I expect that matching textual and voice analysis results will give us richer results. I expect to see statistically and economically significant positive (negative) abnormal returns to positive (negative) sentiment in the text along with audio features that is associated with it. Therefore, I propose the following null hypotheses:

H1a: There is no association between the textual sentiment and audio features of earnings conference calls and abnormal stock returns.

H1a predicts that audio features provide information content that is independent of the textual sentiment in the sentences. Another potential benefit of audio information for investors is that it can assist them in interpreting the information in the text. Specifically, investors may perceive textual information as more credible if audio analysis support or justify these measures. Therefore, I propose the following null hypotheses:

H1b: There is no effect of Audio Features on the association between the textual sentiment and abnormal stock returns.

The accounting literature shows that trading volume is driven by differences in the interpretation of public signals, differences in private information-based beliefs, or the strength of the signal itself (Bamber et al. 2011). There is a body of research that examined the relationship between earnings releases and trading activity (Bamber et al. 2011). There is generally an increase in trading activity around earnings releases and this increase is positively associated with the size of the earnings surprise and the stock return (Beaver 1968; Atiase and Bamber 1994; Lobo and Tung 1997). The findings suggest a strong link between earnings surprise and trading activity. For firms that report bad news, trading volume and contemporaneous returns are negatively associated. In other words, when a firm reports good news, an increase in trading volume is typically accompanied by an increase in the stock returns. When a firm reports bad news, an increase in trading volume is typically accompanied by a decrease in stock returns. This suggests that trading volume and contemporaneous returns are closely related, and that investors react differently to positive and negative earnings surprises.

Easley and O'Hara (1992) shows that a strong positive or negative signal can lead to a large number of orders leading to price changes as investors update their beliefs. Healy et al. (1999) show that expanded disclosure of information about a company's financial performance and operations positively impact investor perceptions and decisions, resulting in higher valuations of the company's stock and increased liquidity and interest in the stock from institutions and analysts. Blankespoor et al. (2014) also finds that distribution of company-generated news on Twitter is linked to improved market liquidity and a reduction in information asymmetry, particularly for less visible firms that benefit more from this additional communication channel. The study also found a positive relationship between the dissemination of news on Twitter and liquidity as measured by volume.

Therefore, I propose the following null hypotheses:

H2a: There is no association between the textual sentiment and audio features of earnings conference calls and abnormal stock volume.

H2b: There is no effect of Audio Features on the association between the textual sentiment and abnormal stock volume.

Conference calls allow companies to share important information with their stakeholders, as well as provide an opportunity for analysts to ask questions and get a more in-depth understanding of the company's performance. During the MD session, managers give their perspective on the company's quarter performance and any voluntary disclosures they choose in the presentation part of the call usually following a scripted presentation. This allows the management to present information in a structured and controlled manner. In the M&D session, analysts and other stakeholders are able to gain more insight by questioning the management's perspective and gaining additional information. This session is usually more spontaneous and unscripted. They also present the information verbally, which can be informative to the market due to the meaning conveyed by verbal cues (Mayew et al. 2012). Matsumoto et al. (2011) show that both parts of the call contain incremental information beyond what was included in the press release and the M&D segment is more informative than the presentation segment.

During the MD session, managers use a pre-written text for their speech, their voice may be less expressive and less variation in tone, which means that audio features will be less pronounced as compared to the M&D session. However, during the M&D session, where questions may be unexpected or harsh, I anticipate that investors will have a stronger reaction to vocal cues. Additionally, managers may have received training on their speech, so they may be more aware of their speech patterns during the M&D session. As a result, managers may use softer tones to communicate negative news in order to minimize the effect on the financial market.

H3a: There is no association between the textual sentiment and audio features of earnings conference calls during the presentation and M&D section and abnormal stock return.

H3b: There is no association between the textual sentiment and audio features of earnings conference calls during the presentation and M&D section and abnormal stock volume.

H3c and (H3d): There is no effect of Audio Features on the association between the textual sentiment of earnings conference calls during the presentation and M&D section and abnormal stock return(volume).

SECTION 4
RESEARCH DESIGN AND SAMPLE SELECTION

4.1 Data Selection

I obtain quarterly earnings conference calls of U.S. public companies' audio and transcripts produced by S&P Capital IQ between 2017 and 2020 for which Compustat, CRSP, I/B/E/S data and NYSE TAQ data available. NYSE Trade and Quote (TAQ) database has milliseconds intraday transactions data. There are 45,373 earnings calls audio and transcripts available between 2017 and 2020. S&P Capital IQ quarterly earnings conference call transcripts have the date and starting time of the call. In addition, the start and end of Presentation and "Question and Answer" sections of the transcripts are marked inside the text. I use python natural language toolkit to analyze the text.

Table 1
Sample Selection

Panel A: Sample selection criteria	# of firms	# of conference calls	# Sentences
Conference call transcripts and audios, 2017-2020	3658	45,373	16,498,496
Retain: firms with GVKEY	3623	45,007	16,370,888
Retain: firms have financial information to calculate control variables	3281	41,286	15,033,267
Retain: sentences longer than 1 second for audio analysis	3262	41,246	14,299,110
Final sample	3262	41,246	14,299,110

This panel summarizes our sample selection. The sample covers the period of 2017–2020.

I use the *Financial Phrase Bank* for sentiment analysis which developed by (Malo et al., 2014a). *Financial Phrase Bank* consists of 3,448 English sentences selected randomly from financial news found on *LexisNexis* database. Each sentence of the phrase bank was annotated by 5 to 8 participants with background in finance and business. The annotators assigned labels (negative, neutral, and positive) to each sentence according to their assessment of how the information in the text might affect the stock price. I run a validation test predicting sentiments in *Financial Phrase Bank*. My validation tests achieves 83.5% accuracy.

4.2 Research Design

I use the intraday returns and volume data during the earnings calls. I use Python and the Kaldi toolkit, which is a C++-based speech recognition toolkit, to process earnings call transcripts. Specifically, I use the force alignment technique to generate a time-aligned transcript, which involves obtaining timestamps for each word in the transcript. I then use these timestamps to calculate intraday returns for each sentence in the transcript. Figure 1 provides an overview of the analysis.

Figure 1: Multi modal analysis

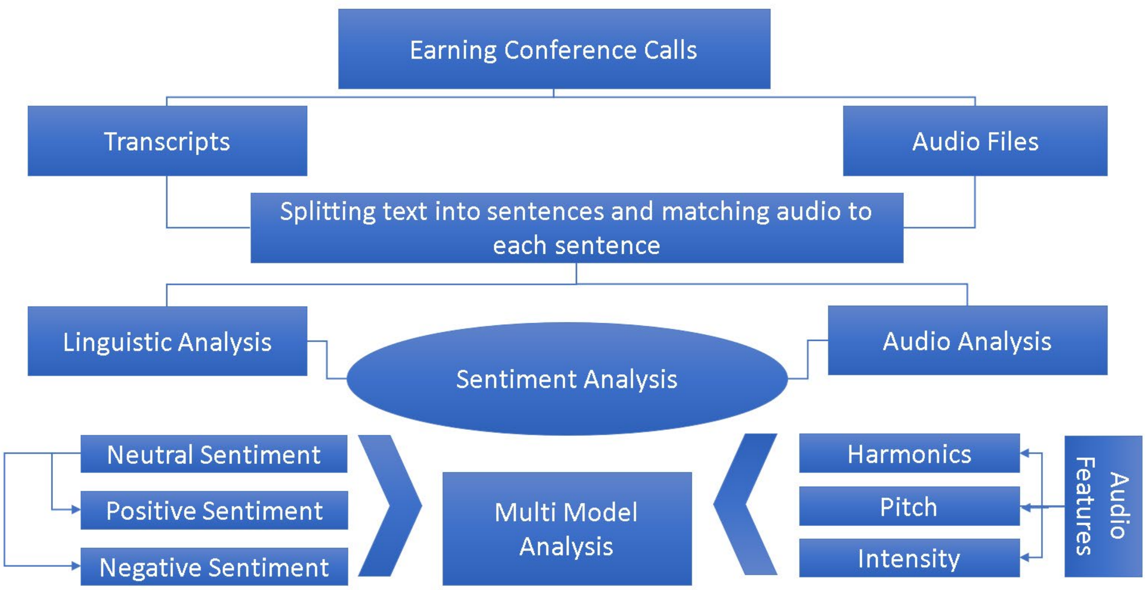


Table 2, Panel A presents summary statistics for the textual data and audio properties such as the number of sentences, words, and characters in earnings call transcripts. I analyze the presentation and M&D part separately. Panel B shows the audio length for per earnings call and per sentence in the earnings call. The mean number of sentences in the earnings calls is larger in the presentation section than in the M&D section. The mean number of words per sentence in the M&D section is less than the mean number of words per sentence in the Presentation section. Overall, managers use less words when answering analysts' questions. Findings are similar for audio length per sentence. The mean length of sentences in the M&D section is marginally shorter than the mean length of sentences in the presentation section.

Table 2
Descriptive Statistics

Panel A: Summary Statistics

Variables	**N**	**Mean**	**SD**	**p10**	**Median**	**p90**
AbnRet	14,299,110	0.000003	0.002146	-0.000183	0	0.000184
AbnVol	14,299,110	0.000003	0.00003	-0.000001	0	0.000003
SenChar	14,299,110	117.256572	69.554927	44	103	208
SenWord	14,299,110	19.965647	11.57896	8	18	35
SenDur	14,299,110	7.398378	5.72305	2.15	6.3	13.749999
hnr	14,299,110	-0.000033	0.000532	-0.000169	-0.000009	0.000096
Jitter loc	14,299,110	0.025386	0.007381	0.017749	0.024405	0.033685
Jitter rap	14,299,110	0.011608	0.00423	0.007297	0.010974	0.016402
F0	14,299,110	147.936825	35.836673	111.885029	140.628651	194.10692
stdevF0	14,299,110	62.712243	37.991525	18.356718	56.581173	115.016763
Shimmer	14,299,110	0.140102	0.027524	0.107529	0.137957	0.175412
Shimmer apq3	14,299,110	0.060897	0.015722	0.042601	0.059434	0.080824
zc	14,299,110	0.088156	0.02702	0.058915	0.084567	0.120747
spectral energy	14,299,110	945.769336	123.058521	814.6874	934.802361	1083.889959
Rms	14,299,110	0.053744	0.043122	0.017334	0.041939	0.101848
Energy	14,299,110	0.052383	0.146645	0.002334	0.017558	0.116577

This panel reports descriptive statistics for the variables

4.2.1 Textual Sentiment measure

I measure the textual sentiment tone in the presentation and M&D portion of the earnings call. Sentiment analysis assumes various forms, from models of polarity (i.e., positive, negative, and neutral) to those of feelings and emotions (i.e., angry, happy, and sad), and even models that identify intentions (i.e., interested versus not interested). At the end of 2018, Google opened source code for a new technique for Natural Language Processing (NLP) called Bidirectional Encoder Representations from Transformers (*BERT*) to the public. Single directional models only looks at left-to-right, right-to-left or combined left-to-right and right-to-left. *BERT* is bidirectional. The paper's (Devlin et al. (2018) results show bidirectionally trained model has a deeper sense of language context and flow compared to single-direction language models.

I use *Financial Phrase Bank* from (Malo et al., 2014b) to base my sentiment scoring. The dataset consists of a list of sentences categorized as positive, negative or neutral according to the information available in the given sentence by 5 to 8 annotators. I use 3,448 sentences that 75% of annotators agree on its categorization. I set aside 20% of all sentences as test and 20% of the remaining as validation set. I further train the *BERT* base model with *Financial Phrase Bank* classifications. I use python natural language processing tool kit to split the text into sentences. The new model assigns negative, neutral, or positive Textual Sentiment prediction to each sentence.

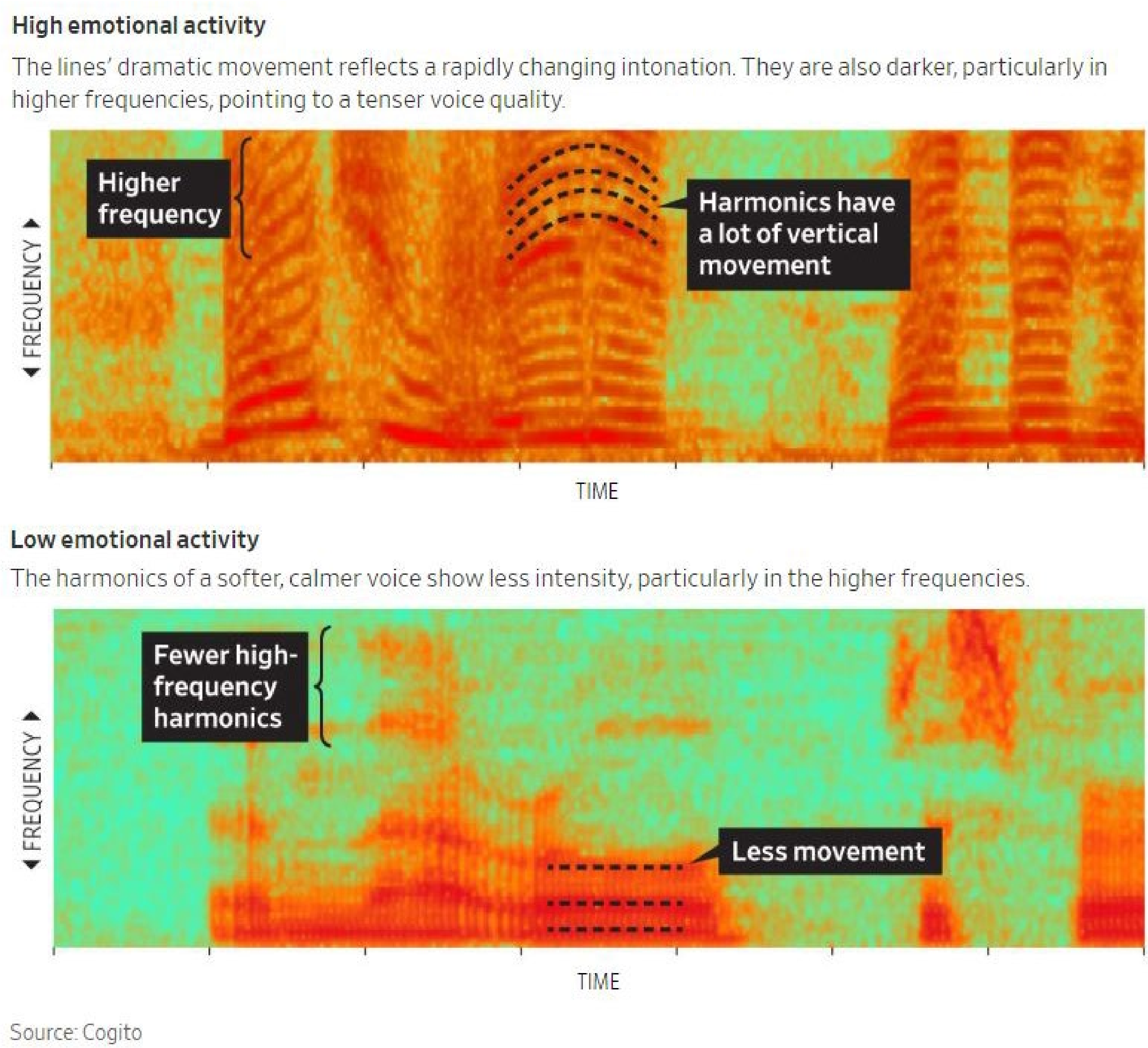
Table 2, Panel A presents summary statistics for textual sentiment variable. As expected, the mean number of negative sentences is significantly lower than both the mean number of positive and neutral sentences. Both the mean number of negative sentences is lower in the M&D section compared to the presentation section. Since the *BERT* is trained using written or pre-prepared data, it might be difficult for the model to pick up the sentiment in unprepared or spoken context.

4.2.2 Audio Features

Audio feature extraction is the process of extracting features from audio signals, which can then be used for tasks such as classification, clustering, and pattern recognition. I use Librosa and Praat(parselmouth) for audio feature extraction. Since I have timestamps for each individual sentence, I use Python and *FFMPEG,* to extract the audio file of each sentence from the original sentence. I divide the audio signals into short-term windows(frames), then calculate the 10 audio features for each frame and find the means for each sentence. Borrowing from and adding to (Mayew and Venkatachalam 2012), I use the following audio features: Mean fundamental frequency (F0), standard deviation of fundamental frequency (F0Std), Jitter , Jitter relative average perturbation (Jitter_rap), Shimmer, Shimmer three-point amplitude perturbation quotient (Shimmer_apq3), zero crossing rate (ZC), spectral energy, loudness (RMS-root mean square), energy and mean harmonic-to-noise (HNR) ratio.

Table 2 presents summary statistics for audio features. Figure 2 shows the representation of how the harmonics and intensity of the audio changes by high and low emotional activity.

Figure 2: High Emotional and Low Emotional Activity sound recordings



http://webreprints.djreprints.com/56102.html

4.3 Abnormal Returns

I examine the effect of textual sentiment and Audio Features during conference calls on stock return while controlling for industry and year fixed effects using the following model:

$$AbnRet_{i,j} = \beta_0 + \beta_1 TextSentiment_{i,j} + \beta_2 AudioFeatures_{i,j} + \varepsilon \qquad (1)$$

where $AbnRet_{i,t}$ is defined as abnormal return for firm *i* as the difference between the return for firm *i* for each sentence j and the market return during that period. *TextSentiment* and AudioFeatures are sentiment and AudioFeatures for firm *i* for sentence j. I also include fiscal year and industry fixed effects in the regressions and cluster standard errors at year-month level. I expect that if additional information is disclosed during the conference calls, the coefficient of *TextSentiment* to be both significant and positively related to the abnormal returns.

I examine whether the Audio Features *have an effect on the association between the textual sentiment and abnormal stock returns* while controlling for industry and year fixed effects using the following model:

$$\begin{aligned} AbnRet_{i,j} = \;& \beta_0 + \beta_1 TextSentiment_{i,j} + \beta_2 AudioFeatures_{i,j} + \\ & \beta_3 TextSentiment_{i,j} * AudioFeatures_{i,j} + \varepsilon \qquad (2) \end{aligned}$$

4.4 Abnormal Trading Volume

I calculate Abnormal Volume as the firm's total shares traded during the sentence period divided by total shares outstanding, minus the average shares traded over the last 5 business days during the same sentence period divided by total shares outstanding (Blankespoor et al. 2019).

I examine the effect of textual sentiment and Audio Features during conference calls on volume while controlling for industry and year fixed effects using the following model:

$$AbnVol_{i,j} = \beta_0 + \beta_1 TextSentiment_{i,j} + \beta_2 AudioFeatures_{i,j} + \varepsilon \qquad (3)$$

I examine whether the Audio Features *have an effect on the association between the textual sentiment and abnormal stock volume* while controlling for industry and year fixed effects using the following model:

$$\begin{aligned} AbnVol_{i,j} = \; & \beta_0 + \beta_1 TextSentiment_{i,j} + \beta_2 AudioFeatures_{i,j} + \\ & \beta_3 TextSentiment_{i,j} * AudioFeatures_{i,j} + \varepsilon \qquad (4) \end{aligned}$$

I run cross-sectional regressions using Eq. (1) and Eq. (2) with AbnRet and *AbnVol* separately. A positive and significant sentiment variable using AbnRet shows that as the sentiment increases the abnormal return increases. However, the volume might increase with the extreme sentiment. Therefore, as the sentiment becomes positive or negative, I expect an increase in *AbnVol* in Eq. (2). A detailed description of each independent variable can be found in Appendix 1.

SECTION 5

EMPIRICAL ANALYSIS

I hypothesize that combination of audio and textual analysis of earnings conference calls better explains stock returns when compared to textual only analysis of earnings conference calls. My main test examines how audio features and textual sentiment explains abnormal stock returns and volume during intraday earnings calls. Table 3 shows that there is a negative association between negative text sentiment and abnormal returns. In addition, I find that shimmer, zero crossing and spectral energy helps explain abnormal returns.

In column (2) and (4), I test how textual sentiment and audio features explains abnormal stock returns and abnormal trading volume. In column (3) and (5), I include interaction of text sentiment and audio features.

Column (2) shows that negative text sentiment (Neg) elicit abnormal return of (-0.0000789210). I find three the audio features are associate with the abnormal return. Shimmer_apq3 and spectral energy are associated with positive abnormal return. Column (3) shows the results of whether the Audio Features have an effect on the association between the textual sentiment and abnormal stock return. I find that harmonic is associated with a decrease in abnormal return for both the positive and negative text sentiment. Jitter, spectral energy, Energy and shimmer_apq3 are associated with a decrease in abnormal return for negative text sentiment. Shimmer_apq3 is associated with a decrease in abnormal return for positive text sentiment. Zero crossing is associated with an increase in abnormal return for both positive and negative text sentiment.

I expect that abnormal trading volume to increase significantly for extremely positive or extremely negative. I find that positive (Pos) and negative text sentiment (Neg) are both positively associated with abnormal volume. I also find that Jitter_rap, standard deviation of pitch(stdevF0) and energy are positively associated with abnormal volume. Jitter, pitch (F0) and shimmer and loudness (RMS) are negatively associated with abnormal volume. Column (5) shows the results of whether the Audio Features have an effect on the association between the textual sentiment and abnormal volume. Harmony is associated with an increase in abnormal return for both positive and negative text sentiment. Jitter and Energy are associated with a decrease in abnormal return for negative text sentiment.

While the evidence is consistent with market participants eliciting more (less) information from the textual data and audio features, I have no evidence regarding how investors' react to the presentation and M&D section of the earnings calls. The following subsections deals with the presentation and M&D sections separately.

Table 3 Earnings Conference Call Full Duration -- Abnormal Return and Abnormal Volume

	(2)	(3)	(4)	(5)
Variables	AbnRet	AbnRet	AbnVol	AbnVol
Textual Sentiment				
Pos	-0.0000347207	0.0004757130	0.0000003771***	0.0000009754
	(-0.91)	(1.27)	(4.53)	(0.74)
Neg	-0.0000789210***	0.0005329785**	0.0000003114***	-0.0000014943
	(-3.66)	(2.03)	(3.03)	(-1.43)
hnr	0.0109474896	0.0165437449*	0.0001696710	0.0001590678
	(1.62)	(1.86)	(1.48)	(1.63)
Jitter	0.0118194144	0.0237852241**	-0.0001387983***	-0.0001269026***
	(1.11)	(2.24)	(-4.92)	(-4.23)
Jitter_rap	-0.0033710348	-0.0258630092	0.0001993300***	0.0001737259***
	(-0.16)	(-1.31)	(4.04)	(3.49)
F0	0.0000004548	0.0000008466	-0.0000000100***	-0.0000000092***
	(0.70)	(1.15)	(-9.06)	(-10.05)
stdevF0	-0.0000007640	-0.0000012726	0.0000000141***	0.0000000141***
	(-0.82)	(-1.10)	(5.46)	(5.43)
Shimmer	-0.0044281102	-0.0055049157	-0.0000074010***	-0.0000080386***
	(-1.39)	(-1.37)	(-3.43)	(-5.12)
Shimmer_apq3	0.0084300247*	0.0108383292*	-0.0000039128	-0.0000023925
	(1.75)	(1.79)	(-0.83)	(-0.63)
zc	-0.0016096198**	-0.0021714928**	-0.0000045566	-0.0000054275
	(-2.27)	(-2.15)	(-0.87)	(-0.93)
spectral_energy	0.0000006082***	0.0000007014**	0.0000000013	0.0000000013
	(2.90)	(2.53)	(0.60)	(0.54)
Rms	-0.0005117037	-0.0006263830	-0.0000193281***	-0.0000223682***
	(-0.69)	(-0.71)	(-4.77)	(-7.16)
Energy	0.0001953665	0.0003836985	0.0000075149***	0.0000096436***
	(1.16)	(1.20)	(4.30)	(6.72)
Pos × hnr		-0.0168083950*		0.0000124083
		(-1.86)		(0.16)
Neg × hnr		-0.0310698830*		0.0001661125**
		(-1.71)		(2.14)
Pos × Jitter		-0.0579952900		-0.0000292583
		(-1.55)		(-1.49)
Neg × Jitter		-0.0232823062**		-0.0001436054***
		(-2.48)		(-3.28)
Pos × Jitter_rap		0.1162335586		0.0000740869**
		(1.24)		(2.31)
Neg × Jitter_rap		0.0253534226		0.0003034570***
		(1.25)		(2.79)
Pos × F0		-0.0000020091		-0.0000000037**
		(-1.32)		(-2.10)
Neg × F0		-0.0000009092		-0.0000000031
		(-1.31)		(-1.04)
Pos × stdevF0		0.0000022490		-0.0000000021
		(1.35)		(-1.22)
Neg × stdevF0		0.0000008507		0.0000000001
		(0.74)		(0.03)
Pos × Shimmer		0.0052622705		0.0000036622
		(1.24)		(0.79)
Neg × Shimmer		0.0049212140		0.0000054573
		(1.15)		(0.83)
Pos × Shimmer_apq3		-0.0135222609*		-0.0000062951
		(-1.85)		(-0.82)
Neg × Shimmer_apq3		-0.0087228958		-0.0000228460
		(-1.33)		(-1.59)
Pos × zc		0.0021512796*		0.0000051949
		(1.68)		(1.31)
Neg × zc		0.0019899864*		0.0000000688
		(1.87)		(0.02)
Pos × spectral_energy		-0.0000003270		-0.0000000008
		(-0.62)		(-0.50)

Neg × spectral_energy		-0.0000006463**		0.0000000033*
		(-2.26)		(1.92)
Pos × Rms		-0.0001754300		0.0000076481
		(-0.14)		(1.37)
Neg × Rms		0.0013119198**		0.0000001156
		(2.04)		(0.04)
Pos × Energy		-0.0003233251		-0.0000046884**
		(-0.98)		(-2.34)
Neg × Energy		-0.0005484643*		-0.0000018049*
		(-1.68)		(-1.76)
Intercept	-0.0004857262**	-0.0005913722*	0.0000060058***	0.0000059893***
	(-2.08)	(-2.00)	(4.56)	(4.03)
Observations	14299110	14299110	14299110	14299110
R^2	0.0000211462	0.0000230292	0.0026461601	0.0026787063

*Notes: t-values are in parentheses. *** p<.01, ** p<.05, * p<.1*
This table reports OLS estimation of Equation (1,2,3 and 4) using a sample of 41,246 earnings conference calls.
See Appendix A for definitions of all variables.

5.1 Analysis of Presentation Section

I test the explanatory power of audio emotion and text sentiment measures on abnormal stock returns during the presentation section. Matsumoto, Pronk, and Roelofsen (2011) find that informativeness of the M&D discussion section is more than the presentation section. The start and end of the presentation section is tagged in the transcripts. Table 4 shows that textual sentiment and audio features does not explain abnormal stock returns but explain abnormal volume.

Table 4 Earnings Conference Call Presentation Section -- Abnormal Return and Abnormal Volume

	(2)	(3)	(4)	(5)
Variables	AbnRet	AbnRet	AbnVol	AbnVol
Textual Sentiment				
Pos	0.0000398599	-0.0003864791	0.0000005782***	0.0000004916
	(0.83)	(-1.05)	(6.54)	(0.75)
Neg	-0.0000154211	0.0001028027	0.0000006322***	0.0000002584
	(-0.76)	(0.58)	(5.70)	(0.27)
hnr	-0.0075574690	-0.0118517075	0.0002880522	0.0002799389
	(-1.31)	(-1.21)	(1.49)	(1.46)
Jitter	-0.0192888657	0.0007272305	-0.0001106980***	-0.0000817614***
	(-0.94)	(0.39)	(-4.67)	(-3.75)
Jitter_rap	0.0492443272	-0.0055836507	0.0001795809***	0.0001157354***
	(0.93)	(-0.79)	(3.88)	(2.98)
F0	-0.0000008548	-0.0000003532	-0.0000000064***	-0.0000000037***
	(-1.14)	(-1.45)	(-3.82)	(-2.79)
stdevF0	0.0000007952	0.0000002578	0.0000000073***	0.0000000066***

	(0.86)	(0.68)	(2.82)	(3.86)
Shimmer	0.0013458743	0.0010965355	-0.0000024913	-0.0000008173
	(1.48)	(1.07)	(-0.57)	(-0.27)
Shimmer_apq3	-0.0032237346	-0.0017118757	-0.0000094782	-0.0000111115
	(-1.24)	(-1.12)	(-1.05)	(-1.59)
zc	-0.0004139030	0.0000228355	-0.0000018357	-0.0000041312
	(-1.11)	(0.08)	(-0.47)	(-1.33)
spectral_energy	0.0000003175	0.0000000770	0.0000000007	0.0000000004
	(1.09)	(0.75)	(0.86)	(0.63)
Rms	-0.0003861785	0.0003702581	-0.0000135947**	-0.0000184518***
	(-0.54)	(0.75)	(-2.65)	(-4.84)
Energy	0.0000179643	0.0000270682	0.0000046491**	0.0000070435***
	(0.39)	(0.53)	(2.50)	(4.66)
Pos × hnr		0.0112047568		0.0000110608
		(1.10)		(0.23)
Neg × hnr		0.0030941841		0.0000789622
		(0.17)		(0.78)
Pos × Jitter		-0.0519052524		-0.0000493104*
		(-0.95)		(-1.99)
Neg × Jitter		-0.0000723349		-0.0001236719***
		(-0.01)		(-3.88)
Pos × Jitter_rap		0.1448416331		0.0001223793**
		(1.02)		(2.57)
Neg × Jitter_rap		0.0045516998		0.0002323256***
		(0.43)		(3.75)
Pos × F0		-0.0000015766		-0.0000000070***
		(-0.72)		(-3.59)
Neg × F0		0.0000002257		-0.0000000029
		(0.68)		(-1.13)
Pos × stdevF0		0.0000014555		0.0000000012
		(0.63)		(0.45)
Neg × stdevF0		-0.0000002643		-0.0000000014
		(-0.67)		(-0.38)
Pos × Shimmer		0.0009454484		-0.0000036250
		(0.42)		(-0.69)
Neg × Shimmer		-0.0021230850		-0.0000024624
		(-1.38)		(-0.40)
Pos × Shimmer_apq3		-0.0053141281		0.0000032571
		(-0.73)		(0.34)
Neg × Shimmer_apq3		0.0034706014		0.0000037936
		(1.60)		(0.35)
Pos × zc		-0.0008822239		0.0000053654
		(-0.91)		(1.50)
Neg × zc		-0.0004416241		0.0000047830
		(-0.74)		(0.88)
Pos × spectral_energy		0.0000006168		0.0000000005
		(0.91)		(0.85)
Neg × spectral_energy		-0.0000000370		0.0000000011
		(-0.32)		(1.04)
Pos × Rms		-0.0019722381		0.0000083958
		(-1.14)		(1.55)
Neg × Rms		-0.0003593169		-0.0000006396
		(-0.71)		(-0.17)
Pos × Energy		0.0000430404		-0.0000037022*
		(0.47)		(-2.00)
Neg × Energy		-0.0000759974		-0.0000001359
		(-1.15)		(-0.13)
Intercept	-0.0002180108	-0.0000433483	0.0000046198***	0.0000047521***
	(-1.12)	(-0.37)	(6.48)	(7.84)
Observations	5497201	5497201	5497201	5497201
R^2	0.0000087678	0.0000138479	0.0024517396	0.0024870061

*Notes: t-values are in parentheses. *** p<.01, ** p<.05, * p<.1*

This table reports OLS estimation of Equation (1,2,3 and 4) using a sample of 41,246 earnings conference calls presentation section.

See Appendix A for definitions of all variables.

5.2 Analysis of M&D Section

Column (3) shows that abnormal returns are positively associated with Positive text sentiment.

Table 5 Earnings Conference Call M&D Section -- Abnormal Return and Abnormal Volume

	(2)	(3)	(4)	(5)
Variables	AbnRet	AbnRet	AbnVol	AbnVol
Textual Sentiment				
Pos	-0.0000410243	0.0004782925***	0.0000008678***	0.0000017283
	(-1.34)	(3.84)	(12.23)	(1.03)
Neg	-0.0000549009	-0.0000868347	0.0000007998***	-0.0000051817
	(-1.67)	(-0.52)	(6.23)	(-1.11)
hnr	0.0222870389**	0.0018458797	0.0000991616	0.0001102445
	(2.04)	(0.63)	(1.14)	(1.44)
Jitter	0.0272923471**	-0.0006309697	-0.0001269523***	-0.0001276034***
	(2.24)	(-0.36)	(-3.74)	(-3.66)
Jitter_rap	-0.0283571677	-0.0012824964	0.0001737095***	0.0001745090***
	(-1.26)	(-0.20)	(2.98)	(3.03)
F0	0.0000012741	0.0000003779	-0.0000000090***	-0.0000000093***
	(1.39)	(1.27)	(-7.99)	(-8.94)
stdevF0	-0.0000017271	-0.0000001665	0.0000000143***	0.0000000148***
	(-1.32)	(-1.37)	(5.10)	(4.92)
Shimmer	-0.0066704627	0.0000330469	-0.0000070819***	-0.0000084342***
	(-1.48)	(0.07)	(-3.32)	(-5.01)
Shimmer_apq3	0.0127905422*	0.0014130800	-0.0000022217	-0.0000001765
	(1.91)	(1.33)	(-0.55)	(-0.05)
zc	-0.0024015835**	0.0000979464	-0.0000050612	-0.0000058431
	(-2.10)	(0.27)	(-0.74)	(-0.81)
spectral_energy	0.0000007095**	0.0000000408	0.0000000011	0.0000000013
	(2.32)	(0.99)	(0.39)	(0.44)
Rms	-0.0006495219	0.0009083644	-0.0000241233***	-0.0000231082***
	(-0.60)	(0.99)	(-8.72)	(-7.69)
Energy	0.0004493459	-0.0000232737	0.0000107340***	0.0000107250***
	(1.27)	(-0.54)	(8.01)	(7.33)
Pos × hnr		-0.0045735415		-0.0000894899
		(-0.61)		(-1.37)
Neg × hnr		-0.0298758008		0.0001611398**
		(-1.05)		(2.05)
Pos × Jitter		-0.0091584897		0.0000445236
		(-1.58)		(1.37)
Neg × Jitter		-0.0086353832		-0.0001577430
		(-1.27)		(-1.18)
Pos × Jitter_rap		0.0171761137		-0.0000805939
		(1.63)		(-1.46)
Neg × Jitter_rap		0.0158112176		0.0003950410
		(1.01)		(1.25)
Pos × F0		-0.0000007567**		0.0000000047**
		(-2.65)		(2.21)
Neg × F0		-0.0000002185		-0.0000000049
		(-0.35)		(-0.98)
Pos × stdevF0		0.0000002517		-0.0000000062**
		(0.77)		(-2.19)
Neg × stdevF0		-0.0000010333		0.0000000075
		(-0.82)		(1.57)
Pos × Shimmer		-0.0020903027*		0.0000123735**
		(-1.99)		(2.07)
Neg × Shimmer		0.0011770006		0.0000139544
		(0.46)		(1.31)
Pos × Shimmer_apq3		0.0023242392		-0.0000146848

		(1.33)		(-1.52)
Neg × Shimmer_apq3		0.0021505456		-0.0000627482*
		(0.48)		(-1.99)
Pos × zc		0.0002332561		0.0000104561*
		(0.45)		(1.87)
Neg × zc		-0.0006950418		-0.0000081557
		(-0.67)		(-1.37)
Pos × spectral_energy		-0.0000001832*		-0.0000000028
		(-1.87)		(-1.19)
Neg × spectral_energy		-0.0000000349		0.0000000092
		(-0.10)		(1.33)
Pos × Rms		-0.0006951741*		-0.0000082390**
		(-1.76)		(-2.32)
Neg × Rms		0.0010544985		-0.0000075125
		(0.94)		(-1.50)
Pos × Energy		0.0001916687**		0.0000008185
		(2.48)		(0.54)
Neg × Energy		-0.0002692154		0.0000001391
		(-0.90)		(0.07)
Intercept	-0.0005937538*	-0.0001451237	0.0000061645***	0.0000061085***
	(-1.84)	(-0.95)	(3.41)	(3.22)
Observations	8801909	8801757	8801909	8801909
R^2	0.0000378778	0.0002293513	0.0032368751	0.0032540576

*Notes: t-values are in parentheses. *** p<.01, ** p<.05, * p<.1*

This table reports OLS estimation of Equation (1,2,3 and 4) using a sample of 41,246 earnings conference calls M&D section.

See Appendix A for definitions of all variables.

SECTION 6

CONCLUSION AND DISCUSSION

Earnings conference calls provide managers an important tool to inform the investors about the operating environment of companies and expectations for future periods. These conferences contain both qualitative and quantitative information. Qualitative textual information help investors to assess quantitative information. Audio-based qualitative information such as audio features speech can also be an important source of information in investors' decision-making process. I use machine learning models to analyze and understand the textual disclosures. I conduct a sentence based analysis with the associated audio features during the earnings conference calls. I employ *BERT* language model of (Devlin et al., 2018) for textual sentiment analysis. Audio Features can complement investor sentiment obtained from textual data. I find that both text sentiment and audio features are informative for the investors.

REFERENCES

Akçay, M. B., and K. Oğuz. 2020. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication* 116: 56–76.

Arora, S., Y. Liang, and T. Ma. 2017. A Simple but Tough-to-Beat Baseline for Sentence Embeddings. In *International Conference on Learning Representations*.

Atiase, R. K., and L. S. Bamber. 1994. Trading volume reactions to annual accounting earnings announcements. The incremental role of predisclosure information asymmetry. *Journal of Accounting and Economics*.

El Ayadi, M., M. S. Kamel, and F. Karray. 2011. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition* 44 (3): 572–587.

Bamber, L. S., O. E. Barron, and D. E. Stevens. 2011. Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research. *Contemporary Accounting Research* 28 (2): 431–471.

Banse, R., and K. R. Scherer. 1996. Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology* 70 (3): 614–636.

Beaver, W. H. 1968. The Information Content of Annual Earnings Announcements. *Journal of Accounting Research* 6: 67.

Blankespoor, E., E. Dehaan, J. Wertz, and C. Zhu. 2019. Why Do Individual Investors Disregard Accounting Information? The Roles of Information Awareness and Acquisition Costs. *Journal of Accounting Research* 57 (1): 53–84.

Blankespoor, E., G. S. Miller, and H. D. White. 2014. The Role of Dissemination in Market Liquidity: Evidence from Firms' Use of Twitter™. *The Accounting Review* 89 (1): 79–112.

Blau, B. M., J. R. DeLisle, and S. M. K. Price. 2015. Do sophisticated investors interpret earnings conference call tone differently than investors at large? Evidence from short sales. *Journal of Corporate Finance* 31: 203–219.

Bochkay, K., J. Hales, and S. Chava. 2020. Hyperbole or reality? Investor response to extreme language in earnings conference calls. *Accounting Review* 95 (2): 31–60.

Ten Bosch, L. 2003. Emotions, speech and the ASR framework. *Speech Communication* 40 (1–2): 213–225.

Brown, G. W., and M. T. Cliff. 2004. Investor sentiment and the near-term stock market.

Journal of Empirical Finance* 11 (1): 1–27.

Brown, S. V., and J. W. Tucker. 2011. Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research* 49 (2): 309–346.

Burgoon, J., W. J. Mayew, J. S. Giboney, A. C. Elkins, K. Moffitt, B. Dorn, M. Byrd, and L. Spitzley. 2016. Which Spoken Language Markers Identify Deception in High-Stakes Settings? Evidence From Earnings Conference Calls. *Journal of Language and Social Psychology* 35 (2): 123–157.

Caffi, C., and R. W. Janney. 1994. Toward a pragmatics of emotive communication. *Journal of Pragmatics* 22: 325–373.

Cahn, J. E. 1990. The Generation of A ect in Synthesized Speech.

Cambria, E., D. Das, S. Bandyopadhyay, and A. Feraco. 2017. Affective Computing and Sentiment Analysis: 1–10.

Chatterjee, A., U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal. 2019. Understanding Emotions in Text Using Deep Learning and Big Data. *Computers in Human Behavior* 93 (April 2018): 309–317.

Cho, H., and V. Muslu. 2021. How do firms change investments based on MD&A disclosures of peer firms? *Accounting Review* 96 (2): 177–204.

Corive, R., E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine* 18 (1): 32–80.

D'Mello, S. K., and J. Kory. 2015. A Review and Meta-Analysis of Multimodal Affect Detection Systems. *ACM Computing Surveys (CSUR)* 47 (3).

Datcu, D., and L. J. M. Rothkrantz. 2015. Semantic Audiovisual Data Fusion for Automatic Emotion Recognition. *Emotion Recognition: A Pattern Analysis Approach*: 411–435.

Davis, A. K., J. M. Piger, and L. M. Sedor. 2012. Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research* 29 (3): 845–868.

Dellaert, F., T. Polzin, and A. Waibel. Recognizing emotion in speech. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96* 3: 1970–1973.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Easley, D., and M. O'Hara. 1992. Time and the Process of Security Price Adjustment. *The Journal of Finance* 47 (2): 577–605.

Frankel, R., W. J. Mayew, and Y. Sun. 2010. Do pennies matter? Investor relations consequences of small negative earnings surprises. *Review of Accounting Studies* 15 (1): 220–242.

Goodell, J. W., S. Kumar, P. Rao, and S. Verma. 2023. Emotions and stock market anomalies: A systematic review. *Journal of Behavioral and Experimental Finance* 37: 100722.

Goudbeek, M., and K. Scherer. 2010. Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America* 128 (3): 1322.

Guo, L., F. Shi, and J. Tu. 2016. Textual analysis and machine leaning: Crack unstructured data in finance and accounting. *The Journal of Finance and Data Science* 2 (3): 153–170.

Han, S., J. S. Lerner, and D. Keltner. 2007. Feelings and consumer decision making: The appraisal-tendency framework. *Journal of Consumer Psychology* 17 (3): 158–168.

Healy, P. M., A. P. Hutton, and K. G. Palepu. 1999. Stock Performance and Intermediation Changes Surrounding Sustained Increases in Disclosure. *Contemporary Accounting Research* 16 (3): 485–520.

Hobson, J. L., W. J. Mayew, and M. Venkatachalam. 2012. Analyzing Speech to Detect Financial Misreporting. *Journal of Accounting Research* 50 (2): 349–392.

Huang, A. H., H. Wang, and Y. Yang. 2022. FinBERT : A Large Language Model for Extracting Information from Financial Text† . *Contemporary Accounting Research*.

Juslin, P. N., and P. Laukka. 2001. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion (Washington, D.C.)* 1 (4): 381–412.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2): 263–291.

Kearney, C., and S. Liu. 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis* 33 (Cc): 171–185.

Kothari, S. P., X. Li, and J. E. Short. 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *Accounting Review* 84 (5): 1639–1670.

Kreiman, J., and D. Sidtis. 2011. Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*.

Lazarus, R. S. 1991. Progress on a cognitive-motivational-relational theory of emotion. *The American psychologist* 46 (8): 819–834.

Lee, W. Y., C. X. Jiang, and D. C. Indro. 2002. Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance* 26 (12): 2277–2299.

Lerner, J. S., J. H. Goldberg, and P. E. Tetlock. 1998. Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin* 24 (6): 563–574.

Lerner, J. S., and D. Keltner. 2000. Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition and Emotion* 14 (4): 473–493.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2–3): 221–247.

Li, F. 2010. The information content of forward- looking statements in corporate filings-A naïve bayesian machine learning approach. *Journal of Accounting Research* 48 (5): 1049–1102.

Lobo, G. J., and S. Tung. 1997. Relation between predisclosure information asymmetry and trading volume reaction around quarterly earnings announcements. *Journal of Business Finance and Accounting* 24 (6): 851–867.

Loughran, T., and B. Mcdonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance* 66 (1): 35–65.

Loughran, T., and B. Mcdonald. 2016. Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research* 54 (4): 1187–1230.

Matsumoto, D., M. Pronk, and E. Roelofsen. 2011. What makes conference calls useful? The information content of managers' presentations and analysts' discussion sessions. *Accounting Review* 86 (4): 1383–1414.

Mayew, W. J., V. M., and M. Venkatachalam. 2012. The power of voice: Managerial affective states and future firm performance. *Journal of Finance* 67 (1): 1–43.

Mayew, W. J., C. A. Parsons, and M. Venkatachalam. 2013. Voice pitch and the labor market success of male chief executive officers. *Evolution and Human Behavior* 34 (4): 243–248.

Mayew, W. J., M. Sethuraman, and M. Venkatachalam. 2020. Individual Analysts' Stock Recommendations, Earnings Forecasts, and the Informativeness of Conference Call Question and Answer Sessions. *The Accounting Review* 95 (6): 311–337.

Mayew, W. J., and M. Venkatachalam. 2012. The power of voice: Managerial affective states and future firm performance. *Journal of Finance* 67 (1): 1–43.

Mohammad, S. M. 2016. Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. *Emotion Measurement*: 201–237.

Murray, I. R., and J. L. Arnott. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America* 93 (2): 1097–1108.

Muslu, V., S. Radhakrishnan, K. R. Subramanyam, and D. Lim. 2015. Forward-looking MD&A disclosures and the information environment. *Management Science* 61 (5): 931–948.

Owren, M. J., and J.-A. Bachorowski. 2007. Measuring emotion-related vocal acoustics. In *Handbook of emotion elicitation and assessment.*, 239–266. Series in affective science. New York, NY, US: Oxford University Press.

Pang, B., and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135.

Poria, S., E. Cambria, N. Howard, G.-B. Huang, and A. Hussain. 2016. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing* 174: 50–59.

Price, S. M. K., J. S. Doran, D. R. Peterson, and B. A. Bliss. 2012. Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking and Finance* 36 (4): 992–1011.

Price, S. M. K., M. J. Seiler, and J. Shen. 2017. Do Investors Infer Vocal Cues from CEOs During Quarterly REIT Conference Calls? *Journal of Real Estate Finance and Economics* 54 (4): 515–557.

Qin, Y., and Y. Yang. 2019. What You Say and How You Say It Matters: Predicting Stock Volatility Using Verbal and Vocal Cues. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*: 390–401.

Ravi, K., and V. Ravi. 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems* 89 (November): 14–46.

Rennekamp, K. M., M. Sethuraman, and B. A. Steenhoven. 2022. Engagement in earnings conference calls. *Journal of Accounting and Economics* 74 (1): 101498.

Scherer, K. R. 1999. Appraisal theory. In *Handbook of cognition and emotion.*, 637–663. Hoboken, NJ, US: John Wiley & Sons Ltd.

Scherer, K. R., R. Banse, H. G. Wallbott, and T. Goldbeck. 1991. Vocal cues in emotion encoding and decoding. *Motivation and Emotion* 15 (2): 123–148.

Smales, L. A. 2015. Time-variation in the impact of news sentiment. *International Review of Financial Analysis* 37: 40–50.

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

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and

I. Polosukhin. 2017. Attention Is All You Need. *Advances in Neural Information Processing Systems* 2017-December: 5999–6009.

Vogt, T., and E. André. 2005. Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. *IEEE International Conference on Multimedia and Expo, ICME 2005* 2005: 474–477.

Wang, X., S. Liu, J. Liu, J. Chen, J. Zhu, and B. Guo. 2016. TopicPanorama: a full picture of relevant topics. *IEEE Trans Vis Comput Graphics* 22 (12): 2508–2521.

Weiss, K., T. M. Khoshgoftaar, and D. D. Wang. 2016. A survey of transfer learning. *Journal of Big Data* 3 (1): 1–40.

Wigglesworth, R. 2020. Robo-surveillance shifts tone of CEO earnings calls | Financial Times. *Financial Times*, December 5.

Zuckerman, M., B. M. Depaulo, and R. Rosenthal. 1981. Verbal and Nonverbal Communication of Deception. *Advances in Experimental Social Psychology* 14 (C): 1–59.

Appendix 1: Variables Descriptions

Variable	Definition
Text Sentiment	Text sentiment (Neutral, Positive and Negative)
Abnormal Return	Abnormal return for firm i as the difference between the return for sentence j and the market proxy (SPY -SP500 index ETF) return
Abnormal Volume	Abnormal volume for firm i as the ratio of volume during sentence j and average volume during sentence j for the past 5 days divided by total shares outstanding
Pitch - F0 (fundamental frequency)	Lowest periodic cycle of the acoustic signal
StdevF0 – standard deviation of pitch (F0)	Standard deviation of pitch (F0)
RMS - Loudness	Measures of energy in the acoustic signal
Jitter - Pitch irregularity	Frequency instability of F0
Jitter_rap	Relative average perturbation, meaning the average absolute difference between a period and the average of it and its two neighbours, divided by the average period
Shimmer - Loudness irregularity	Amplitude instability of F0
Shimmer_apq3	This is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its neighbours, divided by the average amplitude.
ZC – Zero Crossing Rate	Rhe rate at which a signal changes from positive to zero to negative or from negative to zero to positive.
Energy	Amplitude of the sound.
Timbre - Spectral Energy	Relative energy in different frequency bands
HNR (harmonics to noise ratio) Voice quality	Mean ratio of quasi-periodic to non-periodic signals across time segments

Table 2
Descriptive Statistics

Panel B: Summary Statistics for the Presentation Section - Neutral Textual Sentiment

Variables	N	Mean	SD	p10	Median	p90
AbnRet	2774909	0.000002699	0.00219018	-0.00014888	0	0.00014995
AbnVol	2774909	0.000002256	0.00002676	-3.432E-07	0	0.000001326
SenChar	2774909	116.46341	68.073464	46	104	201
SenWord	2774909	18.972852	10.367811	8	17	32
SenDur	2774909	6.940159	5.358153	2.152438	5.970692	12.61
hnr	2774909	-0.00003054	0.0005375	-0.00016292	-0.00000794	0.0000987
Jitter	2774909	0.02594184	0.0066297	0.01874045	0.02519873	0.03376227
Jitter_rap	2774909	0.01188412	0.00380837	0.00779676	0.01138755	0.01641781
F0	2774909	154.48322	35.130459	117.5889	146.88972	203.69339
stdevF0	2774909	60.952219	35.363394	20.917501	54.444684	110.33076
Shimmer	2774909	0.14072514	0.02637794	0.10873067	0.13921067	0.17488729
Shimmer_apq3	2774909	0.06100703	0.01529168	0.04260468	0.05992236	0.08068554
zc	2774909	0.08610848	0.02499613	0.05823364	0.08267419	0.11822869
spectral_e~y	2774909	934.73466	116.08657	808.29722	925.82506	1066.6778
Rms	2774909	0.05621144	0.04403969	0.01900598	0.04407699	0.10525687
Energy	2774909	0.05197187	0.1359475	0.00267453	0.01804748	0.11542527

This panel reports descriptive statistics for the variables for the Presentation Section - Neutral Textual Sentiment.

Panel C: Summary Statistics for the Presentation Section - Positive Textual Sentiment

Variables	N	Mean	SD	p10	Median	p90
AbnRet	2341494	0.000001016	0.00235324	-0.00019735	0	0.00020233
AbnVol	2341494	0.00000303	0.00002949	-5.215E-07	0	0.000003072
SenChar	2341494	153.45108	68.806566	78	143	239
SenWord	2341494	24.496867	10.568369	13	23	38
SenDur	2341494	9.702097	5.291552	4.559999	8.909999	15.51
hnr	2341494	-0.00002821	0.00050142	-0.00014701	-0.000007507	0.00009078
Jitter	2341494	0.02618029	0.00593378	0.01980334	0.02553362	0.03309235
Jitter_rap	2341494	0.01198517	0.00344754	0.00829679	0.01155419	0.01607095
F0	2341494	149.71705	30.043174	117.89783	144.27988	189.58224
stdevF0	2341494	62.448012	33.014911	24.899185	56.624547	108.23285
Shimmer	2341494	0.14222566	0.02426991	0.11289702	0.1406298	0.17400701
Shimmer_apq3	2341494	0.06176007	0.014042	0.04487117	0.060723	0.08005499
zc	2341494	0.08467524	0.02250822	0.05842142	0.08185921	0.11509146
spectral_e~y	2341494	929.76301	102.15283	814.07478	923.86126	1050.5958
Rms	2341494	0.05639429	0.04329161	0.02014435	0.04435063	0.10416046
Energy	2341494	0.07310961	0.20347784	0.00495024	0.0275975	0.16231925

This panel reports descriptive statistics for the variables for the Presentation Section - Positive Textual Sentiment.

Panel D: Summary Statistics for the Presentation Section - Negative Textual Sentiment						
Variables	**N**	**Mean**	**SD**	**p10**	**Median**	**p90**
AbnRet	429652	0.00000601	0.00242438	-0.00021983	0	0.00022304
AbnVol	429652	0.0000031	0.00003054	-6.526E-07	0	0.000003152
SenChar	429652	150.10168	66.790581	78	140	234
SenWord	429652	24.225622	10.457699	12	23	38
SenDur	429652	9.778174	5.499556	4.7	8.960001	15.57861
hnr	429652	-0.00002423	0.00049978	-0.00014158	-0.000006599	0.00008784
Jitter	429652	0.02600211	0.0059844	0.0195413	0.02534044	0.03303039
Jitter_rap	429652	0.01192963	0.00345827	0.00821897	0.01148729	0.0160605
F0	429652	148.19955	30.556606	116.19468	142.14044	189.78635
stdevF0	429652	61.525249	33.215515	23.851186	55.470158	107.72976
Shimmer	429652	0.14225088	0.02444803	0.11264633	0.14065348	0.17436795
Shimmer_apq3	429652	0.06187155	0.01409394	0.04485837	0.0608355	0.08028834
zc	429652	0.08411452	0.02245479	0.05796694	0.08125174	0.11456995
spectral_e~y	429652	926.79737	101.86694	811.03017	921.2291	1047.2139
Rms	429652	0.0535938	0.04102773	0.01914621	0.04245224	0.09841208
Energy	429652	0.06639239	0.15587297	0.00457599	0.02560094	0.14593692

This panel reports descriptive statistics for the variables for the Presentation Section - Negative Textual Sentiment.

Panel E: Summary Statistics for the M&D Section - Neutral Textual Sentiment						
Variables	**N**	**Mean**	**SD**	**p10**	**Median**	**p90**
AbnRet	7461177	0.000002817	0.00201472	-0.00017964	0	0.00017948
AbnVol	7461177	0.000003073	0.00002944	-6.229E-07	0	0.000003252
SenChar	7461177	101.79689	64.104855	39	85	186
SenWord	7461177	18.099331	11.542696	7	15	33
SenDur	7461177	6.472233	5.683211	1.81	5.11	12.6267
hnr	7461177	-0.00003593	0.00052565	-0.00018069	-0.00001054	0.00009804
Jitter	7461177	0.02498359	0.00812812	0.01691832	0.02364952	0.03411728
Jitter_rap	7461177	0.01138603	0.00464132	0.00686465	0.01055225	0.01661212
F0	7461177	146.37446	38.322465	109.12784	138.05333	194.26684
stdevF0	7461177	63.336891	40.523139	15.201964	57.275261	119.20613
Shimmer	7461177	0.1390369	0.0291654	0.10500837	0.13631844	0.1763801
Shimmer_apq3	7461177	0.06042951	0.0165695	0.04157589	0.0586111	0.08123891
zc	7461177	0.09091549	0.0295833	0.05966209	0.08704524	0.12473856
spectral_e~y	7461177	958.84068	133.77171	819.02423	945.43071	1106.9326
Rms	7461177	0.05196979	0.04289599	0.01573934	0.04024509	0.10015407
Energy	7461177	0.04393016	0.12503116	0.00173511	0.01391985	0.0974326

This panel reports descriptive statistics for the variables for the M&D Section - Neutral Textual Sentiment.

Panel F: Summary Statistics for the M&D Section - Positive Textual Sentiment

Variables	N	Mean	SD	p10	Median	p90
AbnRet	1238088	0.000004639	0.00227613	-0.00026065	0	0.00026191
AbnVol	1238088	0.000004281	0.00003461	-7.567E-07	0	0.000005849
SenChar	1238088	132.00265	74.856825	52	117	229
SenWord	1238088	23.214059	13.098905	9	21	40
SenDur	1238088	8.728882	5.940269	2.96	7.530001	15.78
hnr	1238088	-0.00003356	0.00061668	-0.00016337	-0.000009299	0.00009665
Jitter	1238088	0.02497726	0.0069852	0.0178103	0.02396877	0.03296111
Jitter_rap	1238088	0.01154367	0.0040421	0.00743931	0.01090608	0.01619892
F0	1238088	140.44889	31.452172	108.75071	134.59458	178.95854
stdevF0	1238088	63.7643	38.144737	17.969629	57.979563	116.45267
Shimmer	1238088	0.1404762	0.02644343	0.10915226	0.13814911	0.17494669
Shimmer_apq3	1238088	0.06147579	0.01495276	0.04396972	0.05999907	0.08082438
zc	1238088	0.08441556	0.0228647	0.05796059	0.08161541	0.11462341
spectral_e~y	1238088	930.45295	106.12143	809.6066	924.04761	1056.4879
Rms	1238088	0.05409022	0.04248698	0.01857586	0.04243085	0.10121736
Energy	1238088	0.06009889	0.15526079	0.00318621	0.02091889	0.13328914

This panel reports descriptive statistics for the variables for the M&D Section - Positive Textual Sentiment.

Panel G: Summary Statistics for the M&D Section - Negative Textual Sentiment

Variables	N	Mean	SD	p10	Median	p90
AbnRet	182119	0.00000643	0.00230002	-0.00029255	0	0.00029191
AbnVol	182119	0.000004017	0.00003147	-0.000000961	0	0.000005872
SenChar	182119	119.61918	64.269948	51	107	203
SenWord	182119	21.162015	11.261312	9	19	36
SenDur	182119	8.044887	5.513058	2.98822	7.000001	14.21
hnr	182119	-0.0000303	0.000553	-0.00016379	-0.000009012	0.00009636
Jitter	182119	0.02452866	0.00697333	0.01731622	0.0235436	0.03252854
Jitter_rap	182119	0.01128473	0.00405039	0.00716285	0.01065281	0.01593546
F0	182119	139.59549	31.477481	108.04797	133.79375	177.92877
stdevF0	182119	62.983718	37.799912	17.320115	57.377588	114.73793
Shimmer	182119	0.13932752	0.02651794	0.1081128	0.13702337	0.17367018
Shimmer_apq3	182119	0.06106597	0.0149682	0.04361517	0.05956674	0.0803547
zc	182119	0.08605437	0.02313381	0.05896455	0.08349521	0.11652325
spectral_e~y	182119	933.06181	105.99936	811.27805	927.77334	1058.4448
Rms	182119	0.05275903	0.04145363	0.01813286	0.0414798	0.09820103
Energy	182119	0.05293937	0.13676478	0.00293738	0.01888525	0.117034

This panel reports descriptive statistics for the variables for the M&D Section - Negative Textual Sentiment.